

Biogenesis and stability of germline small RNAs in *C. elegans*

by

Allison Chelsa Billi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Human Genetics)
in the University of Michigan
2013

Doctoral Committee:

Assistant Professor John K. Kim, Chair
Professor Sally A. Camper
Professor Eric R. Fearon
Assistant Professor Sundeep Kalantry
Professor John V. Moran

© Allison Chelsa Billi 2013

DEDICATION

To my parents, Jack Billi and Sheryl Hirsch, and my brother, Andrew, for not resenting my perpetual tardiness, not protesting when I brought my laptop to the lakehouse to work, and not teasing me for failing to open it once I was there with them.

ACKNOWLEDGMENTS

I am very grateful to my colleagues in the Kim lab for their invaluable support both experimental and otherwise. I am profoundly indebted to Vishal Khivansara, who has graciously spent countless hours teaching me the techniques and art of molecular biology. Ting Han has been an endless source of scientific knowledge and experience. Through their instruction and examples, these two colleagues have motivated me to achieve more than I thought possible in the last several years. Mallory Freeberg's creative ideas, attention to detail, and formidable expertise in bioinformatics have made our collaborations productive and enjoyable. It has also been a pleasure to collaborate with Amelia Alessi, Tony Chun, and Amanda Day. I was fortunate to begin my research career under the guidance of Miriam Meisler, who taught me that a great mentor is honest, demanding, and fiercely dedicated to promoting the scientific growth of his or her students. I would like also to thank my dissertation committee of Sally Camper, Eric Fearon, Sundeep Kalantry, and John Moran for their insight and advice. Finally, I would like to express my deepest appreciation to my committee chair and mentor, John Kim, who has infected me with his inexhaustible scientific curiosity and equipped me with the tools to pursue it.

PREFACE

As the two first-author publications represented in this dissertation address entirely distinct areas of *C. elegans* small RNA biology, I have enclosed for an introduction a complete literature review of *C. elegans* endogenous small interfering RNA (endo-siRNA) and Piwi-interacting RNA (piRNA) biology. This introduction, Chapter One, is in preparation for submission to WormBook as a solicited review and includes data published after the acceptance of the included manuscripts that address key unanswered questions.

Chapter Two of this dissertation describes the identification and characterization of the functional HEN1 ortholog of *C. elegans*, HENN-1. Taking advantage of the unique proliferation of Argonaute proteins and small RNA classes in *C. elegans*, I have illuminated conserved mechanisms for selective stabilization of small RNAs that have defied explanation in previous animal studies of HEN1. This work was published in *PLoS Genetics* in 2012.

Chapter Three of this dissertation describes my collaborative work with bioinformatician Mallory Freeberg defining paradigm-breaking mechanisms for piRNA expression in *C. elegans*. I used *C. elegans* transgenesis to address longstanding questions in the field, demonstrating that, unlike other animal piRNAs, *C. elegans* piRNAs are expressed independently from tiny, autonomous transcriptional units. Our investigations further revealed that the upstream sequence motif unique to the nematode piRNA is required for piRNA expression and directs restricted male or female germline enrichment through variation at a single nucleotide position. This work was published in *PLoS Genetics* in 2013.

A discussion of future research prospects concludes the dissertation as Chapter Four.

Appended are a first-author review of *C. elegans* piRNA function (published in *Genome Biology* in 2012), a co-second author study identifying small RNA pathway factors through phylogenetic signatures (published in *Nature* in 2013), and a mid-author study describing a conserved role for MORC ATPases in gene silencing (published in *Science* in 2012).

TABLE OF CONTENTS

ii	DEDICATION
iii	ACKNOWLEDGMENTS
iv	PREFACE
viii	LIST OF FIGURES
xii	LIST OF TABLES
xiii	LIST OF APPENDICES
xiv	LIST OF ABBREVIATIONS AND ACRONYMS
xvi	ABSTRACT
1	CHAPTER ONE: Introduction to <i>Caenorhabditis elegans</i> germline small RNAs
1	INTRODUCTION
3	MAIN TEXT
65	CONTRIBUTIONS OF THE DISSERTATION
68	REFERENCES
78	CHAPTER TWO: The <i>Caenorhabditis elegans</i> HEN1 ortholog, HENN-1, methylates and stabilizes select subclasses of germline small RNAs
78	ABSTRACT
79	AUTHOR SUMMARY
80	INTRODUCTION
83	RESULTS
106	DISCUSSION
110	MATERIALS AND METHODS
116	ACKNOWLEDGMENTS
117	SUPPLEMENT
140	REFERENCES

145	CHAPTER THREE: A conserved upstream motif orchestrates autonomous, germline-enriched expression of <i>Caenorhabditis elegans</i> piRNAs
145	ABSTRACT
146	AUTHOR SUMMARY
147	INTRODUCTION
149	RESULTS
171	DISCUSSION
175	MATERIALS AND METHODS
182	ACKNOWLEDGMENTS
183	SUPPLEMENT
197	REFERENCES
201	CHAPTER FOUR: Future directions
201	MAIN TEXT
219	REFERENCES
223	APPENDICES

LIST OF FIGURES

86	2.1. Methylation of 21U RNAs Requires <i>C. elegans</i> HEN1 Ortholog HENN-1
90	2.2. HENN-1 Stabilizes 21U RNAs
92	2.3. HENN-1 Selectively Methylates ERGO-1 Class 26G RNAs in an ERGO-1-dependent Manner
94	2.4. ERGO-1 is Required for Methylation of 26G RNAs
97	2.5. HEN1 Stabilizes ERGO-1 Class, but Not ALG-3/ALG-4 Class, 26G RNAs
100	2.6. The <i>henn-1</i> Mutant Exhibits Opposite RNAi Sensitivity Phenotypes in Soma and Germline
103	2.7. HENN-1 is Broadly Expressed in <i>C. elegans</i> Germline
117	2.S1. Alignment of HEN1 Ortholog
119	2.S2. <i>C02F5.6</i> Alleles and Transgenes
120	2.S3. Methylation Status of Additional Small RNAs
121	2.S4. Diverse 21U RNAs Exhibit HENN-1 Dependence in Early Development and Adulthood
123	2.S5. miRNAs Do Not Exhibit HENN-1 Dependence
124	2.S6. HENN-1 Dependence of Substrate-dependent Secondary siRNAs
125	2.S7. <i>henn-1</i> Contributes to Robust Fertility at Elevated Temperatures
126	2.S8. Many ERGO-1 Class 26G RNAs Exhibit HENN-1 Dependence throughout Development
128	2.S9. ALG-3/ALG-4 Class 26G RNAs Do Not Exhibit HENN-1 Dependence
130	2.S10. The <i>henn-1(tm4477)</i> Mutant Does Not Exhibit Significant Upregulation of ERGO-1 Class 26G RNA Target mRNAs
131	2.S11. The <i>henn-1(tm4477)</i> Mutant Exhibits a Mild but General Somatic Eri Phenotype
132	2.S12. The <i>henn-1(tm4477)</i> Mutant Exhibits a General Germline Rde Phenotype
133	2.S13. HENN-1 is Broadly Expressed in the Germline and Soma
135	2.S14. Comparison of <i>C. elegans</i> Argonautes
151	3.1. Over 70% of 21U RNAs show distinct germline enrichment

155	3.2. Variation in the core upstream motif correlates with 21U RNA germline enrichment
158	3.3. A transgenic synthetic 21U RNA shows characteristics of endogenous 21U RNAs
162	3.4. A 5' cytidine in the core upstream motif promotes male germline expression pattern of 21UR-synth
166	3.5. 21U RNA sequences are specified by the genomic positions of upstream core motifs
170	3.6. 21U RNAs represent independent transcriptional units
183	3.S1. Computational identification of male and female germline-enriched 21U RNAs
184	3.S2. Female 21U RNAs are preferentially abundant in embryo
186	3.S3. 21U RNAs target significantly non-overlapping sets of genes
187	3.S4. Transgenic array expression varies across transgenes
189	3.S5. 21U RNAs are specifically immunoprecipitated with PRG-1 complexes
190	3.S6. 21U RNA expression requires a 5' genomic thymidine
191	3.S7. RNA polymerase II occupancy at 21U RNA loci is below background level
193	3.S8. Model of 21U RNA expression
201	4.1. Argonaute cladistics reflect selective methylation
202	4.2. Argonaute rescue transgene diagrams
209	4.3. Female 26G RNAs are specifically depleted in the <i>tkt-1Δ</i> mutant
211	4.4. The pentose phosphate pathway
216	4.5. Candidate small RNA pathway factor mutant screening results
234	B.1. Phylogenetic profiling analysis shows correlated conservation patterns of <i>C. elegans</i> proteins
236	B.2. Phylogenetic clusters of candidate small RNA pathway proteins
238	B.3. Select phylogenetic clusters enriched with hits from proteomic and functional genomic small RNA screens
240	B.4. Inactivation of genes implicated in RNAi pathways re-animates transgenes that are silenced by RNAi
244	B.S1. Phylogenetic profiles of the 50 proteins mostly correlated with
246	B.S2. Identification of proteins that cluster phylogenetically with known small RNA co-factors or with hits from a set of small RNA genetic and biochemical screens

- 247 B.S3. Overlap of genes and known small RNA factors between different screens without (A-B) and with (C-D) taking into account the phylogenetic clustering
- 248 B.S4. Proteins assigned to the spliceosome by KEGG pathway analysis
- 249 B.S5. Relation between the average number of introns per gene in each species and the presence or absence of Argonaute proteins
- 250 B.S6. Inactivation of splicing factors that are implicated in the RNAi pathways reanimates transgenes targeted by RNAi
- 251 B.S7. Receiver operating characteristic (ROC) analysis
- 277 C.1. Mutations of two MORC homologs induce *SDC::GFP* and TE overexpression
- 279 C.2. DNA methylation is not impaired in *atmorc1* and *atmorc6* mutants
- 282 C.3. AtMORC1 and AtMORC6 are required for maintenance of chromatin architecture and form nuclear bodies near chromocenters, and *morc-1* is involved in gene silencing in *C. elegans*
- 290 C.S1. Schematic representation of the *SDC::GFP* construct, and Schematic of AtMORC1 and AtMORC6 proteins showing the GHKL and S5 ATPase domains together with putative Coiled-Coil (CC) domains
- 291 C.S2. Mapping of *cmt3* #7, wt #67 and *cmt3* #49 mutations by bulk segregant analysis coupled with whole genome re-sequencing
- 292 C.S3. Characterization of *atmorc1* and *atmorc6* EMS and T-DNA mutant alleles
- 294 C.S4. Similar sets of TEs and protein-coding genes are upregulated in *atmorc1* and *atmorc6*
- 295 C.S5. DNA methylation at all protein-coding genes and transposons is not altered in *atmorc1* and *atmorc6*
- 296 C.S6. H3K9me2 is not altered in *atmorc1* and *atmorc6*
- 297 C.S7. Small RNA accumulation is unaltered in *atmorc1* and *atmorc6*
- 298 C.S8. AtMORC1 and AtMORC6 are required for chromocenter condensation
- 299 C.S9. H3K9me2 and DAPI staining show similar chromocenter condensation patterns in nuclei defined in Fig. S8 as nuclei showing wild type, intermediate or decondensed chromocenters in *cmt3-11 atmorc1-3*
- 300 C.S10. H3K9me2 and DAPI staining show similar chromocenter condensation patterns in nuclei defined in Fig. S8 as nuclei showing wild type, intermediate or decondensed chromocenters in *atmorc6-1*
- 301 C.S11. H3K9me2 and DAPI staining show similar chromocenter condensation patterns in nuclei defined in Fig. S8 as nuclei showing wild type, intermediate or decondensed chromocenters in *atmorc1-3 atmorc6-1*

- 302 C.S12. Loci upregulated in *atmorc1* and *atmorc6* mostly localize to pericentromeric regions
- 303 C.S13. Interaction matrix of the *atmorc6-1* genome from Hi-C analysis
- 304 C.S14. Myc tagged AtMORC1 and AtMORC6 complement the EMS mutant lines
- 305 C.S15. AtMORC6 bodies are adjacent to chromocenters

LIST OF TABLES

137	2.S1. Oligonucleotides for Northern Blot Analysis
138	2.S2. Small RNA Sequences for Taqman Probe Design
139	2.S3. Primers for RT-qPCR
160	3.1. Descriptions of transgenic alleles and features of the transgenes
194	3.S1. Descriptions of small RNA sequencing libraries used in this study
195	3.S2. Welch's <i>t</i> -test p-values for all abundance comparisons between 21U RNAs with different core motifs
203	4.1. Assays for 26G RNA biogenesis and function
212	4.2. <i>C. elegans</i> PPP genes (Fig. 4.4) and small RNA pathway evidence
252	B.S1. The <i>C. elegans</i> phylogenetic profile database
252	B.S2. Top siRNA pathway candidates and experimental tests of informatic predictions
253	B.S3. Overlap between positives in each of the functional genomic and proteomic screens and with the lists of known siRNA and miRNA pathway proteins
254	B.S4. Phylogenetic Clustering of hits from small RNA functional genomic screens
255	B.S5. Top miRNA pathway candidates by Bayesian analysis
255	B.S6. RNA interference defects after gene inactivations of <i>C. elegans</i> orthologues of known splicing factors
255	B.S7. Genome-wide transgene desilencing screen positives
256	B.S8. The validated siRNA and the miRNA pathway proteins
307	C.S1. Lists of Transposable elements (TEs) 4'fold upregulated in both EMS and T-DNA <i>atmorc6</i> and <i>atmorc1</i> mutant alleles
308	C.S2. Lists of Protein coding genes 4-fold upregulated in both EMS and T-DNA <i>atmorc6</i> and <i>atmorc1</i> mutant alleles
309	C.S3. Lists of Transposable elements (TEs) and protein-coding genes (PCGs) that were 4-fold upregulated in the <i>atmorc1-3 atmorc6-1</i> double mutant
311	C.S4. Sequences of primers used in this study

LIST OF APPENDICES

223	APPENDIX A: piRNAs and siRNAs collaborate in <i>Caenorhabditis elegans</i> genome defense
223	ABSTRACT
223	MAIN TEXT
228	ACKNOWLEDGMENTS
229	REFERENCES
230	APPENDIX B: Identification of small RNA pathway genes using patterns of phylogenetic conservation and divergence
230	ABSTRACT
231	MAIN TEXT
242	METHODS SUMMARY
243	ACKNOWLEDGMENTS
244	SUPPLEMENT
271	REFERENCES
274	APPENDIX C: MORC family ATPases required for heterochromatin condensation and gene silencing
274	ABSTRACT
275	MAIN TEXT
283	ACKNOWLEDGMENTS
283	MATERIALS AND METHODS
290	SUPPLEMENT
312	REFERENCES

LIST OF ABBREVIATIONS AND ACRONYMS

AU: arbitrary units

β e: beta-elimination

csRNA: capped small RNA

C. elegans: *Caenorhabditis elegans*

Chip-seq: chromatin immunoprecipitation sequencing

dsDNA: double-stranded DNA

dsRNA: double-stranded RNA

E: embryo

endo-: endogenous

Ekl: enhancement of *ksr-1* lethality (phenotype)

Eri: enhanced RNA interference (phenotype; complex: ERI)

exo-: exogenous

GFP: green fluorescent protein

H3K9me(2/3): histone H3 lysine 9 (di-/tri-)methylation

HENN-1: HEN1 of nematode

Him: high incidence of male (phenotype)

hr: hour

kb: kilobase(s)

miRNA: microRNA

MORC: Microrchidia

MRS: max ratio score

Mut: mutator (phenotype; complex: MUT)

NLS: nuclear localization signal

Nrde: nuclear RNA interference defective (phenotype, pathway)

nt: nucleotide

piRNA: Piwi-interacting RNA

Pol II: RNA polymerase II
PPP: pentose phosphate pathway
qPCR: quantitative polymerase chain reaction
Rde: RNA interference defective (phenotype; complex: RDE)
RdRP: RNA-dependent RNA polymerase
RIP: RNA immunoprecipitation
RISC: RNA-induced silencing complex
RNAe: RNA-induced epigenetic silencing
RNAi: RNA interference
RPM: reads per million
RT-qPCR: real-time quantitative polymerase chain reaction
SD: standard deviation
SEM: standard error of the mean
siRNA: small interfering RNA
ssRNA: single-stranded RNA
TE: transposable element
ts: temperature-sensitive
TSS: transcription start site
UTR: untranslated region
Wago: worm-specific Argonaute
WT: wild-type

ABSTRACT

Across the animal kingdom, small, noncoding RNAs preserve and promote fertility by engaging Argonaute effector proteins to silence deleterious genetic elements. Generated in germline and inherited into progeny, endogenous small interfering RNAs (endo-siRNAs) and Piwi-interacting RNAs (piRNAs) regulate vast suites of gametic and zygotic genes, yet remarkably little is known about how they are regulated. With an expanded repertoire of small RNA classes, *Caenorhabditis elegans* provides an ideal model for investigating how animals drive epigenetic inheritance of fertility-preserving germline small RNAs.

The conserved methyltransferase HEN1 methylates small RNAs to prevent their degradation. Methylation of germline small RNAs enhances accumulation, promoting robust inheritance into progeny. All plant small RNAs are methylated, but animal HEN1 methylates only some small RNAs. The mechanisms of selective methylation were unknown. I identified the functional *C. elegans* ortholog of HEN1 and demonstrated that it methylates all piRNAs but only select subclasses of endo-siRNAs. I further found that particular endo-siRNAs are methylated in maternal, but not paternal, germlines. Through genetic and biochemical analyses, I showed that small RNA methylation status is likely dictated by the associated Argonaute. This established selective expression of divergent Argonautes as a novel mechanism for differentially stabilizing germline small RNAs, with significant implications for preferential inheritance of maternal epigenetic information.

piRNAs are essential for animal fertility, but their expression mechanisms are poorly characterized. In collaboration with bioinformatician Mallory Freeberg, I showed that *C. elegans* male and female germlines express distinct piRNA subsets that evolve independently and differ in inheritance. A common sequence motif lies upstream of

nematode piRNA loci. We discovered that this motif varies significantly between male and female piRNAs. Using a novel transgenic approach, I established that *C. elegans* piRNAs represent thousands of tiny, autonomous transcriptional units, rivaling coding genes in number. I further demonstrated that the upstream motif is required for piRNA expression and that variation at a single nucleotide position within this motif orchestrates selective male versus female germline enrichment and inheritance of piRNAs.

These and additional included studies define novel factors and mechanisms involved in regulation of germline small RNAs and transgenerational transmission of their crucial epigenetic information.

CHAPTER ONE: Introduction to *Caenorhabditis elegans* germline small RNAs

INTRODUCTION

C. elegans endogenous small non-coding RNAs are subdivided into microRNAs, endogenous small interfering RNAs (endo-siRNAs), and Piwi-interacting RNAs (piRNAs). All three types of small RNAs bind to Argonaute effector proteins, recognize target transcripts exhibiting partial or perfect complementarity, and direct target regulation that is primarily inhibitory in nature. Yet microRNAs, endo-siRNAs, and piRNAs vary greatly in biogenesis mechanisms, specific protein cofactors, and even effector function. The details of *C. elegans* microRNA biology have been previously described [1] and will be addressed in this review only insofar as they intersect with endo-siRNA and piRNA biology. Similarly, a thorough examination of exogenously-derived siRNAs and exogenous RNAi (exo-RNAi) is beyond the scope of this review; however, as *C. elegans* exo-RNAi engages a downstream endo-RNAi amplification pathway shared by primary endo-siRNAs and piRNAs, only the mechanisms of primary exo-siRNA biogenesis are excluded.

The transcript silencing capacity of antisense RNA was first described in *C. elegans* over two decades ago [2]. The effective interfering agent was subsequently determined to be double-stranded RNA, and its incredible potency suggested the existence of a catalytic or amplification mechanism engaged by exogenous dsRNA [3]. This dsRNA is processed into primary exo-siRNAs [4,5], increasing the ratio of trigger to target, but still to a degree insufficient to explain the potency of exo-RNAi. Further studies of the small RNA effector populations during exo-RNAi in *C. elegans* revealed that primary exo-siRNAs are not the ultimate effectors of interference; rather, they trigger vigorous production of secondary siRNAs by RNA-dependent RNA polymerases (RdRPs),

amplifying the signal and focusing the interference response on expressed sequences [6]. Before the discovery of endogenous silencing pathways in *C. elegans*, isolation of mutations causing deficiencies in both exo-RNAi and endogenous silencing phenomena such as transposon silencing suggested a shared mechanism [7-9]. Mutations in core molecular machinery of RNAi also were found to result in phenotypes indicating essential endogenous roles. Loss of DCR-1, the sole *C. elegans* ortholog of Dicer, results in profound cell fate specification defects and germline abnormalities leading to lethality and sterility [10-12]; RNAi-mediated depletion of Argonautes PRG-1 and PRG-2 impairs germline stem cell maintenance [13]; and the RdRP EGO-1 is required for germline development [14].

The discovery of the first microRNA, *lin-4*, in *C. elegans* suggested that endogenous products also initiate transcript regulation through antisense mechanisms [15]. It was not until a decade later that other endogenous small RNAs were identified in *C. elegans* [16,17], and only in 2006 did deep sequencing first reveal the incredible diversity of *C. elegans* small RNAs [18]. Among the species identified were a large pool of 5' guanosine antisense small RNAs identified as endo-siRNAs that appeared to represent distinct 26- and 22-nucleotide (nt) subpopulations, later determined to correspond to primary and secondary endo-siRNAs, respectively [18-23]. Subsequent dissection of these 26G and 22G RNA populations identified unique subgroups with largely overlapping biogenesis requirements but that engage distinct effector pathways. Most notably, the 22G RNAs are subdivided into the WAGO 22G RNAs and CSR 22G RNAs; the former represent secondary siRNAs that effect target silencing, whereas the latter are a class of siRNAs complementary to germline-expressed transcripts that do not silence target genes but rather promote their proper organization during mitosis [24,25]. Also identified in the initial deep sequencing dataset were the 21U RNAs, 5' uridine 21-nt small RNAs later determined to represent the piRNAs of *C. elegans* [26,27]. This review summarizes the literature contributing to our current understanding of the *C. elegans* 26G RNAs, WAGO 22G RNAs, CSR 22G RNAs, and 21U RNAs, discussing mechanisms of triggering, biogenesis, and effector function, where known. Whereas microRNAs are required for diverse developmental and physiological processes in the

soma, endo-siRNAs and 21U RNAs serve as the guardians of the immortal germline, constituting a complex, interconnected, and tremendously robust system for surveillance of the *C. elegans* genome.

MAIN TEXT

Features and targets of 26G RNAs

Early deep sequencing of *C. elegans* small RNAs revealed a distinct population of 26 nt species that are largely anti-sense to annotated genes and therefore classified as endo-siRNAs [18]. Like endo-siRNAs sequenced earlier in *C. elegans* [28], these 26 nt species show a 5' guanosine bias and are thus termed 26G RNAs [18,20,29-31]. 26G RNAs are 5' monophosphorylated [18,20,30,31] and enriched for adenosine and guanosine across their lengths [31]. 26G RNAs are quite enriched in male and female germlines [20], where they comprise two distinct subpopulations that are temporally isolated and bound by unique, germline-specific effector complexes. 26G RNAs in the spermatogenic gonad are bound by redundant Argonautes ALG-3 and ALG-4, whereas 26G RNAs in oogenic gonad are bound by the ERGO-1 Argonaute [19,20,23]. ERGO-1 class 26G RNAs are also highly abundant in embryo [20,30] and perdure through early larval development [31].

Little is known about how transcripts are selected for targeting by 26G RNAs. 26G RNAs map primarily to protein-coding genes with a strong antisense bias, although some target unannotated loci [18,20,30]. They are transcribed from spliced mRNA templates by the RNA-dependent RNA polymerase RRF-3 [19,20,31,32], as indicated by the sequencing of rare species spanning exon-exon junctions [18,20,31] and by the loss of complementary 26G RNAs upon nonsense-mediated decay of template mRNA [20].

Targets of ALG-3/4 class 26G RNAs are highly enriched for transcripts classified as spermatogenesis-enriched [19,20,32,33], explaining earlier observations connecting endo-siRNAs primarily with regulation of sperm function genes [18,34]. ALG-3/4 class 26G RNAs map across the lengths of their target sequences, but preferentially target transcript 5' and 3' termini; transcripts exhibiting higher 5' UTR targeting appear to be more efficiently silenced [19].

Although biogenesis of ERGO-1 class 26G RNAs initiates in oogenesis, their targets are depleted of germline-intrinsic transcripts [20,33]; accordingly, ERGO-1 class 26G RNAs primarily regulate zygotic targets throughout development [20,31]. The majority of abundant ERGO-1 class 26G RNAs map to within 5 Mb of chromosome ends at more gene-poor regions [23,35]. Unlike ALG-3/4 class 26G RNAs, ERGO-1 class 26G RNAs are excluded from the first ~100 nt of target transcripts [23]. Roughly half of ERGO-1 class 26G RNAs map to coding loci or pseudogenes and half to loci that likely correspond to unannotated transcription units [23]. Many of their targets represent ancient duplications, suggesting ERGO-1 class 26G RNAs may buffer expression of rapidly expanding gene families [23,35]. A large-scale proteomics dataset [36] shows peptides corresponding to less than 20% of ERGO-1 class 26G RNA targets, but 54% of all annotated coding genes, indicating that very few of these targets represent functional, coding loci [35].

Biogenesis of 26G RNAs: The ERI complex

The ERI complex mediates 26G RNA biogenesis and is named for the enhanced *RNAi* (Eri) phenotype that is associated with compromise of ERGO-1 class 26G RNA function. Loss of ERI complex members also results in temperature-sensitive (ts) sterility at 25°C due to defective spermatogenesis as well as a *high incidence of male progeny* (Him) phenotype, indicative of increased X chromosome nondisjunction, due to compromised ALG-3/4 class 26G RNA function. The Eri and ts sterile phenotypes are explained in greater detail below. At the heart of the ERI complex is a core RdRP

module consisting of RRF-3, DRH-3, and ERI-5. Associated with this module are DCR-1, accessory factors ERI-1b, ERI-3, RDE-4, and possibly the RNA phosphatase PIR-1. Analyses of DCR-1, ERI-1b, and ERI-5 complexes immunopurified from gravid adult and embryo have revealed interactions with each other member of the ERI complex [37,38], and the ~850 kD estimated mass of the ERI complex is similar to the summed mass of these proteins, ~810 kD [38]. The ERI complex shares only DCR-1 and RDE-4 with the RDE (*RNAi defective*) complex, which mediates processing of dsRNA trigger to initiate the *exo-RNAi* pathway [39]. DCR-1 and ERI-1 also both interact with ERI-9, another factor implicated in ERGO-1 class 26G RNA biogenesis [38], suggesting that the ERI complex includes other transient interactors.

ERI complex factors show some interdependence and hierarchy of assembly. Both ERI-5 and ERI-3 bind the N-terminal helicase domain of DCR-1, and both are required for ERI-1b to associate robustly with DCR-1. Within the RdRP module, ERI-5 protein requires RRF-3 for full accumulation and association with DRH-3 and DCR-1. DRH-3 does not interact directly with DCR-1, so ERI-5 and RRF-3 recruit the RdRP module to DCR-1, independently of ERI-1b and ERI-3. ERI-5 is proposed to tether the RdRP module to DCR-1 to potentiate 26G RNA biogenesis [38]. The association between the RdRP module and DCR-1 does not require dsRNA substrate production, as the complex assembles normally with mutant, catalytically inactive RRF-3 [38].

While the other factors are required for biogenesis of 26G RNAs, loss of ERI-5 merely attenuates their expression due to partial compensation by paralog EKL-1. RNAi-mediated knockdown of *ekl-1* does not affect 26G RNA levels in wild type, but completely depletes them in the *eri-5* mutant background [38]. EKL-1 interacts with DRH-3 [24] and substitutes for ERI-5 in the RdRP module, but it does not interact with DCR-1 [38]. Untethered to DCR-1, this EKL-1 RdRP module still produces 26G RNAs of normal genomic distribution and in a DCR-1-dependent manner, but 26G RNA accumulation is impaired [38].

RRF-3: RRF-3 is an RNA-dependent RNA polymerase (RdRP) [40] required for 26G RNA biogenesis [19,20,23,31,32,37,41,42]. The *rrf-3* transcript is classified as germline-intrinsic [33], but an *rrf-3::gfp* transcriptional fusion transgene shows expression in many different larval and adult cell types [43]. *C. elegans* RdRPs and homologs catalyze primer-independent synthesis of antisense siRNAs [21,22,44,45] (see RRF-1 below). The RdRP activity of RRF-3 is necessary for 26G RNA accumulation and function [42], suggesting that RRF-3 directly transcribes 26G RNAs. Each 26G RNA is likely generated as an independent transcript, as homologous RdRPs exhibit very low processivity [44,45], and 26G RNAs show irregular phasing and occasional overlap [18,23,35]. The N terminus of RRF-3 is divergent and may confer target specificity to the ERI RdRP module [38], as DRH-3 is also involved in biogenesis of other endo-siRNAs. In addition to loss of 26G RNAs, *rrf-3* mutants show depletion of dependent secondary siRNAs, called WAGO 22G RNAs, whose biogenesis is triggered by 26G RNA targeting [19,23,31,32,37,41]. Loss of RRF-3 results in the characteristic sperm-origin ts sterile, Him, and Eri phenotypes associated with compromise of the 26G RNA pathway [37,42,46,47].

DRH-3: The conserved *Dicer-related helicase* genes include two homologs, *drh-1* and *drh-3*, and a probable pseudogene, *drh-2*. These genes are named for the similarity of their encoded DExD/H box helicase domains to that of DCR-1 and likely act as helicases upon dsRNA intermediates during biogenesis of siRNAs [39]. The functions of DRH-1 and DRH-3 in siRNA biogenesis complexes differ. DRH-1 interacts with DCR-1, RDE-4, and Argonaute RDE-1 within the RDE complex to process primary exo-siRNAs from dsRNA trigger [38,39]. DRH-1 interacts directly with DCR-1 but does not participate in an RdRP module, whereas DRH-3 does not interact directly with DCR-1 but appears to represent an essential component of all *C. elegans* RdRP modules [24,25,37,44,48]. Thus, loss of DRH-3 results in loss of 26G RNAs [23], although many other populations of small RNAs are also lost. Accordingly, the phenotypes associated with loss of DRH-3 are pleiotropic, with more severe phenotypes masking defects attributable to loss of 26G RNAs, and are discussed below with the relevant WAGO and CSR 22G RNA pathways.

ERI-5: ERI-5 is a protein containing two Tudor domains that putatively tethers the ERI complex RdRP module to DCR-1 [37,38]. In vitro, recombinant ERI-5 binds recombinant DCR-1 but fails to bind recombinant RRF-3, suggesting that RRF-3 may be modified in vivo to generate Tudor domain-binding sites [38]. Loss of ERI-5 results in the sperm-origin ts sterility and Eri phenotype of ERI complex mutants [37]. However, due to redundancy with EKL-1, 26G RNAs are only depleted twofold and dependent WAGO 22G RNAs ninefold in *eri-5* mutant embryo, with modest impairment of 26G RNA target regulation [37,38].

DCR-1: Dicer contains a helicase domain, a PAZ domain, and dual RNase III motifs and is a conserved member of a family of RNase III nucleases that cleave dsRNA [49]. In *C. elegans* extracts, DCR-1 dices long dsRNA into 23 bp duplexes [11]. Immunopurified DCR-1 complexes digest dsRNA processively into siRNAs in the presence of ATP but terminate processing when ssRNA is encountered, demonstrating the specificity for dsRNA substrate [11]. *dcr-1* encodes the sole Dicer homolog of *C. elegans* and acts in microRNA, primary exo-siRNA, and primary endo-siRNA biogenesis [11,20,50,51]. Loss of DCR-1 results in sterility, with abnormal oocyte formation and absence of fertilization [11,50,51], as well as heterochronic phenotypes due to defects in microRNA processing [11,50]. The *dcr-1* transcript is classified as germline-intrinsic [33], and maternal inheritance of DCR-1 somewhat ameliorates somatic *dcr-1* null mutant phenotypes. The *dcr-1* mutant germline, however, shows an *RNAi-defective* (Rde) phenotype and impaired transgene silencing [11,50,51]. Although 26G RNAs are transcribed by RRF-3, they do not show the signature 5' triphosphate of unprimed synthesis by an RdRP [45]. Rather, 26G RNAs show a 5' monophosphate characteristic of Dicer products, and indeed DCR-1 is required for their biogenesis [20], as well as accumulation of dependent WAGO 22G RNAs [37,41].

DCR-1 helicase domain: The helicase domain of DCR-1 appears to play a specific role in biogenesis of primary siRNAs. The *dcr-1(mg375)* allele encodes a missense mutation in the helicase domain that does not affect ERI complex formation but abrogates biogenesis of 26G RNAs and dependent 22G RNAs [42,52]. The *dcr-1(mg375)* helicase

mutant shows sperm-origin ts sterile, Him, and Eri phenotypes, but is viable, fertile, and does not show heterochronic defects, indicating that DCR-1 helicase activity is dispensable for microRNA production [42,52]. The RNAi sensitivity of the *dcr-1(mg375)* helicase mutant is less enhanced than that of other ERI complex mutants [42], suggesting that the helicase activity of DCR-1 may enhance primary exo-siRNA production from dsRNA triggers. This would be consistent with the theory that the helicase domain also functions as a translocase, allowing DCR-1 to catalyze multiple, processive cleavage events before dissociation from a dsRNA substrate [52]. Study of the function of DCR-1 in cell-free embryo extracts has provided some insight into the specific contribution of the helicase domain to small RNA biogenesis [53]: The *C. elegans* Dicer homolog appears to measure from the 3' terminus of dsRNA substrate. DCR-1-mediated cleavage of substrates with 3'-overhanging termini does not require helicase activity and generates 21-23 nt products. This accounts for the intact microRNA levels in the *dcr-1(mg375)* helicase mutant and the observed length of mature microRNAs, which are cleaved from precursors with 3' overhangs. In contrast, cleavage of blunt or 5'-overhanging substrates is impaired in *dcr-1(mg375)* helicase mutant extract. This may be because blunt and 5'-overhanging termini engage the helicase domain to unwind the dsRNA and enable processive cleavage without dissociation: DCR-1 cleavage of blunt or 5'-overhanging termini, but not 3'-overhanging termini, is highly processive, yielding cleavage products from internal dsRNA regions. This processive cleavage requires both ATP and the helicase activity. Cleavage of short (~40 bp) blunt-ended dsRNA yields a 26 nt small RNA with a 22-23 nt passenger strand. Processing of longer (~100 bp) blunt-ended dsRNA in the presence of ATP yields both 26 and 27 nt species, with subsequent internal cleavages yielding ~23 nt duplexes with 3' overhangs. Failure to produce 26 nt species from internal cleavages strongly supports independent processing of 26G RNAs from short RdRP products rather than sequential cleavage of a long dsRNA precursor. Finally, in embryo extract, dsRNA substrates are cleaved with similar efficiency by DCR-1 regardless of the 5' nt identity, indicating that the 5' guanosine bias of 26G RNAs and other endo-siRNAs is not imposed by preferential DCR-1 processing [53].

ERI-1b: *eri-1* encodes two isoforms of a conserved RNase that contains a DEDDh-like 3' to 5' exonuclease and a SAP/SAF-box domain [54]. In vitro, ERI-1 partially degrades siRNAs with 2-nt 3' overhangs, but not ssRNA or siRNA internally hybridized to a long RNA [54]. In vivo, ERI-1 is required for 26G RNA biogenesis and degrades the 3' end of the 5.8S rRNA, which pairs with the 5' end of the 25-28S rRNA. Loss of ERI-1 results in sperm-origin ts sterile, Him, and Eri phenotypes with loss of 26G RNAs and dependent 22G RNAs [19,20,31,37,54,55]. The significance of 5.8S rRNA processing is unknown. The *eri-1* transcript is classified as germline-intrinsic [33], but its expression is not restricted to germline tissues [54]. Both ERI-1 isoforms are cytoplasmically localized, and dsRNA exposure does not change expression or localization [54,55]. Whereas either ERI-1 protein isoform is capable of rescuing 5.8S rRNA processing in vivo, *eri-1* mutant phenotypes and 26G RNA accumulation are only rescued by expression of ERI-1b [54,55]. ERI-1a may be insufficient for rescue because only ERI-1b interacts with DCR-1 [37], possibly via its extended C-terminal domain [54,55]. ERI-1 exonuclease activity is required for 26G RNA accumulation and rRNA processing [55], but the precise function of ERI-1b in 26G RNA biogenesis is unclear. Possibly, ERI-1b recognizes 3' stem-loop structures in mRNA targets and removes excess nucleotides to generate a suitable RNA substrate for RRF-3-dependent synthesis [37]. Alternatively, RRF-3 may synthesize dsRNA with short 3' overhangs that must be processed by ERI-1b to create blunt termini that engage DCR-1-mediated production of 26 nt species [53].

ERI-3: ERI-3 is a protein without identifiable domains encoded by a germline-intrinsic transcript [33,37]. *eri-3* is encoded in an operon with *taf-6.1* and can be expressed as a single polypeptide or as a fusion protein with TAF-6.1, which is also detected in immunopurified DCR-1 complexes [37]. Loss of *eri-3* results in the characteristic phenotypes of ERI complex compromise [37], and ERI-3 may serve only to recruit ERI-1b to DCR-1.

RDE-4: RDE-4 contains two dsRNA-binding motifs and binds long dsRNA preferentially in vitro without specificity for sequence or overhang structure [39,56]. Although the *rde-4* transcript is classified as oogenesis-enriched [33], RDE-4 protein mediates exo-RNAi in

both germline and soma by interacting with DCR-1, DRH-1, and the Argonaute RDE-1 [57]. In vitro and in vivo studies of recombinant RDE-4 have dissected the contributions of the constituent domains to RDE-4 function: The C-terminal domain of RDE-4 directs its dimerization in solution and is required for its function in siRNA production, possibly by activating DCR-1, but dispensable for dsRNA binding [56,58]. The 3' dsRNA-binding motif is important for dsRNA binding and mediates interaction with DCR-1 along with the linker between the motifs [58,59]. RDE-4 binds dsRNA cooperatively [56,58] and promotes its accumulation in vivo [39]. During exo-RNAi, RDE-4 binds long trigger dsRNA in vivo, but not mRNA or amplified siRNAs [39]. RDE-4 is required for primary exo-siRNA and dependent 22G RNA production, but its absence can be partially bypassed by injection of prepared 24-25 nt siRNA duplexes [39,60]. RDE-4 is required for full production of ERGO-1 class 26G RNAs and dependent 22G RNAs [23,37,41,52]. Although its role in ALG-3/4 class 26G RNA production has not been explicitly tested, RDE-4 appears to be required for detection of an endo-siRNA corresponding to ALG-3/4 class 26G target mRNA *ssp-16* [52], and microarray profiling indicates that mRNA levels of *ssp-16* are elevated in *rde-4* mutant adult [61]. In spite of this, loss of RDE-4 does not appear to result in the Him phenotype and highly penetrant sperm-origin sterility associated with compromise of ALG-3/4 class 26G RNAs: while RDE-4 is required for full fertility, the likely null *rde-4(ne299)* mutant appears capable of propagation at elevated temperature (25°C), albeit with some embryonic lethality and developmental delay and arrest [52,59]. Perhaps loss of RDE-4 does not fully abrogate 26G RNA biogenesis, as target desilencing is less profound in *rde-4* than *eri-1* mutant embryo [62]. The function of RDE-4 in 26G RNA biogenesis remains somewhat unclear. The preference shown by RDE-4 for binding of long dsRNA may promote exo-RNAi by aiding release of dsRNA siRNA duplexes after DCR-1 processing [56], but argues against a role for RDE-4 in binding and stabilizing the likely very short dsRNA 26G RNA precursor.

PIR-1: Although PIR-1 was initially identified as a member of the ERI complex that is also required for exo-RNAi [37], no data have since been reported indicating a role for PIR-1 in either pathway.

Biogenesis of 26G RNAs: Other factors implicated in 26G RNA biogenesis

ERI-9: ERI-9 is a worm-specific RNA transferase [42,63] that interacts with DCR-1 [37] and is required only for ERGO-1 class 26G RNA and dependent 22G RNA accumulation [42]. ERI-9 is encoded by an oogenesis-enriched transcript [33], and its loss results in the Eri phenotype associated with loss of ERGO-1 class 26G RNAs, but neither the Him phenotype nor the sperm-origin ts sterility associated with loss of ALG-3/4 class 26G RNAs [42].

ERI-6/7: ERI-6/7 is a helicase protein required only for ERGO-1 class 26G RNA and dependent 22G RNA accumulation [35,64]. The ERI-6/7 protein is encoded by antiparallel *eri-6* and *eri-7* pre-mRNAs that are trans-spliced to generate a fusion mRNA; these two genes constitute a single, contiguous gene in *C. briggsae* and the *C. elegans* CB4856 isolate [64]. Although *eri-7* is classified as oogenesis-enriched [33], transcriptional fusion reporters indicate that these genes are also somatically expressed [64]. ERI-6/7 is predominantly cytoplasmically localized, suggesting RNA helicase function [64]. Loss of ERI-6/7 results in an Eri phenotype [64], but only moderately decreased fertility [35]. ALG-3/4 class 26G RNAs and all other major classes of *C. elegans* small RNAs are not decreased by loss of ERI-6/7 [35].

MUT-16: MUT-16 is a worm-specific protein with proline-rich and glutamine/asparagine-rich regions [65] that functions critically in Mutator foci in the WAGO 22G RNA pathway [66] and will therefore be further discussed below. Loss of MUT-16 results in severe depletion of ERGO-1 class 26G RNAs and dependent 22G RNAs, but not ALG-3/4 class 26G RNAs [62]. Some *C. elegans* laboratory strains, such as the *dcr-1(mg375)* helicase mutant, contain a *mut-16(mg461)* mutation; this allele fails to express somatic MUT-16 fully, potentially compromising somatic 26G RNA production and complicating interpretation of experimental results pertaining to the ERGO-1 class 26G RNA pathway or somatic WAGO pathway-dependent gene silencing [62].

Other Mutator (MUT) proteins: ERGO-1 class 26G RNAs also require MUT-2, MUT-7, and MUT-15 and show minor dependence on MUT-8 and MUT-14 [62]. These proteins are recruited by MUT-16 to form Mutator foci in the germline for WAGO 22G RNA amplification and target silencing (see below) [66]. ALG-3/4 class 26G RNAs are modestly decreased by loss of MUT-7, but appear intact upon loss of the other Mutator foci MUT proteins [62]. The dependence of ERGO-1 class 26G RNAs on this set of factors suggests that Mutator foci may also serve as amplification centers for primary endo-siRNA biogenesis.

26G RNA Argonaute loading

Consistent with the embryo extract studies showing DCR-1-mediated cleavage of 26 nt/23 nt duplexes from blunt-ended transcripts [53], deep sequencing libraries show evidence of shorter passenger strands antisense to 26G RNAs. Most commonly, the 5' nt of the sense read maps to nt 23 of the 26G RNA, corresponding to a 3 nt 3' overhang for the 26G RNA [18,35,67]. While these sense fragments may be temporarily stabilized in the dsRNA duplex, the 3' ends of sense reads are highly variable, suggesting 3' to 5' degradation [35,67].

The ERGO-1 and ALG-3/4 Argonautes are required for accumulation of their respective 26G RNA populations [19,20,23,31]. While many of the Argonautes encoded by *C. elegans* do not show conservation of the three critical catalytic residues mediating slicer activity, all of the Argonautes that bind primary small RNAs (ERGO-1, ALG-3/4, PRG-1/2, RDE-1, ALG-1/2, and, putatively, CSR-1) show intact catalytic triads in their RNase H-related PIWI domains [68]. Although the slicer activity of an Argonaute refers generally to the ability to catalyze target cleavage, it appears also to play a critical role in effector complex maturation through passenger strand removal. The catalytic activity of RDE-1, the Argonaute that binds primary exo-siRNAs, is required only for efficient removal of the passenger strand; thus, the catalytic mutant of RDE-1 shows only a partial Rde phenotype attributable to impaired target mRNA interaction [69]. A putative

ERGO-1 catalytic mutant exhibits an Eri phenotype [35]. This indicates that catalytic activity is required for ERGO-1-mediated triggering of WAGO 26G RNAs (see below), but 26G RNA and passenger strand levels have not been assessed. The *ergo-1* mutant shows vastly depleted ERGO-1 class 26G RNA levels but twofold increased passenger strand levels [35], supporting a role for ERGO-1 slicer activity in liberating the passenger strand from a 26G RNA dsRNA intermediate. In comparison, both mature ERGO-1 class 26G RNAs and passenger strands are depleted by loss of ERI-1 or ERI-6/7, suggesting function upstream of 26G RNA duplex formation [35]. Similar analyses of catalytically inactive ALG-3 or ALG-4 have not been reported.

ALG-3/4: The highly homologous Argonautes ALG-3 and ALG-4 redundantly bind and stabilize 26G RNAs generated in spermatogenic germline and mediate their effector functions [19,20]. The *alg-3* transcript is classified as spermatogenesis-enriched [33], and *alg-3* mRNA and protein are enriched in male worms and depleted in female worms [19,20]. In the spermatogenic germline, ALG-3 expression begins in postpachytene spermatocytes, showing cytoplasmic localization with enrichment in P granules [19]. After spermatogenesis, ALG-3 is detected only in the spermatheca, where it is confined to residual bodies after mature, postmeiotic spermatids have budded off [19]. The fertility of the *alg-3* or *alg-4* single mutants does not differ significantly from that of wild type [19,20], reflecting their redundancy, but loss of both ALG-3 and ALG-4 impairs fertility at 20°C and results in the characteristic sperm-origin ts sterility at 25°C [19,20]. The RNAi sensitivity of the single and double mutants is wild type [20].

ERGO-1: The Argonaute ERGO-1 binds and stabilizes 26G RNAs generated in oogenic germline, embryo, and likely beyond and mediates their effector function [20,23,31,42]. The *ergo-1* transcript is classified as oogenesis-enriched [33], and ERGO-1 is nearly absent from L3 and L4 larva and young adult [23], paralleling decreased detection of ERGO-1 class 26G RNAs during these stages [20,70]. In the hermaphrodite oogenic germline, ERGO-1 expression begins at pachytene exit and persists into embryo, showing cytoplasmic localization throughout [70]. The *ergo-1* mutant exhibits the

characteristic Eri phenotype associated with loss of ERGO-1 class 26G RNAs, but only a minor fertility defect and no Him phenotype [20,42,68].

HENN-1: HENN-1 is the *C. elegans* ortholog of HEN1 methyltransferase [71]. HENN-1 catalyzes 2'-O-methylation of the 3' terminus of small RNAs associated with Argonautes of the PIWI clade [70,72], namely, ERGO-1 class 26G RNAs and 21U RNAs [18,70,72,73]. Direct interaction between HENN-1 and either ERGO-1 or PRG-1 has not been demonstrated [70,73], and recombinant HENN-1 is capable of methylating RNA oligomers in the presence of S-adenosyl methionine [73]. Nevertheless, HENN-1 appears to methylate small RNAs only following Argonaute binding, as loss of ERGO-1 results in loss of methylation for the rare residual ERGO-1 class 26G RNAs [70]. The *henn-1* transcript is classified as germline-intrinsic [33], and mRNA and protein are detected at all stages in both germline and soma with strongest expression in germline and embryo [70,73]. HENN-1 is detected throughout male and female germlines, with proximal oocytes showing cytoplasmic and intense nucleoplasmic signal; the nucleoplasmic enrichment is lost upon fertilization [70]. During sperm maturation, HENN-1 becomes enriched in residual bodies, suggesting possible exclusion from mature spermatids [70]. In embryo, HENN-1, like ERGO-1, is abundant and diffusely cytoplasmic [70,73]. HENN-1-mediated methylation is critical for ERGO-1 class 26G RNA stability and inheritance into embryo; in the absence of HENN-1, ERGO-1 class 26G RNAs show size heterogeneity [70,73] and increased levels of non-templated nucleotide additions [72]. However, this trimming and tailing activity is not limited to unmethylated 26G RNAs. Analysis of trimming and tailing rates by small RNA class reveals that the methylated 21U RNAs show the lowest frequency and ALG-3/4 class 26G RNAs the highest, but ERGO-1 class 26G RNAs show a trimming and tailing rate nearly as high as that of ALG-3/4 class 26G RNAs [72]. 21U RNAs show significantly decreased perdurance in the absence of HENN-1, but their initial accumulation is less severely affected than that of ERGO-1 class 26G RNAs [70,72,73]. The relevance of HENN-1 to 21U RNA accumulation and stability are discussed further below. It is unclear how HENN-1 affects 22G RNA levels. Global levels of both WAGO and CSR 22G RNAs are decreased in the absence of HENN-1 by ~30% without major changes in

size or frequency of addition of non-templated nucleotides [73], contributing to decreased detection of 22G RNAs dependent upon methylated primary siRNAs [70,72]. The effects on target mRNAs are difficult to interpret, as global mRNA analysis shows general downregulation of germline-expressed genes upon loss of HENN-1; however, ERGO-1 class 26G RNA target mRNAs are not downregulated, hinting at impaired target silencing [73], and another report identifies significant upregulation of several ERGO-1 class 26G RNA targets [72]. It is possible that the general depletion of germline mRNAs causes decreased 22G RNA levels due to decreased template-dependent synthesis, but another explanation may be that mRNA levels of critical 22G RNA pathway factors such as PPW-2 and MUT-7 are decreased by loss of HENN-1 [73]. This may also explain a curious phenotype of *henn-1* mutants: while the soma shows an Eri phenotype, presumably due to somatic ERGO-1 class 26G RNA depletion, the germline is Rde [70,73]. Loss of HENN-1 also results in slightly decreased fertility at 25°C and a mild Him phenotype [70,73].

26G RNA effector function

Loss of factors involved in ERGO-1 and/or ALG-3/4 class 26G RNA biogenesis or accumulation results in desilencing of the relevant population of complementary target mRNAs [19,20,23,31,32,37,41,42]. Factors that play lesser roles in accumulation such as ERI-5 or Mutator foci MUT proteins show less pronounced silencing defects [37,62]. While several factors have been identified that appear to contribute exclusively to ERGO-1 class 26G RNA accumulation [35,42,70,72-74], ALG-3 and ALG-4 remain the only factors known to be specifically required for accumulation of ALG-3/4 class 26G RNAs [19,20]. 26G RNAs repress target mRNA expression in their cognate cell types; loss of 26G RNAs does not result in ectopic expression within other cell types of the native male or female germline or inappropriate target expression in the opposite germline [20].

During 26G RNA biogenesis, DCR-1 likely catalyzes the cleavage of the mRNA template in generating the dsRNA intermediate [53]. 26G RNA Argonaute recognition may also trigger target cleavage. However, neither is likely to represent a significant means of target silencing, as WAGO 22G RNAs triggered by 26G RNA targeting are the major effectors of the 26G RNA endo-RNAi pathway. These secondary siRNAs, discussed in depth below, are strictly required for 26G RNA target silencing [19,23,31], as well as silencing of other small RNA pathway targets (see WAGO 22G RNAs below). It is not known precisely how 26G RNA RISC association triggers 22G formation, but target transcripts show 26G RNA-dependent accumulation of WAGO 22G RNAs [19,23,31], and many 22G and 26G RNAs originate from the same 5' nt [31].

22G RNAs amplify the silencing signal of 26G RNAs and increase the perdurance of the repression. ERGO-1 class 26G RNA levels peak in embryo and decline significantly throughout larval development [20,23,31]; however, ERGO-1 class 26G RNAs primarily target transcripts that are not germline-intrinsic [20,33], and triggering the production of WAGO 22G RNAs enables their silencing influence to persist through larval development [20,23,31]. Similarly, 26G RNAs and ALG-3/4 are depleted in mature sperm, but WAGO 22G RNA Argonaute WAGO-1 is abundant in mature sperm [19]. This precludes significant inheritance of paternal 26G RNAs, but evidently not dependent 22G RNAs: an ALG-3/4 class 26G RNA-dependent 22G RNA can be detected even at the L1 larval stage [20].

Analysis of 26G RNA-dependent 22G RNAs further suggests that 26G RNAs target transcripts in trans as well as in cis. The X-cluster [28] describes a region of the X chromosome from which abundant 22Gs, but none or very few 26Gs, are generated [23]. Accumulation of X-cluster 22G RNAs requires ERGO-1 [68] and a particular ERGO-1 class 26G RNA derived from the *K02E2.11* transcript, which shares no other significant homology with the X-cluster [72]. This particular 26G RNA shows multiple putative target sites throughout the X-cluster. None is perfectly complementary, and the best sites contain a minimum of three adjacent, central mismatches and two wobble

pairs [72]. This indicates that 26G RNA targeting of mRNA targets to initiate 22G RNA production in trans is mismatch-tolerant, suggesting immense targeting capacity.

Sperm-origin ts sterility due to loss of ALG-3/4 class 26G RNAs

As indicated above, compromise of ALG-3/4 class 26G RNA pathway activity results in sterility at 25°C due to defective spermatogenesis and spermiogenesis [19,20,42]. The Him phenotype also results from X chromosome missegregation during compromised spermatogenesis. *eri-1*, *eri-3*, and *rrf-3* mutants do not show maternal rescue of this fertility defect, unlike the Eri phenotype [74]. 26G RNA and dependent 22G RNA production is lost in *eri-1*, *eri-3*, *eri-5*, and *rrf-3* mutants at both elevated and permissive temperatures [37], indicating that it is not siRNA production that is sensitive to temperature but rather that these siRNAs are only required for functional spermatogenesis at elevated temperatures. The temperature-sensitive period for fertility in the absence of ALG-3/4 class 26G RNAs coincides with spermatogenesis at the L4 larval stage [19,32,42].

The defects occurring during spermatogenesis at 25°C in the absence of 26G RNAs have been detailed in *rrf-3* and *dcr-1(mg375)* helicase mutants as well as in the *alg-3*; *alg-4* double mutant, which lacks only ALG-3/4 class 26G RNAs [19,32,42]. The phenotypes are highly similar, suggesting that compromise of only the male germline-expressed class of 26G RNAs is sufficient to cause these defects; consistent with this, loss of ERGO-1 ERI-9, or ERI-6/7 results on only very mild decreases in fertility at 25°C [20,35].

The spermatogenesis defects associated with compromised ALG-3/4 class 26G RNA target silencing are as follows [19,32,42]: Mutant gonads show delayed and decreased production of sperm-like nuclei. Early spermatogenesis appears grossly wild type, with normal germ cell number and morphology in mitotic and meiotic regions. Defects are first evident in primary spermatocytes, which may show nuclear abnormalities such as

chromosome bridges. Some late-stage spermatocytes also show chromatin bridges and arrest as multinucleate masses. Cytokinesis, if completed, produces spermatids with large, misshapen, or multiple nuclei. These defects may arise from delayed progression through spermatogenesis, as older males generate more normal spermatids. Spermatocytes that develop without obvious morphological abnormalities produce smaller numbers of residual bodies and sperm. Mutant sperm show abnormal wreaths of microtubules surrounding nuclei. In vitro, activation of these mutant sperm is impaired; many fail to form pseudopods and instead show long spike structures and impaired motility. In vivo, mutant sperm fail to localize to the spermatheca. Although transfer by mating is successful, mutant sperm are rapidly expelled through the vulva.

Precisely how ALG-3/4 class 26G RNAs promote thermotolerant fertility is unknown. It is unlikely that desilencing of a single target results in these sperm development and motility defects, as individual depletion of 68 spermatogenesis-enriched transcripts desilenced in *eri-1* or *rrf-3* mutants failed to suppress the ts sterility and EMS mutagenesis failed to identify any suppressors among several million genomes screened [42]. Many of these defects are observed with variable penetrance at 20°C, but nearly all sperm show the abnormal spike structures during activation at 25°C, correlating with the penetrant ts sterility. Many ALG-3/4 class 26G RNAs target transcripts encoding major sperm proteins [19,20,34,42], which assemble into filamentous fibers in the pseudopod. One function of 26G RNAs may be to promote male fertility by limiting excess accumulation of major sperm proteins. Additionally, *rrf-3* and *eri-1* mutant hermaphrodites cultured at 23°C generate embryos with significant spindle structure abnormalities including tripolar spindles or male pronuclei with supernumerary microtubule asters [32]. These defects, as well as the abnormal microtubule wreaths in spermatids [32] and the chromosome segregation defects in spermatocytes [19,32,42] suggest a role for ALG-3/4 class 26G RNAs in regulating microtubule organization during spermatogenesis. Alternatively, these microtubule defects may reflect abnormalities in chromosome complements resulting from dysfunctional spermatogenesis.

Enhanced RNAi sensitivity due to loss of ERGO-1 class 26G RNAs

In *C. elegans*, exo-RNAi and endo-RNAi engage a common downstream pathway to effect target silencing: the WAGO 22G RNAs. The earliest evidence for intersection between these two pathways was the identification of mutants that show both Rde and Mutator (Mut) phenotypes [57,75], indicating that common mechanisms mediate exo-RNAi and germline transposon silencing. Discovery of distinct primary and secondary phases of exo-RNAi first prompted the hypothesis that loss of RRF-3 might release limiting, common cofactors for function in the amplification phase of RNAi [46]. The WAGO Argonautes appear to represent these limiting factors in RNAi: loss of one can impair sensitivity to exogenous RNAi [68,76], and overexpression of one enhances accumulation of 22G RNAs and RNAi sensitivity [68]. Accordingly, loss of ERGO-1 class 26G RNAs results in an Eri phenotype due to decreased competition for WAGO Argonaute occupancy by secondary siRNAs [24,37,41,68], in line with an earlier observation that loss of ERI-1 increases accumulation of exo-RNAi-triggered siRNAs [54]. The converse is also true. In the absence of exogenous dsRNA, RDE-1 scavenges diverse small RNAs and triggers endogenous somatic 22G RNA production [24,61,77]; loss of RDE-1 and these dependent endo-siRNAs enhances accumulation of an ERGO-1-dependent somatic 22G RNA [68].

Loss of ERGO-1 class 26G RNAs, but not ALG-3/4 class 26G RNAs, results in the Eri phenotype of ERI complex mutants [20]. Only ERGO-1 class 26G RNAs are inherited by offspring to generate abundant 22G RNAs and effect target silencing [19,23,31]. Eri mutants show strong maternal, but not paternal, rescue of Eriness in the soma [74], consistent with the Eri phenotype arising due to liberation of WAGO Argonautes in the absence of maternally inherited ERGO-1 class 26G RNAs and dependent 22G RNAs. It is possible that upregulation of target mRNAs encoding factors such as helicases and dsRNA-binding proteins may also contribute to the Eri phenotype associated with loss of ERGO-1 class 26G RNAs [35].

22G RNAs comprise two major small RNA pathways

Early capture of *C. elegans* small RNAs using a protocol insensitive to 5' structure indicated that a majority of antisense small RNAs are 22 nt in length [28]. Protocols selective for 5' monophosphorylated species failed to recover these sequences effectively [18,78], revealing a 5' structure different from that of the 26G RNAs. Subsets of these small RNAs were subsequently shown to exhibit a 5' triphosphate [21] and a 3' hydroxyl [21,22,24]. Finally, 22 nt siRNAs in *C. elegans* show a prominent 5' guanosine bias [18,24,28], distinguishing these as 22G RNAs.

As a whole, 22G RNAs target about ~50% of the annotated coding genome, with most 22G RNAs targeting unique genome sequences [24]. 22G RNAs mapping antisense to mRNAs show enrichment primarily at transcript 3' ends, consistent with RdRP engagement at the mRNA 3' terminus, but some also show enrichment at 5' ends [24]. Most 22G RNAs are germline expressed and deposited into embryo, potentially coinherited with their mRNA targets [24]. The 22G RNAs comprise two distinct classes of small RNAs that are synthesized by similar RdRP modules but engage unique pathways mediated by nonoverlapping Argonautes and cofactor proteins to effect entirely distinct outcomes. The WAGO 22G RNAs bind semiredundant Argonautes of the worm-specific WAGO clade to mediate silencing of certain protein-coding genes, transposons, pseudogenes, and cryptic loci through both transcriptional and post-transcriptional mechanisms [24,79-86]. The CSR 22G RNAs bind CSR-1, another WAGO Argonaute, and target germline-expressed genes to fulfill an essential role in promoting chromosome segregation [25]. Thus, 22G RNAs target both silent and expressed loci genome-wide to maintain the germline.

The 22G RNA RdRP module: RdRP (RRF-1/EGO-1), DRH-3, and EKL-1

RdRP: RdRPs RRF-1 and EGO-1 both contribute to biogenesis of 22G RNAs. The *ego-1; rrf-1* double mutant lacks 22G RNAs synthesized de novo in the germline [24]. EGO-

1 and RRF-1 proteins show greater than 50% sequence identity [40] and collaborate to generate WAGO 22G RNAs, but EGO-1 alone is required for CSR 22G RNAs [24,25]. Accordingly, the *ego-1* transcript is classified as germline-intrinsic [33], and *ego-1* mRNA and protein are enriched in young adult, adult, and embryo but nearly absent during larval development [25,40,87]. EGO-1 protein is also enriched in the nuclear fraction of cell lysates [88]. In contrast, a transgene expressing GFP::RRF-1 that rescues the *rrf-1* mutant phenotype shows robust somatic expression, with prominent cytoplasmic and weak nuclear localization [44]. Despite their different expression patterns, the *rrf-1* gene is encoded directly downstream of *ego-1* in an operon [89].

WAGO RdRP: Because exo-RNAi engages WAGO 22G RNAs to mediate transcript knockdown [21,22,46], the RNAi sensitivity phenotypes associated with loss of RRF-1 versus EGO-1 reveal their respective contributions to the WAGO pathway. RRF-1 is required in somatic tissues for 22G RNA accumulation and therefore exo-RNAi [23,31,46], whereas no role for EGO-1 is detected during RNAi in the soma [40,46]. Accordingly, 22G RNAs mapping to ERGO-1 class 26G RNA targets are largely RRF-1-dependent [23]. In the germline, however, EGO-1 and RRF-1 are partially redundant for WAGO 22G RNA biogenesis. The *rrf-1* mutant germline is sensitive to exo-RNAi of germline transcripts [46], revealing compensation by EGO-1. This explains why loss of RRF-1 does not recapitulate the ts sterility and other germline phenotypes associated with total compromise of WAGO 22G RNA target silencing. Loss of EGO-1 results in a partial germline Rde phenotype [40]. Interestingly, injection of dsRNA enhances the germline defects of an *ego-1* null mutant [40], suggesting the possibility that RRF-1 may be recruited, although insufficiently, to CSR 22G RNA biogenesis in the absence of EGO-1. EGO-1 and phenotypes associated with its loss are discussed further with CSR 22G RNAs below.

RRF-1: In a cell-free system for analyzing secondary siRNA production triggered by exo-RNAi, RRF-1 accounts for 90% of RdRP activity [44]. Secondary siRNAs generated by immunopurified GFP::RRF-1 complexes are complementary to a supplied RNA template, and production is inhibited by addition of RNA chain elongation terminator.

RRF-1 also exhibits terminal transferase activity in cell extracts. 93% of RdRP products generated by GFP::RRF-1 complexes show a 5' guanosine. This indicates that the 5' guanosine bias of 22G RNAs, and perhaps also 26G RNAs, reflects an RdRP preference for initiation of transcription with GTP. GFP::RRF-1 complexes show very low processivity when incubated with ribonucleotides and mRNA template, generating products 22-23 nt in length; this length is not imposed by DCR-1-mediated cleavage, as DCR-1 is not recovered in GFP::RRF-1 complexes. DRH-3 immunopurifies with GFP::RRF-1 and is required for RdRP activity. GFP::RRF-1 complexes fail to generate RdRP products from long dsRNA template, but polyadenylated and non-polyadenylated mRNAs are equally effective templates *in vitro*; in cell extract, the non-polyadenylated template stimulates more robust RdRP activity. Neither in cell extract nor *in vitro* do GFP::RRF-1 complexes catalyze extension of an RNA primer complementary to a template, arguing against primer-dependent synthesis. In contrast, GFP::RRF-1 complexes successfully incorporate labeled GTP as the 5' nucleotide of RdRP products *in vitro*, demonstrating unprimed synthesis [44]. The 5' ends of these RdRP products are also sensitive to capping, suggesting these products show the 5' triphosphate of 22G RNAs generated *in vivo* [21]. Unprimed synthesis also occurs *in vivo*, as secondary siRNAs triggered by a mismatch-containing primary siRNA show perfect complementarity to the endogenous target [22]. Furthermore, the rare secondary siRNAs that are generated 3' to the trigger sequence could not originate via extension of *exo*-siRNA-derived primers [21].

DRH-3: As mentioned above, DRH-3 is an essential component of all known *C. elegans* RdRP modules [24,25,37,44,48]. DRH-3 is detected at all developmental stages and is not restricted to germline [25], but the *drh-3* transcript is classified as germline-intrinsic [33], and *drh-3* mRNA levels are threefold higher in young adult and adult than in larva [48]. The most severe phenotypes associated with loss of DRH-3 result from compromise of the CSR 22G RNA pathway [25] and will be discussed below. Nevertheless, partial loss of DRH-3 function results in phenotypes and target upregulation attributable to loss of WAGO 22G RNAs: point mutations in the helicase domain yield viable *drh-3* hypomorphic mutants that show variable germline and

somatic Rde, Mut, ts sterile, and Him phenotypes as well as embryonic lethality possibly due to reduced CSR 22G RNA accumulation [24]. A sterile *drh-3* null mutant is insensitive to RNAi in germline but not soma [37], presumably due to rescue by maternal inheritance. Deep sequencing of this mutant using a 5' monophosphate-independent protocol shows dramatically reduced endo-siRNA reads mapping to protein-coding genes, pseudogenes, repetitive elements, and unannotated loci but intact microRNAs and 21U RNAs [24]. Residual 22G RNAs detected in the *drh-3* null mutant map primarily to the 3' ends of transcripts, suggesting that the RdRP module may be loaded at transcript 3' ends, with DRH-3 possibly acting to promote sequential initiation of polymerization [24]. Proposed roles for DRH-3 within the RdRP module include relaxing template secondary structure and promoting 22G RNA dissociation from the transcript for Argonaute loading [24].

EKL-1: EKL-1 is a Tudor domain protein [90] paralogous to ERI-5. Like ERI-5, EKL-1 interacts with DRH-3, but only ERI-5 interacts DCR-1 [38]. This is consistent with 22G RNA biogenesis occurring through a DCR-1-independent mechanism [24,42,44]. The *ekl-1* transcript is classified as germline-intrinsic [33], but EKL-1 is detected at all developmental stages and does not show germline restriction [25]. Like RRF-1/EGO-1 and DRH-3, EKL-1 is also required for 22G RNA biogenesis [24] and exhibits phenotypes characteristic of compromise of CSR and WAGO 22G RNA pathways. Loss of EKL-1 results in sterility as well as somatic and germline defects in exo-RNAi and related phenomena [25,90-92].

Triggering of WAGO 22G RNAs

WAGO 22G RNAs serve as the critical amplification pathway upon which most primary small RNAs in *C. elegans* converge. Accordingly, WAGO 22G RNAs are generated from and target transcripts targeted by 26G RNAs, primary exo-siRNAs, RDE-1-scavenged small RNAs, and 21U RNAs. The requirements for triggering of WAGO 22G RNAs by each of these primary small RNA types are addressed below.

Secondary siRNAs in *C. elegans* were first identified during exo-RNAi as a population of small RNAs that are antisense to a target mRNA but not derived from the initial dsRNA trigger [46]. However, the existence of some means of amplification had been hypothesized several years prior. It had been noted that a few molecules of dsRNA are sufficient to deplete a cellular pool of target mRNAs, arguing against simple stoichiometric interaction and indicating the involvement of a catalytic and/or amplification mechanism [93,94]. Although trigger dsRNA cleavage generates some measure of amplification, this was recognized as still insufficient to explain the potent silencing ability of exogenous dsRNA [46]. The essential contribution of RdRP activity to this amplification was subsequently established by demonstrating complete insensitivity of the *rrf-1* mutant soma to injection of prepared siRNA duplexes [46].

WAGO 22G RNAs are generated from spliced transcripts targeted by the triggering primary small RNAs [21]. RdRP initiation appears to be non-random, however, as 22G RNAs show some phasing across a target transcript, and similar sets of 22G RNAs are produced in different transgenic lines [22]. 22G RNA biogenesis shows a limited degree of spreading from the target site of the primary small RNA [46]. During exo-RNAi, spreading occurs primarily 5' of the primary RNA target site with respect to the sense of the target transcript [21,22,46]. ERGO-1 class 26G RNAs rarely target the first ~100 nt of a transcript, but 22G RNAs show robust coverage of this interval at these same targets [23]. Similarly, 21U RNAs trigger 22G RNA production primarily 5' of the target site [95]. This may have implications for transgenic construct design, as transgenes encoding foreign sequence 3' to endogenous sequence are likelier to evade silencing by 21U RNAs [86]; presumably, 5' spread of 22G RNA biogenesis from foreign to endogenous sequence could be detrimental, resulting in selection for evaders. The range of the spreading is relatively modest and may vary by primary small RNA abundance. Exo-RNAi appears to generate secondary siRNAs with a range of 100-180 nt 5' to the region targeted by dsRNA [96]. 21U RNAs trigger production of 22G RNAs within the surrounding 40-100 nt [95,97].

Triggering by 26G RNAs: As described above, 26G RNAs trigger robust production of WAGO 22G RNAs at target sites, and these secondary siRNAs are essential for target silencing [19,23,31]. However, loss of 26G RNAs does not significantly deplete 22G RNA levels [23,24,31,32], indicating the existence of other primary triggers of WAGO 22G RNA biogenesis.

Triggering by primary exo-siRNAs: Competition between endo- and exo-RNAi implicates primary exo-siRNAs and the associated Argonaute RDE-1 in triggering 22G RNAs as well [37,41,46,68]. Exo-RNAi triggers production of antisense secondary siRNAs that share all of the features and dependencies of endogenously triggered WAGO 22G RNAs [21,22,24,44,46,98]. Expression of a single hairpin-derived 22 nt siRNA (22siR) targeting the *unc-22* mRNA 3' UTR reveals some of the requirements for RDE-1-triggered 22G RNA biogenesis [22]: A 22siR showing perfect target site complementarity depletes *unc-22* mRNA levels by 50% and generates abundant secondary siRNAs upstream of the target site. A mutated 22siR bearing target site mismatches at positions 10-12 binds the *unc-22* mRNA but does not trigger secondary siRNA biogenesis. Single mismatches at position 11 or position 21 weaken target silencing, but do not alter the range of secondary siRNAs triggered. The lack of an effect of mismatch at the putative cleavage site (position 11) on the distribution of secondary siRNAs suggests that RDE-1-mediated target cleavage is not necessary for triggering secondary siRNA production [22], consistent with a restricted role RDE-1 cleavage activity in passenger strand removal [69].

Triggering by RDE-1-scavenged small RNAs: In the absence of exogenous dsRNA, RDE-1 binds a variety of DCR-1 products; among these are microRNAs and dsRNA-derived siRNAs cleaved from endogenous hairpins or bidirectionally transcribed genomic regions [77]. Some of these scavenged siRNAs initiate significant 22G RNA production at target loci: the *Y47H10A.5* transcript generates abundant 22G RNAs triggered by miR-243-loaded RDE-1, for which a perfectly complementary target site 22 nt in length appears on *Y47H10A.5* transcript [24,61,77]. Like primary exo-siRNAs, miR-243-mediated triggering of these 22G RNAs requires RDE-4, in addition to RDE-1 [24].

In contrast, the chief microRNA pathway Argonaute ALG-1 does not trigger secondary siRNA biogenesis at target sites [99]; this may explain why microRNAs fail to elicit dramatic target knockdown despite the vast abundances of individual microRNAs relative to individual endo-siRNAs. It is important to note that 22G RNAs themselves are rarely bound by RDE-1 [46], prohibiting triggering of further siRNA biogenesis from aberrant tertiary targets in trans. Similarly, 22G RNAs bound by WAGO Argonautes also do not appear to trigger further amplification or spreading at target sites: a sensor transgene with a single target site for an abundant 22G RNA shows confined loss of secondary siRNAs at the target site when the triggering 22G RNA is removed [72].

Triggering by 21U RNAs: A fourth population of primary small RNAs converging on the WAGO 22G RNAs are 21U RNAs, the piRNAs of *C. elegans*. These small RNAs are bound by and require the PIWI clade Argonaute PRG-1 for accumulation and targeting [26,27]. 21U RNAs constitute a library of mismatch-tolerant sequences that are depleted for targeting of protein-coding genes but capable of silencing non-self transcripts by triggering WAGO 22G RNAs [79,86,95,97]. 21U RNA properties, biogenesis, and function are discussed in depth below. Initial studies identified 21U RNA-dependent secondary siRNAs [26,27], but only recently were these established as WAGO 22G RNAs and the targeting requirements reported. 21U RNAs trigger 22G RNA biogenesis at engineered sensor targets with up to two target site mismatches and at endogenous targets with up to three or four mismatches [95,97]. Seed site pairing, which plays an important role in microRNA-mediated target repression, may also be important for 21U RNA targeting. One study found that nearly perfect pairing is required at positions 2-8, with a maximum of one G-U wobble pair across the interval, to trigger 22G RNA biogenesis [95]; however, another study found that transgenic 21U RNA sensor transcripts are effectively targeted regardless of the position of dinucleotide mismatches across the length of the 21U RNA [97]. Interestingly, 21U RNAs seem less effective at stimulating a secondary siRNA response, as less than 5% of target sites exhibiting perfect complementarity to 21U RNAs show unambiguous triggering of 22G RNAs [95,97]. This may be in part due to the low abundances of many 21U RNA species, as levels of 22G RNAs triggered correlate with both 21U RNA abundance and degree of

complementarity [95,97]. Recombinant PRG-1 shows some slicing activity in vitro, but this activity is dispensable for triggering 22G RNA production as well as target silencing [95,97]. Also in support of a cleavage-independent mechanism, 21U RNA target site mismatches at positions 10 and 11 do not abrogate 22G RNA triggering [95,97].

Germline amplification of WAGO 22G RNAs in perinuclear Mutator foci

Screens for factors involved in transposon silencing in the germline (mutators) [65,75,100], exo-RNAi [57], and related phenomena [90,91,101] recurrently identified members of a group of proteins implicated in formation of Mutator foci. These perinuclear processing compartments interact with RRF-1 and are required for WAGO 22G RNA amplification and target silencing in the germline [66]. The six proteins that assemble to form Mutator foci are MUT-16, MUT-7, MUT-8/RDE-2, MUT-2/RDE-3, MUT-15, and MUT-14. Consistent with their critical, non-redundant role in the WAGO 22G RNA pathway, loss of any of these proteins results in mutator (*Mut*), *ts* sterile, *Him*, and *Rde* phenotypes [24,65,66,75,98,100,102,103], although a *mut-16* mutant carrying a silent missense mutation shows only a germline *Rde* phenotype [66,98].

MUT-16 is a worm-specific protein containing proline-rich and glutamine/asparagine-rich regions [65]. A translational MUT-16::GFP fusion protein is expressed broadly in cytoplasm and nuclei [65]. MUT-16 is also required for accumulation of ERGO-1 class 26G RNAs, whereas other Mutator foci components are not strictly required [62]. MUT-7 is a 3' to 5' exonuclease [75] that is conserved in animals [66]. A *mut-7::gfp* transcriptional fusion shows expression in many different cell types in larva and adult [43]. MUT-8/RDE-2 contains no known domains and may only exist to recruit MUT-7 [66]. MUT-2/RDE-3 is a beta-nucleotidyltransferase; mutations of conserved residues at the active site produce *Rde* phenotypes, suggesting polymerase activity is required for function [102]. MUT-2/RDE-3 expressed in frog oocytes shows no polymerase activity, unlike homolog CDE-1 [104]. The *mut-2/rde-3* transcript is classified as germline-

intrinsic [33]. MUT-15 contains no known domains. MUT-14 is a DEAD box RNA helicase [98].

Dynamics of Mutator foci: Fluorescently tagged translational fusions of these proteins reveal the dynamics of Mutator foci throughout development [66]: In male and hermaphrodite germlines, Mutator foci are perinuclear, adjacent to P granules, and most prominent in the mitotic region and transition zone. By meiotic diakinesis, they begin to show cytoplasmic localization. In somatic cells, Mutator foci components are cytoplasmically localized but still required for efficient WAGO 22G RNA amplification and silencing function, as most of the component proteins are required for exo-RNAi in both germline and soma [24,65,66,75,91,102,103]. Possibly, lower somatic expression of these proteins may prevent focal recruitment to the nuclear periphery. Failure to assemble Mutator foci in somatic tissues likely decreases the effectiveness of the components in promoting genome surveillance, as transposons efficiently silenced in germline can be active in the soma [105]. Maternal presence of Mutator foci appears to be critical for transposon suppression in offspring: *mut-7* heterozygous mutants show a Mut phenotype with maternal, but not paternal, inheritance of the mutation, and paternally inherited *mut-7* mutant alleles require several generations of homozygosity to reach maximal transposition activation, whereas maternal *mut-7* alleles show maximal activation in the first homozygous generation [75].

Assembly of Mutator foci: Mutator foci may associate directly with the nuclear pore complex [66], much like P granules [106]. However, Mutator foci are not required for P granule formation, nor are P granules required for Mutator focus assembly [66]. Consistent with the many associations of the nuclear pore with RNA processing bodies, nuclear pore complex proteins routinely emerge from screens for factors involved in RNAi [65,91]. Mutator focus assembly is a regulated process: MUT-16, which contains a Q/N-rich domain that may mediate protein-protein interactions, recruits MUT-2/RDE-3, MUT-15, MUT-14, and MUT-8/RDE-2, which itself recruits MUT-7 [66,103]. In the absence of MUT-16, the other five components do not interact [66]. Intriguingly, formation of Mutator foci appears not to require the presence of WAGO 22G RNAs, as

loss of DRH-3, EKL-1, or EGO-1 and RRF-1 simultaneously or alone does not affect their formation, nor does loss of 26G RNAs, CSR 22G RNAs, or several individual WAGO Argonautes [66].

Function of Mutator foci: Mutator foci serve as amplification centers for WAGO 22G RNAs in the germline [66]. *mut-15*, *mut-16*, and *mut-2/rde-3* mutants show defects in secondary siRNA accumulation in response to exo-RNAi [22], and deep sequencing of *mut-2*, *mut-7*, and *mut-16* mutants reveals general defects in accumulation of WAGO 22G RNAs [24,62]. Somatic X-cluster 22G RNAs are depleted in *mut-15*, *mut-2/rde-3*, *mut-7*, and *mut-16* mutants [62,66]. A *mut-14* mutant shows intact X-cluster 22G RNAs, but loss of germline 22G RNAs [66]. Therefore, the germline and somatic Rde phenotypes of Mutator foci mutants reflect their respective capacities to support 22G RNA biogenesis. In Mutator foci mutants, however, loss of WAGO 22G RNAs is incomplete, and depletion does not appear proportionate for all species. For example, the targets showing greatest 22G RNA depletion upon loss of MUT-16, MUT-2, or MUT-7 are those that generate very high levels of 22G RNAs [24,62,66]. Furthermore, MUT-2/RDE-3 is required for full target knockdown during dsRNA feeding, but not when dsRNA is transgenically expressed [102]. These data suggest that Mutator foci are most important for heavy siRNA amplification and are less critical when primary small RNA triggers are abundant. The mechanisms by which Mutator foci promote WAGO 22G RNA amplification are not known, although recruitment of RdRP to target transcripts is likely involved. Germline RRF-1 localizes exclusively to Mutator foci, whereas EGO-1 localizes both to P granules and Mutator foci [25,66], reflecting the dual function of EGO-1 in the WAGO and CSR 22G RNA pathways [24]. The adjacency of Mutator foci and P granules suggests possible interaction between the two [66]. P granules contain ALG-3 and PRG-1 [19,26,107], two Argonautes that require WAGO 22G RNA-mediated amplification to effect target silencing [19,24,27,95,97]. When a nascent mRNA passes through the nuclear pore, the presence of the hydrophobic P granule may retard its diffusion to enhance recognition by regulatory molecules such as loaded Argonautes [108]. Once a target transcript enters a P granule and is recognized by a primary small RNA Argonaute, the adjacent Mutator focus may engage the transcript for secondary

siRNA amplification. Notably, CSR 22G RNAs do not accumulate to high levels or trigger gene silencing [24,25] and do not require Mutator foci for amplification, as a *mut-16* mutant does not show depletion of 22G RNAs targeting a validated set of CSR 22G RNA pathway targets [62].

Additional amplification of WAGO 22G RNAs by RDE-10/RDE-11

The RDE-10/RDE-11 protein complex amplifies 22G RNAs and potentiates degradation of the targets that serve as their templates [109,110]. RDE-10 and RDE-11 are worm-specific genes [110]. The *rde-10* transcript is classified as mixed oogenesis-somatic, whereas *rde-11* is germline intrinsic [33]. WormBase reports RDE-10 as containing a Maelstrom domain, which adopts an RNase H fold that may confer nuclease or RNA-binding functionality. RDE-11 contains a RING-type zinc finger domain [110]. Mutation of a key zinc-coordinating residue in RDE-11 abrogates its interaction with RDE-10 in vitro and results in failure to *rde-11* mutant phenotypes (see below) [110]. Thus, zinc binding facilitates the formation of the RDE-10/RDE-11 complex, possibly by promoting folding of RDE-11.

RDE-10/RDE-11 amplifies 22G RNAs triggered by exo-RNAi and ERGO-1: The RDE-10/RDE-11 complex appears to play a significant role in amplification of WAGO 22G RNAs triggered by exo-RNAi [109,110]. The absence of RDE-10 or RDE-11 results in a partial germline and somatic Rde phenotype, but no fertility defect or Him phenotype [109,110]. The exo-RNAi deficiency of these mutants is dosage-sensitive and results from impaired accumulation of secondary WAGO 22G RNAs: loss of RDE-10/RDE-11 depletes exo-RNAi-triggered 22G RNAs by ~80% [109,110]. ERGO-1 class 26G RNAs also variably require RDE-10/RDE-11 for maximal amplification of secondary siRNAs and target silencing. ERGO-1 is detected in immunopurified RDE-10 complexes [109,110], and embryonic 22G RNAs targeting ERGO-1 26G RNA targets are depleted by more than half upon loss of RDE-10 [109]. 22G RNAs are most abundant in the germline [24,31], and some primary small RNAs are restricted to germline tissues

[19,26,27,107]. The more pronounced effects of RDE-10/RDE-11 on *exo*-RNAi and ERGO-1-mediated silencing therefore suggest that the RDE-10/RDE-11 complex may be particularly important in somatic 22G RNA amplification. Indeed, the somatic X-cluster and *Y47H10A.5* 22G RNAs triggered by ERGO-1 and RDE-1, respectively, showed strong dependence on RDE-10 and RDE-11 for accumulation [109]. In comparison, accumulation of a germline-enriched, ERGO-1-independent 22G RNA targeting *T01A4.3* [37] does not require RDE-10/11 [109], nor apparently do many other WAGO 22G RNAs, microRNAs, 21U RNAs, 26G RNAs, primary *exo*-siRNAs, or CSR 22G RNAs [109,110]. Perhaps the RDE-10/RDE-11 complex supports WAGO 22G RNA amplification in the soma, in the absence of perinuclear Mutator foci.

RDE-10/RDE-11-mediated target degradation: In addition to amplifying 22G RNAs, the RDE-10/RDE-11 complex may directly promote target deadenylation and degradation. RDE-10 binds mRNAs targeted by *exo*-RNAi, even in the absence of RDE-11 [110]. This binding activity is lost in the absence of RDE-1, but intact in the absence of RRF-1, indicating that primary *exo*-siRNAs are guiding association of RDE-10/RDE-11 with mRNA targets independently of secondary 22G RNAs [110]. After recognition, RDE-10-bound mRNAs are deadenylated and degraded in an RDE-11-dependent manner [110]. Targets undergoing RDE-10/RDE-11-mediated degradation retain 5' caps, providing a transiently stable substrate for continued secondary siRNA production [110]. Thus, the RDE-10/RDE-11 complex may promote secondary siRNA biogenesis by binding *exo*-RNAi targets to promote RRF-1 binding as deadenylation and degradation occur [110]. The RDE-10/RDE-11-mediated degradation pathway likely acts in parallel to target silencing mediated by somatic nuclear Argonaute NRDE-3 and other WAGO Argonautes, as many NRDE-3-bound 22G RNAs are intact in the absence of RDE-10/RDE-11 [109,110]. In somatic *exo*-RNAi in particular, the RDE-10/RDE-11 complex and NRDE-3 may represent the two main pathways for target silencing: *rde-10*, *rde-11*, and *nrde-3* mutants all show partial sensitivity to *unc-22* dsRNA feeding, but loss of RDE-10 or RDE-11 in a *nrde-3* mutant background results in complete insensitivity to *unc-22* dsRNA and a defect in *unc-22* mRNA knockdown as profound as observed upon complete loss of *exo*-RNAi in the *rde-1* mutant [110].

RSD-2, RSD-6, and/or HAF-6 may potentiate RDE-10/RDE-11 function: RSD-2 and RSD-6 are worm-specific proteins that contain no annotated domains [111]. The *rsd-6* transcript is classified as germline-enriched [33]. HAF-6 is a half-molecule ATP-binding cassette transporter detected with a reticular pattern in intestinal, muscle, and germline cells, with perinuclear localization in germline [112]. Like *rde-10* and *rde-11* mutants, *rsd-2*, *rsd-6*, and *haf-6* mutants also show dosage-sensitive Rde phenotypes [111-113] due to defects in WAGO 22G RNA accumulation [109,113]. These three proteins also regulate accumulation and target silencing of a subset of WAGO 22G RNAs, many of which are ERGO-1- and RDE-10/RDE-11-dependent [109]. RSD-2, RSD-6, and HAF-6 are all required for accumulation of X-cluster 22G RNAs [109]. The shared dosage-sensitive Rde phenotype and functional overlap of RDE-10/RDE-11 and RSD-2, RSD-6, and HAF-6 suggests the possibility of a common complex. Supporting this, RSD-2 interacts with both RDE-10 and RSD-6 [109,111]. However, loss of RSD-2 also results in a Mut phenotype at 25°C, and loss of RSD-6 causes ts sterile and Him phenotypes [113]. These more severe WAGO 22G RNA pathway phenotypes indicate that RSD-2 and RSD-6 may play more critical or additional roles in WAGO 22G RNA accumulation or function. The perinuclear germline localization of HAF-6 [112] suggests that HAF-6 could also interact with Mutator foci, and there is some evidence for genetic interaction between HAF-6 and Mutator foci components: double heterozygous *mut-8/rde-2(+/-)*; *mut-7(+/-)* mutants show impaired exo-RNAi; heterozygosity for a *haf-6* mutation in either a *mut-8/rde-2(+/-)* or *mut-7(+/-)* heterozygous background also results in defects in exo-RNAi [112].

WAGO 22G RNA Argonaute binding

Early analysis of *C. elegans* Argonautes revealed significant expansion of the protein family [50]. A large subset of these cluster phylogenetically into a WAGO, or worm-specific Argonaute, clade that is roughly equally distant from the conserved PIWI and AGO clades [68]. Included in these are CSR-1, the Argonaute that mediates the CSR 22G RNA pathway [25], and related protein C04F12.1. Also variably included is the

divergent RDE-1, the first *C. elegans* Argonaute to be linked to RNAi [57]. The other 12 Argonaute proteins in this clade constitute the WAGO Argonautes of the WAGO 22G RNA pathway. These 12 Argonautes bind and stabilize WAGO 22G RNAs semiredundantly to mediate target silencing. Simultaneous loss of all 12 Argonautes in the MAGO12 Argonaute mutant results in penetrant ts sterile, Him, and Rde phenotypes and massive depletion of 22G RNAs that are not dependent on CSR-1 [24,25]. The ts sterility of the MAGO12 mutant is only partially rescuable by wild type sperm, likely because loss of 21U RNAs results in ts fertility defects in female germline as well as male [107]. These WAGO-dependent 22G RNAs are also globally depleted in *mut-7* and *mut-2/rde-3* mutants [24], reflecting the contribution of Mutator foci components to WAGO 22G RNA accumulation [66].

The 12 WAGO Argonautes of the cognate 22G RNA pathway are clustered into three subclades referred to by phylogenetic branch: Branch 1 comprises WAGO-1 (R06C7.1), WAGO-2 (F55A12.1), WAGO-3 (PPW-2), WAGO-4 (F58G1.1), WAGO-5 (ZK1248.7); Branch 2 comprises WAGO-6/8 (SAGO-2/F56A6.1), WAGO-7 (PPW-1), WAGO-8/6 (SAGO-1/K12B6.1); and Branch 3 comprises WAGO-9 (HRDE-1), WAGO-10 (T22H9.3), WAGO-11 (Y49F6A.1), WAGO-12 (NRDE-3) [24]. *wago-1*, *wago-2*, *wago-3/ppw-2*, and *wago-5* transcripts are classified as germline-intrinsic [33]. *wago-4* and *wago-9/hrde-1* transcripts are classified as oogenesis-enriched, whereas *wago-10* is spermatogenesis-enriched [33]. The *wago-12/nrde-3* transcript is classified as mixed oogenesis-somatic [33]. The others are not assigned germline enrichment classifications [33]. WAGO-12/NRDE-3 and WAGO-9/HRDE-1 will be discussed in greater detail below in the context of the nuclear RNAi pathway.

While loss of single WAGO Argonautes can cause mild defects in exo-RNAi sensitivity and germline transposon silencing [65,68,76], simultaneous loss of multiple WAGO Argonautes within the same phylogenetic branch produces more significant impairment. The MAGO strain lacks Branch 2 and WAGO-4 Argonautes and shows a somatic and germline Rde phenotype and a ts fertility defect [68]. The Quadruple mutant lacks all of the Branch 1 Argonautes but WAGO-1 and shows an Rde phenotype [24]. Similarly,

WAGO Argonautes bind and stabilize WAGO 22G RNAs semiredundantly. In exo-RNAi of the *unc-22* transcript, Branch 2 Argonautes bind secondary siRNAs redundantly, and overexpression of any Branch 2 Argonaute rescues the RNAi defect of the MAGO strain, which otherwise fails to accumulate *unc-22* secondary siRNAs [68]. Still, particular 22G RNAs in germline and soma show different dependencies upon different WAGO Argonautes. The Quadruple mutant, while showing a germline Rde phenotype, still generates wild type levels of a germline *F37D6.3* 22G RNA, whereas loss of the remaining Branch 1 Argonaute, WAGO-1, severely depletes this species [24]. Loss of all five Branch 1 Argonautes in the Quintuple mutant results in dramatic reduction of germline 22G RNAs, whereas a MAGO+2 mutant lacking Branch 2, WAGO-3, and WAGO-4 Argonautes [24] shows near wild type levels of germline 22G RNAs, but loss of a somatic *Y47H10A.5* 22G RNA triggered by RDE-1 [24].

WAGO 22G RNA-mediated post-transcriptional gene silencing

Early evidence suggested that exo-RNAi occurred primarily through a post-transcriptional silencing mechanism. Injection of dsRNA segments corresponding to intron and promoter sequences does not elicit interference [93], and injection of dsRNA targeting a transcript encoded in an operon generally does not decrease levels of transcripts encoded downstream of the target within the operon appreciably [114]. Little progress has been made in understanding cytoplasmic, post-transcriptional target silencing by WAGO 22G RNAs. It is unlikely to involve Argonaute-mediated target cleavage, as WAGO Argonautes lack the catalytic triad residues required for slicer activity [68]; furthermore, target cleavage induced by introduction of 5' triphosphorylated siRNAs into cell lysate is predominantly mediated by CSR-1 [44], which is not involved in the WAGO 22G RNA pathway [24,25]. In the germline, post-transcriptional target silencing may occur at P granules, as GFP::WAGO-1 is expressed in the germline and localizes to perinuclear P granules [24], like Argonautes ALG-3, PRG-1, and CSR-1 [19,25,26,107]. It is possible that RDE-10/RDE-11-mediated target deadenylation and degradation represents the main post-transcriptional mechanism for target silencing

[110]. Earlier data connecting the nonsense-mediated decay (NMD) pathway with exo-RNAi likely resulted from the presence of the *mut-16(mg461)* mutation in the mutant strains of *smg-2*, *smg-5*, and *smg-6* genes, which encode NMD pathway proteins [62]. This explanation is consistent with an initial report that failed to demonstrate a role for the SMG proteins in exo-RNAi [114], although a transgenic system for studying transcriptional gene silencing showed impairment of repression upon depletion of *smg-5* by RNAi [101].

WAGO 22G RNA-mediated transcriptional gene silencing: The nuclear RNAi pathway

Early evidence for transcriptional gene silencing: Chromatin-associated factors MES-2, MES-3, MES-6, and MES-4 are required for germline silencing of high-copy arrays [115], and silenced arrays carry histone modifications consistent with a heterochromatinized state [116]. A screen for factors involved in cosuppression, silencing in trans of an endogenous locus in response to a cognate transgenic array, identified numerous factors that regulate chromatin structure and transcription [90]. In this system, silencing occurs at the transcriptional level, as relieving cosuppression through loss of MUT-16 or MUT-8/RDE-2 results in proportional increases in levels of both spliced and unspliced transcripts [90]. Another transgene was found to show transcriptional silencing upon dsRNA feeding that occurs through a mechanism that requires both RNAi factors and proteins involved in chromatin regulation [101]. Transcriptional silencing of this transgene correlates with decreased RNA polymerase (Pol) II occupancy as well as increased H4 histone acetylation. Furthermore, this transgene shows spontaneous silencing upon loss of RRF-3 [101], suggesting convergence of endo- and exo-RNAi upon a common transcriptional silencing pathway.

Early evidence for heritable and multigenerational gene silencing: The effects of dsRNA injection are often observed in progeny, but rarely heritable beyond the first

generation [93,114], suggesting the existence of an epigenetic heritable interfering agent of limited longevity. However, hermaphrodites injected with dsRNA targeting transcripts expressed in the maternal germline show multigenerational inheritance of silencing mediated by a dominant factor that can be inherited through either sperm or oocyte [117]. Similarly, a germline-expressed transgene transcriptionally silenced by cosuppression takes several generations to recover expression after relief of cosuppression [90]. In contrast, transcriptional silencing of a transgene expressed in the soma is not inherited [101]. These data suggest the existence of two different forms of inherited RNAi: single-generation inheritance of somatic RNAi and multigenerational inheritance of germline RNAi.

NRDE-3 and HRDE-1 WAGO Argonautes mediate nuclear RNAi: Two WAGO Argonautes, WAGO-12/NRDE-3 and WAGO-9/HRDE-1, mediate somatic and germline nuclear RNAi, respectively, through what is referred to as the *Nrde* (Nuclear RNAi-defective) pathway. Although both NRDE-3 and HRDE-1 were reported to be completely dispensable for RNAi-mediated knockdown of *pos-1* or *let-2* [68], these Branch 3 WAGO Argonautes appear to be the sole mediators of transcriptional gene silencing. As WAGO Argonautes, neither has an intact catalytic triad conferring slicer activity [68], but each contains a bipartite nuclear localization signal (NLS) and shows predominantly nuclear localization [79,80,84]. NRDE-3 and HRDE-1 are triggered by WAGO 22G RNA binding to enter the nucleus and associate with nascent pre-mRNA targets, where they recruit NRDE-2, NRDE-1, and NRDE-4 to inhibit Pol II elongation and deposit the repressive H3K9me3 chromatin mark (see below) [79-86,118]. The two Argonautes appear to use common silencing mechanisms and machinery [79-81,85], but their protein expression patterns are distinct: a rescuing GFP::NRDE-3 translational fusion protein shows expression in most somatic cells after the ~80-cell embryo stage [84], whereas GFP::HRDE-1 is expressed in male and female germ cells [80,86]. Consequently, NRDE-3 mediates somatic nuclear RNAi, and HRDE-1 mediates germline nuclear RNAi.

NRDE-3 collaborates with other WAGO Argonautes to silence somatic targets:

NRDE-3-mediated silencing is not always a significant component of the silencing response triggered by exo-RNAi or primary endogenous small RNAs [84,118]. While NRDE-3 likely does not play an important role in silencing of cytoplasmic mRNAs [84], it is required for exo-RNAi of somatic nuclear-localized mRNAs, including polycistronic pre-mRNAs such as *lin-15a-lin15b* and *lir-1-lin-26* [84]. Loss of *nrde-3* also results in increased pre-mRNA and mRNA levels of select endogenous targets [84]; however, a majority of NRDE-3 targets show far less upregulation in the absence of NRDE-3 than in the absence of RRF-1/EGO-1, DRH-3, or EKL-1 [24]. This indicates that other somatic Argonautes such as SAGO-1 and SAGO-2 [68] collaborate with NRDE-3 to mediate WAGO 22G-RNA target silencing, potentially through the RDE-10/RDE-11 complex [109,110]. Accordingly, loss of NRDE-3 abrogates somatic nuclear RNAi entirely, but does not result in a classical somatic Rde phenotype [84]. NRDE-3 is also required for single-generation inheritance of somatic target silencing triggered by exo-RNAi; in progeny, NRDE-3 promotes continued accumulation of secondary WAGO 22G RNAs to reestablish H3K9me3 marks (see below) [82].

HRDE-1 transgenerationally silences diverse targets to promote germline

immortality: In contrast to NRDE-3, HRDE-1 engages the *Nrde* pathway in germ cells to direct silencing that can be inherited over many generations (also called RNAe for *RNA-induced epigenetic silencing*) [79,80,83,85,86,97]. This transgenerational silencing can be initiated by exo-RNAi to establish transcript knockdown that is stable for several generations without additional trigger exposure [79,80,83]. Endogenous small RNA pathways also robustly engage germline nuclear RNAi. 22G RNAs bound by HRDE-1 and WAGO-1 largely overlap [86], affirming that nuclear and cytoplasmic WAGO Argonautes share a common set of siRNA cofactors and targets. These include WAGO 22G RNAs triggered by 21U RNAs and 26G RNAs that maintain germline integrity. In particular, 21U RNAs encode an epigenetic memory of non-self critical for genome surveillance [79,86,95,97]. These fertility-promoting small RNAs associate with HRDE-1 to initiate transgenerational silencing of targets that pose a threat to germline integrity; accordingly, loss of HRDE-1 or downstream *Nrde* factors causes progressive target

desilencing and emergence of pleiotropic germline defects including a Him phenotype, production of nonfunctional gametes, and failure to generate mature oocytes or sperm in late generations at 25°C [80]. Effectively, loss of HRDE-1 results in a Mrt, or *mortal* germline, phenotype. While this phenotype is shared by mutants of downstream Nrde pathway members (NRDE-2, NRDE-1, and NRDE-4), it is not observed with loss of the somatic nuclear Argonaute NRDE-3 [80]. After engaging HRDE-1 to initiate transgenerational silencing, the triggering primary small RNA may not be required, as suggested by the dispensability of PRG-1 for maintenance of transgene silencing initiated by 21U RNAs [79,85,86,95]. For silencing of some targets, however, HRDE-1 and the Nrde pathway collaborate with other, likely cytoplasmic WAGO Argonautes: some transgenes are incompletely silenced by HRDE-1 and show increased desilencing upon loss of additional Argonautes including WAGO-10 [86]. Additional requirements for transgenerational silencing are discussed further below.

NRDE-3 target recognition

Like other WAGO Argonautes, NRDE-3 binds WAGO 22G RNAs. Immunopurified NRDE-3 complexes contain primarily 5' triphosphorylated ~22 nt endo-siRNAs [84]. Consistent with production by a cytoplasmic RdRP acting on spliced transcripts, NRDE-3-associated siRNAs map to exonic sequences, even when RNAi is triggered by dsRNA containing introns [84]. NRDE-3 primarily binds somatic 22G RNAs triggered by ERGO-1 class 26G RNAs, such as *E01G4.5* 22G RNAs, but also binds secondary siRNAs during exo-RNAi targeting a somatic transcript [31,84,118]. Due to redundancy among the WAGO Argonautes, loss of NRDE-3 does not significantly impair stability of these 22G RNAs [84].

Upon binding a cytoplasmic WAGO 22G RNA, NRDE-3 translocates to the nucleus to execute target silencing [84]. Nuclear localization is required for NRDE-3 function: an NLS-defective NRDE-3 mutant protein that binds siRNAs at wild type levels but does not redistribute to the nucleus fails to rescue *nrde-3* mutant phenotypes [84]. siRNA

binding is required for translocation to the nucleus, as mutating the NRDE-3 PAZ domain to abrogate siRNA binding also results in failure to localize to the nucleus [84]. Similarly, loss of siRNA cofactors abolishes nuclear localization. As NRDE-3 binds primarily somatic 22G RNAs triggered by ERGO-1 class 26G RNAs, GFP::NRDE-3 is cytoplasmic in *eri-1*, *ergo-1*, *mut-7*, *rde-4*, *mut-2*, *eri-9*, *eri-6*, and *dcr-1(mg375)* helicase mutants [35,42,84]. The nuclear localization of GFP::NRDE-3 is rescued upon introduction of dsRNA in an *eri-1* mutant, but not an *rde-4* mutant, which cannot initiate exo-RNAi [84]; this suggests that siRNA binding is necessary and sufficient for nuclear redistribution of nuclear Argonautes.

After translocation to the nucleus, the loaded WAGO 22G RNA directs NRDE-3 to associate with unspliced target pre-mRNA [84]. NRDE-3 targeting recognition appears to require very high sequence complementarity [72]: A reporter transgene with a single target site for an X-cluster 22G RNA is efficiently silenced in a wild type background and shows primarily NRDE-3-dependent silencing. Silencing is abrogated by mismatches at positions 1-3, 4-5, 12-14, or deletion of a nucleotide at position 4. Silencing is impaired by mismatches at positions 9-11 or deletion or insertion of a nucleotide at position 13. Mismatches at position 13 or 20-22 are tolerated.

The downstream nuclear RNAi machinery

The proteins NRDE-2, NRDE-1, and NRDE-4 mediate the transcription silencing effects of nuclear Argonautes. After NRDE-3 binds a WAGO 22G RNA, translocates to the nucleus, and recognizes a nascent pre-mRNA target, NRDE-3 associates with NRDE-2 to recruit NRDE-1 to the transcript and direct its deposition on chromatin through a mechanism that requires NRDE-4 [81,84,118]. NRDE-2, NRDE-1, and NRDE-4 are also engaged by HRDE-1 to mediate nuclear RNAi in the germline [79,80,85]. Accordingly, loss of any of these factors results in the same Mrt phenotype observed with loss of HRDE-1 [80], as well as loss of sensitivity to nuclear exo-RNAi [81,118].

NRDE-2: NRDE-2 is a predominantly nuclear-localized, evolutionarily conserved protein encoding a domain of unknown function, a Serine/Arginine-rich domain, and a HAT-like domain [118]. Immunopurified NRDE-2 complexes contain NRDE-3, but only when the NRDE-3 NLS is intact, indicating an exclusively nuclear interaction [118]. NRDE-2 is dispensable for NRDE-3 to transport siRNAs from cytoplasm to nucleus and for NRDE-3 to associate with the target pre-mRNA [118]. Within the nucleus, NRDE-3 recruits NRDE-2 to the targeted nascent transcript. It is interesting to note that siRNA counts appear to be elevated twofold by loss of NRDE-2 [83]; possibly, failure to silence target transcripts in the nucleus via the Nrde pathway results in increased cytoplasmic transcript accumulation, enhancing WAGO 22G RNAs production by RdRPs.

NRDE-1: NRDE-1 is a nematode-specific, nuclear-localized protein containing no obvious protein domains [81]. NRDE-1 is recruited by NRDE-2/NRDE-3 to pre-mRNA near the target site, where it is required for nuclear RNAi-mediated inhibition of transcription elongation; a *nrde-1* mutant fails to show 5' transcription inhibition of *lin-15b* upon RNAi and survives RNAi targeting of the *lir-1* polycistron in a sensitizing *eri-1* mutant background [81]. NRDE-1 also associates with chromatin at the targeted genomic locus in a NRDE-4-dependent manner and links nuclear RNAi to chromatin regulation by promoting H3K9 trimethylation at genomic sites targeted by WAGO 22G RNAs.

NRDE-4: NRDE-4 is a nematode-specific protein with a predicted bipartite NLS and no other obvious domains [81]. The *nrde-4* transcript is classified as germline-intrinsic [33]. NRDE-4 does not recruit NRDE-3, NRDE-2, or NRDE-1 to pre-mRNA but is required for recruitment of NRDE-1 to chromatin and therefore for transcriptional inhibition and H3K9 trimethylation [81].

Nuclear RNAi inhibits transcriptional elongation and drives H3K9 trimethylation

Nuclear RNAi terminates transcription and promotes establishment of a repressed chromatin state at genomic loci targeted by exo-RNAi as well as multiple loci targeted by endo-siRNAs and 21U RNAs [80,81,83,85,86]. Nuclear RNAi requires interaction with target pre-mRNA, as targeting 5' to the 5' UTR does not result in silencing [83]. NRDE-3, NRDE-2, and NRDE-1 bind to pre-mRNA at the target site but are not detected in association with pre-mRNA 5' to the target site, presumably due to co-transcriptional splicing of nascent transcripts [81,118,119]. Nrde targeting inhibits transcription elongation, as Pol II occupancy increases near the target site but not at initiation sites [118]. Nuclear run-on assays show Pol II inhibition occurring ~2 kb downstream of the target site, correlating with decreased Pol II occupancy by chromatin immunoprecipitation 3' to the target site [81,118]. Because transcription is inhibited, NRDE-3, NRDE-2, and NRDE-1 do not bind pre-mRNA sequences encoded 3' to the target site; however, Nrde factors appear to remain associated with pre-mRNA fragments after silencing [81,118].

Nrde pathway targeting results in enrichment of H3K9 trimethylation at genomic target sites that correlates with the transcription inhibition [81], consistent with the established repressive role of H3K9 methylation. Whereas NRDE-3 and NRDE-2 recruit NRDE-1 to the RNA target near or 5' to the target site, NRDE-1 requires NRDE-4 for chromatin interaction and associates with chromatin predominantly 3' to the site of RNAi [81]. Accordingly, H3K9 trimethylation occurs across the gene but peaks 3' to the target site [81]. The chromatin response to nuclear RNAi occurs at the subgenic level, as shifting the RNAi trigger site shifts the interval of H3K9me3 deposition [83]. During exo-RNAi-triggered nuclear RNAi, secondary siRNAs accumulate mainly at exonic regions proximal to the dsRNA trigger site, but both exons and introns show high H3K9me3 levels after targeting [83].

Although many of the studies of NRDE-3 function in nuclear RNAi were conducted under exo-RNAi, NRDE-3-mediated nuclear RNAi is also important for transcriptional silencing triggered by a subset of endogenous small RNAs in the soma [81]. As mentioned above, NRDE-3 binds somatic WAGO 22G RNAs, many of which are

triggered by ERGO-1 class 26G RNAs [31,84,118]. 22G RNAs targeting the ERGO-1 target *E01G4.5* are abundantly bound by NRDE-3 [84] and direct chromatin modification at this locus through the somatic nuclear RNAi pathway [81]. Loss of members of the ERI complex, NRDE-3, or downstream *Nrde* pathway factors results in a subtle but detectable increase in *E01G4.5* transcription inhibition, increased transcript levels, and depletion of H3K9me3 marks at the *E01G4.5* locus [81]. The strong influence of NRDE-3 on H3K9me3 deposition at the *E01G4.5* locus is uncommon, and levels of many ERGO-1 targets show little or no effect of loss of the somatic *Nrde* pathway [81], consistent with significant overlap between cytoplasmic and nuclear WAGO Argonautes in silencing of somatic targets [24]. Nonetheless, in the absence of ERGO-1 class 26G RNAs, the somatic *Nrde* pathway can be heavily engaged by *exo*-RNAi, as indicated by the emergence of RNAi phenotypes for polycistronic transcripts in an *eri-1* mutant [54,120]. In fact, *exo*-RNAi of some monocistronic transcripts also engages NRDE-3 in the absence of ERGO-1 class 26G RNAs, as *exo*-RNAi targeting *unc-73* and *dpy-13* results in less severe phenotypes in an *eri-1; nrde-3* double mutant than in an *eri-1* mutant [84]. While loss of ERGO-1 class 26G RNAs does sensitize the somatic nuclear RNAi system for phenotypic analysis, an *eri-1* or equivalent mutant background is not necessary for the transcription- and chromatin-level effects of nuclear RNAi [81].

NRDE-3 directs single-generation epigenetic inheritance of silencing

Although NRDE-3 acts primarily in somatic cells, the silencing effects of somatic nuclear RNAi are transmissible for one generation [82]. *Exo*-RNAi in the parental generation generates secondary siRNAs abundantly bound by NRDE-3; inheriting progeny also show high levels of NRDE-3-bound secondary siRNAs from embryo through adulthood, although these are not inherited into F2 embryos [82]. In the absence of NRDE-1 or NRDE-4, these secondary siRNAs are still inherited and associate with NRDE-3, but levels of NRDE-3-bound secondary siRNAs are not maintained throughout development [82]. This indicates that somatic nuclear RNAi is not required for inheritance of secondary siRNAs but is required for continued expression of these siRNAs in progeny.

H3K9me3 marks established in the parental generation at RNAi target loci are lost in progeny during the embryo stage, possibly due to the absence of NRDE-3 in early embryogenesis [82,84]. Silencing is then reestablished through NRDE-3 and downstream Nrde factors during development and strengthened to exceed the H3K9me3 levels in the parental generation [82]. The mechanism is likely the same in progeny as in the parental generation, as NRDE-3 siRNA binding and nuclear translocation are still required for inherited silencing [82]. However, the increased H3K9me3 deposition in progeny suggests that the inheritance of a silencing signal may further promote the ability of nuclear RNAi to repress chromatin, possibly by enhancing coupling of siRNA generation and H3K9me3 deposition in progeny. Importantly, this reestablishment of silencing requires the activity of NRDE-3 in progeny: homozygous *nrde-3* mutant progeny of heterozygous *nrde-3* hermaphrodites subjected to exo-RNAi do not show inheritance of silencing [82]. The converse has not been tested; it is not known whether somatic nuclear RNAi is required in the parental generation for RNAi inheritance and chromatin repression in progeny. Furthermore, it is as yet unclear whether somatic nuclear RNAi is required for establishment or maintenance of transcriptional silencing during RNAi inheritance in the soma.

Somatic nuclear RNAi in the parental generation directs siRNA expression and H3K9 trimethylation in inheriting progeny through one of two possible mechanisms. NRDE-3-bound secondary siRNAs may be directly deposited into germ cells to direct embryonic silencing in association with other WAGO Argonautes, then re-bind NRDE-3 to promote maintenance of siRNA expression through nuclear RNAi during development. Alternatively, a different silencing signal such as primary small RNAs may be deposited in germ cells to promote expression of siRNAs in progeny that bind NRDE-3 and re-engage nuclear RNAi [82]. In support of the latter mechanism, NRDE-3-bound WAGO 22G RNAs do not appear to initiate their own amplification at target sites [72], suggesting that primary small RNA trigger may be required for secondary siRNA re-amplification in progeny. One likely possibility is that primary exo-siRNAs are directly inherited from hermaphrodites exposed to dsRNA. Indeed, *rde-1* transcript is abundant in hermaphrodite germline and detected at high levels in embryos of all stages

(NEXTDB clone 26b3), suggesting that primary exo-siRNAs may be inherited bound to RDE-1 to trigger NRDE-3-mediated nuclear RNAi in inheriting progeny. Such a mechanism would explain why somatic nuclear RNAi is heritable for only a single generation.

HRDE-1 directs multigenerational epigenetic inheritance of silencing

Like NRDE-3, HRDE-1 engages the *Nrde* downstream machinery to direct transcriptional silencing and H3K9me3 deposition at loci targeted by exo-RNAi and endogenous small RNAs; however, HRDE-1 binds germline WAGO 22G RNAs and engages the downstream *Nrde* pathway to direct transgenerational silencing of germline transcripts [79,80,83,85,86,97]. Loss of HRDE-1, NRDE-2, NRDE-1, or NRDE-4 results in the *Mrt* phenotype described above [80], likely due to desilencing of germline targets such as transposons that pose a threat to genome integrity when active. Further testing of mutants for various upstream and downstream 21U RNA, 26G RNA, and WAGO 22G RNA pathway factors is necessary to determine the respective contributions of these pathways to promoting germline immortality. Curiously, the *Mrt* phenotype of *hrde-1* and *nrde-2* mutants manifests only at 25°C, whereas *nrde-1* and *nrde-4* mutants are *Mrt* at 25°C and 20°C [80], suggesting the chromatin machinery of the *Nrde* pathway may have additional roles in germline maintenance.

Multigenerational inheritance of transcriptional silencing requires germline expression of the targeted transcript [101]. Most likely, an mRNA template must be present in the germline for RdRP to generate siRNAs capable of propagating silencing across each generation; this may account for the limited perdurance of NRDE-3-mediated somatic nuclear RNAi [82]. In contrast, HRDE-1-mediated nuclear RNAi in the germline results in continued siRNA production and silencing over multiple generations that maintain H3K9 trimethylation [79,80,83,85,86]. Accordingly, factors such as MUT-7 and RDE-3 that are required for 22G RNA accumulation are also required for maintenance of transgenerational silencing [85,86]. Screens using transgenerationally silenced exo-

RNAi and 21U RNA sensor transgenes have identified a set of chromatin factors necessary for maintenance of transgenerational RNAi: the chromo domain protein HPL-2 (semiredundantly with HPL-1), Polycomb complex protein MES-3, Trithorax complex protein MES-4, SET domain SET-32, and putative H3K9 methyltransferase SET-25 [79,86]. These proteins may be recruited by NRDE-1 to mediate H3K9 methylation or otherwise promote a chromatin state favorable for continued siRNA biogenesis to maintain nuclear RNAi.

In the parental generation exposed to dsRNA, HRDE-1 is dispensable for germline silencing; however, HRDE-1 is required in the F1 generation for inheritance of parental germline silencing, in the F2 generation for inheritance from the F1, and so on, as indicated by a failure of homozygous *hrde-1* mutant offspring to inherit germline transgene silencing from a *hrde-1* heterozygous hermaphrodite [80]. This result differs somewhat from another study, which shows that loss of HRDE-1 or MUT-7 reactivates a silenced locus only in the second generation of *hrde-1* homozygosity [85], suggesting that epigenetic silencing may be established in the parental germline.

Surprisingly, the germline nuclear RNAi pathway appears to require cytoplasmic WAGO Argonautes for initiation: during exo-RNAi targeting the oogenesis-enriched *smg-1* transcript [33], the target site fails to become enriched for H3K9me3 marks in the MAGO mutant strain [83]. This definitively indicates that the nuclear and cytoplasmic WAGO Argonautes do not compete for siRNA cofactors. Rather, cytoplasmic WAGO Argonautes promote nuclear RNAi through an unknown mechanism. It is not yet established whether WAGO Argonautes bind 22G RNAs irreversibly or reversibly; if the latter, perhaps HRDE-1 appropriates WAGO 22G RNAs from the cytoplasmic Argonautes for nuclear silencing. NRDE-3 does not appear to be required for establishment or maintenance of transgenerational silencing [80,85], nor is a sensitizing mutation in endo-RNAi factors such as ERI-1 or RRF-3 required [83]. Other WAGO Argonautes do, however, contribute variably to HRDE-1 target silencing, as some 21U RNA sensor transgenes show dependence on additional Argonautes for complete silencing [86,97].

While establishment of transgenerational silencing requires factors involved in the formation of the initial trigger, such as RDE-1 in the case of primary exo-siRNAs [80,83] or PRG-1 in the case of 21U RNAs [79,85,86,95,97], maintenance of transgenerational silencing does not [79,80,85,86,95]. Once established, HRDE-1-mediated nuclear silencing is stable for multiple generations and can be inherited from either the male or female germline as a dominant signal, although silencing may show a generational lag when inherited from the male [79,85,86]. This silencing in trans is mediated by a diffusible agent, namely, secondary 22G RNAs, capable of repressing homologous DNA regardless of whether the primary small RNA target site is present [85]. These 22G RNAs map further upstream of the initiating 21U RNA target site than observed during early, PRG-1-dependent silencing [97], suggesting that transgenerational RNAi may entail additional spreading during maintenance [85]. 22G RNAs whose expression is stimulated at transgenerationally silenced 21U RNA targets to maintain silencing may not correspond precisely to the species triggered by 21U RNA targeting, as the latter appear to be PRG-1-dependent and are lost when transgenerationally silenced loci are introduced into a *prg-1* mutant background [79].

After exposure to dsRNA, secondary siRNA accumulation and association with HRDE-1 decrease progressively across generations [80]. Transgenerational silencing initiated by endogenous small RNAs shows a similar effect, as endogenous targets of HRDE-1 silencing show progressive loss of H3K9me3 enrichment and increased expression over successive generations in a *hrde-1* mutant background [80]. This suggests that endogenous primary small RNA triggers may be required for periodic reinforcement of established transgenerational silencing. It is intriguing to speculate what function the nuclear RNAi pathways serve. While nuclear RNAi may simply provide a heritable signal to direct silencing of aberrant and potentially dangerous transcripts, it has also been suggested that targeting of these endogenous loci by nuclear RNAi may also fulfill a larger function in regulating chromatin dynamics.

Features and targets of CSR 22G RNAs

Those 22G RNAs that are not bound by WAGO Argonautes are the CSR 22G RNAs. Named for their sole Argonaute effector, CSR-1, these 22G RNAs do not effect target silencing but rather associate with chromatin to promote proper association of targets within the holocentric chromosomes of *C. elegans* [25]. CSR 22G RNAs are produced in the germline by the same DRH-3- and EKL-1-containing RdRP module as WAGO 22G RNAs, but require EGO-1 and not RRF-1 for accumulation [24,25]. Although CSR 22G RNAs share the features and biogenesis machinery of the secondary siRNAs engaged by WAGO Argonautes, there is no evidence that CSR 22G RNAs are themselves secondary siRNAs.

CSR 22G RNAs largely target germline-expressed genes. Immunopurified CSR-1 complexes are enriched for a set of 22G RNAs that are antisense to over 4,000 protein-coding genes, whereas microRNAs, 21U RNAs, and 22G RNAs targeting repetitive sequences, pseudogenes, and intergenic or unannotated loci are depleted [25]. Similarly, comparison of deep sequencing libraries generated from *csr-1* and *ego-1* mutants [25] with *drh-3* and *ekl-1* mutants [24] reveals that CSR-1 and EGO-1 are required for accumulation of 22G RNAs antisense to protein-coding genes, many of which also require DRH-3 and EKL-1 [25]. Most CSR 22G RNAs are depleted in the absence of germline [24], consistent with the germline-restricted expression pattern of EGO-1 [25,40,87,88]. Less than 1% of 22G RNAs recovered in immunopurified CSR-1 complexes target repetitive elements and pseudogenes [25].

The CSR 22G RNA pathway molecular machinery

The 22G RNA RdRP module is discussed in detail above; the following sections focus on the features and functions of EGO-1, DRH-3, and EKL-1 only as they pertain to the CSR 22G RNA pathway. In situ hybridization indicates that concentrations of *ego-1*, *drh-3*, *ekl-1*, and *csr-1* transcripts are highest in gonad and early embryo [121], in line with other transcript and protein localization data [25,33,40,48,87,88]. DRH-3 and EKL-1

are not germline-restricted, as they also function in WAGO 22G RNA biogenesis in germline and soma [24,25].

CSR-1: The CSR-1 Argonaute clusters phylogenetically within the WAGO clade [68], but does not bind WAGO 22G RNAs [24,25]. Nevertheless, loss of CSR-1 results in partially defective germline and early zygotic transgene silencing [68], although this may be due to defective localization of P granules in the absence of CSR-1 [25]. CSR-1 contains an intact catalytic triad [68]. In a cell-free system, addition of single-stranded, 5'triphosphorylated 23 nt siRNA and target mRNA results in significant target cleavage that is 90% depleted in *csr-1* null mutant adult lysate; substitution of 5' monophosphorylated siRNA decreased cleavage efficiency considerably [44]. Recombinant CSR-1 is also capable of catalyzing robust cleavage in vitro when incubated with single-stranded 5' triphosphorylated, but not 5' monophosphorylated, siRNA and target [44]. Like the slicer activity of RISC in other organisms [122,123], CSR-1-mediated cleavage in cell lysate requires Magnesium ions and is abolished by mutation of the two nucleotides flanking the target cleavage site [44]. Although these in vitro findings demonstrate the catalytic competence of CSR-1, no direct function for its cleavage activity has been demonstrated. The *csr-1* gene encodes two isoforms. Both are expressed at all developmental stages, but *csr-1* mRNA and protein are most enriched in young adult, gravid adult, and embryo [25]. The larger CSR-1 protein isoform is expressed throughout larval development and is not germline restricted [25]. CSR-1 does not require EGO-1, DRH-3, or EKL-1 for stability, nor does any of those proteins require CSR-1 or another member of the RdRP module for stability [25].

Phenotypes associated with CSR 22G RNA pathway compromise

EGO-1, DRH-3, EKL-1, and CSR-1 represent the main components of the CSR 22G RNA pathway [25]. Earlier phenotypic analyses and focused or genome-wide RNAi screens have identified an overlapping set of phenotypes associated with depletion of these proteins that include embryonic lethality, larval arrest, and sterility [37,40,68,124-

133]. These four factors were first implicated in a common pathway by their shared enhancement of *ksr-1* lethality (Ekl) phenotype, which describes a failure to specify excretory duct cell fate in late embryogenesis [92]. Transcripts whose loss produces an Ekl phenotype are expressed in maternal germline to regulate embryonic gene expression [92].

Germline phenotypes: Mutation or RNAi-mediated depletion of *ego-1*, *drh-3*, *ekl-1*, or *csr-1* results in common mitotic and meiotic defects leading to sterility [37,68,90,124,129]. In general, loss of these factors results in underproliferation of the germline, with nuclei of abnormal shape and size [25,37,87,88,121]. *ego-1* mutants show moderately reduced germ cell number, enlargement of the transition zone with large, diffuse transition zone nuclei, reduction of the pachytene zone, and overall delayed gametogenesis [40,87,124]. The spermatogenesis-to-oogenesis transition is also delayed, resulting in production of excess sperm [40]. Sperm produced by an *ego-1* mutant are incapable of fertilization, and oocytes are small, nonfunctional, and show unpaired homologous chromosomes [40,124]. Like *ego-1* mutants, *drh-3*, *ekl-1*, and *csr-1* mutant germlines show protracted transition zones that include some large nuclei with diffuse chromosome morphology [37,121]. These mutants also show aberrantly segregated chromosomes in diakinesis oocyte nuclei, suggesting defects in pairing, synapsis, or recombination [25,48,121]. *drh-3* and *ekl-1* mutants produce sperm with abnormally large and variably sized nuclei, indicating impaired chromatin condensation or segregation; a similar phenotype is observed in an *ego-1*; *csr-1* double mutant [121]. These phenotypes suggest pervasive chromosome segregation defects.

Embryonic phenotypes: Loss of EGO-1, DRH-3, EKL-1, or CSR-1 results in embryonic lethality [37,68,92]. Fertilized *ego-1* mutant oocytes produce eggshells, but arrest as balls of 20-50 cells without undergoing morphogenesis or gastrulation [40]. Similarly, RNAi-mediated depletion of *ego-1*, *drh-3*, *ekl-1*, or *csr-1* results in production of embryos that show normal prophase chromosome condensation, poor metaphase alignment, chromosomal bridging at anaphase, and bisection of lagging chromosomes by the cleavage furrow at cytokinesis during every cell division; these defects cause

accumulation of abnormally shaped nuclei with non-wild type chromosomal complements and result in embryonic lethality due to arrest between gastrulation and early morphogenesis [25,37,48,68]. *drh-3* hypomorphic mutants and a *csr-1* mutant partially rescued by transgenic CSR-1 expression both produce dead embryos arrested at various points in embryogenesis that contain many nuclei with abnormal DNA content [24,25]. A Him phenotype is also observed in these mutants, confirming impairment of chromosome segregation [25].

Localization and assembly of the CSR 22G RNA complex

Germline localization: In the germline, EGO-1, DRH-3, and CSR-1, but not EKL-1, localize to P granules and promote or maintain P granule structure and association with the nuclear periphery [25]. As oocytes mature, EGO-1 is lost from P granules, but some DRH-3 and CSR-1 continue to associate with P granules in germ cells throughout the life cycle. DRH-3 recruits EGO-1 to the P granule, which recruits CSR-1 in an EKL-1-dependent manner. This assemblage then drives P granule association with the nuclear periphery, possibly by binding mRNA targets as they exit through nuclear pores; such a mechanism would be consistent with loss of P granule perinuclear localization in states of low transcription such as in early germ cells or oocytes [25]. Mutation of *ego-1*, *drh-3*, *ekl-1*, or *csr-1* disrupts perinuclear localization of P granules [25,87], possibly contributing to the Rde phenotypes exhibited by these mutants [24,37,40,68,90,91]. These four factors may also promote normal nuclear pore structure. In wild type germlines, P granules are detected near nuclear pore clusters, although not every nuclear pore is associated with a P granule [106]. Pachytene chromosomes are not located adjacent to nuclear pores and may not be able to attach to the nuclear envelope in regions of high nuclear pore density [106]. Because the nuclear envelope assembles around chromatin, the chromatin may therefore influence distribution of nuclear pores [134]. Loss of EGO-1 results in a patchy distribution of nuclear pore marker nucleoporin in mitotic and transition zones that might reflect abnormal chromatin dynamics [87]. In sperm and oocytes, CSR-1 is detected in chromatin fractions [135], consistent with

immunofluorescence data indicating that CSR-1 and EGO-1 are enriched in mature oocyte nuclei, with CSR-1 associating with diakinetically chromosomes [25].

Embryo localization: EGO-1, DRH-3, EKL-1, and CSR-1 localize to mitotic chromosomes in embryo [25]: In embryonic cells, all four factors show cytoplasmic localization. During mitosis, however, they show nuclear enrichment beginning at prophase. As embryonic mitotic chromosomes condense, EGO-1, DRH-3, and EKL-1 become enriched along the length of each chromosome, but CSR-1 remains nuclear. At the metaphase plate, DRH-3 and CSR-1 localize to chromosomes in a pattern similar to that of cohesins, whereas EKL-1 and, to a lesser degree, EGO-1 localize to chromosomes in a pattern like that of kinetochore proteins. EKL-1 stays associated with chromosomes during anaphase, while the others are more difficult to detect. The factors assemble on chromosomes in a hierarchical order, with DRH-3 recruiting EGO-1 and EKL-1, which in turn recruit CSR-1 [25]. In embryo, CSR-1 pathway components are found to associate directly with chromatin in a CSR 22G RNA-dependent manner, supporting a direct role for CSR-1 and associated 22G RNAs in promoting chromosome segregation: CSR-1 shows enrichment at CSR 22G RNA target loci, but not WAGO 22G RNA target loci, by chromatin immunoprecipitation; this association is lost in a *drh-3* mutant [25]. As CSR 22G RNA target loci are distributed relatively uniformly along chromosomes, CSR-1 associates with chromatin genome-wide to influence chromosome segregation [25].

CSR 22G RNA effector function

Although CSR-1 shows cleavage activity in cell lysate [44], CSR-1 does not appear to silence CSR 22G RNA targets: these targets are expressed in germline, oocyte, and embryo and do not show changes in expression upon loss of CSR-1 or DRH-3 [24,25]. Rather, EGO-1, DRH-3, EKL-1, and CSR-1 localize to chromosomes to promote proper organization and alignment of metaphase chromosomes and proper orientation of kinetochores to opposing spindle poles [25]. Loss of these factors disrupts chromosome

condensation and cohesion: depletion of any by RNAi results in highly disorganized condensin and cohesin protein loading onto mitotic chromosomes in embryo [25]. Loss of these factors also disrupts kinetochore assembly. Whereas inner centromeric proteins HCP-3/CENP-A and HCP-4/CENP-C should be poleward-localized on both sides of metaphase plate, embryos depleted of CSR 22G RNA pathway factors show dramatically disorganized loading of these proteins on metaphase chromosomes; outer kinetochore protein KLP-7/MCAK and spindle checkpoint protein BUB-1 are also misloaded onto mitotic chromosomes [25]. Thus, the CSR 22G RNA pathway directly influences chromosome organization to promote proper segregation. The X chromosome is depleted of genes expressed in the germline [33] and therefore of CSR 22G RNA targets. Nevertheless, the X chromosome usually segregates correctly, suggesting that other mechanisms may govern X chromosome segregation [25].

The CSR 22G RNA pathway also regulates H3K9me2 distribution. H3K9me2 becomes enriched on unpaired chromosomes and high-copy extrachromosomal arrays during meiosis [116,136]. Loss of *EGO-1*, *DRH-3*, *EKL-1*, or *CSR-1* results in reduced H3K9me2 accumulation on unpaired chromosomes such as the male X chromosome and ectopic H3K9me2 accumulation on paired and synapsed chromosomes [88,121]. Whereas wild type male pachytene nuclei show a single focus of H3K9me2 labeling, *ego-1*, *drh-3*, *ekl-1*, or *csr-1* mutant nuclei with normal pachytene morphology lack a single strong focus of H3K9me2 labeling, and some show multiple bright H3K9me2 foci with higher overall levels of H3K9me2 [88,121]. Mutant nuclei, however, may be large, morphologically abnormal, and polyploid, with diffuse chromosome morphology and multiple H3K9me2 foci and high overall H3K9me2 [121]. This demonstrates a role for CSR 22G RNAs in specifying chromatin modification. However, as CSR 22G RNAs target protein-coding genes expressed in germline [24,25], it is likely that CSR-1 and associated 22G RNAs are excluding repressive H2K9me2 marks from target genomic loci rather than depositing them.

Another function of CSR 22G RNAs may be to define areas of germline transcription. This is consistent with a proposed role for CSR 22G RNAs in acting as a self-

recognition system in opposition to 21U RNA-directed silencing of DNA perceived as non-self; targeting of transcripts by CSR-1 may therefore represent a licensing process that helps to maintain expression of bona fide germline transcripts [79,86]. As CSR 22G RNAs are generated in germline through the activity of RdRP EGO-1, CSR 22G RNA targets are robustly expressed in maternal germline and may therefore engage CSR-1 as nascent transcripts during gametogenesis; this may define CSR-1 chromatin domains that could perdure through embryogenesis independent of transcription [25]. Such a mechanism would account for the prominent localization of CSR-1 to chromatin in mature oocytes [25]. Presumably, CSR-1 establishes euchromatic genomic regions and defines the boundaries of adjacent kinetochores, potentially by mediating deposition of activating chromatin marks and HCP-3/CENP-A incorporation, respectively. In embryo, CENP-A is loaded at low density onto about half of the genome, inverse to regions transcribed in germline and early embryo, and defines centromeric regions [137]. Pre-existing CENP-A nucleosomes are not necessary to guide recruitment of new CENP-A nucleosomes, as CENP-A is not transmitted by sperm during fertilization and is unloaded and reloaded during oogenic meiotic prophase [137]. Germline transcription of a genomic region may exclude CENP-A incorporation in progeny, with reinforcement by zygotic transcription in early embryogenesis. Domains targeted by the CSR 22G RNA pathway and regions enriched for the centromeric histone variant CENP-A appear to be mutually exclusive [25]; therefore, the CSR 22G RNA pathway may serve to transmit patterns of germline transcription to early embryo. Alternatively, H3K36 methylation may fulfill this role, as its pattern also correlates inversely with CENP-A occupancy [137,138].

Finally, the CSR pathway also appears to play a critical role in maturation of replication-dependent core histone mRNAs. Processing of replication-dependent core histone pre-mRNAs requires endonucleolytic cleavage in the 3'UTR just downstream of the stem-loop that in many organisms is directed by the U7 snRNA [139]. Nematodes, however, lack the U7 snRNA [140], and CSR 22G RNAs may fulfill this role in its stead [141]: A majority of histone genes (66%) are depleted by loss of CSR-1. Immunopurified CSR-1 complexes are enriched for 22G RNAs targeting all core histone genes, and EGO-1-

dependent 22G RNAs map to histone pre-mRNAs at the 3' end sites of mature histone mRNAs [141]. CSR-1 binds these 22G RNAs to target histone pre-mRNAs and either directly cleave them or facilitate their cleavage. Loss of CSR 22G RNA pathway factors results in unprocessed histone mRNAs accumulation, with a corresponding decrease in histone production that could explain the phenotypes of these mutants [141]. For example, histone biogenesis defects in cultured cells result in S-phase extension and cell cycle delay [142,143] reminiscent of the extended transition zone observed in *ego-1* mutant germlines [40]. Additionally, RNAi-mediated knockdown of a single core histone gene causes chromosome segregation defects and sterility similar to the phenotypes of CSR 22G RNA pathway mutants [131,144]. Consistent with this explanation, overexpression of transgenic histone pre-mRNAs that do not require cleavage rescues lethality associated with *csr-1* or *ego-1* depletion [141]. Phenotypes similar to those shown by CSR 22G RNA pathway mutants also result upon depletion of the stem-loop-binding protein (SLBP), CDL-1. SLBP is required for histone mRNA processing, stabilization, and efficient translation [139]. In the germline, CDL-1::GFP shows nuclear enrichment in developing oocytes, much like CSR-1 [25,141]. RNAi-mediated knockdown of *cdl-1* results in accumulation of unprocessed histone mRNA, severely depletes histone levels, and causes adult sterility and chromosome segregation defects in early embryo [141,144,145]. In contrast, RNAi-mediated knockdown of condensin component *smc-4* does not result in histone processing and accumulation defects, showing that this is not a general feature associated with compromised chromosome segregation [141]. These results suggest that CSR-1 and CDL-1 may collaborate to promote histone maturation in oocytes, ensuring that stores of histones will be sufficient to sustain embryonic division until zygotic transcription begins. It is interesting to note that the sterility associated with compromise of the CSR 22G RNA pathway may be independent of the P granule defects, as loss of *cdl-1* does not disrupt P granules [141]. Finally, intestinal cell endoreduplication also requires high levels of histone proteins [146]. Possibly related to somatic expression observed for the longer *csr-1* isoform [25], CSR-1 protein is detected in adult intestinal nuclei [141]. These cells show depletion of histone H2B upon RNAi knockdown of *csr-1*, suggesting that CSR-1 function in histone processing may not be restricted to the germline [141].

CDE-1 regulates the CSR 22G RNA pathway

The CSR 22G RNA pathway requires the conserved beta-nucleotidyltransferase CDE-1 for uridylation of 22G RNAs to restrict their accumulation [147]. CDE-1 is encoded by a germline-enriched transcript [33] that is detected by in situ hybridization throughout the entire gonad but not in other tissues [147]. In the male or hermaphrodite spermatogenic germline, CDE-1 localizes to bright granules in close proximity to the condensing DNA of mature sperm. In the oogenic germline, CDE-1 localization is cytoplasmic with perinuclear enrichment. Within the embryonic P lineage, CDE-1 localizes to perinuclear granules, colocalizing mainly with PGL-1; in mitotic cells, CDE-1 shows EGO-1-dependent localization to the outer edges of condensing chromosomes at prometaphase and the poleward sides of chromosomes at metaphase, with expression detectable through anaphase. CDE-1 immunopurifies with EGO-1 in embryo extract, but CSR-1 is not required for EGO-1 accumulation. Loss of EGO-1 results in dispersed localization of CDE-1 over the whole metaphase plate and a CDE-1 halo surrounding the DNA in condensing sperm. Depletion of CSR-1 results in very faint but correct CDE-1 localization to the metaphase plate, but no association of CDE-1 granules with DNA in sperm. Loss of EGO-1 or CSR-1 does not affect CDE-1 localization to P-granules, and EGO-1 and CSR-1 show normal localization in the absence of CDE-1. CDE-1 protein is not detected in intestine [147].

CDE-1 is a catalytically active nucleotidyltransferase. Immunopurified CDE-1 complexes preferentially catalyze the addition of uridines to small RNA 3' ends in vitro [147], consistent with earlier findings expressing CDE-1 in *Xenopus* oocytes [104]. Uridine tails added in vitro are short, suggesting low processivity, and can be blocked by 2'-O-methylation of the substrate small RNA 3' end [147]. In vivo, CDE-1 uridylates a portion of CSR 22G RNAs. Endo-siRNAs, but not microRNAs, show addition of one or more untemplated 3' uridines that are lost in the absence of CDE-1 in adult; however, some endo-siRNAs in young adult still show untemplated uridines in a *cde-1* mutant, suggesting redundant terminal uridylation activity at earlier stages [147]. It is not clear whether CDE-1 uridylates 22G RNAs already bound by CSR-1; however, no association

between the two proteins is detected in embryo extract [147]. Uridylation by CDE-1 appears to destabilize CSR 22G RNAs, as those frequently found uridylated in wild type are increased upon loss of CDE-1, whereas levels of microRNAs, 21U RNAs, and other antisense small RNAs are not [147]. Accordingly, immunopurification of CSR-1 complexes from a *cde-1* mutant recovers many more siRNAs, although the distribution of targets does not change [147]. This suggests that CDE-1 acts to reduce the half-life of CSR 22G RNAs, ensuring that CSR-1 is only partially loaded, and, indeed, CSR 22G RNAs are rarer by far than WAGO 22G RNAs [24,25]. Their modest accumulation may also explain why CSR-1 22G RNA targets are not silenced, as only the more abundantly WAGO 22G RNA-targeted loci are appreciably silenced [24,25]. CDE-1-mediated uridylation and partial loading of CSR-1 may serve to restrict CSR 22G RNAs to the CSR 22G RNA pathway, as loss of CDE-1 results in erroneous gene silencing likely due to misloading of CSR 22G RNAs onto WAGO 22G RNA Argonautes [147]. This may explain the phenotypes of *cde-1* mutants, which show defects in cosuppression, increased Tc1 transposon activity, and partial insensitivity to RNAi in the germline [90,147]. In the absence of CDE-1-mediated uridylation, excess CSR 22G RNAs compete with WAGO 22G RNAs for Argonaute binding, compromising exo-RNAi in the germline. Similarly, Tc1-targeting siRNAs are decreased, possibly due to decreased stabilization by available WAGO Argonautes, resulting in Tc1 activation and increased dsDNA breaks and apoptosis [147].

Loss of CDE-1 leads to meiotic and embryonic mitotic segregation defects with resulting Him phenotype, presumably due to erroneous targeting of CSR 22G RNA targets by WAGO Argonautes [147]. Germline nuclei do not show mitotic missegregation, but diakinetid oocytes often show univalents, suggesting impaired pairing of homologous chromosomes during meiosis [147]. *cde-1* mutant embryos show metaphase plate disorganization similar to that observed in *csr-1* and *ego-1* mutant embryos, with abnormal CENP-A deposition, and occasional severe spindle defects and polar body retention [25,40,68,147]. In analysis of *cde-1* mutant hermaphrodite x wild type male hybrid embryos, both parental chromosomes are affected by major segregation defects, indicating that missegregation is not limited to the parental germline and persists into

embryonic development [147]. Consistent with the collaborative but antagonistic functions of CDE-1 and CSR 22G RNA pathway factors, the embryonic lethality of *cde-1* mutants is less severe than that of *csr-1* and *ego-1* mutants [147]. CDE-1 also regulates processing of pre-let-7 through uridylation [148] and may act in additional roles in other small RNA pathways.

21U RNA features and loci

The following pertains to canonical or “Type 1” 21U RNAs. “Type 2” 21U RNAs are associated with transcription start sites (TSSs) of protein-coding mRNAs and other Pol II transcripts and are discussed below.

Early deep sequencing of *C. elegans* small RNAs identified a highly diverse population of 21 nt small RNAs with a 5' uridine bias and no other common sequence features [18]. These 21U RNAs are 5' monophosphorylated [18] and terminally methylated by HENN-1 methyltransferase [18,70,72,73]. Unlike endo-siRNAs, 21U RNAs are not generated from mRNA templates by RdRPs but rather transcribed directly from genomic loci primarily clustered in two gene-depleted regions of chromosome IV [18,149]. 21U RNA loci are vastly depleted for overlap with exons. Those that do overlap show no sense or antisense bias [18]. 21U RNA loci do not show a prominent strand bias and overlap less than expected by chance [18], consistent with their independent transcription from short precursors [150], [149]. Although the larger genomic clusters of 21U RNA loci show synteny across rhabditids, the 21U RNAs themselves show no sequence conservation [18,151]. This lack of conservation suggests that evolutionary pressure has maximized the diversity of 21U RNA sequences, in line with their purported role as a genetic immune system capable of selectively targeting non-self transcripts [79,86,95]. Consistent with their free evolution, the sequence content of a 21U RNA appears to be irrelevant for expression, with the exception of the 5' uridine [150].

21U RNAs are robustly expressed in the germline [26,27] and detected in early embryo, but their levels decline across development [30]. A majority of 21U RNAs show differential expression in male and female germlines [152] [150]. These male and female germline-enriched subpopulations of 21U RNAs direct the production of nonoverlapping 22G RNAs that target distinct transcripts. Targets of male germline-enriched 21U RNAs are depleted for transcripts enriched in spermatogenesis but not oogenesis [33], suggesting selection against evolution of male-expressed 21U RNAs that target male germline genes [152] [150]. Female germline-enriched 21U RNAs do not show depletion of spermatogenesis-enriched transcripts and are paradoxically enriched for oogenesis-enriched transcripts. While the significance of this is uncertain, the distinct enrichment patterns suggest that male and female germline-enriched 21U RNAs are subject to different evolutionary pressures [152] [150]. Female germline-enriched 21U RNAs are preferentially detected in embryo [152] [150], similar to the pattern observed for ERGO-1 class 26G RNAs [20].

21U RNA targets

21U RNAs target transcripts by directing Argonaute PRG-1 to imperfectly complementary sites with up to three, or possibly even four, mismatches [95,97]. 21U RNA targets are depleted of protein-coding transcripts but not pseudogene or transposon transcripts [97]. As 21U RNA targeting triggers local production of WAGO 22G RNAs [27,95], 21U RNA targets are enriched for WAGO 22G RNA targets and depleted for CSR 22G RNA targets [95]. As 21U RNAs are capable of triggering transgenerational silencing that persists for tens of generations in the absence of PRG-1 and additional 21U RNA trigger [79,85], identification of target transcripts is challenging. Targets can be approximated as those transcripts that show robust 21U RNA-dependent 22G RNA accumulation or those that show upregulation upon loss of PRG-1, although the latter are more likely to report a specific subset of targets whose silencing is more heavily weighted toward cytoplasmic than nuclear RNAi. Possibly, the rapid evolution of 21U RNA sequences and the profoundly stable silencing exerted by

21U RNAs renders definition of a specific target set irrelevant, as many previously silenced targets have likely been eroded through genetic drift. So may be the eventual fate of sequences not protected from 21U RNA targeting by a self-recognition system such as CSR 22G RNAs are proposed to represent [79,86].

21U RNA biogenesis

21U RNAs are not transcribed by RdRP complexes, nor does their biogenesis appear to require DCR-1 or any of the factors directly involved in endo-siRNA biogenesis [26,27]. Rather, 21U RNAs are encoded as independent Pol II transcriptional units [150] [149]. In *C. elegans* and other rhabditids, 21U RNA loci show a common upstream sequence arrangement of two motifs, the large and small motifs, that together drive the regulated expression of 21U RNAs in the germline [18,149,153] [150]. In *C. elegans* and *C. briggsae*, 21U RNA loci distribution shows a pattern of enrichment at the start and end of full-length DNA transposon loci in sense and antisense orientations, respectively, suggesting that 21U RNA upstream motifs may serve as traps for triggering 21U RNA biogenesis against recent integrations [97].

The 21U RNA large motif and Forkhead family transcription factors: The large motif, occupying ~60-25 bp upstream of the 21U RNA locus, consists of A/T-richness and a central 8 bp core motif of consensus sequence CTGTTTCA [18]. The A/T-richness of this region repels nucleosomes, resulting in nucleosome depletion upstream of 21U RNA locus; accordingly, the genomic 21U RNA clusters of chromosome IV show general depletion of nucleosomes [153]. The core motif is conserved in other rhabditids, but more divergent in *Pristionchus pacificus*; however, the spacing is conserved, suggesting mechanistic conservation [151] [152]. The core motif is required for expression of the downstream 21U RNA, as its deletion or mutation results in loss of the associated 21U RNA [153] [150]. A subset of Forkhead family transcription factors, including at least FKH-3, FKH-4, FKH-5, and UNC-130, binds this core motif to promote 21U RNA expression: 21U RNA upstream regions immunopurify with UNC-130::GFP,

and recombinant Forkhead proteins preferentially bind dsDNA probes containing the core motif octamer [153]. An *unc-130* mutant shows ~50% depletion of individual 21U RNA species; similar depletion is observed upon partial, simultaneous RNAi knockdown of *fkh-3/4/5* [153]. It is not yet understood whether these Forkhead proteins are semiredundantly required for expression of 21U RNAs or individually contribute to expression. Consistent with the core motif acting as a promoter element for independent transcription of 21U RNAs, the spacing between core motifs and 21U RNA loci affects 21U RNA expression, as does the sequence content. Plotting 21U RNA abundances by spacer length reveals a strong correlation, with spacer lengths of 37-40 nt showing the highest abundances and tailing down to either side [18] [150]. The tolerance in the spacer length affords expression of multiple, genomically miniclustered 21U RNAs in association with the same core motif [154] [150]. Core motifs that more closely match the consensus sequence, and in particular those that contain the central GTTTC, are correlated with higher abundance 21U RNAs and are more likely to be associated with 21U RNA miniclusters than solitary 21U RNAs [26] [150]. Miniclustered 21U RNA loci necessarily lie at different genomic distances from the core motif, and their respective abundances recapitulate the correlation between optimal spacer length and expression level [150]. The sequence content of the core motif also influences germline enrichment. Core motifs whose first nucleotide is cytidine are preferentially enriched in male germline, whereas female germline-enriched 21U RNAs show no preference at this motif position and in general show core motifs with poorer consensus match [150]. A simple explanation for differential germline expression of 21U RNAs according to their motifs would be selective binding by Forkhead transcription factors; however, neither UNC-130 nor FKH-3/4/5 appears to be selectively required for 21U RNAs expressed in male germline or showing a cytidine-initial motif (A.C.B, M.A.F., and J.K.K., unpublished data), suggesting that other factors confer germline enrichment.

The 21U RNA small motif and precursor: The small motif shows a YRNT consensus sequence with the thymidine situated at position 1 of the 21U RNA locus and the YR pair at 3 and 2 nt upstream, resembling the initiator element required for Pol II transcription initiation [18,149]. Global profiling of endogenous capped small RNAs

(csRNAs) reveals that the R position of the small motif corresponds to the first nucleotide of the 21U RNA precursor [149]. These csRNA precursors show blocked 5' ends consistent with Pol II-mediated capping and an average length of 26 nt, corresponding to the m⁷GpppAmNm cap-2 structure upstream and ~3 nt downstream that are removed during 21U RNA maturation [149]. The abundances of csRNA precursors correlate significantly with abundances of their cognate 21U RNAs, supporting their status as precursors [149]. A very small proportion of 21U RNAs show evidence of longer, >70 nt capped sequences [153]; however, these are unlikely to represent csRNA/21U RNA precursors, as their abundances do not correlate with abundances of cognate 21U RNAs [149]. The importance of the YR consensus match is indicated by the respective abundances of 21U RNAs encoded at loci staggered by 1 nt. Only one 21U RNA of each pair can show a YRNT motif; those that do are tenfold more abundant than their non-YRNT sister species [149].

21U RNA precursor processing and Argonaute loading

In association with an upstream large motif or at a TSS, the YR pair of the YRNT small motif appears sufficient to direct csRNA expression, whereas the thymidine is required for 21U RNA accumulation. The accumulation of csRNAs, but not mature 21U RNAs, associated with YNRV small motifs suggests that the 5' uridine of 21U RNAs is dispensable for transcription but plays a role in mature 21U RNA stabilization or processing [149]. In support of this, mutation of the thymidine to G or A abrogates accumulation of a cognate 21G or 21A species [150]. Furthermore, comparison of other *C. elegans* isolates with YRNT motifs where the N2 isolate shows YRNV reveals that substitution for a genomic thymidine results in accumulation the associated 21U RNA [150] [149]. The presence of a uridine at position 3 of the csRNA may be required for removal of the m⁷GpppAmNm cap-2 structure to generate the mature 21U RNA 5' end; however, recombinant PRG-1 shows preferential target cleavage in association with 5' uridine ssRNA [97], suggesting PRG-1 may purify candidate small RNAs processed from csRNAs by selectively binding and stabilizing those with 5' uridines. Identification

of decapping machinery responsible for csRNA 5' end processing will address this question definitively. Trimming of the csRNA 3' end to achieve the mature 21 nt length may occur after PRG-1 binding. While the vast majority of small RNA deep sequencing reads corresponding to 21U RNA loci are 21 nt in length, a small portion show extensions, with a significant bias toward 3' extensions [154]. These 3' extended reads may represent biogenesis intermediates and support the hypothesis that 3' trimming occurs after 5' processing. In silkworm, PIWI Argonaute-bound small RNAs are trimmed after Argonaute loading by a 3' to 5' exonuclease until the length is sufficiently short for anchoring of the 3' end in the Argonaute [155]. While the 26 nt length of 26G RNAs is determined by DCR-1 processing, both 22G RNAs and 21U RNAs are DCR-1-independent [26,27,42,44]. Perhaps a common exonuclease trims 22G RNAs and 21U RNAs.

PRG-1: PRG-1 was first identified as a factor required for germline stem cell proliferation and/or self-renewal: injection of antisense *prg-1* RNA into the germline results in significant shortening of the mitotic proliferative zone, reduced sperm production, and decreased production of offspring [156]. Loss of PRG-1 results in decreased brood size at 20°C and sterility at 25°C [68]. This is likely due to defective sperm activation and fertilization. Spermatocytes are present at 25°C, but mature spermatids are rare; those that do form fail to produce pseudopodia upon activation in vitro [107]. *prg-1* mutant fertility is rescued to wild type levels by mating to wild type males at 20°C, but still shows an average of half the wild type brood size when out-crossed at 25°C [107], suggesting compromised female fertility as well. PRG-1 is one of three PIWI clade Argonautes encoded by the *C. elegans* genome; ERGO-1 is highly divergent, whereas PRG-1 and PRG-2 show 91% amino acid identity and are likely the result of a recent gene duplication [27,68,156]. Despite their nearly identical sequences, only PRG-1 is required for 21U RNA accumulation [26,27]. Expression of *prg-1* mRNA is highly germline-restricted and detected primarily in young adult and gravid adult [26,27]. PRG-1 protein is abundant in young adult, gravid adult, and embryo, suggesting maternal inheritance [26]. Throughout the germline, PRG-1 associates consistently with P granules, with perinuclear localization in mitotic and meiotic zones, loss of expression

during spermatid maturation, and dissociation from the nuclear periphery in mature oocytes [26,107]. Embryonic PRG-1 is also associated with P granules in the P lineage [26]. PRG-1 is required for stabilizing 21U RNAs, and therefore accumulation of 21U RNAs parallels PRG-1 protein expression [26,27,33]. PRG-1 binds 21U RNAs to target transcripts that are often imperfectly complementary by triggering biogenesis in trans of WAGO 22G RNAs [27,95,97]. These 21U RNA-dependent 22G RNAs can trigger transgenerational silencing by binding HRDE-1 and engaging the germline nuclear RNAi pathway [79,85,86]. Loss of PRG-1 results in target derepression due to failure to trigger dependent 22G RNAs; in particular, levels of Tc3 transposase increase and WAGO 22G RNAs targeting Tc3 decrease [24,26,27]. However, transcripts upregulated by loss of PRG-1 may not represent the full set of targets silenced by 21U RNAs, as transgenerational silencing engaged by 21U RNA-dependent WAGO 22G RNAs can be effectively maintained in the absence of 21U RNA trigger [79,85,86,95]. This, in addition to the fact that PRG-1-mediated targeting may not trigger WAGO 22G RNA biogenesis as effectively as other primary small RNA triggers [95,97], likely explains the subtle effect of 21U RNA loss upon global transcript expression [26]. The catalytic triad of PRG-1 is intact, and recombinant PRG-1 is capable of cleaving target RNA when incubated with complementary 5' uridine ssRNA; cleavage is abrogated by mutating a key catalytic residue or by introducing a mismatch at the cleavage site [97]. Nevertheless, introduction of catalytically inactive PRG-1 appears to rescue *prg-1* mutant 21U RNA accumulation, 22G RNA triggering, and target silencing, although the rescue of fertility may be imperfect [95,97]. This suggests that, much like ERGO-1 class 26G RNAs, 21U RNAs silence targets primarily by triggering biogenesis of WAGO 22G RNAs that engage cytoplasmic and nuclear Argonautes.

HENN-1: HENN-1 protein and loss of function phenotypes are described in greater detail above. HENN-1 terminal methylates 21U RNAs universally [70,72,73]. Although HENN-1 does not obviously localize to P granules [70] and stable interaction between PRG-1 and HENN-1 could not be detected [73], 21U RNA methylation presumably occurs after Argonaute loading, in analogy to ERGO-1, and 3' end trimming [70]. HENN-1 is detected in a larger complex of ~100 kD, suggesting that HENN-1 may act as a

complex or associate with another factor, such as perhaps a trimmer exonuclease [73]. HENN-1 is required for robust 21U RNA inheritance into embryo and perdurance in offspring [70,72], but loss of HENN-1 shows a lesser effect on 21U RNA accumulation in the adult germline [70,73]. In contrast, ERGO-1 class 26G RNAs are quite significantly affected [70,72,73]. This discrepancy does not appear to be due to differences in targeting complementarity preferentially triggering 26G RNA degradation, however, as the presence of a perfectly complementary 21U RNA target does not appear to stimulate 21U RNA trimming and tailing upon loss of methylation [73]. Furthermore, both classes of 26G RNAs show roughly equivalent frequency and 21U RNAs the lowest of trimming and tailing among major small RNA classes in a wild type background [72], suggesting that terminal methylation in *C. elegans* may not simply exist to oppose trimming and tailing. Nevertheless, loss of HENN-1 does result in increased levels of 20 nt reads corresponding to 21U RNAs [73]. The proportion of a particular 21U RNA species that is trimmed is highly correlated across different *henn-1* mutant libraries, and 21U RNAs that show higher trimming frequencies are more likely to be depleted in *henn-1* mutant libraries, indicating that 21U RNA trimming is related to greater instability and varies by species [73]. The presence of HENN-1 appears to enhance 21U RNA-mediated silencing of a reporter transgene, but this may be an indirect consequence of decreased expression of factors involved in the WAGO 22G RNA pathway [73]. Loss of HENN-1 results in extremely modest, if any, effect on levels of Tc3 transposase [70,72,73] and no increase in Tc3 transposition [73].

TSS-associated 21U RNAs

Mapping of csRNAs reveals that a majority are not associated with upstream core CTGTTTCA sequences; rather, many of these map bidirectionally to TSSs of protein-coding and other Pol II transcripts [149]. Like chromosome IV csRNAs associated with large motifs, these TSS-associated csRNAs are also processed into mature 21U RNAs when encoded by YRNT motifs [149]. The abundance of csRNAs at a particular promoter correlates with the abundance of longer capped reads originating from that

promoter, suggesting that these csRNAs are products of Pol II initiation at active promoters and may reflect promoter-proximal pausing [149]. The significance of these TSS-associated 21U RNAs is as yet unknown. Much as Pol II initiation at canonical TSSs produces both csRNAs and longer pre-mRNA transcripts, a small fraction of 21U RNA loci show both csRNAs and corresponding longer reads [149,153]. This suggests that promoter elements and 21U RNA upstream motifs may both promote Pol II association and transcription initiation, but 21U RNA loci generally lack sequence elements or chromatin configuration necessary to promote further transcript elongation. csRNA distribution does reveal differences in 21U RNA and canonical promoters, as bidirectional transcription is not reported to occur at 21U RNA loci. 21U RNAs are required for full fertility [26,27,68,107], so it seems possible that csRNAs mapping to canonical promoters are incidental byproducts of selection for Pol II activity that produces 21U RNA precursors. However, promoter-associated csRNAs are observed in other metazoan genomes [157,158], suggesting rather that 21U RNA production may originally have arisen incidental to this more ancient process of unknown significance.

CONTRIBUTIONS OF THE DISSERTATION

C. elegans germline small RNA regulatory mechanisms are a unifying theme of this dissertation, but the mechanisms investigated in Chapters Two through Four are too diverse for adequate representation in anything but a thorough review of endo-siRNA and piRNA biology. Therefore, this introduction provides a comprehensive review of biogenesis, regulation, and function of the diverse classes of *C. elegans* germline small RNAs. The most significant findings presented in Chapters Two (Billi et al., *PLoS Genet* 2012 [70]) and Three (Billi et al., *PLoS Genet* 2013 [150]) of this dissertation are included in the review where relevant. This is intended to demonstrate how these publications contributed to the current body of knowledge within the larger context of *C. elegans* small RNA literature without placing special emphasis on the research of the author.

Importantly, competing, coordinated, and subsequent studies are also included. Billi et al., *PLoS Genet* 2012 was published in coordination with Montgomery et al., *PLoS Genet* 2012 [72], with Kamminga et al., *PLoS Genet* 2012 [73] published shortly thereafter. Using unique approaches to investigate the functional HEN1 ortholog of *C. elegans*, HENN-1, the three publications draw many distinct yet complementary conclusions. In particular, our study examines the mechanisms governing selective methylation of small RNAs catalyzed by animal orthologs of HEN1. Previous crystal structure and modeling analyses of human Argonaute PAZ domains indicated that the PAZ domain of Piwi clade Argonautes is capable of accommodating, and indeed shows a mild binding preference for, methylated small RNA termini, whereas the Ago clade Argonaute PAZ domain cannot accommodate the bulky terminal methyl group [159]. This observation led the authors to conclude that the divergent Piwi Argonaute PAZ domain targets terminally methylated ssRNAs, while the Ago Argonaute PAZ domain targets dsRNAs with non-methylated 2-nt terminal overhangs. This conclusion suggested that methylation precedes Argonaute binding, and small RNAs are selected for methylation by an unknown mechanism and subsequently sorted into Piwi or Ago Argonautes according to the differing binding preferences. Our study shows that methylation of ERGO-1 class 26G RNAs is lost in the absence of ERGO-1, revealing that Argonaute binding probably occurs prior to small RNA methylation. By establishing this temporal relationship, this study reveals that the Argonautes themselves likely dictate which small RNAs are methylated by permitting or prohibiting HEN1-mediated methylation of associated small RNAs. Our study also reports that 26G RNAs are robustly bound by ERGO-1 in the absence of HENN-1 and thus terminal methylation, further supporting the new model. This mechanistic insight explains selective methylation of 26G RNAs in oocyte but not sperm and identifies differential expression of divergent Argonautes in male and female germlines as a novel mechanism for directing selective methylation, and thus robust inheritance, of small RNAs in female germline. Other results exclusive to Billi et al., *PLoS Genet* 2012 include endogenous ERGO-1 and HENN-1 protein expression patterns in germline and embryo and examination of accumulation and inheritance of numerous and diverse small RNAs across development.

In 2012, a number of studies emerged describing the function of *C. elegans* piRNAs, marking the first significant advance in understanding of *C. elegans* piRNA biology in nearly half a decade; however, no progress had yet been made in understanding *C. elegans* piRNA biogenesis and expression mechanisms. During review of Billi et al., *PLoS Genet* 2013, a competing study (Cecere et al., *Mol Cell* 2012 [153]) was published that identified forkhead family proteins as putative *C. elegans* piRNA transcription factors. The results reported in Cecere et al., *Mol Cell* 2012 could not be reproduced during revisions of our study. This inconsistency was not included in the revised manuscript, as the two studies examined largely distinct questions in *C. elegans* piRNA biology and drew predominantly distinct conclusions. Following final submission of our study, another study was published (Gu et al., *Cell* 2012 [149]) that uses an entirely distinct approach to examine *C. elegans* piRNA biogenesis. The conclusions of this study complement and greatly inform our results, yet our study presents evidence supporting many unique conclusions. Most significantly, the transgenic approach applied in Billi et al., *PLoS Genet* 2013 reveals that *C. elegans* piRNA loci represent tiny, autonomous transcriptional units. Additionally, more than half of *C. elegans* piRNAs were found to show specific male or female germline enrichment, an observation corroborated by Shi et al., *Genome Res* 2013 [152], another contemporary study. These piRNA subsets target distinct sets of genes that show different properties, suggesting that male and female germline-enriched piRNA subpopulations evolve independently in response to unique evolutionary pressures. Quite intriguingly, our study also shows that specific enrichment in male or female germline is heavily influenced by variation at a single position within an 8-nt cis regulatory element that lies upstream of each nematode piRNA. These significant findings address longstanding questions in the field of *C. elegans* piRNA biology and, along with Gu et al., *Cell* 2012, distinguish the 21U RNAs as unique among piRNAs.

REFERENCES

1. Fischer SE (2010) Small RNA-mediated gene silencing pathways in *C. elegans*. *The International Journal of Biochemistry & Cell Biology* 42: 1306-1315.
2. Fire A, Albertson D, Harrison SW, Moerman DG (1991) Production of antisense RNA leads to effective and specific inhibition of gene expression in *C. elegans* muscle. *Development* 113: 503-514.
3. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, et al. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391: 806-811.
4. Zamore PD, Tuschl T, Sharp PA, Bartel DP (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101: 25-33.
5. Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409: 363-366.
6. Sijen T, Fleenor J, Simmer F, Thijssen KL, Parrish S, et al. (2001) On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* 107: 465-476.
7. Tabara H, Sarkissian M, Kelly WG, Fleenor J, Grishok A, et al. (1999) The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* 99: 123-132.
8. Ketting RF, Haverkamp TH, van Luenen HG, Plasterk RH (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.
9. Ketting RF, Plasterk RH (2000) A genetic link between co-suppression and RNA interference in *C. elegans*. *Nature* 404: 296-298.
10. Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, et al. (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.
11. Ketting RF, Fischer SE, Bernstein E, Sijen T, Hannon GJ, et al. (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes & development* 15: 2654-2659.
12. Knight SW, Bass BL (2001) A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* 293: 2269-2271.
13. Cox DN, Chao A, Baker J, Chang L, Qiao D, et al. (1998) A novel class of evolutionarily conserved genes defined by *piwi* are essential for stem cell self-renewal. *Genes & Development* 12: 3715-3727.
14. Smardon A, Spoerke JM, Stacey SC, Klein ME, Mackin N, et al. (2000) EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Current biology* : CB 10: 169-178.
15. Lee R, Feinbaum R (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*.
16. Ambros V, Lee RC, Lavanway A, Williams PT, Jewell D (2003) MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Current biology* : CB 13: 807-818.

17. Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, et al. (2003) The microRNAs of *Caenorhabditis elegans*. *Genes & Development* 17: 991-1008.
18. Ruby JG, Jan C, Player C, Axtell MJ, Lee W, et al. (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* 127: 1193-1207.
19. Conine CC, Batista PJ, Gu W, Claycomb JM, Chaves DA, et al. (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* 107: 3588-3593.
20. Han T, Manoharan AP, Harkins TT, Bouffard P, Fitzpatrick C, et al. (2009) 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* 106: 18674-18679.
21. Pak J, Fire A (2007) Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* 315: 241-244.
22. Sijen T, Steiner FA, Thijssen KL, Plasterk RHA (2007) Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science* 315: 244-247.
23. Vasale JJ, Gu W, Thivierge C, Batista PJ, Claycomb JM, et al. (2010) Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. *Proc Natl Acad Sci U S A* 107: 3582-3587.
24. Gu W, Shirayama M, Conte D, Jr., Vasale J, Batista PJ, et al. (2009) Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. *Mol Cell* 36: 231-244.
25. Claycomb JM, Batista PJ, Pang KM, Gu W, Vasale JJ, et al. (2009) The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. *Cell* 139: 123-134.
26. Batista PJ, Ruby JG, Claycomb JM, Chiang R, Fahlgren N, 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.
27. Das PP, Bagijn MP, Goldstein LD, Woolford JR, Lehrbach NJ, 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.
28. Ambros V, Lee RC, Lavanway A, Williams PT, Jewell D (2003) MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr Biol* 13: 807-818.
29. Asikainen S, Heikkinen L, Wong G, Storvik M (2008) Functional characterization of endogenous siRNA target genes in *Caenorhabditis elegans*. *BMC genomics* 9: 270.
30. Stoeckius M, Maaskola J, Colombo T, Rahn HP, Friedlander MR, et al. (2009) Large-scale sorting of *C. elegans* embryos reveals the dynamics of small RNA expression. *Nature methods* 6: 745-751.
31. Gent JI, Lamm AT, Pavelec DM, Maniar JM, Parameswaran P, et al. (2010) Distinct phases of siRNA synthesis in an endogenous RNAi pathway in *C. elegans* soma. *Molecular cell* 37: 679-689.

32. Gent JI, Schvarzstein M, Villeneuve AM, Gu SG, Jantsch V, et al. (2009) A *Caenorhabditis elegans* RNA-directed RNA polymerase in sperm development and endogenous RNA interference. *Genetics* 183: 1297-1314.
33. Reinke V, Gil IS, Ward S, Kazmer K (2004) Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. *Development* 131: 311-323.
34. Asikainen S, Storvik M, Lakso M, Wong G (2007) Whole genome microarray analysis of *C. elegans* rrf-3 and eri-1 mutants. *FEBS Letters* 581: 5050-5054.
35. Fischer SE, Montgomery TA, Zhang C, Fahlgren N, Breen PC, et al. (2011) The ERI-6/7 helicase acts at the first stage of an siRNA amplification pathway that targets recent gene duplications. *PLoS Genetics* 7: e1002369.
36. Schrimpf SP, Weiss M, Reiter L, Ahrens CH, Jovanovic M, et al. (2009) Comparative functional analysis of the *Caenorhabditis elegans* and *Drosophila melanogaster* proteomes. *PLoS biology* 7: e48.
37. Duchaine TF, Wohlschlegel JA, Kennedy S, Bei Y, Conte D, Jr., et al. (2006) Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. *Cell* 124: 343-354.
38. Thivierge C, Makil N, Flamand M, Vasale JJ, Mello CC, et al. (2012) Tudor domain ERI-5 tethers an RNA-dependent RNA polymerase to DCR-1 to potentiate endo-RNAi. *Nature structural & molecular biology* 19: 90-97.
39. Tabara H, Yigit E, Siomi H, Mello CC (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.
40. Smardon A, Spoerke JM, Stacey SC, Klein ME, Mackin N, et al. (2000) EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Current biology* : CB 10: 169-178.
41. Lee RC, Hammell CM, Ambros V (2006) Interacting endogenous and exogenous RNAi pathways in *Caenorhabditis elegans*. *RNA* 12: 589-597.
42. Pavelec DM, Lachowiec J, Duchaine TF, Smith HE, Kennedy S (2009) Requirement for the ERI/DICER complex in endogenous RNA interference and sperm development in *Caenorhabditis elegans*. *Genetics* 183: 1283-1295.
43. Hunt-Newbury R, Viveiros R, Johnsen R, Mah A, Anastas D, et al. (2007) High-throughput in vivo analysis of gene expression in *Caenorhabditis elegans*. *PLoS biology* 5: e237.
44. Aoki K, Moriguchi H, Yoshioka T, Okawa K, Tabara H (2007) In vitro analyses of the production and activity of secondary small interfering RNAs in *C. elegans*. *EMBO J* 26: 5007-5019.
45. Makeyev EV, Bamford DH (2002) Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Molecular Cell* 10: 1417-1427.
46. Sijen T, Fleenor J, Simmer F, Thijssen KL, Parrish S, et al. (2001) On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* 107: 465-476.
47. Simmer F, Tijsterman M, Parrish S, Koushika SP, Nonet ML, et al. (2002) Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. *Curr Biol* 12: 1317-1319.

48. Nakamura M, Ando R, Nakazawa T, Yudazono T, Tsutsumi N, et al. (2007) Dicer-related drh-3 gene functions in germ-line development by maintenance of chromosomal integrity in *Caenorhabditis elegans*. *Genes to cells : devoted to molecular & cellular mechanisms* 12: 997-1010.
49. Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409: 363-366.
50. Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, et al. (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.
51. Knight SW, Bass BL (2001) A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science* 293: 2269-2271.
52. Welker NC, Pavelec DM, Nix DA, Duchaine TF, Kennedy S, et al. (2010) Dicer's helicase domain is required for accumulation of some, but not all, *C. elegans* endogenous siRNAs. *RNA* 16: 893-903.
53. Welker NC, Maity TS, Ye X, Aruscavage PJ, Krauchuk AA, et al. (2011) Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode. *Molecular Cell* 41: 589-599.
54. Kennedy S, Wang D, Ruvkun G (2004) A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* 427: 645-649.
55. Gabel HW, Ruvkun G (2008) The exonuclease ERI-1 has a conserved dual role in 5.8S rRNA processing and RNAi. *Nature structural & molecular biology* 15: 531-533.
56. Parker GS, Eckert DM, Bass BL (2006) RDE-4 preferentially binds long dsRNA and its dimerization is necessary for cleavage of dsRNA to siRNA. *RNA* 12: 807-818.
57. Tabara H, Sarkissian M, Kelly WG, Fleenor J, Grishok A, et al. (1999) The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* 99: 123-132.
58. Parker GS, Maity TS, Bass BL (2008) dsRNA binding properties of RDE-4 and TRBP reflect their distinct roles in RNAi. *Journal of molecular biology* 384: 967-979.
59. Blanchard D, Parameswaran P, Lopez-Molina J, Gent J, Saynuk JF, et al. (2011) On the nature of in vivo requirements for *rde-4* in RNAi and developmental pathways in *C. elegans*. *RNA biology* 8: 458-467.
60. Parrish S, Fire A (2001) Distinct roles for RDE-1 and RDE-4 during RNA interference in *Caenorhabditis elegans*. *RNA* 7: 1397-1402.
61. Welker NC, Habig JW, Bass BL (2007) Genes misregulated in *C. elegans* deficient in Dicer, RDE-4, or RDE-1 are enriched for innate immunity genes. *RNA* 13: 1090-1102.
62. Zhang C, Montgomery TA, Gabel HW, Fischer SE, Phillips CM, et al. (2011) *mut-16* and other mutator class genes modulate 22G and 26G siRNA pathways in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America* 108: 1201-1208.
63. Zhuang JJ, Hunter CP (2012) RNA interference in *Caenorhabditis elegans*: uptake, mechanism, and regulation. *Parasitology* 139: 560-573.

64. Fischer SE, Butler MD, Pan Q, Ruvkun G (2008) Trans-splicing in *C. elegans* generates the negative RNAi regulator ERI-6/7. *Nature* 455: 491-496.
65. Vastenhouw NL, Fischer SE, Robert VJ, Thijssen KL, Fraser AG, et al. (2003) A genome-wide screen identifies 27 genes involved in transposon silencing in *C. elegans*. *Current biology : CB* 13: 1311-1316.
66. Phillips CM, Montgomery TA, Breen PC, Ruvkun G (2012) MUT-16 promotes formation of perinuclear mutator foci required for RNA silencing in the *C. elegans* germline. *Genes & development* 26: 1433-1444.
67. Warf MB, Shepherd BA, Johnson WE, Bass BL (2012) Effects of ADARs on small RNA processing pathways in *C. elegans*. *Genome research* 22: 1488-1498.
68. Yigit E, Batista PJ, Bei Y, Pang KM, Chen CC, et al. (2006) Analysis of the *C. elegans* Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* 127: 747-757.
69. Steiner FA, Okihara KL, Hoogstrate SW, Sijen T, Ketting RF (2009) RDE-1 slicer activity is required only for passenger-strand cleavage during RNAi in *Caenorhabditis elegans*. *Nature structural & molecular biology* 16: 207-211.
70. Billi AC, Alessi AF, Khivansara V, Han T, Freeberg M, et al. (2012) The *Caenorhabditis elegans* HEN1 ortholog, HENN-1, methylates and stabilizes select subclasses of germline small RNAs. *PLoS genetics* 8: e1002617.
71. Park W, Li J, Song R, Messing J, Chen X (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr Biol* 12: 1484-1495.
72. Montgomery TA, Rim YS, Zhang C, Downen RH, Phillips CM, et al. (2012) PIWI associated siRNAs and piRNAs specifically require the *Caenorhabditis elegans* HEN1 ortholog henn-1. *PLoS genetics* 8: e1002616.
73. Kamminga LM, van Wolfswinkel JC, Luteijn MJ, Kaaij LJ, Bagijn MP, et al. (2012) Differential Impact of the HEN1 Homolog HENN-1 on 21U and 26G RNAs in the Germline of *Caenorhabditis elegans*. *PLoS genetics* 8: e1002702.
74. Zhuang JJ, Hunter CP (2011) Tissue specificity of *Caenorhabditis elegans* enhanced RNA interference mutants. *Genetics* 188: 235-237.
75. Ketting RF, Haverkamp TH, van Luenen HG, Plasterk RH (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.
76. Tijsterman M, Okihara KL, Thijssen K, Plasterk RH (2002) PPW-1, a PAZ/PIWI protein required for efficient germline RNAi, is defective in a natural isolate of *C. elegans*. *Current biology : CB* 12: 1535-1540.
77. Corrêa RL, Steiner FA, Berezikov E, Ketting RF (2010) MicroRNA-Directed siRNA Biogenesis in *Caenorhabditis elegans*. *PLoS Genet* 6: e1000903.
78. Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, et al. (2003) The microRNAs of *Caenorhabditis elegans*. *Genes & development* 17: 991-1008.
79. Ashe A, Sapetschnig A, Weick EM, Mitchell J, Bagijn MP, et al. (2012) piRNAs Can Trigger a Multigenerational Epigenetic Memory in the Germline of *C. elegans*. *Cell* 150: 88-99.
80. Buckley BA, Burkhart KB, Gu SG, Spracklin G, Kershner A, et al. (2012) A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. *Nature*.

81. Burkhart KB, Guang S, Buckley BA, Wong L, Bochner AF, et al. (2011) A pre-mRNA-associating factor links endogenous siRNAs to chromatin regulation. *PLoS genetics* 7: e1002249.
82. Burton NO, Burkhart KB, Kennedy S (2011) Nuclear RNAi maintains heritable gene silencing in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America* 108: 19683-19688.
83. Gu SG, Pak J, Guang S, Maniar JM, Kennedy S, et al. (2012) Amplification of siRNA in *Caenorhabditis elegans* generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. *Nature genetics* 44: 157-164.
84. Guang S, Bochner AF, Pavelec DM, Burkhart KB, Harding S, et al. (2008) An Argonaute transports siRNAs from the cytoplasm to the nucleus. *Science* 321: 537-541.
85. Luteijn MJ, van Bergeijk P, Kaaij LJ, Almeida MV, Roovers EF, et al. (2012) Extremely stable Piwi-induced gene silencing in *Caenorhabditis elegans*. *The EMBO journal* 31: 3422-3430.
86. Shirayama M, Seth M, Lee HC, Gu W, Ishidate T, et al. (2012) piRNAs Initiate an Epigenetic Memory of Nonself RNA in the *C. elegans* Germline. *Cell* 150: 65-77.
87. Vought VE, Ohmachi M, Lee MH, Maine EM (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.
88. Maine EM, Hauth J, Ratliff T, Vought VE, She X, et al. (2005) EGO-1, a putative RNA-dependent RNA polymerase, is required for heterochromatin assembly on unpaired dna during *C. elegans* meiosis. *Current biology : CB* 15: 1972-1978.
89. Mangone M, Manoharan AP, Thierry-Mieg D, Thierry-Mieg J, Han T, et al. (2010) The landscape of *C. elegans* 3'UTRs. *Science* 329: 432-435.
90. Robert VJ, Sijen T, van Wolfswinkel J, Plasterk RH (2005) Chromatin and RNAi factors protect the *C. elegans* germline against repetitive sequences. *Genes Dev* 19: 782-787.
91. Kim JK, Gabel HW, Kamath RS, Tewari M, Pasquinelli A, et al. (2005) Functional genomic analysis of RNA interference in *C. elegans*. *Science* 308: 1164-1167.
92. Rocheleau CE, Cullison K, Huang K, Bernstein Y, Spilker AC, et al. (2008) The *Caenorhabditis elegans* ekl (enhancer of ksr-1 lethality) genes include putative components of a germline small RNA pathway. *Genetics* 178: 1431-1443.
93. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, et al. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391: 806-811.
94. Montgomery MK, Fire A (1998) Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression. *Trends in genetics : TIG* 14: 255-258.
95. Lee HC, Gu W, Shirayama M, Youngman E, Conte D, Jr., et al. (2012) *C. elegans* piRNAs Mediate the Genome-wide Surveillance of Germline Transcripts. *Cell* 150: 78-87.
96. Alder MN, Dames S, Gaudet J, Mango SE (2003) Gene silencing in *Caenorhabditis elegans* by transitive RNA interference. *RNA* 9: 25-32.

97. Bagijn MP, Goldstein LD, Sapetschnig A, Weick EM, Bouasker S, et al. (2012) Function, targets, and evolution of *Caenorhabditis elegans* piRNAs. *Science* 337: 574-578.
98. Tijsterman M, Ketting RF, Okihara KL, Sijen T, Plasterk RH (2002) RNA helicase MUT-14-dependent gene silencing triggered in *C. elegans* by short antisense RNAs. *Science* 295: 694-697.
99. Steiner FA, Hoogstrate SW, Okihara KL, Thijssen KL, Ketting RF, et al. (2007) Structural features of small RNA precursors determine Argonaute loading in *Caenorhabditis elegans*. *Nature structural & molecular biology* 14: 927-933.
100. Collins J, Saari B, Anderson P (1987) Activation of a transposable element in the germ line but not the soma of *Caenorhabditis elegans*. *Nature* 328: 726-728.
101. Grishok A, Sinskey JL, Sharp PA (2005) Transcriptional silencing of a transgene by RNAi in the soma of *C. elegans*. *Genes & development* 19: 683-696.
102. Chen CC, Simard MJ, Tabara H, Brownell DR, McCollough JA, et al. (2005) A member of the polymerase beta nucleotidyltransferase superfamily is required for RNA interference in *C. elegans*. *Current biology* : CB 15: 378-383.
103. Tops BB, Tabara H, Sijen T, Simmer F, Mello CC, et al. (2005) RDE-2 interacts with MUT-7 to mediate RNA interference in *Caenorhabditis elegans*. *Nucleic Acids Research* 33: 347-355.
104. Kwak JE, Wickens M (2007) A family of poly(U) polymerases. *RNA* 13: 860-867.
105. Emmons SW, Yesner L (1984) High-frequency excision of transposable element Tc 1 in the nematode *Caenorhabditis elegans* is limited to somatic cells. *Cell* 36: 599-605.
106. Pitt JN, Schisa JA, Priess JR (2000) P granules in the germ cells of *Caenorhabditis elegans* adults are associated with clusters of nuclear pores and contain RNA. *Developmental biology* 219: 315-333.
107. Wang G, Reinke V (2008) A *C. elegans* Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. *Current biology* : CB 18: 861-867.
108. Sheth U, Pitt J, Dennis S, Priess JR (2010) Perinuclear P granules are the principal sites of mRNA export in adult *C. elegans* germ cells. *Development* 137: 1305-1314.
109. Zhang C, Montgomery TA, Fischer SE, Garcia SM, Riedel CG, et al. (2012) The *Caenorhabditis elegans* RDE-10/RDE-11 complex regulates RNAi by promoting secondary siRNA amplification. *Current biology* : CB 22: 881-890.
110. Yang H, Zhang Y, Vallandingham J, Li H, Florens L, et al. (2012) The RDE-10/RDE-11 complex triggers RNAi-induced mRNA degradation by association with target mRNA in *C. elegans*. *Genes & development* 26: 846-856.
111. Tijsterman M, May RC, Simmer F, Okihara KL, Plasterk RH (2004) Genes required for systemic RNA interference in *Caenorhabditis elegans*. *Current biology* : CB 14: 111-116.
112. Sundaram P, Echaliier B, Han W, Hull D, Timmons L (2006) ATP-binding cassette transporters are required for efficient RNA interference in *Caenorhabditis elegans*. *Molecular biology of the cell* 17: 3678-3688.
113. Han W, Sundaram P, Kenjale H, Grantham J, Timmons L (2008) The *Caenorhabditis elegans* *rsd-2* and *rsd-6* genes are required for chromosome

- functions during exposure to unfavorable environments. *Genetics* 178: 1875-1893.
114. Montgomery MK, Xu S, Fire A (1998) RNA as a target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America* 95: 15502-15507.
 115. Kelly WG, Fire A (1998) Chromatin silencing and the maintenance of a functional germline in *Caenorhabditis elegans*. *Development* 125: 2451-2456.
 116. Kelly WG, Schaner CE, Dernburg AF, Lee MH, Kim SK, et al. (2002) X-chromosome silencing in the germline of *C. elegans*. *Development* 129: 479-492.
 117. Grishok A, Tabara H, Mello CC (2000) Genetic requirements for inheritance of RNAi in *C. elegans*. *Science* 287: 2494-2497.
 118. Guang S, Bochner AF, Burkhart KB, Burton N, Pavelec DM, et al. (2010) Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. *Nature* 465: 1097-1101.
 119. Moore MJ, Proudfoot NJ (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* 136: 688-700.
 120. Boshier JM, Dufourcq P, Sookhareea S, Labouesse M (1999) RNA interference can target pre-mRNA: consequences for gene expression in a *Caenorhabditis elegans* operon. *Genetics* 153: 1245-1256.
 121. She X, Xu X, Fedotov A, Kelly WG, Maine EM (2009) Regulation of heterochromatin assembly on unpaired chromosomes during *Caenorhabditis elegans* meiosis by components of a small RNA-mediated pathway. *PLoS Genetics* 5: e1000624.
 122. Elbashir SM, Lendeckel W, Tuschl T (2001) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes & development* 15: 188-200.
 123. Schwarz DS, Tomari Y, Zamore PD (2004) The RNA-induced silencing complex is a Mg²⁺-dependent endonuclease. *Current biology* : CB 14: 787-791.
 124. Qiao L, Lissemore JL, Shu P, Smardon A, Gelber MB, et al. (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.
 125. Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M, et al. (2000) Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* 408: 325-330.
 126. Maeda I, Kohara Y, Yamamoto M, Sugimoto A (2001) Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Current biology* : CB 11: 171-176.
 127. Zipperlen P, Fraser AG, Kamath RS, Martinez-Campos M, Ahringer J (2001) Roles for 147 embryonic lethal genes on *C.elegans* chromosome I identified by RNA interference and video microscopy. *The EMBO Journal* 20: 3984-3992.
 128. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, et al. (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421: 231-237.
 129. Simmer F, Moorman C, van der Linden AM, Kuijk E, van den Berghe PV, et al. (2003) Genome-wide RNAi of *C. elegans* using the hypersensitive *rff-3* strain reveals novel gene functions. *PLoS biology* 1: E12.

130. Rual JF, Ceron J, Koreth J, Hao T, Nicot AS, et al. (2004) Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library. *Genome research* 14: 2162-2168.
131. Sonnichsen B, Koski LB, Walsh A, Marschall P, Neumann B, et al. (2005) Full-genome RNAi profiling of early embryogenesis in *Caenorhabditis elegans*. *Nature* 434: 462-469.
132. Fernandez AG, Gunsalus KC, Huang J, Chuang LS, Ying N, et al. (2005) New genes with roles in the *C. elegans* embryo revealed using RNAi of ovary-enriched ORFeome clones. *Genome research* 15: 250-259.
133. Lehner B, Crombie C, Tischler J, Fortunato A, Fraser AG (2006) Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nature Genetics* 38: 896-903.
134. Vasu SK, Forbes DJ (2001) Nuclear pores and nuclear assembly. *Current opinion in cell biology* 13: 363-375.
135. Chu DS, Liu H, Nix P, Wu TF, Ralston EJ, et al. (2006) Sperm chromatin proteomics identifies evolutionarily conserved fertility factors. *Nature* 443: 101-105.
136. Bean CJ, Schaner CE, Kelly WG (2004) Meiotic pairing and imprinted X chromatin assembly in *Caenorhabditis elegans*. *Nature Genetics* 36: 100-105.
137. Gassmann R, Rechtsteiner A, Yuen KW, Muroyama A, Egelhofer T, et al. (2012) An inverse relationship to germline transcription defines centromeric chromatin in *C. elegans*. *Nature* 484: 534-537.
138. Rechtsteiner A, Ercan S, Takasaki T, Phippen TM, Egelhofer TA, et al. (2010) The histone H3K36 methyltransferase MES-4 acts epigenetically to transmit the memory of germline gene expression to progeny. *PLoS Genetics* 6.
139. Marzluff WF, Wagner EJ, Duronio RJ (2008) Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nature reviews Genetics* 9: 843-854.
140. Davila Lopez M, Samuelsson T (2008) Early evolution of histone mRNA 3' end processing. *RNA* 14: 1-10.
141. Avgousti DC, Palani S, Sherman Y, Grishok A (2012) CSR-1 RNAi pathway positively regulates histone expression in *C. elegans*. *The EMBO Journal* 31: 3821-3832.
142. Ideue T, Hino K, Kitao S, Yokoi T, Hirose T (2009) Efficient oligonucleotide-mediated degradation of nuclear noncoding RNAs in mammalian cultured cells. *RNA* 15: 1578-1587.
143. Wagner EJ, Berkow A, Marzluff WF (2005) Expression of an RNAi-resistant SLBP restores proper S-phase progression. *Biochemical Society transactions* 33: 471-473.
144. Kodama Y, Rothman JH, Sugimoto A, Yamamoto M (2002) The stem-loop binding protein CDL-1 is required for chromosome condensation, progression of cell death and morphogenesis in *Caenorhabditis elegans*. *Development* 129: 187-196.
145. Pettitt J, Crombie C, Schumperli D, Muller B (2002) The *Caenorhabditis elegans* histone hairpin-binding protein is required for core histone gene expression and

- is essential for embryonic and postembryonic cell division. *Journal of cell science* 115: 857-866.
146. McGhee JD (2007) The *C. elegans* intestine. *WormBook : the online review of C elegans biology*: 1-36.
 147. van Wolfswinkel JC, Claycomb JM, Batista PJ, Mello CC, Berezikov E, et al. (2009) CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. *Cell* 139: 135-148.
 148. Heo I, Joo C, Cho J, Ha M, Han J, et al. (2008) Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Molecular Cell* 32: 276-284.
 149. Gu W, Lee HC, Chaves D, Youngman EM, Pazour GJ, et al. (2012) CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as *C. elegans* piRNA precursors. *Cell* 151: 1488-1500.
 150. Billi AC, Freeberg MA, Day AM, Chun SY, Khivansara V, et al. (2013) A Conserved Upstream Motif Orchestrates Autonomous, Germline-Enriched Expression of *Caenorhabditis elegans* piRNAs. *PLoS Genet* 9: e1003392.
 151. de Wit E, Linsen SE, Cuppen E, Berezikov E (2009) Repertoire and evolution of miRNA genes in four divergent nematode species. *Genome research* 19: 2064-2074.
 152. Shi Z, Montgomery TA, Qi Y, Ruvkun G (2013) High-throughput sequencing reveals extraordinary fluidity of miRNA, piRNA, and siRNA pathways in nematodes. *Genome research*.
 153. Cecere G, Zheng GX, Mansisidor AR, Klymko KE, Grishok A (2012) Promoters recognized by forkhead proteins exist for individual 21U-RNAs. *Molecular Cell* 47: 734-745.
 154. Kato M, de Lencastre A, Pincus Z, Slack FJ (2009) Dynamic expression of small non-coding RNAs, including novel microRNAs and piRNAs/21U-RNAs, during *Caenorhabditis elegans* development. *Genome Biology* 10: R54.
 155. Kawaoka S, Izumi N, Katsuma S, Tomari Y (2011) 3' end formation of PIWI-interacting RNAs in vitro. *Molecular Cell* 43: 1015-1022.
 156. Cox DN, Chao A, Baker J, Chang L, Qiao D, et al. (1998) A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes & development* 12: 3715-3727.
 157. Nechaev S, Fargo DC, dos Santos G, Liu L, Gao Y, et al. (2010) Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* 327: 335-338.
 158. Rasmussen EB, Lis JT (1995) Short transcripts of the ternary complex provide insight into RNA polymerase II elongational pausing. *Journal of molecular biology* 252: 522-535.
 159. Tian Y, Simanshu DK, Ma JB, Patel DJ (2011) Structural basis for piRNA 2'-O-methylated 3'-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. *Proceedings of the National Academy of Sciences of the United States of America* 108: 903-910.

CHAPTER TWO: The *Caenorhabditis elegans* HEN1 ortholog, HENN-1, methylates and stabilizes select subclasses of germline small RNAs

AUTHORS: Billi AC, Alessi AF, Khivansara V, Han T, Freeberg MA, Mitani S, Kim JK

AUTHOR CONTRIBUTIONS: The initial discovery of differential methylation of 26G RNAs in male and female germlines is credited to TH. ACB and JKK conceived and designed all subsequent experiments. ACB performed the experiments and analyses in Figures 2.1, 2.2, 2.3, 2.4A, 2.5, 2.6, 2.7A,B, 2.S1, 2.S2, 2.S3, 2.S4, 2.S5, 2.S6, 2.S7, 2.S8, 2.S9, 2.S10, 2.S11, 2.S12, 2.S13A-B, 2.S14. ACB generated the transgenic strains analyzed in Figures 2.7C,D,F and 2.S13C-E. AFA performed the experiments in Figures 2.7C-F and 2.S13C-F. VK performed the experiments in Figure 2.4B,C. MF identified species for design of Taqman small RNA assays. SM isolated the *henn-1* deletion strain. ACB wrote the manuscript. ACB and JKK edited the manuscript.

CITATION: Billi AC, Alessi AF, Khivansara V, Han T, Freeberg M, Mitani S, Kim JK. The *Caenorhabditis elegans* HEN1 ortholog, HENN-1, methylates and stabilizes select subclasses of germline small RNAs. *PLoS Genetics* 8: e1002617 (2012).

ABSTRACT: Small RNAs regulate diverse biological processes by directing effector proteins called Argonautes to silence complementary mRNAs. Maturation of some classes of small RNAs involves terminal 2'-O-methylation to prevent degradation. This modification is catalyzed by members of the conserved HEN1 RNA methyltransferase family. In animals, Piwi-interacting RNAs (piRNAs) and some endogenous and exogenous small interfering RNAs (siRNAs) are methylated, whereas microRNAs are not. However, the mechanisms that determine animal HEN1 substrate specificity have yet to be fully resolved. In *C. elegans*, a HEN1 ortholog has not been studied, but there

is evidence for methylation of piRNAs and some endogenous siRNAs. Here, we report that the worm HEN1 ortholog, HENN-1 (*HEN* of Nematode), is required for methylation of *C. elegans* small RNAs. Our results indicate that piRNAs are universally methylated by HENN-1. In contrast, 26G RNAs, a class of primary endogenous siRNAs, are methylated in female germline and embryo, but not in male germline. Intriguingly, the methylation pattern of 26G RNAs correlates with the expression of distinct male and female germline Argonautes. Moreover, loss of the female germline Argonaute results in loss of 26G RNA methylation altogether. These findings support a model wherein methylation status of a metazoan small RNA is dictated by the Argonaute to which it binds. Loss of *henn-1* results in phenotypes that reflect destabilization of substrate small RNAs: dysregulation of target mRNAs, impaired fertility, and enhanced somatic RNAi. Additionally, the *henn-1* mutant shows a weakened response to RNAi knockdown of germline genes, suggesting that HENN-1 may also function in canonical RNAi. Together, our results indicate a broad role for HENN-1 in both endogenous and exogenous gene silencing pathways and provide further insight into the mechanisms of HEN1 substrate discrimination and the diversity within the Argonaute family.

AUTHOR SUMMARY: Small RNAs serve as sentinels of the genome, policing activity of selfish genetic elements, modulating chromatin dynamics, and fine-tuning gene expression. Nowhere is this more important than in the germline, where endogenous small interfering RNAs (endo-siRNAs) and Piwi-interacting RNAs (piRNAs) promote formation of functional gametes and ensure viable, fertile progeny. Small RNAs act primarily by associating with effector proteins called Argonautes to direct repression of complementary mRNAs. HEN1 methyltransferases, which methylate small RNAs, play a critical role in accumulation of these silencing signals. In this study, we report that the 26G RNAs, a class of *C. elegans* endo-siRNAs, are differentially methylated in male and female germlines. 26G RNAs derived from the two germlines are virtually indistinguishable, except that they associate with evolutionarily divergent Argonautes. Our data support a model wherein the methylation status and, consequently, stability of a small RNA is determined by the associated Argonaute. Therefore, selective

expression of Argonautes that permit or prohibit methylation may represent a new mechanism for regulating small RNA turnover. As we observe this phenomenon in the germline, it may be particularly pertinent for directing inheritance of small RNAs, which can carry information not encoded in progeny DNA that is essential for continued transgenerational genome surveillance.

INTRODUCTION

Argonautes are an evolutionarily conserved family of proteins implicated in diverse cellular processes. They function as effector proteins in the RNA-induced silencing complex (RISC), a gene regulatory complex that binds small, non-coding RNAs to target its silencing effects. Small RNAs are broadly segregated into groups that differ in their mechanisms of biogenesis and silencing, as well as in the subsets of Argonaute effectors that bind them. The microRNAs (miRNAs) are highly conserved small RNAs processed from endogenous hairpin precursors that regulate networks of mRNAs primarily through post-transcriptional repression [1,2]. The piRNAs, so named for the Piwi Argonautes that bind them, function predominantly in maintenance of germline integrity, often through repression of repetitive transposable elements. The small interfering RNAs comprise a more heterogeneous group that includes small RNAs derived from cleavage of exogenous double-stranded RNA (exo-siRNAs) or generated endogenously (endo-siRNAs).

Chemical modification has emerged as an important theme in regulation of small RNA function (for a review, see Kim et al., 2010 [3]). Internal editing has been found to occur in select miRNA precursors through the action of ADAR (*adenosine deaminase acting on RNA*) enzymes, with consequences for miRNA processing efficiency, stability, and targeting [4-8]. Some siRNAs generated in fly and mouse also show evidence of editing by ADARs [9,10], but the significance of such internal editing among siRNAs is not yet known. In contrast, terminal editing through 2'-O-methylation, addition of untemplated nucleotides, or exonucleolytic trimming plays a more general role in small RNA

metabolism. These terminal modifications are not unrelated. Evidence in plants and animals suggests that methylation of the 3' terminal nucleotide protects small RNAs from polyuridylation and polyadenylation, signals that direct exonucleolytic degradation [11-16]. Thus, terminal methylation plays an important role in regulating small RNA turnover. Formation of the 2'-O-methyl group is catalyzed by HEN1, a methyltransferase discovered in *Arabidopsis thaliana* that is conserved across metazoa, fungi, viridiplantae, and bacteria [17]. Although plant and animal HEN1 orthologs exhibit 40-50% amino acid similarity in the conserved methyltransferase domain [18], the proteins differ in their substrate specificity. Plant HEN1 acts on small RNAs in duplex and methylates both siRNAs and miRNAs [19-21]. In contrast, animal HEN1 orthologs modify only single-stranded small RNAs [22-24], enabling methylation of small RNAs such as piRNAs, which are not derived from double-stranded RNA intermediates [25-29]. While animal piRNAs appear to be universally methylated [24,26,27,30-32], animal miRNAs are generally not methylated [19,26,31], and the mechanisms by which animal HEN1 orthologs discriminate between substrates are not entirely clear. HEN1 orthologs that catalyze terminal methylation of small RNAs have been characterized in mouse, fish, and fly, among other organisms [15,22-24,33], yet the orthologous methyltransferase in worm [18] has yet to be investigated. With its expanded Argonaute family and diverse small RNA classes, *Caenorhabditis elegans* provides an advantage for studying HEN1 substrate specificity.

Since the discovery of the founding members of the microRNA family in *C. elegans* [1,2,34], many additional classes of small RNAs have been characterized. A large-scale small RNA sequencing effort revealed a class of terminally methylated 21-nucleotide RNAs with 5' uridines [27]. These 21U RNAs were subsequently determined to represent the piRNAs of *C. elegans* based on their germline-specific expression, association with worm Piwi Argonautes PRG-1 and PRG-2, and function in transposon silencing and maintenance of temperature-dependent fertility [35-38]. Also found through small RNA cloning and deep sequencing were populations of 26- and 22-nucleotide RNAs with a 5' preference for guanosine (the 26G RNAs and 22G RNAs, respectively) that constitute the endo-siRNAs of *C. elegans* [27,39]. The 26G RNAs are

primary endo-siRNAs generated in the germline to regulate spermatogenic and zygotic gene expression. They are divided into two non-overlapping subclasses named for the Argonautes that bind them: the ERGO-1 class 26G RNAs, which are generated in the maternal germline and distributed into the embryo, and the ALG-3/ALG-4 class 26G RNAs, which are specific to the male germline and required for sperm function [40-42]. The 22G RNAs are composed of many small RNA classes, all of which are bound by worm-specific Argonautes (Wagos). A large population of 22G RNAs are secondary endo-siRNAs whose production by RNA-dependent RNA polymerases is triggered by the activity of 21U RNAs and 26G RNAs [36,41-43]; however, many other 22G RNAs are independent of these primary small RNAs [44,45]. Secondary siRNAs serve to amplify the signal of primary small RNAs to effect robust silencing. Production of 22G secondary siRNAs is also triggered by exogenously introduced dsRNAs [43,45-47], suggesting convergence of endogenous and exogenous RNAi pathways at the level of the secondary siRNA response.

Among *C. elegans* small RNAs, only 21U RNAs and 26G RNAs are known to be methylated [27,42]; 22G RNAs triggered by either primary endo- or exo-siRNAs appear to be unmethylated [45,46]. Although the significance of worm small RNA methylation is unknown, loss of terminal methylation has been shown to decrease stability of piRNAs in many animal models [15,22,24] and both endo- and exo-siRNAs in fly [22,48]. Methylation may therefore represent an essential step in stabilization of some classes of worm small RNAs.

In this study, we characterize the *C. elegans* *hen1* ortholog, which has been named *henn-1* (*hen* of *nematode*), as the name *hen-1* has already been assigned to an unrelated *C. elegans* gene. We demonstrate that HENN-1 methylates small RNAs bound by Piwi clade Argonautes: the 21U RNAs and the ERGO-1 class 26G RNAs. However, we show that 26G RNAs bound by Ago clade Argonautes ALG-3 and ALG-4 are not methylated and are therefore *henn-1*-independent. Differential methylation of 26G RNAs provides evidence for an existing model [13,22,23,49,50] wherein evolutionarily divergent Argonautes either direct or prohibit HEN1-mediated methylation

of associated small RNAs. In further support of this Argonaute-dictated methylation model, we find that small RNAs are likely methylated after associating with an Argonaute: the Argonaute ERGO-1 is required for 26G RNA methylation, but methylation is not required for ERGO-1 to bind a 26G RNA.

In the *henn-1* mutant, levels of both 21U RNAs and ERGO-1 class 26G RNAs drop precipitously after their deposition into embryo, suggesting that HENN-1-mediated methylation is essential for perdurance of the maternal small RNA load during filial development. Accordingly, the *henn-1* mutant shows enhanced somatic sensitivity to exogenous RNAi, a phenotype associated with loss of ERGO-1 class 26G RNAs. Surprisingly, however, the *henn-1* mutant germline exhibits an attenuated response to RNAi, suggesting that HENN-1 may also function in the exogenous RNAi pathway. Altogether, our study supports a role for HENN-1 in diverse small RNA pathways in *C. elegans* and offers further insight into the mechanisms governing substrate discrimination for animal HEN1 orthologs.

RESULTS

***C02F5.6* Encodes the *C. elegans* HEN1 Ortholog**

To examine small RNA methylation in *C. elegans*, we began by characterizing *C02F5.6*, the gene previously predicted to encode the HEN1 ortholog in worm [18]. This gene, subsequently named *henn-1*, encodes a protein that exhibits significant amino acid similarity across the conserved HEN1 methyltransferase domain relative to established members of the HEN1 family (Figure 2.S1). Although two *henn-1* gene models with differing 3' ends have been proposed, 3'RACE and protein studies using a rabbit polyclonal antibody generated against a common N-terminal HENN-1 epitope detected only the longer isoform (Figure 2.S2A, B).

To facilitate our studies of the function of HENN-1, we isolated and characterized the *henn-1(tm4477)* allele. This allele carries a deletion that encompasses *henn-1* exon four, which encodes 65% of the conserved methyltransferase domain as annotated by Kamminga et al. [15]. Sequencing of the *henn-1(tm4477)* mRNA indicates that loss of exon four activates a cryptic splice donor site in the third intron, resulting in an extended third exon that encodes a premature termination codon (Figure 2.S2B). The *henn-1(tm4477)* mRNA is readily detected by RT-PCR but does not produce a detectable protein product (Figure 2.S2A) or exhibit methyltransferase activity (see below), suggesting that *henn-1(tm4477)* (hereafter, *henn-1*) represents a functional null allele.

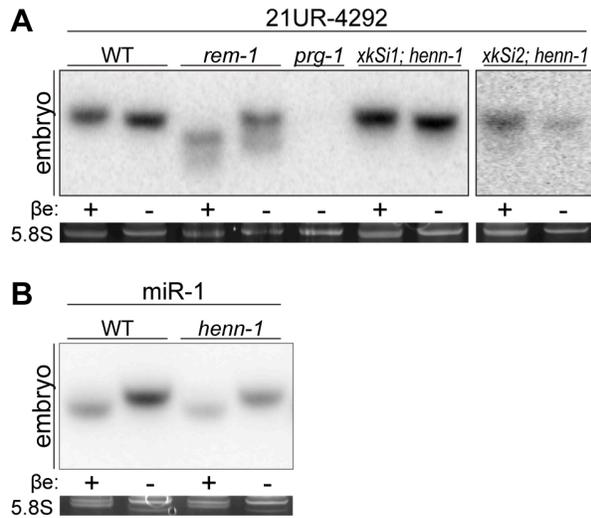
HENN-1 Terminally Methylates and Stabilizes *C. elegans* piRNAs

Like piRNAs in fly [22,23,32], mouse [30,31], and zebrafish [26], the *C. elegans* 21U RNAs are terminally methylated [27], but the factor catalyzing this modification has not yet been identified. To determine if 21U RNA methylation depends on *henn-1*, we assessed methylation status using the β -elimination assay [51]. A small RNA molecule whose terminal nucleotide has been 2'-O-methylated is resistant to this treatment, whereas the cis-diols of an unmodified 3' terminal nucleotide are oxidized by sodium periodate, rendering the nucleotide susceptible to β -elimination under basic conditions. The resulting size difference can be resolved on a polyacrylamide gel to determine methylation status. All 21U RNAs examined were found to be terminally methylated in a *henn-1*-dependent manner (Figures 2.1A, 2.S3A), whereas a control miRNA was not methylated in either wild-type or *henn-1* mutant animals (Figure 2.1B). Although 21U RNAs are still detectable in the *henn-1* mutant, the abundance of the full-length species is visibly decreased for some 21U RNAs; this correlates with the appearance of putative degradation products of unmethylated, unprotected 21U RNAs. To demonstrate that loss of 21U RNA methylation in the *henn-1* mutant is specifically due to the absence of *henn-1*, we used the Mos1-mediated single copy insertion technique [52] to introduce a *henn-1::gfp* transgene driven by the promoter of the polycistronic mRNA that encodes *henn-1* (*xkSi1*) or by the germline-specific *pie-1* promoter (*xkSi2*) into the *henn-1* mutant

(Figure 2.S2C). Both endogenous and germline-specific expression of *henn-1::gfp* restore 21U RNA methylation in the *henn-1* mutant (Figure 2.1A).

Figure 2.1: Methylation of 21U RNAs Requires *C. elegans* HEN1 Ortholog HENN-1.

A) HENN-1 is required for 21U RNA methylation. Endogenous (*xkSi1*) and germline-specific (*xkSi2*) expression of *henn-1::gfp* rescue 21U RNA methylation in *henn-1(tm4477)* mutant embryo. Total embryo RNA of the indicated genotypes was β -eliminated (β e +) or control treated (β e -) and probed for piRNA 21UR-4292. *prg-1(tm872)* lacks 21U RNAs and is included as a negative control. Below, ethidium bromide staining of 5.8S rRNA is shown. Additional 21U RNA northern blots are shown in Figure 2.S3A. B) *C. elegans* miRNAs are unmethylated. Total embryo RNA was probed for miR-1. Variable intensity of 5.8S rRNA bands in embryo indicates unequal loading.



To investigate the relationship between terminal methylation and piRNA accumulation, we used Taqman RT-qPCR to assess 21U RNA levels in wild-type and *henn-1* mutant animals across development at 25°C. Importantly, the Taqman stem-loop RT primer is capable of distinguishing between full-length and terminally degraded small RNAs [53]. For example, the *let-7e* miRNA differs from *let-7a* only in the absence of the final nucleotide and U > G substitution at the ninth nucleotide, a position likely not represented in the stem-loop Taqman primer. Absence of this final nucleotide decreases detection of *let-7e* by the *let-7a* Taqman assay by more than a thousandfold [53]. *henn-1* mutant embryo and early larva show dramatically reduced detection of female germline-enriched piRNA 21UR-1848 (Figure 2.2A), consistent with decreased embryonic detection for some 21U RNAs observed by northern blot (Figures 2.1A, 2.S3A). 21U RNA levels recover to wild-type in late larval stages, coincident with the onset of germline proliferation and de novo 21U RNA biosynthesis; however, in gravid animals at 56 hours, 21UR-1848 levels in the *henn-1* mutant have declined to less than 50% of those observed in wild-type (P=0.0005; two-tailed *t*-test). Eight additional 21U RNAs examined show a similar pattern (Figure 2.S4). These data suggest that *henn-1* is dispensable for piRNA biogenesis but essential for robust inheritance of piRNAs. Parallel analysis of miR-1 and several additional miRNAs across development shows that effects of loss of *henn-1* are specific to its substrates and not due to generalized small RNA dysregulation in the *henn-1* mutant (Figures 2.2B, 2.S5).

HENN-1 Plays a Minor Role in piRNA-mediated Germline Regulation

We next sought to determine the extent to which decreased abundance of piRNAs in the *henn-1* mutant compromises activity of the piRNA pathway. Unlike in fly, where many selfish genetic elements are desilenced in the absence of piRNAs [32], *C. elegans* at present has only a single established molecular readout for piRNA pathway function: increased expression of transposase mRNA from *Tc3*, a Tc1/mariner family transposon [35,36]. Two 21U RNAs have been found to map to *Tc3*, but both map in the sense direction and thus are unlikely to act directly in *Tc3* repression via canonical RNAi

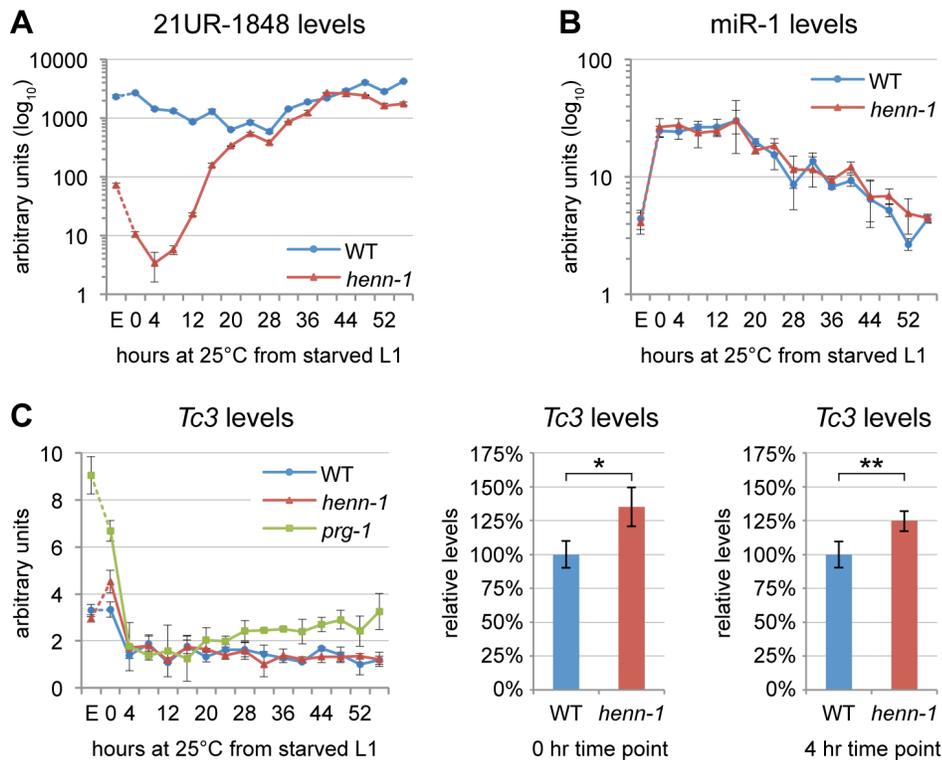
mechanisms [35,36]. Rather, 21U RNAs likely mediate their repressive effects through triggering production of secondary siRNAs, 22G RNAs, that engage worm-specific Argonautes (Wagos) to effect *Tc3* gene silencing [36,45]. We therefore identified a 22G RNA that shows complete antisense complementarity to *Tc3* and can be classified as a Wago-dependent, 21U RNA-dependent secondary siRNA based on its total depletion both in the MAGO12 mutant, which lacks all Wagos, and in the *prg-1(n4357); prg-2(n4358)* double mutant, which lacks piRNAs [36,45]. Levels of this 22G RNA in the *henn-1* mutant are reduced by 44% in embryo but not significantly altered in hatched L1 larva (Figure 2.S6A). This suggests that the low embryonic and early larval levels of 21U RNAs in the *henn-1* mutant are still sufficient to trigger production of secondary siRNAs, although to a lesser degree than in wild-type.

Consistent with the modest effect of loss of *henn-1* on accumulation of piRNA-triggered secondary siRNAs, *henn-1* mutant animals exhibit only a small increase (35% in starved L1 larva, 25% in L1 larva fed for 4 hours at 25°C) in *Tc3* transposase mRNA levels relative to wild-type (Figure 2.2C). This is not unexpected due to the poor coincidence of the time intervals corresponding to piRNA dysregulation in the *henn-1* mutant and *Tc3* sensitivity to 21U RNAs; the *henn-1* mutant shows the greatest disparity in piRNA levels in early larval development, whereas *Tc3* levels are most sensitive to piRNAs in germline and embryo (Figure 2.2A, C). These findings suggest that HENN-1 is not strictly required for piRNA target repression, but contributes to robust silencing of *Tc3*.

In addition to *Tc3* dysregulation, loss of *prg-1* also results in a temperature-sensitive sterility phenotype [38,43]. To determine if the *henn-1* mutant also exhibits a fertility defect, we assessed fertility at 20°C and 25°C. At 20°C, brood size of the *henn-1* mutant does not differ significantly from that of wild-type. In contrast, *henn-1* mutant animals maintained at 25°C exhibit a 25% decrease in brood size relative to wild-type ($P=0.0059$; two-tailed *t*-test) that can be rescued by germline expression of *henn-1::gfp* from the *xkSi2* transgene (Figure 2.S7). The impaired fertility of the *henn-1* mutant is consistent with abnormal fertility phenotypes associated with loss of HEN1 methyltransferase activity in other animals. Loss of *HEN1* in *Tetrahymena thermophila*

depletes Piwi-interacting RNAs called scan RNAs, impairing DNA elimination and, consequently, the viability of progeny [24]. The zebrafish *hen1* mutant fails to maintain a female germline, resulting in an exclusively male population [15]. Nevertheless, we cannot conclude that the temperature-sensitive fertility defect of the *henn-1* mutant is due exclusively to compromise of the 21U RNA pathway.

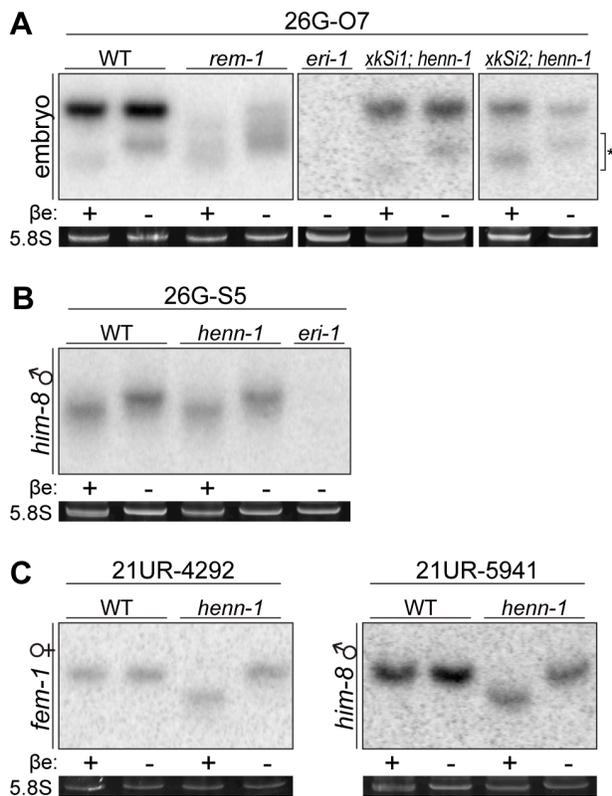
Figure 2.2: HENN-1 Stabilizes 21U RNAs. A) Loss of *henn-1* impairs 21U RNA accumulation in adult, embryo, and early larva. Levels of 21UR-1848 were assayed by Taqman qPCR in embryo and every four hours across development of wild-type and *henn-1(tm4477)* mutant animals at 25°C. Standard deviation is shown for biological triplicates. Taqman qPCR data for eight additional 21U RNAs are shown in Figure 2.S4. B) Effects of loss of *henn-1* are restricted to its small RNA substrates. Levels of miR-1 across development were assayed by Taqman qPCR. Standard deviation is shown for biological triplicates. Additional Taqman qPCR data for miRNAs are shown in Figure 2.S5. C) Loss of *henn-1* impairs *Tc3* transposase silencing primarily in early L1 larva. *Tc3* transposase mRNA levels were assayed by qPCR across development and normalized to mRNA levels of *eft-2*, an abundantly expressed housekeeping gene. *prg-1(tm872)* lacks 21U RNAs and is included as a positive control for *Tc3* upregulation. Significant zero and four hour time points are expanded at right (*: P=0.0251; **: P=0.0250, two-tailed *t*-test). Standard deviation is shown for biological triplicates. E, embryo; hr, hour.



ERGO-1 and ALG-3/ALG-4 Class 26G RNAs Are Differentially Methylated by HENN-1

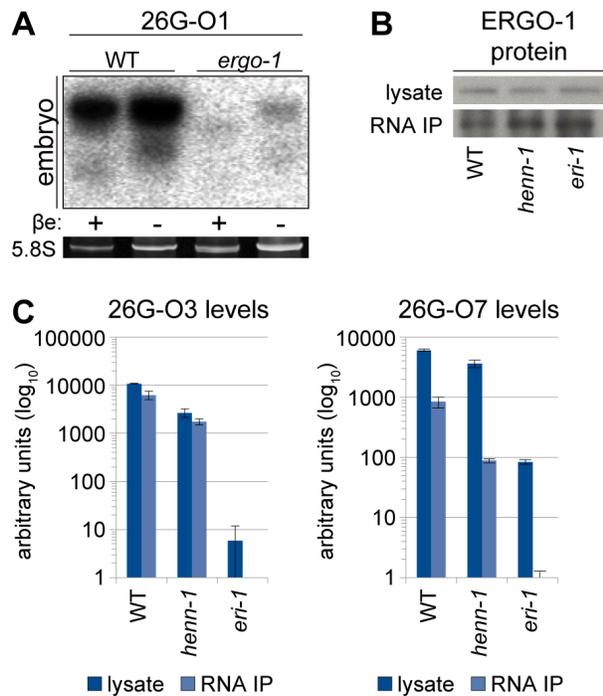
26G RNAs were reported to be methylated in the first *C. elegans* small RNA deep sequencing study [27]. Subsequent studies concluded that the species assessed was an ERGO-1 class 26G RNA [40]. Consistent with these data, we found that ERGO-1 class 26G RNAs, found in female germline and embryo, are methylated. As was the case for piRNAs, this methylation occurs in a *henn-1*-dependent manner (Figures 2.3A, 2.S3B). Surprisingly, however, ALG-3/ALG-4 class 26G RNAs, specific to the male germline, showed no evidence of methylation even in wild-type animals (Figures 2.3B, 2.S3C). One potential explanation for this observation would be that female germline small RNAs are universally methylated, whereas male germline small RNAs are not. To explore this possibility, we assessed 21U RNAs in male and female germlines. Both were methylated (Figure 2.3C), indicating that differential 26G RNA methylation cannot be explained simply by a lack of methyltransferase functionality in the male germline.

Figure 2.3: HENN-1 Selectively Methylates ERGO-1 Class 26G RNAs in an ERGO-1-dependent Manner. A) HENN-1 is required for ERGO-1 class 26G RNA methylation and stability. Total β -eliminated (β e +) or control treated (β e -) embryo RNA of the indicated genotypes was probed for ERGO-1 class 26G RNA 26G-O7. *eri-1(mg366)* lacks 26G RNAs and is included as a negative control. Asterisk indicates signal corresponding to cross-hybridization with unmethylated 22G RNAs. Below, ethidium bromide staining of 5.8S rRNA. Additional ERGO-1 class 26G RNA northern blots are shown in Figure 2.S3B. B) ALG-3/ALG-4 class 26G RNAs are unmethylated. Total *him-8(e1489)* male RNA was probed for ALG-3/ALG-4 class 26G RNA 26G-S5. An additional ALG-3/ALG-4 class 26G RNA northern blot is shown in Figure 2.S3C. C) 21U RNAs are methylated in a HENN-1-dependent manner in both female and male germlines. Total RNA of the indicated genotypes from *fem-1(hc17)* female or *him-8(e1489)* male was probed for female germline-enriched piRNA 21UR-4292 or male germline-enriched piRNA 21UR-5941, respectively.



Because the two classes of 26G RNAs bind unique Argonautes in male and female germlines, we hypothesized that the Argonaute ERGO-1 might direct methylation of 26G RNAs. To address this question, we sought to assess methylation of an ERGO-1 class 26G RNA in the absence of ERGO-1. As 26G RNAs are dramatically depleted in the absence of their respective Argonautes [40], we queried published wild-type and *ergo-1(tm1860)* gravid adult deep sequencing libraries [42] to identify an ERGO-1 class 26G RNA that still accumulates to levels sufficient for visualization by northern blotting in the *ergo-1(tm1860)* mutant. 26G-O1, an extremely abundant ERGO-1 class 26G RNA, is present at roughly 0.5% wild-type levels in the *ergo-1(tm1860)* mutant, but still abundant enough to detect by northern blotting. Consistent with our hypothesis that ERGO-1 is required for 26G RNA methylation, we found that 26G-O1 is unmethylated in the *ergo-1(tm1860)* mutant embryo (Figure 2.4A). We next asked the converse question: Is 26G RNA methylation required for association with ERGO-1? We immunopurified ERGO-1 complexes from wild-type and *henn-1* mutant embryo lysates (Figure 2.4B) and extracted RNA. In both wild-type and *henn-1* mutant samples, ERGO-1 class 26G RNAs are readily detected (Figure 2.4C), indicating that ERGO-1 effectively binds both methylated and unmethylated 26G RNAs. Taken together, these data suggest that 26G RNAs bind ERGO-1 and are subsequently methylated by HENN-1.

Figure 2.4: ERGO-1 is Required for Methylation of 26G RNAs. A) ERGO-1 class 26G RNA 26G-O1 is unmethylated in the absence of ERGO-1. Total embryo wild-type (5 μ g) or *ergo-1(tm1860)* (10 μ g) β -eliminated (β e +) or control treated (β e -) RNA was probed for 26G-O1. B) Anti-ERGO-1 rabbit polyclonal antibody immunoprecipitates ERGO-1 complexes. ERGO-1 complexes were immunopurified from lysates of equalized protein concentration extracted from wild-type, *henn-1(tm4477)* mutant, or *eri-1(mg366)* mutant embryo. Aliquots of lysates and immunoprecipitates (RNA IP) were probed with anti-ERGO-1 antibody. *ergo-1(tm1860)* mutant lysate was run in parallel to ensure specificity of ERGO-1 detection (data not shown). C) ERGO-1 binds methylated and unmethylated 26G RNAs. Taqman RT-qPCR for the indicated ERGO-1 class 26G RNAs was performed on samples described in B. The *eri-1(mg366)* mutant lacks 26G RNAs and serves as a negative control to demonstrate specificity of 26G RNA detection by Taqman assay. Standard deviation is shown for technical duplicates. Results are representative of two independent RNA immunoprecipitation experiments.



To test whether HENN-1-mediated methylation is required to maintain levels of all substrate small RNAs, we assessed ERGO-1 class 26G RNAs for defects in accumulation in the *henn-1* mutant. Loss of *henn-1* has more severe consequences for this class of small RNAs than are observed for 21U RNAs: ERGO-1 class 26G RNA 26G-O3 fails to accumulate to wild-type levels at any stage of development, although the disparity is less pronounced in adulthood, during peak 26G RNA biogenesis (Figure 2.5A). For comparison, we assayed levels of ALG-3/ALG-4 class 26G RNA 26G-S5 across the developmental window during which it is readily detected by Taqman RT-qPCR. Levels of 26G-S5 are similar in the *henn-1* mutant relative to wild-type (Figure 2.5B), consistent with the idea that HENN-1 is required for accumulation of ERGO-1 class 26G RNAs but dispensable for that of ALG-3/ALG-4 class 26G RNAs. Analysis of seven additional ERGO-1 class 26G RNAs and two additional ALG-3/ALG-4 class 26G RNAs corroborated these observations (Figures 2.S8, 2.S9).

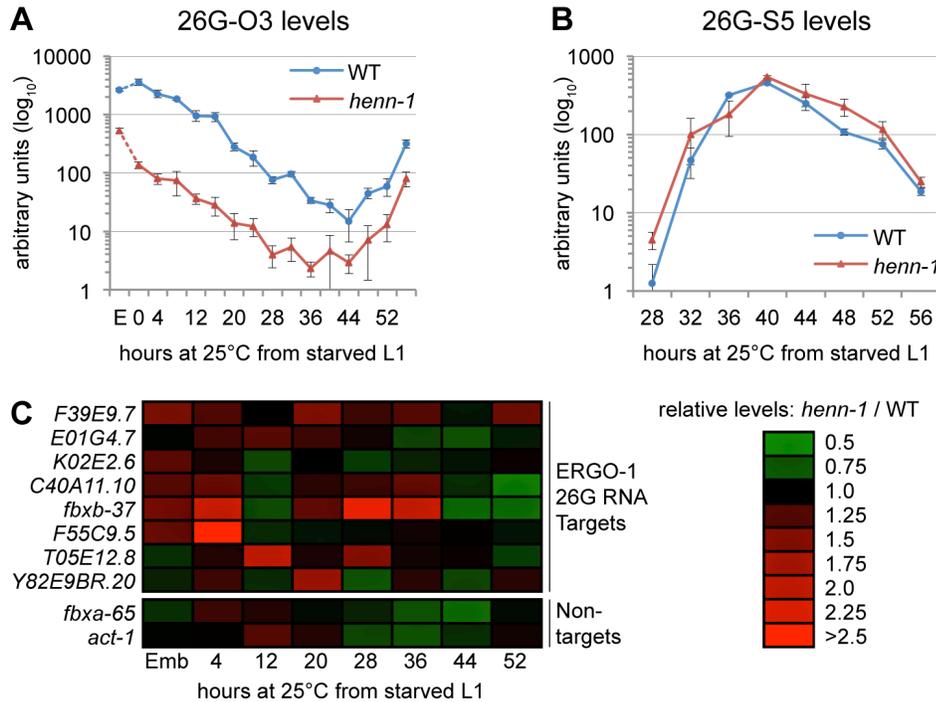
HENN-1 Contributes Minimally to ERGO-1 Class 26G RNA Target Silencing

To determine the effect of loss of *henn-1* on the silencing of ERGO-1 class 26G RNA targets, we assayed levels of a panel of mRNAs targeted by ERGO-1 class 26G RNAs for desilencing in *henn-1* mutant animals. During time points at which ERGO-1 class 26G RNAs are abundant, only modest upregulation of some, but not all, targets was detected; furthermore, no single target shows consistent desilencing in the *henn-1* mutant (Figure 2.5C, S10A). This is not unexpected, however, as the targets themselves vary in both expression and sensitivity to small RNA-mediated silencing across development [40]. To determine the specificity of this effect, two non-targets were examined in parallel. The maximal upregulation for either non-target does not exceed the maximal upregulation observed for any target, suggesting that the upregulation of ERGO-1 class 26G RNA targets in the *henn-1* mutant may be a consequence of 26G RNA depletion (Figures 2.5C, 2.S10B). This connection is supported by our observation that a Wago-dependent and ERGO-1 class 26G RNA-dependent secondary siRNA that presumably enhances target silencing also shows

defects in accumulation in embryo (Figure 2.S6B). The effect is modest, indicating that, as observed for the piRNA pathway, the depleted pool of ERGO-1 class 26G RNAs in the *henn-1* mutant is still sufficient for triggering fairly robust production of secondary siRNAs. Nevertheless, in an accompanying manuscript, Montgomery et al. observe that HENN-1 is required for silencing activity of a similar secondary siRNA upon a sensor transgene [54], suggesting that this pathway may indeed be compromised by loss of *henn-1*.

Figure 2.5: HEN1 Stabilizes ERGO-1 Class, but Not ALG-3/ALG-4 Class, 26G RNAs.

A) Loss of *henn-1* impairs ERGO-1 class 26G RNA accumulation at all stages. Levels of ERGO-1 class 26G RNA 26G-O3 were assayed by Taqman qPCR across development of wild-type and *henn-1(tm4477)* mutant animals at 25°C. Standard deviation is shown for biological triplicates. Taqman qPCR data for seven additional ERGO-1 class 26G RNAs are shown in Figure 2.S8. B) ALG-3/ALG-4 class 26G RNAs are *henn-1*-independent. Levels of ALG-3/ALG-4 class 26G RNA 26G-S5 were assayed across the period of development in which ALG-3/ALG-4 class 26G RNAs are readily detectable. Standard deviation is shown for biological triplicates. Taqman qPCR data for two additional ALG-3/ALG-4 class 26G RNAs are shown in Figure 2.S9. C) Loss of *henn-1* may result in modest, sporadic defects in ERGO-1 class 26G RNA target silencing. Levels of eight target and two non-target mRNAs were assayed across development of wild-type and *henn-1(tm4477)* mutant animals at 25°C and normalized to *eft-2*. Expression in the *henn-1(tm4477)* mutant relative to wild-type is represented according to the red-green color scheme indicated in the right panel. Raw data is shown in Figure 2.S10. E, embryo.



The Soma of the *henn-1* Mutant Exhibits Enhanced Sensitivity to Exogenous RNAi

ALG-3/ALG-4 class 26G RNAs are restricted to the male germline, and their mRNA targets are enriched for genes involved in spermatogenesis [40]. Accordingly, loss of ALG-3/ALG-4 class 26G RNAs results in male-associated sterility at non-permissive temperatures due to defects in sperm activation that are thought to arise from target dysregulation [41]. ERGO-1 class 26G RNAs, in contrast, are dispensable for fertility and target mostly poorly conserved and incompletely annotated genes, many of which reside in duplicated regions of the genome [42]. It is therefore not unexpected that the *ergo-1(tm1860)* mutant, which lacks ERGO-1 class 26G RNAs, exhibits no overt phenotypes that can be traced to target dysregulation. Rather, the *ergo-1(tm1860)* mutant exhibits an enhanced RNAi sensitivity (Eri) phenotype that is attributed to effects of loss of the ERGO-1-dependent small RNAs themselves; presumably, depletion of ERGO-1 class 26G RNAs and dependent secondary siRNAs liberates limiting RNAi factors shared between the endogenous and exogenous RNAi pathways [43,55,56].

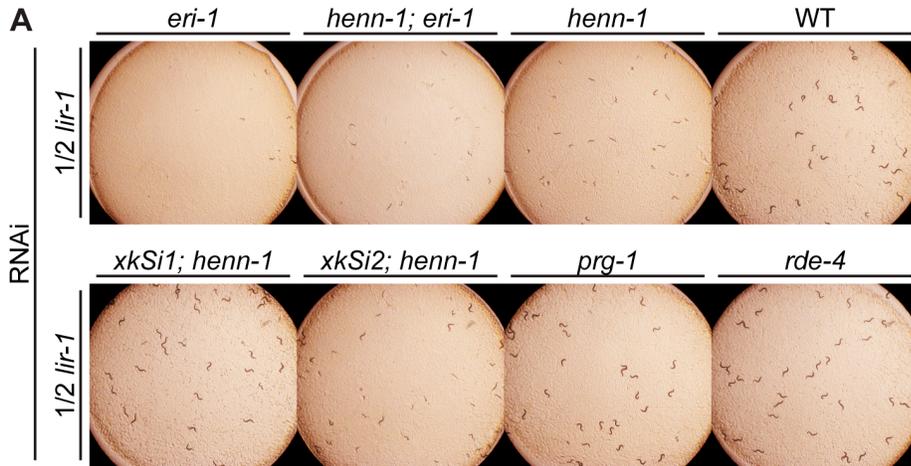
To determine whether loss of *henn-1* depletes ERGO-1 class 26G RNAs sufficiently to produce an Eri phenotype, as observed in the *ergo-1* mutant, we subjected L1 larvae from a panel of strains to feeding RNAi targeting various genes in the soma or germline. In order to expose subtle differences in RNAi sensitivity, we modulated the degree of knockdown, attenuating the dose of dsRNA trigger by diluting the bacterial RNAi clone with a bacterial clone harboring empty vector. RNAi of the somatic gene *lir-1* causes larval arrest and lethality in wild-type animals at full strength, but dilution 1:1 with empty vector largely eliminates the effect. In contrast, the *eri-1(mg366)* mutant, which lacks 26G RNAs, is affected severely by even dilute *lir-1* RNAi. The *henn-1* mutant also shows dramatically increased sensitivity to *lir-1* feeding RNAi relative to wild-type (Figure 2.6A, B). A *henn-1; eri-1* double mutant, however, shows RNAi sensitivity that is virtually identical to that of the single *eri-1* mutant, suggesting that the Eri phenotype of each allele likely stems from the same defect, namely, loss of ERGO-1 class 26G RNAs. While the somatic Eri phenotype of the *henn-1* mutant shows partial rescue by the

germline-specific *henn-1::gfp* transgene *xkSi2*, *henn-1::gfp* expression under the native promoter from transgene *xkSi1* rescues wild type RNAi sensitivity completely in the *henn-1* mutant (Figure 2.6B). These findings suggest that loss of *henn-1* in both germline and soma contributes to the Eri phenotype of the *henn-1* mutant. The *henn-1* mutant exhibits a similar somatic Eri response to RNAi of *dpy-13* and *lin-29* (Figure 2.S11).

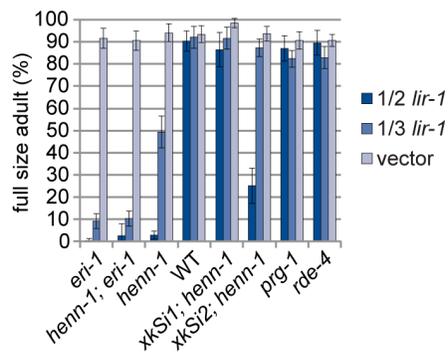
The Germline of the *henn-1* Mutant Exhibits Decreased Sensitivity to Exogenous RNAi

While the somatic Eri phenotype of the *henn-1* mutant was expected, knockdown of genes required for germline development or embryogenesis revealed that, incongruously, the *henn-1* mutant maternal germline exhibits an RNAi defective (Rde) phenotype. Animals subjected to *pos-1* RNAi lay dead embryos because maternally loaded *pos-1* mRNA is required for specifying cell fate of many tissues during embryonic development [57]. On *pos-1* RNAi diluted 1:2 with empty vector (1/3 strength), knockdown in wild-type animals is still sufficiently robust to reduce average brood size to fewer than five offspring per animal. *henn-1* mutant animals at this dilution, however, produce an average brood greater than tenfold that of wild-type, suggesting that loss of *henn-1* confers resistance to RNAi-mediated knockdown of this maternally deposited mRNA (Figure 2.6C). A lesser but statistically significant effect was observed for RNAi of the germline-expressed transcripts *par-1*, *par-2*, *pie-1*, and *glp-1* (Figure 2.S12). Sensitivity to *pos-1* RNAi is effectively rescued by either endogenous or germline-specific expression of *henn-1::gfp*, likely due to the fact that both transgenes are expressed in germline.

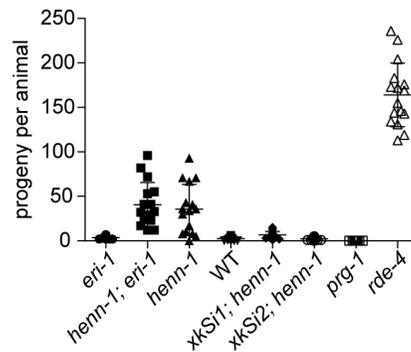
Figure 2.6: The *henn-1* Mutant Exhibits Opposite RNAi Sensitivity Phenotypes in Soma and Germline. A) *henn-1(tm4477)* mutant animals exhibit mildly enhanced somatic RNAi. Animals of the indicated genotype were plated as L1 larvae on *lir-1* feeding RNAi diluted 1:1 with empty vector (1/2 strength) and grown for 70 hours at 20°C. Data is quantified in part B. RNAi sensitivity data for knockdown of two additional somatic transcripts are shown in Figure 2.S11. B) Endogenous expression of *henn-1::gfp* from *xkSi1* rescues somatic RNAi sensitivity. Percent of animals reaching full size on *lir-1* feeding RNAi of the indicated strength at 70 hours is plotted. N = 8 plates of >50 animals per strain. Standard deviation is shown. C) *henn-1(tm4477)* mutant animals exhibit defective germline RNAi. Brood size of animals plated at 20°C as L1 larvae on *pos-1* feeding RNAi diluted 1:2 with empty vector is plotted. N ≥ 13 animals per strain. Mean and standard deviation are shown. RNAi sensitivity data for knockdown of four additional germline transcripts are shown in Figure 2.S12. Alleles used in this figure: *eri-1(mg366)*, *prg-1(tm872)*, *rde-4(ne301)*.



B sensitivity to *lir-1* RNAi



C sensitivity to 1/3 *pos-1* RNAi



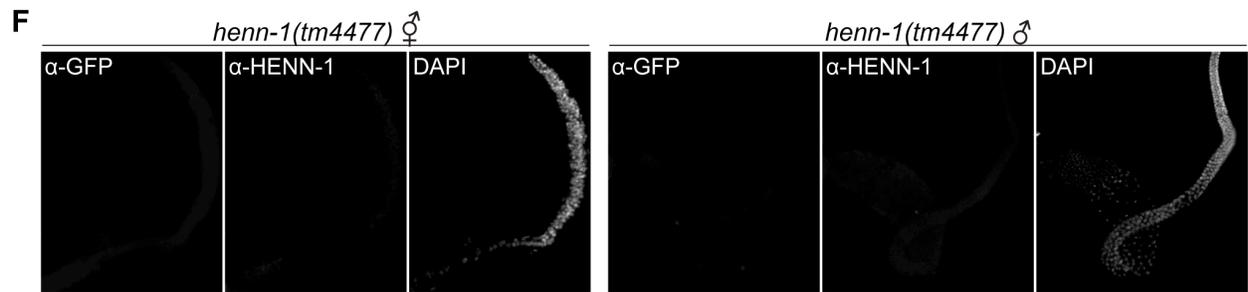
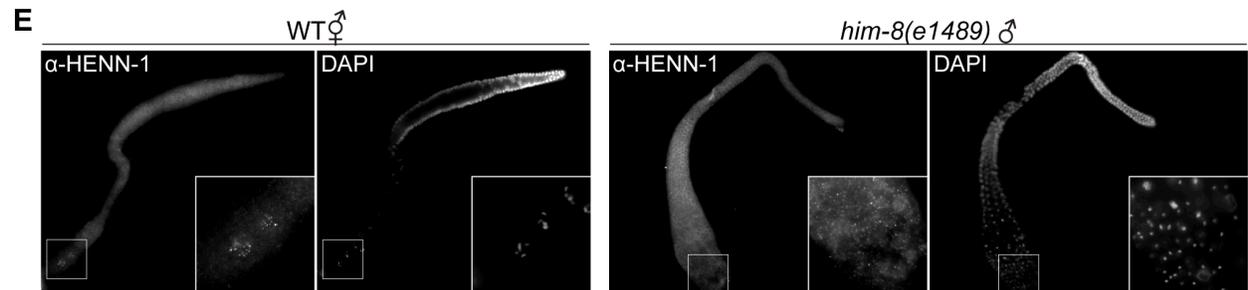
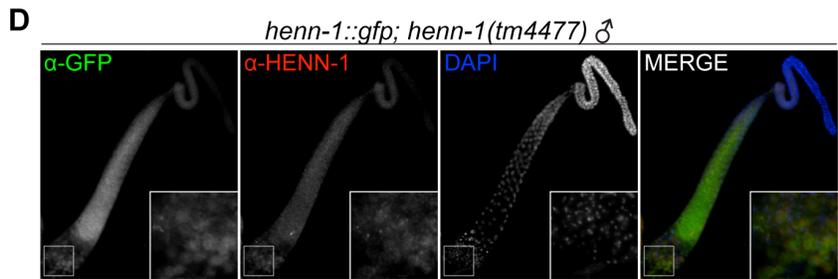
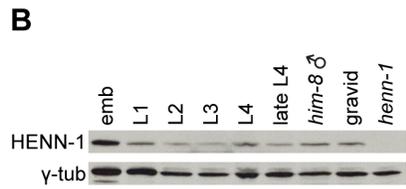
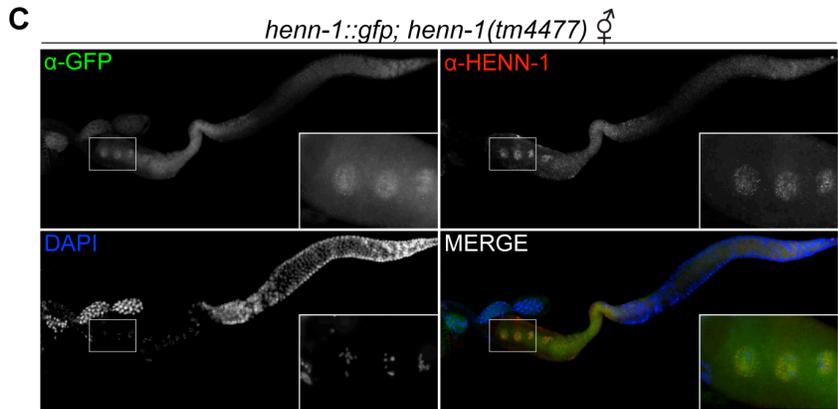
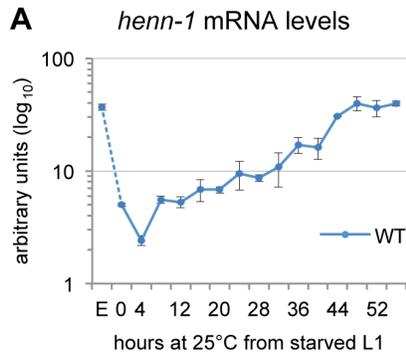
HENN-1 is Expressed in Both Germline and Soma

HEN1 orthologs appear to be restricted to the germline in vertebrates [15,33]; however, we observe phenotypes in both the germline and soma of the *henn-1* mutant that suggest broader activity. To investigate expression of HENN-1 in *C. elegans*, we assessed *henn-1* mRNA and protein levels throughout development. *henn-1* mRNA levels are lowest in young larva and increase as the germline proliferates, peaking in gravid adult (Figures 2.7A, 2.S13A). Germline-deficient *glp-4(bn2)* adult hermaphrodites show approximately a 50% reduction in *henn-1* mRNA levels relative to wild-type (Figure 2.S13B), indicating that *henn-1* mRNA is expressed in both germline and soma. Embryonic levels of *henn-1* are high but decrease rapidly; this pattern suggests that, unlike in zebrafish [15], *henn-1* mRNA may be maternally deposited into the embryo. HENN-1 protein is detectable throughout development and in both hermaphrodite and male adults (Figure 2.7B).

We next assessed the distribution of HENN-1::GFP fusion protein expressed from *xkSi1*, the rescuing *henn-1::gfp* transgene driven by the endogenous promoter, in the *henn-1* mutant background. Although single copy transgene expression levels are too low for direct visualization by fluorescence microscopy, HENN-1::GFP is readily detected using a mouse monoclonal anti-GFP antibody. Whole-mount immunostaining of transgenic L4 larvae reveals that HENN-1::GFP is expressed broadly in diverse somatic tissues and germline (Figure 2.S13C). Non-transgenic larvae show no signal, indicating that detection of HENN-1::GFP is specific. In extruded gonads of *xkSi1; henn-1* hermaphrodites, HENN-1::GFP is detected throughout the germline. Notably, the proximal oocytes show cytoplasmic and intense nucleoplasmic HENN-1::GFP expression (Figure 2.7C). Although nucleoplasmic enrichment is lost following fertilization, HENN-1::GFP is also abundant in embryo, with ubiquitous expression prior to gastrulation (Figure 2.S13D). HENN-1::GFP is also expressed throughout the germline of *xkSi1; henn-1* males (Figure 2.7D). During sperm maturation, we detect enrichment of HENN-1::GFP in residual bodies, but we cannot definitively conclude that it is excluded from sperm (Figure 2.7D, inset). In wild-type animals, studies of

endogenous HENN-1 using the rabbit polyclonal antibody generated against an N-terminal HENN-1 epitope corroborate the above findings, although the signal is more difficult to detect (Figure 2.7E). Staining in the *henn-1* mutant yields no signal for anti-GFP and anti-HENN-1 antibodies (Figure 2.7F); this demonstrates that detection of transgenic and endogenous HENN-1 proteins is specific. Together, these data define an expression pattern consistent with a role for HENN-1 in modifying small RNAs in both male and female germlines as well as in soma.

Figure 2.7: HENN-1 is Broadly Expressed in *C. elegans* Germline. A) The *henn-1* mRNA expression profile is consistent with germline enrichment. Levels of *henn-1* mRNA were assayed throughout development and normalized to *eft-2* mRNA. Standard deviation is shown for biological triplicates. Non-normalized levels are shown in Figure 2.S13A. B) HENN-1 is detected at all stages of development and in male. Lysates from animals of the indicated stages were probed with anti-HENN-1 rabbit polyclonal antibody. C) HENN-1 is abundant in hermaphrodite proximal germline and enriched in proximal oocyte nucleoplasm (inset). Extruded gonads of *xkSi1; henn-1(tm4477)* adult hermaphrodites were stained with anti-GFP mouse monoclonal and anti-HENN-1 rabbit polyclonal antibodies. D) HENN-1 is detectable in male proximal and distal gonad, with enrichment in residual bodies during spermatid maturation (inset). Extruded gonads of *xkSi1; henn-1(tm4477)* adult males were stained with anti-GFP and anti-HENN-1 antibodies. E) Expression of endogenous HENN-1 mirrors expression of HENN-1::GFP from transgene *xkSi1*. Extruded gonads of wild-type animals were stained with anti-HENN-1 antibody. F) Detection of HENN-1 proteins by immunostaining is specific. Extruded gonads of *henn-1(tm4477)* mutant animals were stained with anti-GFP and anti-HENN-1 antibodies. E, embryo.



The 21U RNAs and 26G RNAs appear to be significantly stable only in the presence of their respective Argonaute proteins [35,36,40]; accordingly, the localization patterns of the Argonaute proteins reflect the distribution of the different classes of small RNAs. We therefore wanted to compare the expression patterns of HENN-1 and the 26G RNA-binding Argonautes to determine whether the small RNA substrate specificity of HENN-1 could be explained by differential access to Argonaute-bound small RNAs. ERGO-1, which binds methylated 26G RNAs, is abundant in embryo [42], and its transcript is enriched during oogenesis [58], but its localization has not yet been reported. We assessed the staining pattern of ERGO-1 in hermaphrodite gonad and embryo using a polyclonal antibody generated against a C-terminal ERGO-1 epitope. ERGO-1 expression in the hermaphrodite germline begins at pachytene exit and persists in embryo (Figure 2.S13D, E). ERGO-1 shows cytoplasmic enrichment both in germline and embryo, suggesting that the cytoplasmic pool of HENN-1 may act in methylating 26G RNAs bound by ERGO-1. This interaction may, however, be transient, as we were unable to identify HENN-1 by mass spectrometry of immunopurified ERGO-1 complexes, nor could we detect ERGO-1 in immunopurified HENN-1::GFP complexes by western blot (data not shown). Notably, both HENN-1 and ERGO-1 remain abundant in early embryo (Figure 2.S13D). This is consistent with the proposed existence of a somatic endo-siRNA pathway that promotes continued biosynthesis of ERGO-1 class 26G RNAs after fertilization [59].

We next assessed co-localization of HENN-1 and ALG-3. ALG-3 and its close paralog, ALG-4, bind unmethylated 26G RNAs, and their transcripts are enriched during spermatogenesis [58]. In the male gonad, a rescuing *gfp::alg-3* transgene was reported to express in the proximal male germline, with localization to P granules beginning at late pachytene [41]. During sperm maturation, GFP::ALG-3 is relegated to residual bodies. Dual immunostaining of GFP::ALG-3 and endogenous HENN-1 demonstrates a large region of overlap (Figure 2.S13F), but HENN-1 does not appear to localize to P granules. This does not explain why ALG-3/ALG-4 class 26G RNAs are not methylated, because it is likely that HENN-1 can access P granules transiently: PRG-1 localizes predominantly to P granules [35,37], and the PRG-1-bound piRNAs are methylated.

This is in contrast to zebrafish Hen1, which carries a poorly conserved C-terminal domain (Figure 2.S1) that directs localization of Hen1 to nuage, perinuclear granules similar to *C. elegans* P granules, to methylate piRNAs [15].

DISCUSSION

Differential 26G RNA Methylation Supports an Argonaute-Dictated Methylation Model

We have shown that HENN-1 is essential for methylating select classes of *C. elegans* small RNAs, namely, 21U RNAs and ERGO-1 class 26G RNAs. As is the case in other animals, small RNAs in *C. elegans* that associate with Piwi clade Argonautes require HENN-1 for maintenance of wild-type levels. Ago clade Argonaute-associated microRNAs and ALG-3/ALG-4 class 26G RNAs, in contrast, are HENN-1-independent (Figure 2.S14A). It has been proposed that spatial and temporal regulation of HEN1 ortholog expression may contribute to small RNA substrate specificity in metazoans [24]. However, our immunostaining studies indicate that HENN-1 is coexpressed in the same tissues and subcellular compartments as Argonautes ERGO-1, PRG-1, and ALG-3 and their respective small RNAs (Figures 2.7, 2.S13). Therefore, differences in gross subcellular localization cannot explain the failure of ALG-3/ALG-4 class 26G RNAs to be methylated. Furthermore, although the two subclasses of 26G RNAs are generated in different germlines from non-overlapping targets, their sequences exhibit no obvious distinguishing characteristics that might account for their non-uniform methylation status.

One model of small RNA methylation posits that animal HEN1 orthologs only methylate small RNAs bound by Argonautes [15,22-24,49]. In support of this, work in fly shows that siRNA methylation requires assembly of DmAgo2 RISC [22,50], and in vitro studies using lysate from a silkworm ovary-derived cell line show that methylation of synthetic RNA only occurs after the longer substrate is bound by a Piwi protein and trimmed to piRNA size [60]. This model predicts that all 26G RNAs are bound as unmethylated

species by either ERGO-1 in the female germline or ALG-3/ALG-4 in the male germline and subsequently methylated or not, respectively. This is consistent with our findings *in vivo* that ERGO-1 is required for methylation of 26G RNAs (Figure 2.4A) and associates with 26G RNAs of either methylation status (Figure 2.4C). It has been further proposed that the identity of the Argonaute determines whether bound small RNAs are methylated [22,23,49,50]. An elegant illustration of this is provided by fly miR-277, which associates with both Ago1, the canonical fly miRNA Argonaute, and Ago2, which binds methylated siRNAs [61]. The miR-277 pool contains both methylated and unmethylated species. Depletion of Ago2 in cell culture results in loss of methylated miR-277, whereas Ago1 depletion results in a completely methylated miR-277 population [22]. Similarly, fly hairpin derived hp-esiRNAs sort into Ago1 and Ago2, but accumulate mainly in Ago2 because only hp-esiRNAs bound by Ago2 are methylated and therefore protected against degradation triggered by their extensive target complementarity [50]. In *C. elegans*, the model of Argonaute-dictated methylation can be invoked to explain the disparate methylation of the 26G RNAs: in the male germline, only ALG-3/ALG-4 are expressed, resulting in an unmethylated male 26G RNA population, whereas exclusive expression of ERGO-1 in the female germline and embryo directs methylation of female and zygotic 26G RNAs. This raises the intriguing possibility that selective expression of Argonautes that permit or prevent methylation could represent a new mechanism for differentially regulating small RNA turnover.

It is important to note that our results do not definitively exclude an alternative model wherein 26G RNAs are methylated prior to association with Argonautes and subsequently bound by ALG-3/ALG-4 only if unmethylated or by ERGO-1 only if methylated. In this model, HEN1 would methylate 26G RNAs in both germlines, but degradation of labile unbound siRNAs would result in a purely unmethylated or methylated population of 26G RNAs in male and female germlines, respectively. Because 26G RNAs assessed in embryo are fully methylated (Figures 2.3A, 2.S3B), such a mechanism would require that ERGO-1 exhibit very unfavorable kinetics for association with unmethylated small RNAs. We do not find this to be the case, as ERGO-1 binds some 26G RNAs with similar efficiency when methylated and

unmethylated (Figure 2.4C). Our data therefore provide stronger evidence for a model of Argonaute-dictated methylation of small RNAs.

Possible Advantages for Selective Methylation of Small RNAs

Differential germline expression of Argonautes could have evolved in *C. elegans* because of advantages conferred by selective stabilization of female germline 26G RNAs. Unlike ALG-3/ALG-4 class 26G RNAs, which appear to function exclusively during sperm development [40,41], ERGO-1 class 26G RNAs exert much of their influence during embryonic and larval development, well beyond initiation of their biogenesis in the hermaphrodite germline [40]. Accordingly, their targets are depleted of germline-enriched genes [40,59]. The oocyte contributes the vast majority of the initial zygotic cellular contents; therefore, methylation of 26G RNAs originating in the female germline may ensure robust inheritance and perdurance of primary small RNAs. Methylation of 26G RNAs in the male germline would likely not significantly increase their representation in sperm or zygote, as ALG-3/ALG-4 are relegated to residual bodies during spermatogenesis and exert effects in mature sperm only indirectly through dependent secondary 22G RNAs [41]. Nonetheless, it would be interesting to express ERGO-1 ectopically in sperm and determine whether ALG-3/ALG-4 class small RNAs are methylated. Such a strategy may reveal unexpected consequences related to inappropriate methylation and stabilization of ALG-3/ALG-4 class 26G RNAs.

Role of HENN-1 in the Balance Between Endo- and Exo-RNAi

In the absence of *henn-1*, we show that response to RNAi-mediated knockdown is enhanced for somatic genes (Figures 2.6A,B, 2.S11). This is likely due to destabilization of ERGO-1 class 26G RNAs in the *henn-1* mutant, which reduces competition with primary exo-siRNAs for stimulating secondary siRNA activity mediated by somatic Argonautes such as SAGO-1 and SAGO-2 [43,55]. While germline-specific expression

of *henn-1::gfp* only partially rescues this somatic Eri phenotype, *henn-1* mutant animals rescued with an endogenous *henn-1::gfp* transgene, which drives both somatic and germline expression, show wild-type RNAi sensitivity. Under the model of competing endo- and exo-RNAi pathways, this suggests that HENN-1-mediated methylation of ERGO-1 class 26G RNAs in the germline alone cannot maintain small RNA levels sufficient to sequester an appropriate proportion of the limiting RNAi factors. It is possible that ERGO-1 class 26G RNA biogenesis continues in embryo and larva, as previously suggested [59], and that high concentrations of HENN-1 are necessary for continued stabilization of these small RNAs. Such a model would be consistent with our characterization of the distributions of HENN-1 and ERGO-1, both of which are still detected in abundance in developing embryo (Figure 2.S13D, E).

While the majority of the phenotypes observed in the *henn-1* mutant can be attributed to destabilization of endogenous small RNA substrates, the germline Rde phenotype suggests a role for HENN-1 in exogenous RNAi. It is unclear why HENN-1 is dispensable for robust exogenous RNAi in the soma but required in the germline. While this may be an indirect effect, as suggested in concurrent work by Kamminga et al. [62], one possible explanation is that HENN-1 stabilizes primary exo-siRNAs or dependent 22G secondary siRNAs. There is support in fly for methylation of exo-siRNAs and transgenic hairpin-derived siRNAs [22,63], but this has not yet been demonstrated in *C. elegans*. 22G RNAs triggered by primary exo-siRNAs appear not to be methylated [47], consistent with our and others' observations that Wago-dependent 22G RNAs from diverse endogenous sources are unmethylated (Figures 2.3A, 2.S3B, and [45]). The methylation status of worm primary exo-siRNAs has not been definitively established, although a 22-nucleotide siRNA generated from a transgene encoding a perfect hairpin was not found to be methylated [46].

Structural Differences in Ago and Piwi Clade Argonautes May Dictate HEN1 Substrate Specificity

All Argonautes contain two signature domains, PAZ and Piwi [64]. The Piwi domain, unique to Argonautes, adopts an RNase H-like configuration and serves as the catalytic core of RISC [65,66]. The PAZ domain recognizes and anchors the 3' end of the small RNA [67,68]. Comparison of Piwi and Ago clade Argonautes reveals that Piwi proteins contain a small insertion in their PAZ domains in a loop connecting two β strands [69]. Crystal structures of a human Piwi Argonaute PAZ domain suggest that this insertion results in the formation of a more spacious binding pocket capable of accommodating the 2'-O-methyl group of a piRNA. Interactions between the methyl group and hydrophobic residues lining the pocket confer a threefold to sixfold higher binding affinity for 2'-O-methyl than 2'-OH [69]. In *C. elegans*, only PRG-1/PRG-2 and ERGO-1 show evidence of a PAZ domain insertion (Figure 2.S14B), consistent with their designation as Piwi clade Argonautes and association with methylated small RNAs.

In spite of their shared classification, ERGO-1 exhibits far less homology than PRG-1/PRG-2 to mammalian and insect Piwi proteins (Figure 2.S14A) [43]. Similarly, among worm, fly, and human Argonautes, DmAgo2 and *C. elegans* Argonaute RDE-1 are among the most divergent members of their clades [43]. In fact, so divergent is RDE-1 that its cladistics are ambiguous, with our and other published alignments variably assigning it to each of the three clades (Figure 2.S14A and [43,70]). Both DmAgo2 and RDE-1 bind exo-siRNAs, although only the former has been shown to permit methylation [22]. Interestingly, both lack the insertion found in Piwi Argonaute PAZ domains (Figure 2.S14B). The absence of this insertion in DmAgo2 suggests that it is not required for association with methylated small RNAs, raising the possibility that RDE-1 too may permit methylation of associated small RNAs. If HENN-1 does not methylate RDE-1-bound small RNAs, it is unclear what specific role HENN-1 plays in exo-RNAi in the germline. Nevertheless, its dual functions in endogenous and exogenous RNAi place HENN-1 in the company of DCR-1 and the Wago proteins at the intersection between these two RNAi pathways.

MATERIALS AND METHODS

C. elegans Strains

C. elegans were maintained according to standard procedures. The Bristol strain N2 was used as the standard wild-type strain. The alleles used in this study, listed by chromosome, are: unmapped: *nels23[unc-119(+)] GFP::ALG-3*; LGI: *glp-4(bn2)*, *prg-1(tm872)*; LGII: *xkSi1[PC30A5.3::henn-1::gfp::henn-1 3'UTR cb-unc-119(+)] II*, *xkSi2[Ppie-1::henn-1::gfp::tbb-2 3'UTR cb-unc-119(+)] II*; LGIII: *rde-4(ne301)*, *henn-1(tm4477)*; LGIV: *eri-1(mg366)*, *fem-1(hc17)*, *him-8(e1489)*; LGV: *ergo-1(tm1860)*. The *nels23[unc-119(+)] GFP::ALG-3* strain was generously provided by Craig Mello (University of Massachusetts, Worcester, MA).

RNA Sample Preparation

For embryo samples, L1 larvae were grown at 20°C until gravid. Embryos were isolated using sodium hypochlorite solution; an aliquot of embryos was allowed to hatch overnight at room temperature to determine viability. For male samples, synchronized *him-8(e1489)* L1 larvae were grown at 20°C for 72-75 hours. Males were isolated by filtering through 35 µm mesh [71]. For female samples, synchronized *fem-1(hc17)* L1 larvae were plated and grown at 25°C for 52 hours. For time course samples, synchronized wild-type (N2) and *henn-1(tm4477)* L1 larvae were grown at 25°C until gravid; embryos were extracted and harvested for RNA or hatched overnight at room temperature and then grown at 25°C for the specified number of hours before harvest. The *prg-1(tm872)* time course samples were prepared in the same way, except that animals were grown for the first generation at 20°C to evade temperature-sensitive sterility. Samples were processed by either three rounds of freeze/thaw lysis or two rounds of homogenization for 15 sec using the Tissue Master-125 Watt Lab Homogenizer (Omni International) and the RNA was extracted in TriReagent (Ambion) following the vendor's protocol, with the following alterations: RNA was precipitated in isopropanol for one hour at -80°C; RNA was pelleted by centrifugation at 4°C for 30 min at 20,000 x g; the pellet was washed three times in 75% ethanol; the pellet was resuspended in water.

β-elimination Assay for Small RNAs and Northern Blot Analysis

For detection of small RNAs, 10 or 40 μg of total RNA were β -eliminated as described [51]; control samples were processed in parallel without sodium periodate. Northern blot analysis was performed as described [72]. In brief, 5 or 10 μg of β -eliminated total RNA were resolved on 17.5% or 20% denaturing Urea-PAGE gels (SequaGel, National Diagnostics) and transferred to Hybond-NX membrane (Amersham). 21 and 26 nt synthetic RNAs were run as size markers and visualized in tandem with rRNA by ethidium bromide staining. Pre-hybridization/hybridization and washes were performed at 48°C or 50°C. Oligonucleotides corresponding to the antisense sequences of the small RNAs (Table 2.S1) were synthesized and end-labeled with [α - ^{32}P]-dATP using the miRNA StarFire kit (Integrated DNA Technologies).

RNAi Sensitivity Assay

To test the response to exogenous RNAi, bacterial clones from the Ahringer RNAi library [73] were diluted with bacteria harboring the empty vector *L4440* to achieve a level of RNAi sensitivity that allowed us to differentiate the RNAi responses in the strains examined. To determine *lir-1* RNAi sensitivity, the *lir-1* RNAi bacterial clone diluted with *L4440* bacterial clone at a 1:1 or 1:2 ratio (1/2 or 1/3 strength) was used; >50 L1 larvae were plated per plate and the number of total animals assayed per plate was determined at day two after plating; the percent of animals exhibiting the larval arrest phenotype was determined at 70 hours at 20°C. Sensitivity to RNAi of *dpy-13* and *lin-29* was also assessed using this method, where animals subjected to *dpy-13* RNAi were imaged at 70 hours and those subjected to *lin-29* RNAi were evaluated for the absence of protruding vulva or bursting phenotype. For *pos-1* RNAi, synchronized L1 larvae were singled onto plates with *pos-1* RNAi diluted with empty vector at a 1:2 ratio (1/3 strength) that had been induced overnight at 25°C. Animals were grown at 20°C for six days and progeny were counted. Sensitivity to RNAi of *pie-1*, *par-1*, and *par-2* was assessed similarly at the indicated dilutions with 4 plates of 4 P₀ animals per strain. Sensitivity to *glp-1* RNAi was determined at the indicated dilutions by plating 4 plates of >50 L1 larvae per strain per gene and scoring for the absence of oocytes and embryos in both arms of the germline at 70 hours at 20°C. For all RNAi sensitivity assays, data are representative of at least two independent experiments.

Fertility Assay

To determine brood size, synchronized L1 larvae from gravid adults grown at 20°C or shifted to 25°C for two generations were singled onto plates with OP50 and grown to adulthood at their respective temperatures. Once egg-laying began, animals (N ≥ 13 per strain) were transferred to fresh plates daily until the supply of fertilized eggs was exhausted. Progeny of the singled parents were counted as late larvae/adults. Results are representative of two independent experiments.

Quantitative RT-PCR

Taqman small RNA probes were synthesized by Applied Biosystems (Table 2.S2) [74]. For each reaction, 50 ng of total RNA were converted into cDNA using Multiscribe Reverse Transcriptase (Applied Biosystems). The resulting cDNAs were analyzed by a Realplex thermocycler (Eppendorf) with TaqMan Universal PCR Master Mix, No AmpErase UNG (Applied Biosystems). We could not identify a small RNA whose levels were consistent across development for use in normalization. Therefore, to preserve the developmental profile of each of the small RNA assessed, back transformation was used to calculate relative small RNA levels from qRT-PCR cycle numbers. As a control for RNA quality, miR-1 Taqman assays were run in parallel for all samples excluding the ERGO-1 RNA immunoprecipitation samples, in which miRNAs are absent. For quantification of mRNA levels, 100 ng of total RNA were converted into cDNA with Multiscribe Reverse Transcriptase (Applied Biosystems) following the vendor's protocol with the following changes: 25 units of RT and 7.6 units of RNase OUT (Invitrogen) were used per reaction. cDNAs were analyzed using Power Sybr Green PCR Master Mix (Applied Biosystems) (primers, Table 2.S3). Relative mRNA levels were calculated based on the $\Delta\Delta 2C_t$ method [75] using *eft-2* for normalization. For all qPCR, 40 cycles of amplification were performed; reactions whose signals were not detected were therefore assigned a cycle number of 40. All results presented are the average values of independent calculations from biological triplicates unless indicated. To determine average upregulation of ERGO-1 26G RNA targets in *henn-1* relative to wild-type (Figure 2.5C), the mean was calculated for all of the ratios generated by dividing each *henn-1* biological replicate by each wild-type biological replicate.

3' RACE

3' RACE was performed using the 3' RACE System for Rapid Amplification of cDNA ends (Invitrogen) according to the manufacturer's protocol. *henn-1* gene-specific primer (5' GCAGTATGTCGCCTCCAAGTAGAT 3') was used to amplify *henn-1* 3' ends from cDNA generated from embryo. Product corresponding to only the seven-exon gene model of *henn-1* was observed, consistent with detection of a single protein isoform corresponding to this model on western blot analysis.

Plasmids and Transgenic Strains

The endogenous *henn-1::gfp* reporter construct (*xkSi1*) was generated by introducing the following fragments into pCFJ151: endogenous promoter of the *henn-1*-containing operon CEOP3488 [76] (3.9 kb PCR fragment immediately upstream of the C30A5.3 start codon), *henn-1* genomic coding region (1.8 kb PCR fragment with mutated termination codon), *gfp* coding region (0.9 kb fragment with multiple synthetic introns and termination codon), and *henn-1* endogenous 3'UTR (1.1 kb PCR fragment immediately downstream of *henn-1* termination codon). The germline-only *henn-1::gfp* reporter construct (*xkSi2*) was generated as above with the following substitutions: CEOP3488 operon promoter was replaced with the *pie-1* promoter (2.4 kb PCR fragment immediately upstream of *pie-1* start codon) and *henn-1* endogenous 3'UTR was replaced with the C36E8.4 3'UTR (0.3 kb PCR fragment downstream of C36E8.4). Constructs were cloned into the pCFJ151 vector, confirmed by sequencing, and used to generate single-copy integrated transgenes via the MosSCI technique [52]. Gene fusion products of the expected size were specifically detected by western blot with both anti-HENN-1 and anti-GFP antibodies.

Generation of Antibodies

Synthetic antigenic peptides were conjugated to KLH and each was used to immunize two rabbits (Proteintech). Antisera were subsequently affinity purified using Affi-Gel 15 gel (Bio-Rad). Antigenic peptide sequences are as follows: N-terminal HENN-1 peptide with N-terminal added cysteine (CTYVEAYEQLEIALLEPLDR), C-terminal ERGO-1 peptide (CEVNKDMNVNEKLEGMTFV).

Western Blot Analysis

Proteins immobilized on Immobilon-FL transfer membrane (Millipore) were probed with anti-HENN-1 rabbit polyclonal antibody (1:2000), anti- γ -tubulin rabbit polyclonal antibody (LL-17) (Sigma) (1:2000), or anti-ERGO-1 rabbit polyclonal antibody (1:1000). Peroxidase-AffiniPure goat anti-rabbit IgG secondary antibody was used at 1:10000 (Jackson ImmunoResearch Laboratories) for detection using Pierce ECL Western Blotting Substrate (Thermo Scientific).

Isolation of ERGO-1-associated RNAs

Wild-type, *henn-1*, or *eri-1(mg366)* embryos isolated from gravid adults grown at 20°C were frozen in liquid nitrogen and homogenized with a Mixer Mill MM 400 ball mill homogenizer (Retsch) Homogenates were suspended in lysis buffer (50 mM HEPES (pH 7.4), 1 mM EGTA, 1 mM MgCl₂, 100 mM KCl, 10% glycerol, 0.05% NP-40 treated with a Complete, Mini, EDTA-free Protease Inhibitor Cocktail tablet (Roche Applied Sciences)) and clarified by centrifugation at 12,000 x g for 12 minutes at 4°C. Aliquots of homogenate were reserved as crude lysate for western blot to confirm that immunoprecipitations were performed in lysates of equivalent protein concentration (2 mg/mL). For immunoprecipitations, embryo homogenates were incubated at 4°C for one hour with 75 μ g anti-ERGO-1 rabbit polyclonal antibody conjugated to Dynabeads® Protein A (Invitrogen), after which the beads were washed (500 mM Tris-HCl (pH 7.5), 200 mM KCl, 0.05% NP-40) and associated proteins were eluted with 200 μ L glycine. Three quarters of each eluate were precipitated overnight at 4°C in trichloroacetic acid, pelleted, washed with acetone, and resuspended for western blot analysis. The remaining eluate was treated with 2 mg/ml Proteinase K (Roche) and incubated at 37°C for 30 minutes. RNA was isolated from the eluate by incubation with TriReagent and processed as described above. RNA pellets were resuspended in 10 μ L water and 5 μ L were used for each Taqman RT reaction.

Immunostaining

Primary antibodies were applied according to the following specifications: anti-GFP mouse monoclonal antibody 3E6 (Invitrogen) was diluted 1:1500 to detect HENN-

1::GFP and 1:200 to detect ALG-3::GFP; anti-ERGO-1 rabbit polyclonal was diluted 1:200; and anti-HENN-1 rabbit polyclonal antibody was preabsorbed as described [77] with *henn-1(tm4477)* mutant extract and diluted 1:200. Alexa Fluor 555 goat anti-rabbit IgG and Alexa Fluor 488 goat anti-mouse IgG (Molecular Probes) secondary antibodies were diluted 1:500. All antibodies were diluted in 0.5% bovine serum albumin (Sigma). For immunostaining of gonads and embryos, synchronized gravid hermaphrodites or adult males grown at 20°C were dissected on Superfrost Plus positively charged slides (Fisherbrand) with 27 G x 1/2 inch BD PrecisionGlide needles (Becton, Dickinson and Company) as described by Chan and Meyer in WormBook [78] Protocol 21 with 1.5% paraformaldehyde (Sigma). Slides were incubated with primary antibodies overnight at 4°C and with secondary antibodies for three hours at room temperature. Slides were mounted with VECTASHIELD Mounting Medium with DAPI (Vector Laboratories). For whole-worm immunostaining, synchronized late L4 larvae grown at 20°C were transferred to subbed slides [77] in M9, fixed for six minutes in 1.5% paraformaldehyde, freeze-cracked, and incubated for 15 minutes in ice cold methanol. After fixation, slides were processed as above. Images were captured on an Olympus BX61 epifluorescence compound microscope with a Hamamatsu ORCA ER camera using Slidebook 4.0.1 digital microscopy software (Intelligent Imaging Innovations) and processed using ImageJ.

ACKNOWLEDGMENTS

We thank Patrick Hu, Chi Zhang, and Sylvia Fischer for helpful comments on the manuscript. We thank Carolyn Philips for advice on immunostaining. We also thank the *Caenorhabditis* Genetics Center for *C. elegans* for strains.

SUPPLEMENT

Figure 2.S1: Alignment of HEN1 Orthologs. A) *C. elegans* HENN-1 bears the conserved HEN1 methyltransferase domain. Protein sequences of HEN1 orthologs from *Caenorhabditis elegans* (NP_741250.1), *Drosophila melanogaster* (NP_610732.1), *Danio rerio* (NP_001017842.1), *Mus musculus* (NP_079999.2), *Homo sapiens* (NP_001096062.1), and *Arabidopsis thaliana* (NP_567616.1) were aligned using T-Coffee [79,80] with default parameters. The resulting multiple sequence alignment was cropped to show the conserved HEN1 methyltransferase domain (underlined in red) and the C terminus. Significant alignment was not observed for the N terminus. B) Conservation of the HEN1 methyltransferase domain of HENN-1 is comparable to that of other orthologs. Percent identity was calculated using ClustalW (version 2.1; <http://www.ebi.ac.uk/Tools/msa/clustalw2/>) [81,82] with default parameters to perform pairwise alignments of the conserved HEN1 methyltransferase domains as defined in Figure 2.S1A.

Figure 2.S2: C02F5.6 Alleles and Transgenes. A) anti-HENN-1 polyclonal antibody recognizes a single ~52 kD HENN-1 isoform in wild-type embryo lysate; no protein product is detected in *henn-1(tm4477)* embryo lysate. B) *C02F5.6 (henn-1)* gene structure showing the encoded N-terminal epitope for generating the anti-HENN-1 rabbit polyclonal antibody, conserved HEN1 domain (pink), and deletion region for the *henn-1(tm4477)* allele (red underline). Aberrant splicing of *henn-1(tm4477)* mRNA is diagrammed below. Activation of a cryptic splice donor site in the *henn-1(tm4477)* mRNA produces a premature termination codon (STOP). C) Diagrams of *xkSi1* (endogenous expression) and *xkSi2* (germline-only expression) *henn-1::gfp* transgenes. Transgenes were inserted as single copies on chromosome II via the MosSCI technique [52].

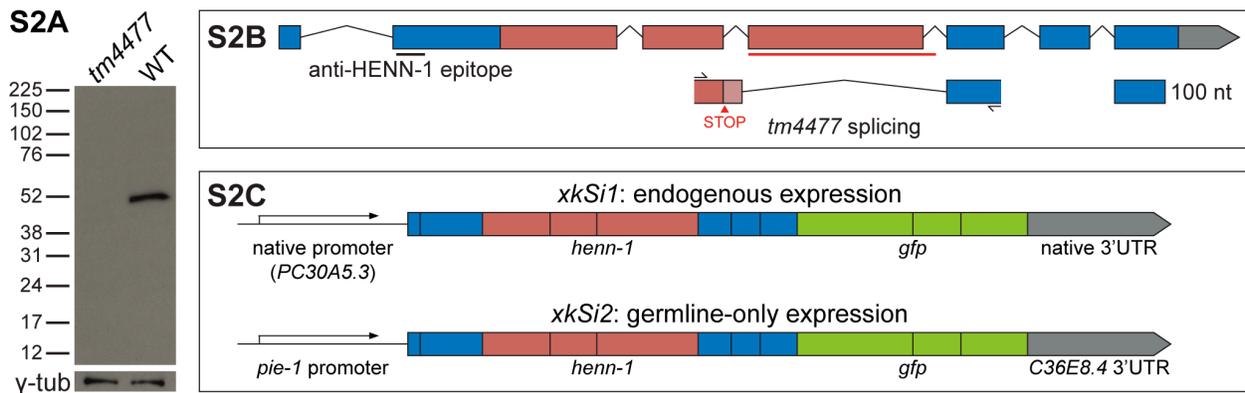


Figure 2.S3: Methylation Status of Additional Small RNAs. A) Additional 21U RNAs show HENN-1-dependent methylation. β -eliminated (β e +) or control treated (β e -) embryo RNA of the indicated genotypes was probed for the specified 21U RNAs. Below, ethidium bromide staining of 5.8S rRNA. *prg-1(tm872)* lacks 21U RNAs and is included as a negative control. B) Additional ERGO-1 class 26G RNAs show HENN-1-dependent methylation in embryo RNA. *eri-1(mg366)* lacks 26G RNAs and is included as a negative control. C) ALG-3/ALG-4 class 26G RNA 26G-S7 shows absence of methylation in *him-8(e1489)* male RNA.

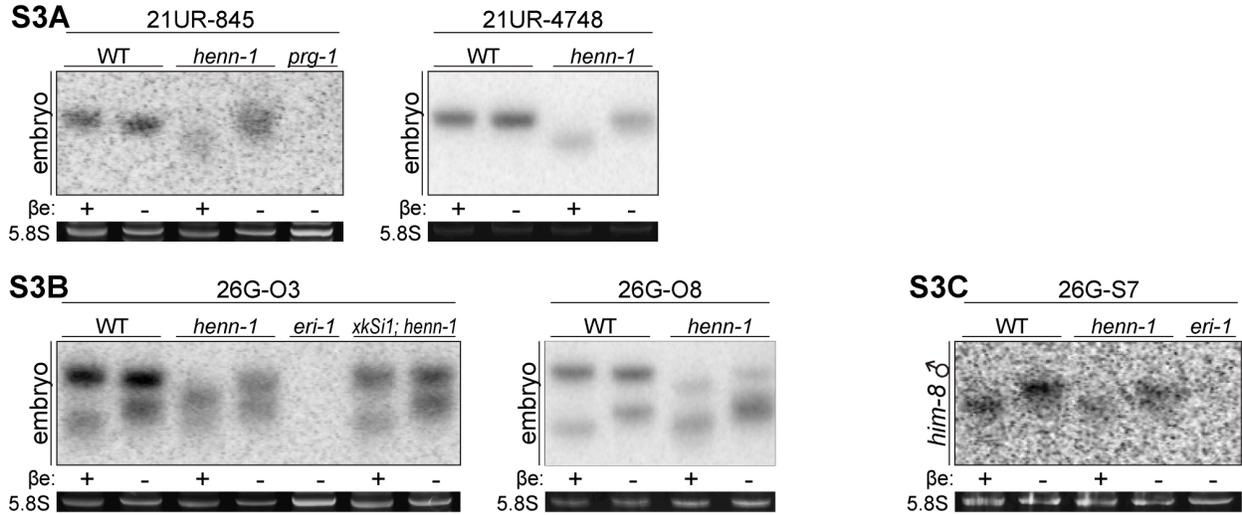
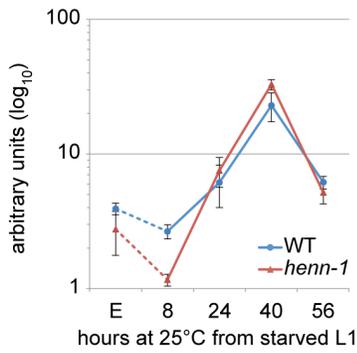
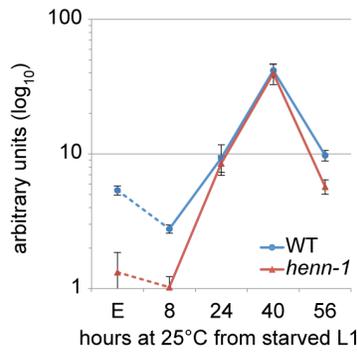


Figure 2.S4: Diverse 21U RNAs Exhibit HENN-1 Dependence in Early Development and Adulthood. A) A panel of additional 21U RNAs exhibit significant defects in accumulation in the *henn-1(tm4477)* mutant. 21U RNA levels were assayed by Taqman qPCR in wild-type and *henn-1(tm4477)* mutant animals at the indicated developmental time points. Standard deviation is shown for biological triplicates. B) 21U RNAs are generally depleted in the *henn-1(tm4477)* mutant relative to wild-type in embryo, early larva, and gravid adult. Abundance in *henn-1(tm4477)* mutant relative to wild-type was calculated for the 21U RNAs shown in A) and Figure 2.2A and the average was plotted for each time point to illustrate the general effect of loss of *henn-1*. C) 21U RNA Taqman assays specifically detect piRNAs. 21U RNA and miRNA levels were assayed in *prg-1(tm872)* mutant embryo biological duplicates. Fold levels relative to wild-type embryo are plotted. E, embryo.

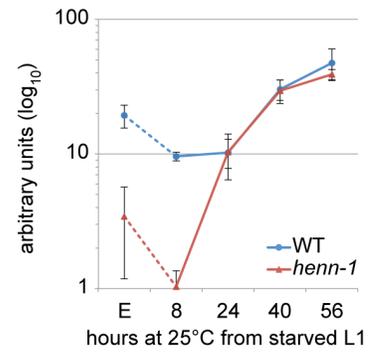
S4A 21UR-845 levels



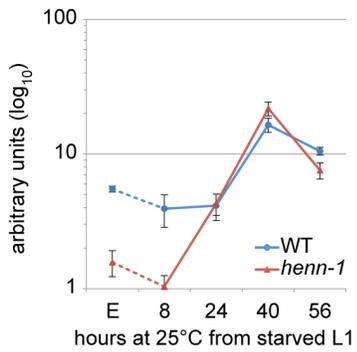
21UR-1063 levels



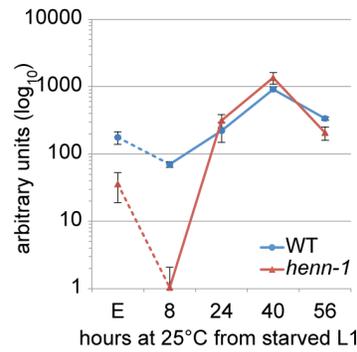
21UR-1267 levels



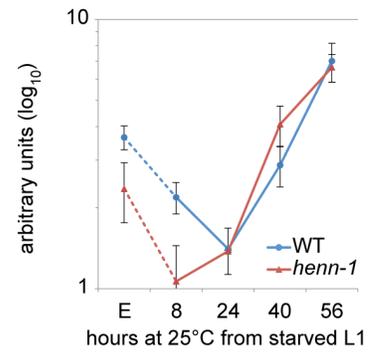
21UR-1343 levels



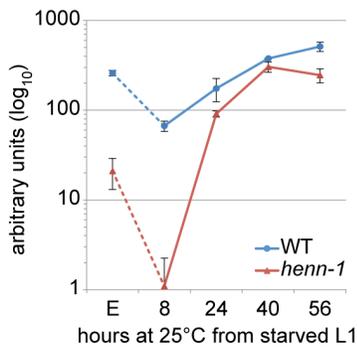
21UR-1832 levels



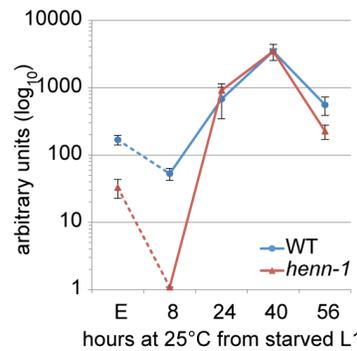
21UR-1838 levels



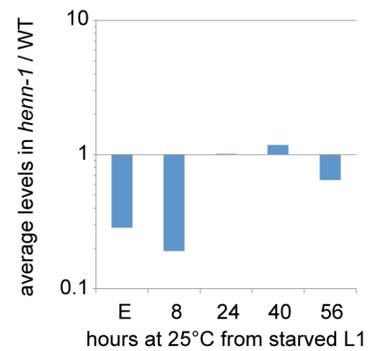
21UR-3129 levels



21UR-5191 levels



S4B 21U RNA levels



S4C levels in embryo

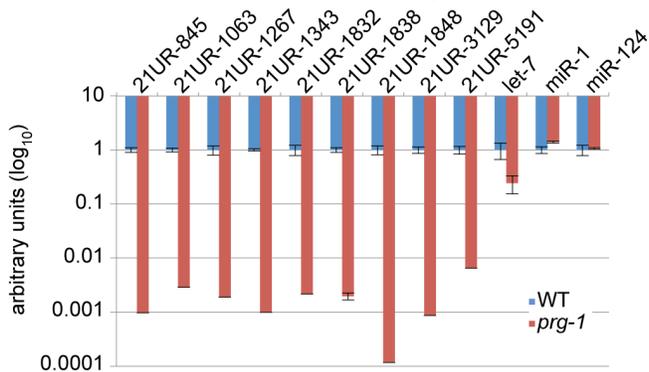


Figure 2.S5: miRNAs Do Not Exhibit HENN-1 Dependence. A) miRNAs are generally unaffected in the *henn-1(tm4477)* mutant. miRNA levels were assayed by Taqman qPCR in wild-type and *henn-1(tm4477)* mutant animals at the developmental time points assessed in Figure 2.S4. Standard deviation is shown for biological triplicates. B) miRNAs are not generally depleted in the *henn-1(tm4477)* mutant relative to wild-type. Abundance in *henn-1(tm4477)* mutant relative to wild-type was calculated for the miRNAs shown in A) and the average was plotted for each time point to illustrate the general effect of loss of *henn-1*. E, embryo.

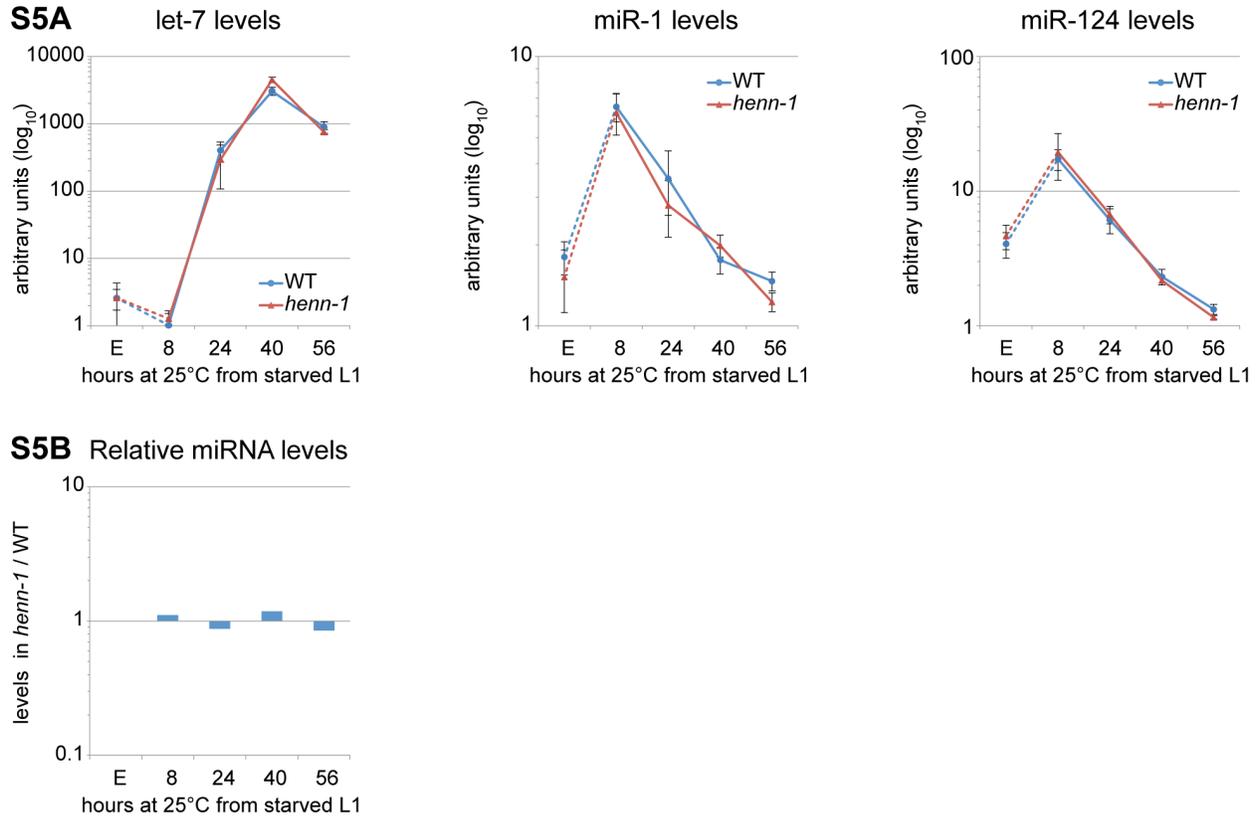


Figure 2.S6: HENN-1 Dependence of Substrate-dependent Secondary siRNAs. A) Levels of a Wago-dependent, 21U RNA-dependent 22G RNA targeting *Tc3* are decreased in *henn-1(tm4477)* mutant embryo ($P=0.0064$; two-tailed *t*-test). Standard deviation is shown for biological triplicates. B) Levels of a Wago-dependent, ERGO-1 class 26G RNA-dependent 22G RNA targeting *E01G4.7* are decreased in *henn-1(tm4477)* mutant embryo ($P=0.044$; two-tailed *t*-test). Standard deviation is shown for biological triplicates.

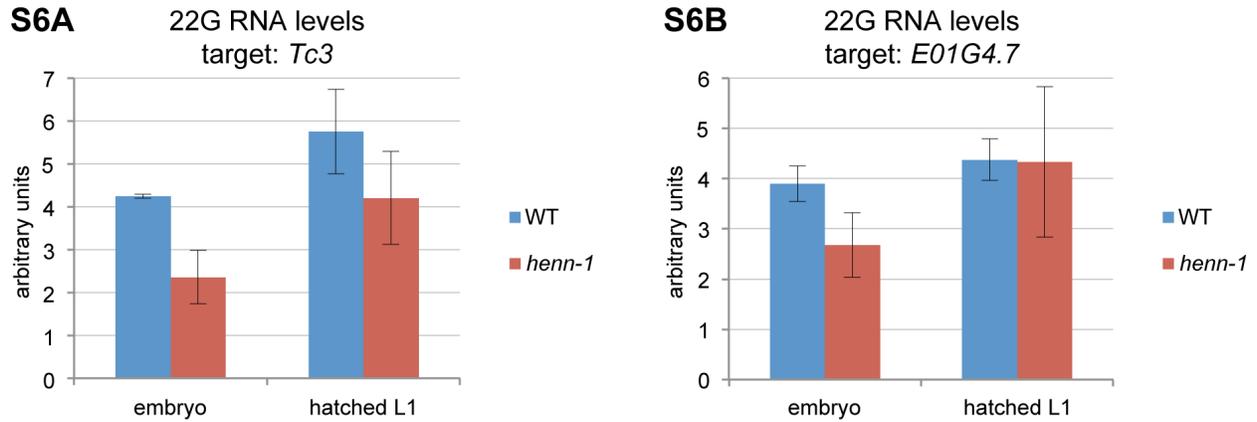


Figure 2.S7: *henn-1* Contributes to Robust Fertility at Elevated Temperatures. *henn-1(tm4477)* mutant animals exhibit a modest fertility defect at 25°C that is rescued by germline-specific expression of *henn-1::gfp* from transgene *xkSi2*. Progeny per animal cultured at 20°C or shifted to 25°C for three generations is plotted for animals of the indicated genotype. Differences between *henn-1(tm4477)* mutant and wild-type or *xkSi2; henn-1(tm4477)* transgenic rescue strain are statistically significant (*: P=0.0059; **: P=0.0130, two-tailed *t*-test). N ≥ 13 animals per strain. Mean and standard deviation are shown.

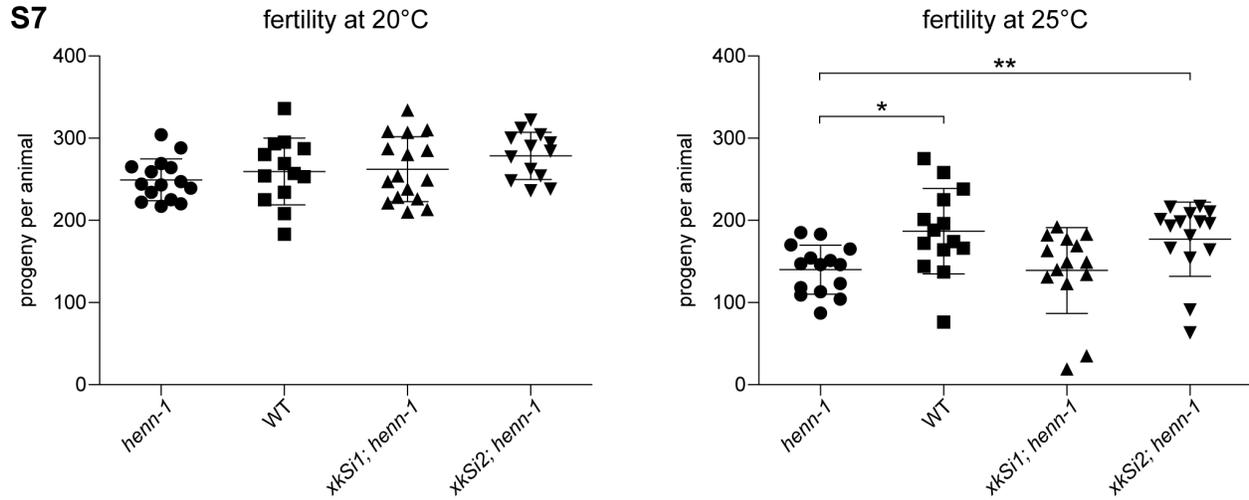
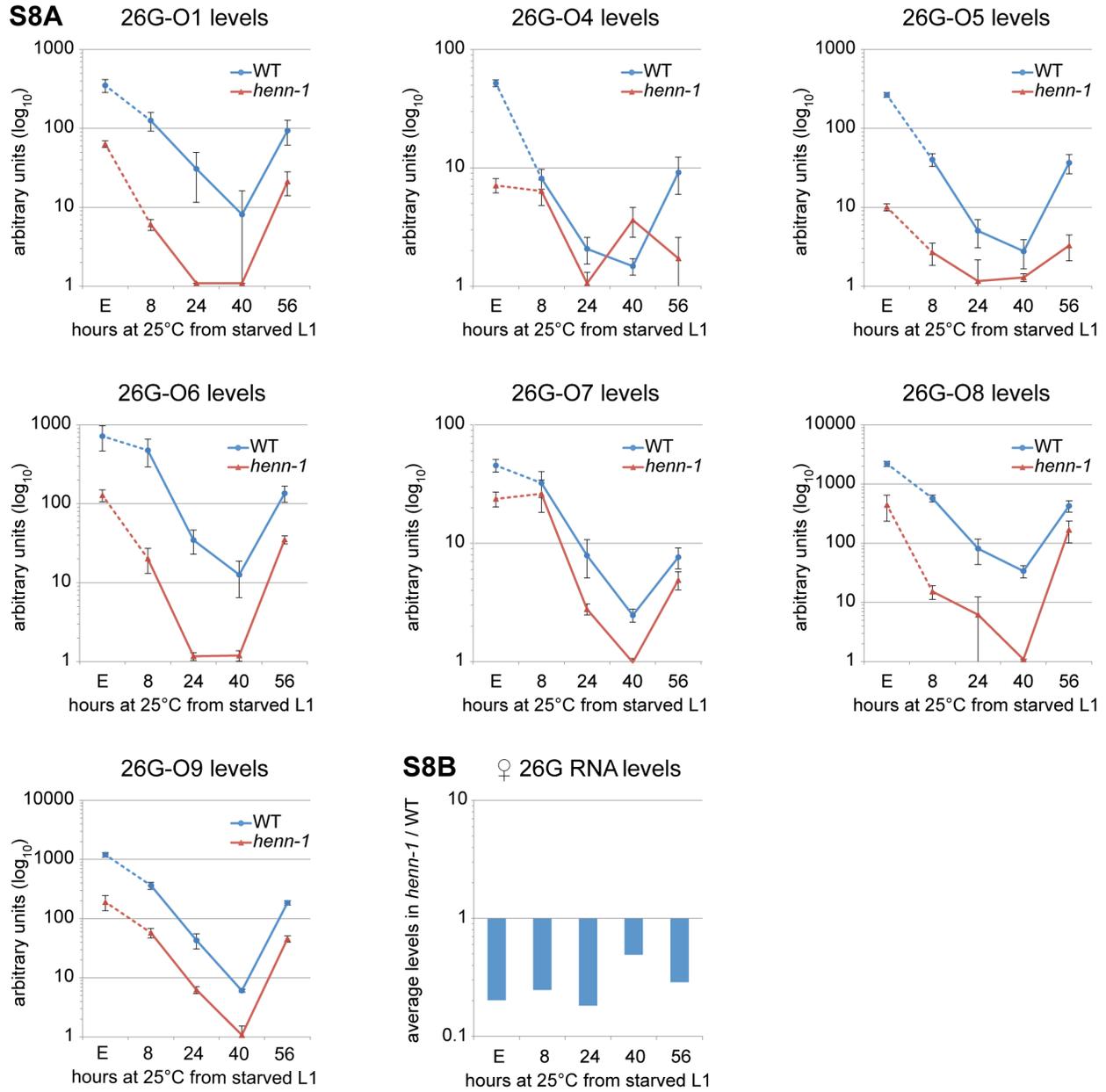


Figure 2.S8: Many ERGO-1 Class 26G RNAs Exhibit HENN-1 Dependence throughout Development. A) A panel of additional ERGO-1 class 26G RNAs exhibit significant defects in accumulation in the *henn-1(tm4477)* mutant. ERGO-1 class 26G RNA levels were assayed by Taqman qPCR in wild-type and *henn-1(tm4477)* mutant animals at the indicated developmental time points. Standard deviation is shown for biological triplicates. B) ERGO-1 class 26G RNAs are generally depleted in the *henn-1(tm4477)* mutant relative to wild-type throughout development. Abundance in *henn-1(tm4477)* mutant relative to wild-type was calculated for the 26G RNAs shown in A) and Figure 2.5A and the average was plotted for each time point to illustrate the general effect of loss of *henn-1*. C) ERGO-1 class 26G RNA Taqman assays specifically detect ERI-1-dependent small RNAs. ERGO-1 class 26G RNA and miRNA levels were assayed in *eri-1(mg366)* mutant embryo biological duplicates. Fold levels relative to wild-type embryo are plotted. E, embryo.

S8A



S8C

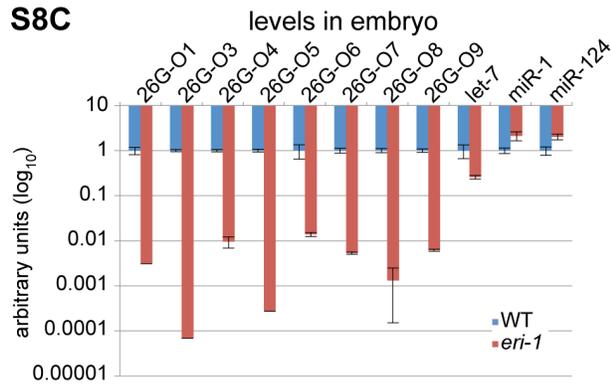


Figure 2.S9: ALG-3/ALG-4 Class 26G RNAs Do Not Exhibit HENN-1 Dependence.

A) Additional ALG-3/ALG-4 class 26G RNAs do not exhibit significant defects in accumulation in the *henn-1(tm4477)* mutant. ALG-3/ALG-4 class 26G RNA levels were assayed by Taqman qPCR in wild-type and *henn-1(tm4477)* mutant animals at the indicated developmental time points. Standard deviation is shown for biological triplicates. B) ALG-3/ALG-4 class 26G RNAs are generally unchanged in the *henn-1(tm4477)* mutant relative to wild-type during their peak expression. Abundance in *henn-1(tm4477)* mutant relative to wild-type was calculated for the 26G RNAs shown in A) and Figure 2.5B and the average was plotted for each time point to illustrate the general effect of loss of *henn-1*. C) ALG-3/ALG-4 class 26G RNA Taqman assays specifically detect ERI-1-dependent small RNAs. ALG-3/ALG-4 class 26G RNA and miRNA levels were assayed in *eri-1(mg366); him-8(e1489)* mutant male biological duplicates. Fold levels relative to wild-type male are plotted. D) miRNAs are generally unaffected in the *henn-1(tm4477)* mutant. miRNA levels were assayed by Taqman qPCR in wild-type and *henn-1(tm4477)* mutant animals at the developmental time points assessed in A. Standard deviation is shown for biological triplicates. B) miRNAs are not generally depleted in *henn-1(tm4477)* mutant relative to wild-type animals. Abundance in *henn-1(tm4477)* mutant relative to wild-type was calculated for the miRNAs shown in D) and the average was plotted for each time point to illustrate the general effect of loss of *henn-1*.

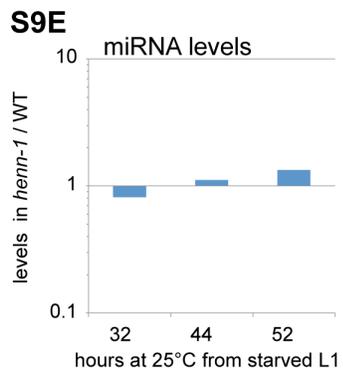
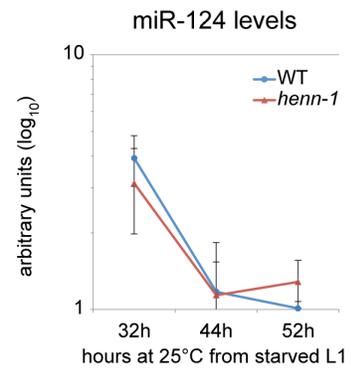
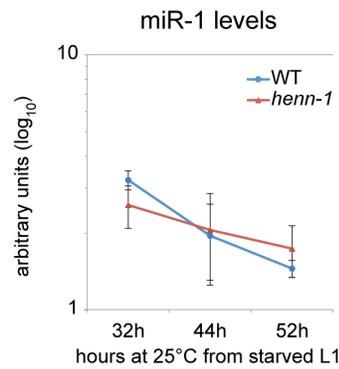
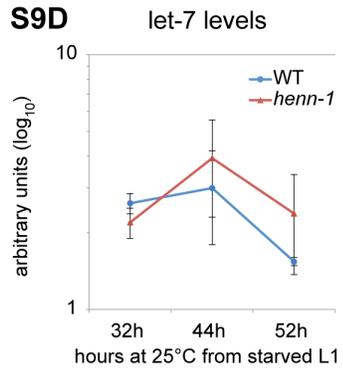
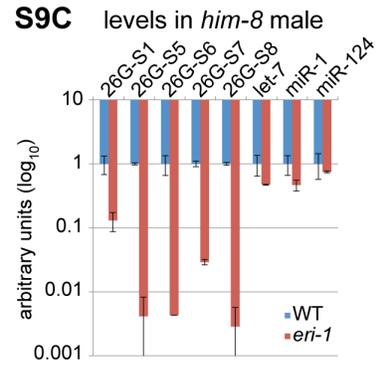
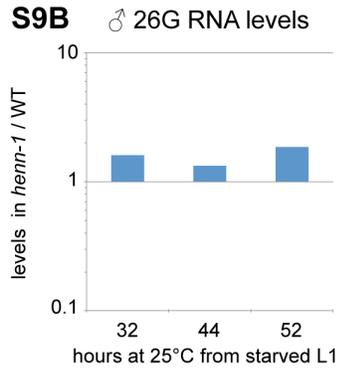
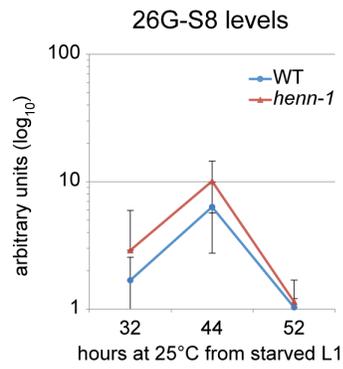
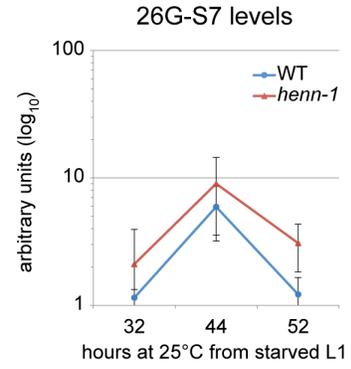
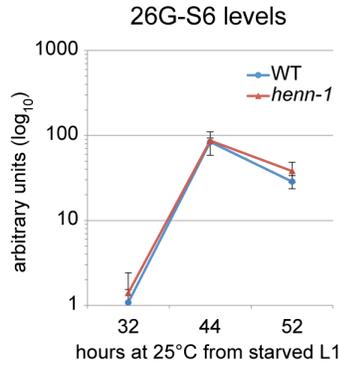
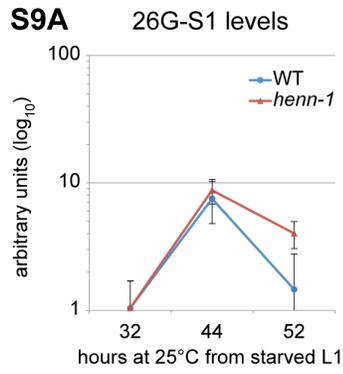


Figure 2.S10: The *henn-1(tm4477)* Mutant Does Not Exhibit Significant Upregulation of ERGO-1 Class 26G RNA Target mRNAs. A) ERGO-1 class 26G RNA target mRNAs show only sporadic HENN-1 dependence. Data is summarized in Figure 2.5C. Levels of eight ERGO-1 class 26G RNA targets were assayed across development of wild-type and *henn-1(tm4477)* mutant animals at 25°C and normalized to mRNA levels of *eft-2*, an abundantly expressed housekeeping gene. Standard deviation is shown for biological triplicates. B) Non-target mRNAs do not show upregulation in the *henn-1(tm4477)* mutant relative to wild-type. Levels of two non-target mRNAs were assayed across development of wild-type and *henn-1(tm4477)* mutant animals at 25°C and normalized to *eft-2*. Standard deviation is shown for biological triplicates. E, embryo.

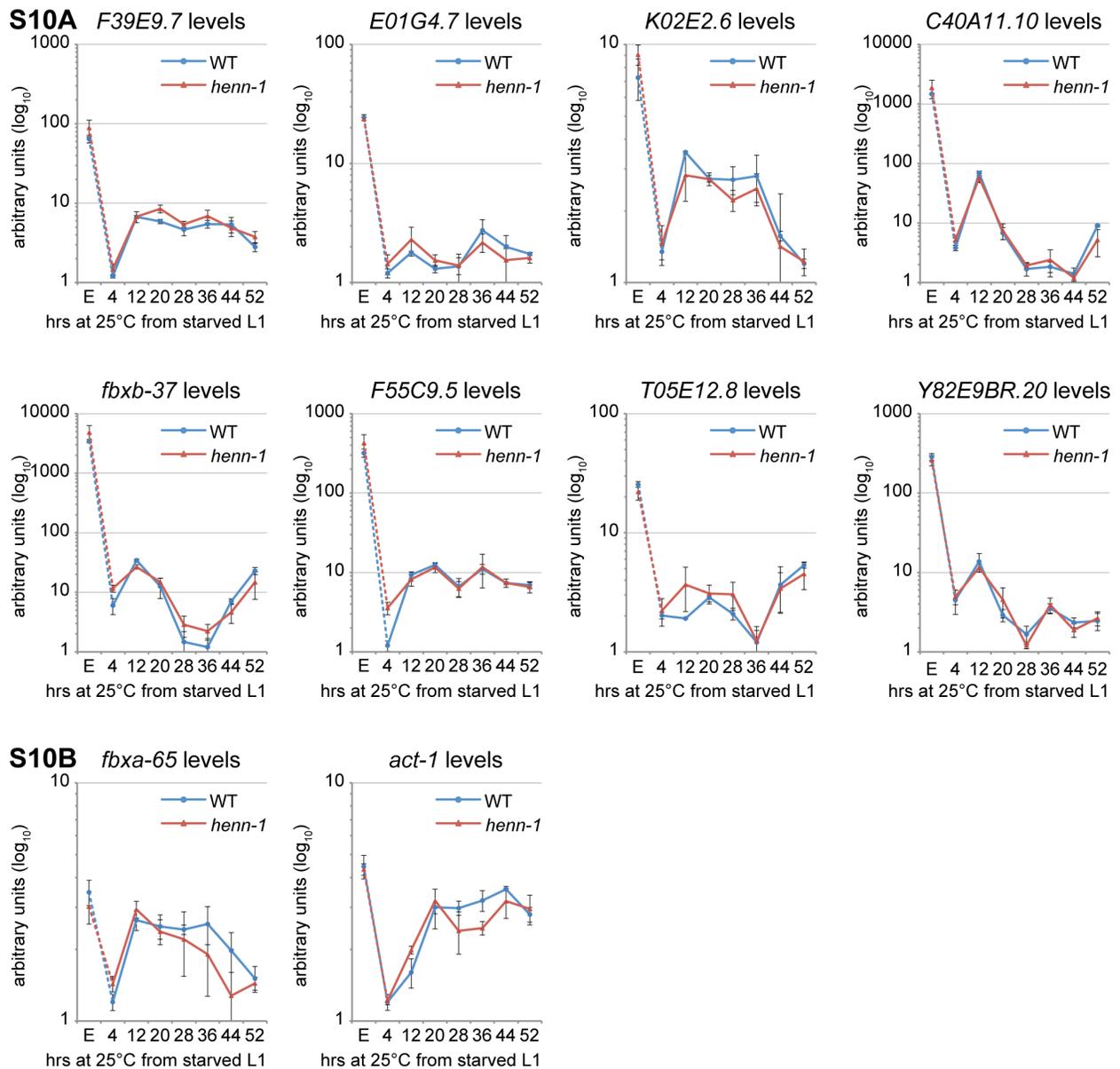
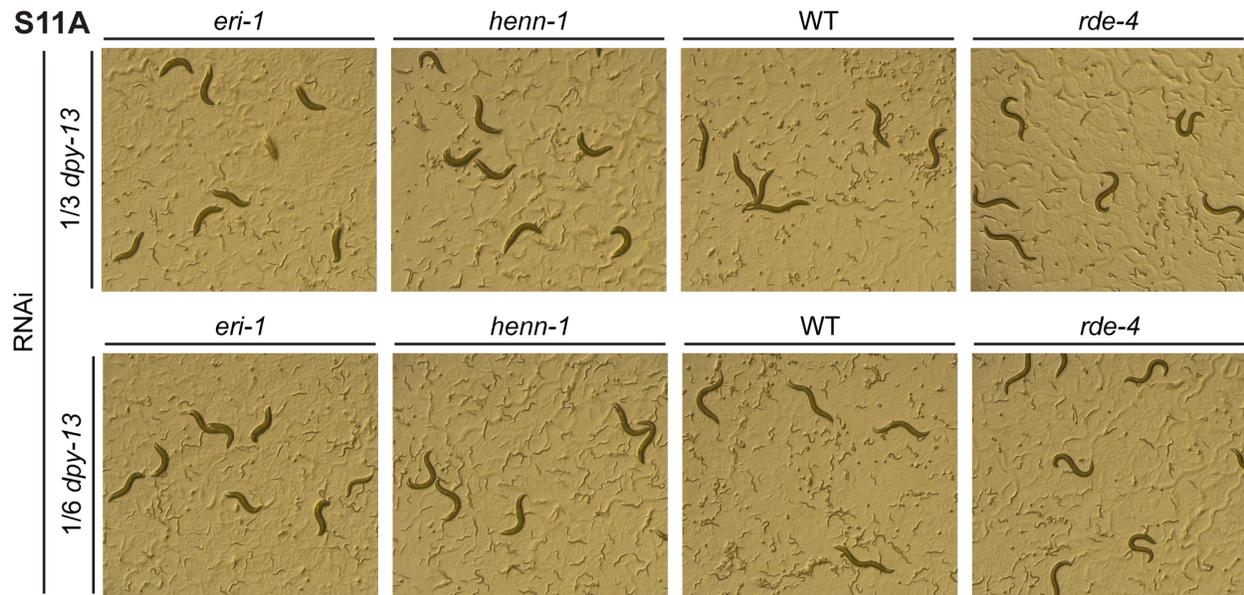


Figure 2.S11: The *henn-1(tm4477)* Mutant Exhibits a Mild but General Somatic Eri Phenotype. A) *henn-1(tm4477)* mutant animals are weakly somatic Eri to RNAi knockdown of *dpy-13*. Animals of the indicated genotypes were plated as L1 larvae on *dpy-13* feeding RNAi diluted 1:2 or 1:5 (1/3 or 1/6 strength) with empty vector and grown for 90 hours at 20°C. *eri-1(mg366)* and *rde-4(ne301)* are included as controls. B) *henn-1(tm4477)* mutant animals are weakly somatic Eri to RNAi knockdown of *lin-29*. Animals of the indicated genotypes were plated as L1 larvae on *lin-29* feeding RNAi diluted 1:2 (1/3 strength) or 1:5 (1/6 strength) with empty vector and grown for 70 hours at 20°C. Percent of animals reaching full size without exhibiting protruding vulva or bursting is plotted. N = 4 plates of >50 animals per strain. Mean and standard deviation are shown.



S11B sensitivity to *lin-29* RNAi

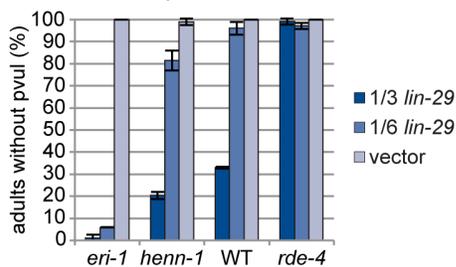


Figure 2.S12: The *henn-1(tm4477)* Mutant Exhibits a General Germline Rde Phenotype. A) *henn-1(tm4477)* mutant animals are Rde to RNAi knockdown of germline genes. Animals of the indicated genotypes were plated as L1 larvae on *par-1*, *par-2*, or *pie-1* feeding RNAi diluted to the indicated strengths with empty vector and grown for 6 days at 20°C. Brood size averaged to the number of P₀ L1s per plate is plotted. N = 4 plates of 4 P₀ animals per strain. Mean and standard deviation are shown. *: P=0.0234; **: P=0.0028; ***: P=0.0151; ****: P=0.0098, two-tailed *t*-test. B) *henn-1(tm4477)* mutant animals are weakly Rde to RNAi knockdown of germline development gene *glp-1*. Animals of the indicated genotypes were plated as L1 larvae on *glp-1* feeding RNAi and grown for 70 hours at 20°C. Percent of animals failing to develop both arms of the germline is plotted. *rde-4(ne301)* is included as a control. N = 4 plates of >50 animals per strain. Mean and standard deviation are shown. †: P=0.0424, two-tailed *t*-test.

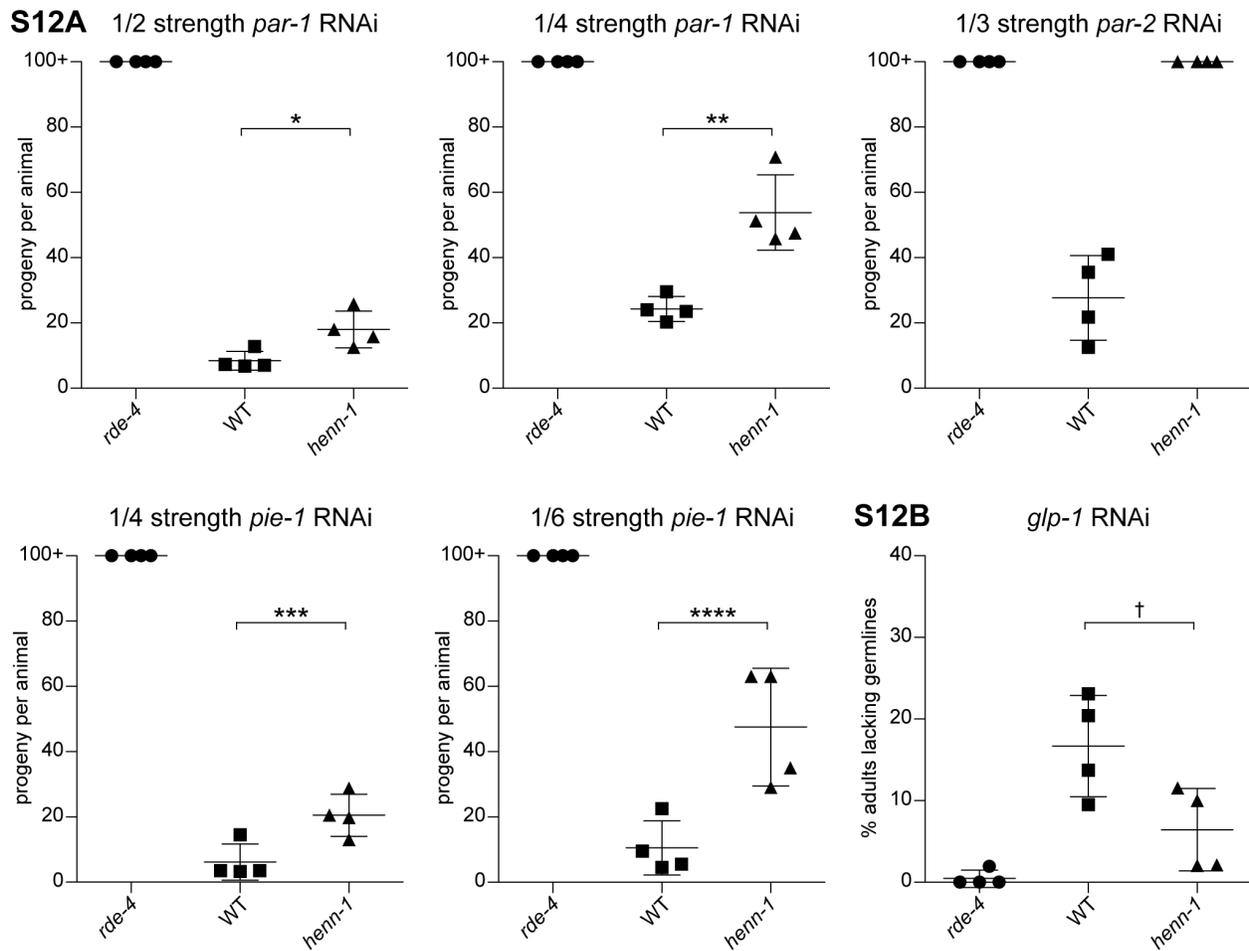


Figure 2.S13: HENN-1 is Broadly Expressed in the Germline and Soma. A) *henn-1* mRNA is highly expressed throughout development. Non-normalized *henn-1* mRNA levels are plotted relative to *eft-2* mRNA levels. The expression profile of *henn-1* is largely unaffected by normalization to *eft-2* (as shown in Figure 2.7A). B) *henn-1* is expressed in germline and soma. Levels of *henn-1* mRNA were assayed in wild-type and *glp-4(bn2)* mutant animals grown for 56 hours at 25°C. C) HENN-1::GFP is broadly expressed in both germline and somatic tissues. HENN-1::GFP was detected in *xkSi1; henn-1(tm4477)* L4 larva but not wild-type control larva using anti-GFP mouse monoclonal antibody. D) ERGO-1 and HENN-1::GFP are generally abundant in early embryo; specificity of anti-ERGO-1 antibody in embryo is shown on right. E) ERGO-1 shows cytoplasmic enrichment in the hermaphrodite proximal germline. Extruded gonads of *xkSi1; henn-1(tm4477)* adult hermaphrodite were stained with anti-GFP and anti-HENN-1 antibodies. Staining of *ergo-1(tm1860)* mutant demonstrates specificity of anti-ERGO-1 antibody (right). F) GFP::ALG-3 expression overlaps with that of HENN-1 (inset: residual bodies). Extruded gonads of *gfp::alg-3* transgenic adult males were stained with anti-GFP and anti-HENN-1 antibodies. E, embryo.

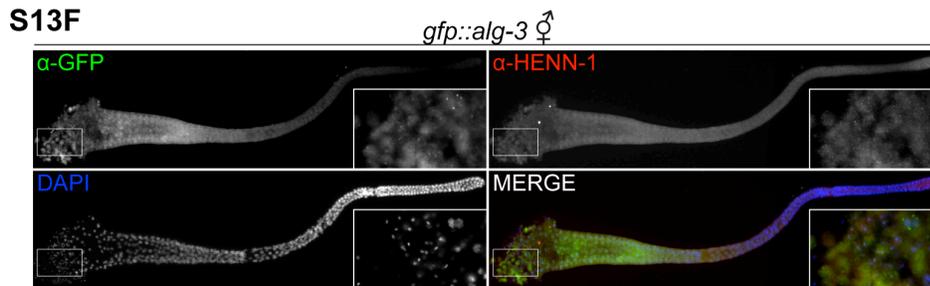
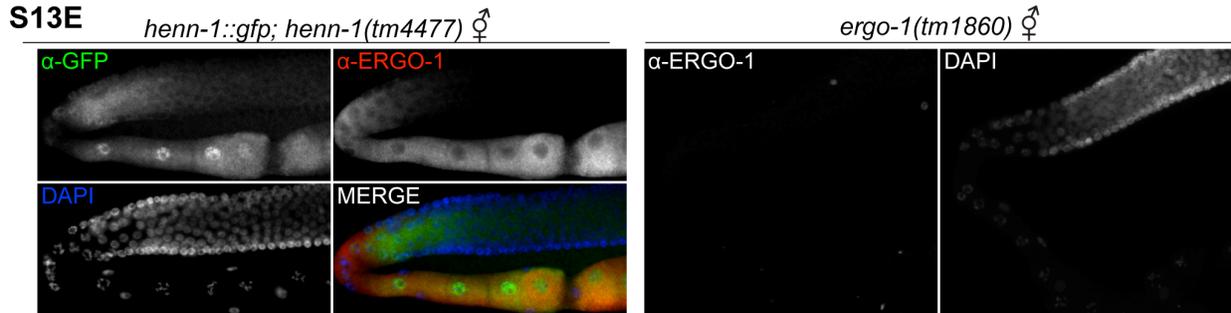
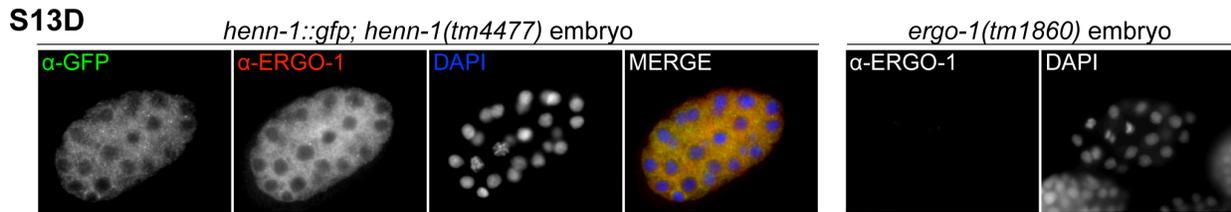
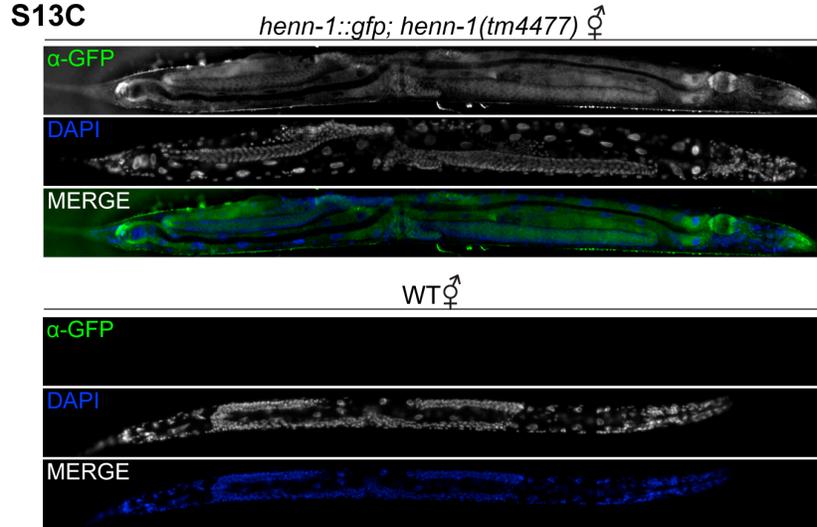
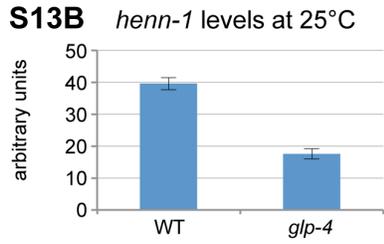
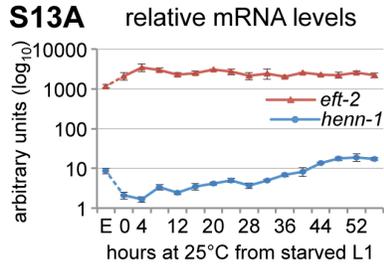
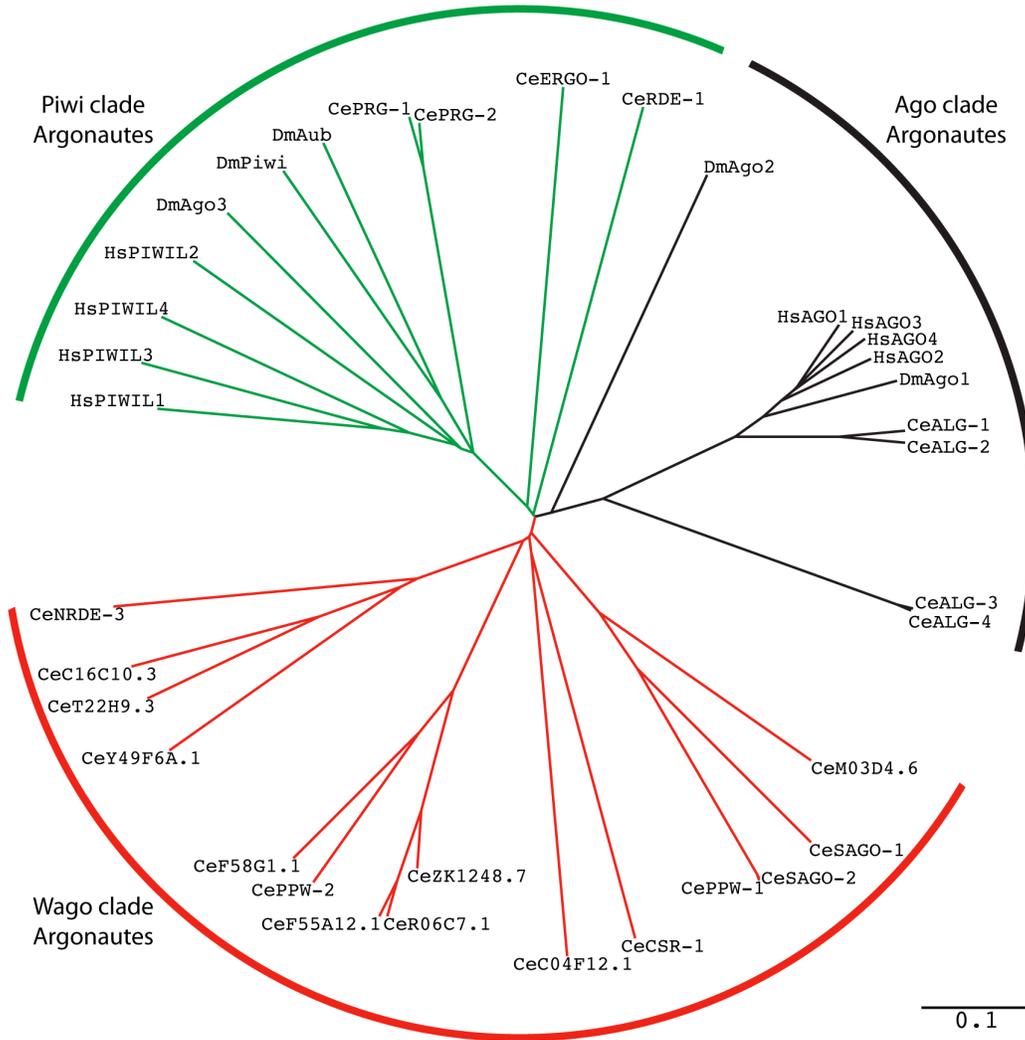


Figure 2.S14: Comparison of *C. elegans* Argonautes. A) Phylogram of human, fly, and worm Argonautes shows divergence of CeERGO-1, CeRDE-1, and DmAgo2 relative to other members of their clades. Multiple sequence alignment of the longest annotated RefSeq protein sequences was performed using ClustalW with default parameters and visualized using Phylodendron (version 0.8d; <http://www.es.embnnet.org/Doc/phylodendron/>). Scale, 0.1: 0.1 substitutions per site. B) Only Piwi clade Argonautes bear the characteristic PAZ domain insertion. Multiple sequence alignment of select Argonautes was performed using ClustalW with default parameters and cropped to show the context of the PAZ domain insertion between strands $\beta 6$ and $\beta 7$ as annotated by Tian et al. [69]. For each Argonaute, methylation status (MET) of associated small RNAs is indicated at right (Yes, methylated; No, not methylated). Sources: HsAGO1, [68,69] and by analogy to mouse [31]; DmAgo1, [19]; CeALG-1, [19]; CeALG-3, this study; HsPIWIL1, [69] and by analogy to mouse [30,31]; DmPiwi, [31,32]; CePRG-1, [27] and this study, DmAgo2, [9,22]; CeERGO-1, [27,42] and this study.

S14A



S14B

		$\beta 5$	$\alpha 3$	$\beta 6$	$\beta 7$	$\beta 8$	MET
HsAGO1	NP_036331.1	307	VECTVAQYFKQKYNLQLKYPHLPCLQVG	-----	QE QKHTYLPLEVCNIVAGQRCIK	352	No
DmAgo1	NP_725341.1	418	VECTVAKYFLDKYRMKLRYPHLPCLQVG	-----	QEHKHTYLPLEVCNIVAGQRCIK	463	No
CeALG-1	NP_510322.2	452	IECTVAKYFYDKYRIQLKYPHLPCLQVG	-----	QE QKHTYLPPEVCNIVPGQRCIK	497	No
CeALG-3	NP_502218.1	440	VVCSVADYFSEKYG-PLKYPKLPCLHVG	-----	PPTRNIFLPMEHCLIDSPQKYNK	484	No
HsPIWIL1	NP_004755.2	346	SEVSFLEYRQYKQYNEITDLKQPVLSVQPKRRRGGTLPGPAMLIPELCYLTG-LTDKM			397	Yes
DmPiwi	NP_476875.1	328	RDISFVEYYLTKYNI RIRDHNQPLLSK-NRDKALKTNASELVVLIPELQVTVG-LNAEM			378	Yes
CePRG-1	NP_492121.1	288	QSITLKEYFKNQYGIETVDDQPIIISSEGKPKQP--GEPPQVSYIVPELCPPTG-LTDEM			337	Yes
DmAgo2	NP_648775.1	682	KKVTIASYFHSRNY-PLKFPQLHCLNVG-----	SSIKSILLPIELCSIEEGQALNR		723	Yes
CeERGO-1	NP_503362.2	485	VELTVADYYLQKYNIRLRYPNLPCVLKKAPEQCG----	NKHSAMPLELVSYIVVPTRYGG		541	Yes
CeRDE-1	NP_741611.1	380	KYDTTLFKIYEENKFFIEFPHLPLVKVKS-----	GAKEYAVPMEHLEVHEKQRYKN		427	?

Table 2.S1: Oligonucleotides for Northern Blot Analysis. Oligonucleotides corresponding to the antisense sequences of small RNAs were synthesized by Integrated DNA Technologies and used for small RNA detection by northern blot.

Small RNA target	Probe sequence
cel-miR-1	5' TACATACTTCTTTACATTCCA /3StarFire/ 3'
21UR-845	5' TTCGAGTTCTTGCTTTCCTGA /3StarFire/ 3'
21UR-4292	5' CCATTCTTTGTCACCCTCGTA /3StarFire/ 3'
21UR-4748	5' TAGCCAGTACTCTACGTTGTA /3StarFire/ 3'
21UR-5941	5' ATTAACCGTTTCGTGCCCGAA /3StarFire/ 3'
26G-O1	5' TTGAAAATAATCTACCGTTTCTGAGC /3StarFire/ 3'
26G-O3	5' AAAAGTATCCGACTTTCGAGTTTGTC /3StarFire/ 3'
26G-O7	5' TTCCACGATCAGAAGGGATGTCCTC /3StarFire/ 3'
26G-O8	5' TGCTGCGAAAAGTGTGGATTTCTAC /3StarFire/ 3'
26G-S5	5' TACCATGTCGCTCACTGCTGATCCAC /3StarFire/ 3'
26G-S7	5' CGATGATCATATTCTACTTCATTTTC /3StarFire/ 3'

Table 2.S2: Small RNA Sequences for Taqman Probe Design. Sequences of the indicated small RNAs were submitted to Applied Biosystems for Taqman small RNA probe design and synthesis.

Small RNA	Sequence (5' to 3')
let-7	TGAGGTAGTAGGTTGTATAGTT
miR-1	TGGAATGTAAAGAAGTATGTA
miR-124	TAAGGCACGCGGTGAATGCCA
21UR-845	TCAGGAAAGCAAGAACTCGAA
21UR-1063	TGAGCGCATTGTATACACTG
21UR-1267	TAGGAACGAAATGAACAAAAT
21UR-1343	TGAAGGAAGAGTACGAACTT
21UR-1832	TTTAACAAATGACGGTAAATC
21UR-1838	TTGTTCTTCGTTCCGGTCCAAA
21UR-1848	TAAAGGCAGAAATTTTATCAAC
21UR-3129	TGTATGTAAAACCTTACGGCA
21UR-5191	TGTAAAAAGTTTTTGTATGTA
22G targeting <i>Tc3</i>	GAATCAGAACCAGTCTGGAGAT
22G targeting <i>E01G4.7</i>	GAGTGACATCCCTTCTGATCGT
26G-O1	GCTCAGAAACGGTAGATTATTTTCAA
26G-O3	GACAAACTCGAAAGTCGGATACTTTT
26G-O4	GAGGGGATAAGAGCTCGTCCGATGGC
26G-O5	GATGGGAATGCAGAAGAAAAGAGGGG
26G-O6	GTAGAAGGATTCATCTGGCATTTCAT
26G-O7	GAGTGACATCCCTTCTGATCGTGGAA
26G-O8	GATGAATCGTCGATAGAAAGACAAAC
26G-O9	GTAGGAAATCCACAGTTTTTCGCAGCA
26G-S1	GCTATGGAGGACGAGAATACATAATT
26G-S5	GTGGATCAGCAGTGAGCGACATGGTA
26G-S6	GACTCTTCGACTTCGGCATTTCGCGGA
26G-S7	GAAGAACGAAAATTTGAAGATGTATA
26G-S8	GAAAATGAAGTAGAATATGATCATCG

Table 2.S3: Primers for RT-qPCR. RT-qPCR primers for detection of the indicated gene targets were synthesized by Integrated DNA Technologies.

Gene target	Forward (5' to 3')	Reverse (5' to 3')
<i>act-1</i>	CCAGGAATTGCTGATCGTATGCAGAA	TGGAGAGGGAAGCGAGGATAGA
<i>C40A11.10</i>	AATGGCTCCTTGAAAAGATCG	TACATTTCCGCCACGTTGAAA
<i>E01G4.7</i>	GCACAAGGTTTCGTTCTTGGTG	AGTGACATCCCTTCTGATCG
<i>eft-2</i>	TGTGTTTCCGGAGTGTGTGT	CCATCGTCGTCTCCGTAAGT
<i>F39E9.7</i>	CCCAGTGGCCCAATTAACG	GCACAAGGTTTCGTTCTTGGTG
<i>F55C9.5</i>	ACCATTGGAGCACGTAAATCAA	GGTCCTAATAATAAAGTTGCGTCG
<i>fbxa-65</i>	ACTTACAAGGATCAAGAAAAGCG	CCTTGACCGCTATTCCGAGAAA
<i>fbxb-37</i>	CATAAGTCCTGGAAGCCATACTCC	ATCTTTCGATACGATGTATGTTCCG
<i>K02E2.6</i>	CAGTGGTACAAGTGGGAGTAAACG	AATTGGCAAGTAACTGATTCCG
<i>rem-1</i>	GGAAGAGGGATGTGTTCAACG	TTCCAGTGCTGATGCGATCATA
<i>ssp-16</i>	GTCATCAAACAACAATGAGTACCG	GCTCCAGCAGTGCGAGTGAT
<i>T05E12.8</i>	TTCCATTTGAGGATTTTGCTACG	ATTATTTGGATGGCAGCCGATG
<i>Tc3</i>	GAGCGTTCACGGAGAAGAAG	AATAGTCGCGGGTTGAGTTG
<i>Y82E9BR.20</i>	CTCCCCTTTCTTGATGTATTG	AGTCCGAACATCAAAGCAG

REFERENCES

1. Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75: 843-854.
2. Wightman B, Ha I, Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* 75: 855-862.
3. Kim Y-K, Heo I, Kim VN (2010) Modifications of small RNAs and their associated proteins. *Cell* 143: 703-709.
4. Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, et al. (2007) Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* 315: 1137-1140.
5. Kawahara Y, Megraw M, Kreider E, Iizasa H, Valente L, et al. (2008) Frequency and fate of microRNA editing in human brain. *Nucleic Acids Research* 36: 5270-5280.
6. Kawahara Y, Zinshteyn B, Chendrimada TP, Shiekhattar R, Nishikura K (2007) RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. *EMBO Rep* 8: 763-769.
7. Yang W, Chendrimada TP, Wang Q, Higuchi M, Seeburg PH, et al. (2006) Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol* 13: 13-21.
8. Hundley HA, Bass BL (2010) ADAR editing in double-stranded UTRs and other noncoding RNA sequences. *Trends in biochemical sciences* 35: 377-383.
9. Kawamura Y, Saito K, Kin T, Ono Y, Asai K, et al. (2008) *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* 453: 793-797.
10. Nejepinska J, Malik R, Filkowski J, Flemr M, Filipowicz W, et al. (2012) dsRNA expression in the mouse elicits RNAi in oocytes and low adenosine deamination in somatic cells. *Nucleic Acids Research* 40: 399-413.
11. Shen B, Goodman HM (2004) Uridine addition after microRNA-directed cleavage. *Science* 306: 997.
12. van Wolfswinkel JC, Claycomb JM, Batista PJ, Mello CC, Berezikov E, et al. (2009) CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. *Cell* 139: 135-148.
13. Ameres SL, Horwich MD, Hung J-H, Xu J, Ghildiyal M, et al. (2010) Target RNA-directed trimming and tailing of small silencing RNAs. *Science* 328: 1534-1539.
14. Ibrahim F, Rymarquis LA, Kim EJ, Becker J, Balassa E, et al. (2010) Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in *Chlamydomonas*. *Proc Natl Acad Sci U S A* 107: 3906-3911.
15. Kamminga LM, Luteijn MJ, den Broeder MJ, Redl S, Kaaij LJ, et al. (2010) Hen1 is required for oocyte development and piRNA stability in zebrafish. *EMBO J* 29: 3688-3700.
16. Li J, Yang Z, Yu B, Liu J, Chen X (2005) Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Curr Biol* 15: 1501-1507.
17. Tkaczuk KL, Obarska A, Bujnicki JM (2006) Molecular phylogenetics and comparative modeling of HEN1, a methyltransferase involved in plant microRNA biogenesis. *BMC Evol Biol* 6: 6.

18. Park W, Li J, Song R, Messing J, Chen X (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr Biol* 12: 1484-1495.
19. Yu B, Yang Z, Li J, Minakhina S, Yang M, et al. (2005) Methylation as a crucial step in plant microRNA biogenesis. *Science* 307: 932-935.
20. Yang Z, Ebright YW, Yu B, Chen X (2006) HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res* 34: 667-675.
21. Vilkaitis G, Plotnikova A, Klimasauskas S (2010) Kinetic and functional analysis of the small RNA methyltransferase HEN1: the catalytic domain is essential for preferential modification of duplex RNA. *RNA* 16: 1935-1942.
22. Horwich MD, Li C, Matranga C, Vagin V, Farley G, et al. (2007) The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr Biol* 17: 1265-1272.
23. Saito K, Sakaguchi Y, Suzuki T, Siomi H, Siomi MC (2007) Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. *Genes Dev* 21: 1603-1608.
24. Kurth HM, Mochizuki K (2009) 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in *Tetrahymena*. *RNA* 15: 675-685.
25. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, et al. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128: 1089-1103.
26. Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, et al. (2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* 129: 69-82.
27. Ruby JG, Jan C, Player C, Axtell MJ, Lee W, et al. (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* 127: 1193-1207.
28. Girard A, Sachidanandam R, Hannon GJ, Carmell MA (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442: 199-202.
29. Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, et al. (2006) Characterization of the piRNA complex from rat testes. *Science* 313: 363-367.
30. Kirino Y, Mourelatos Z (2007) Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. *Nat Struct Mol Biol* 14: 347-348.
31. Ohara T, Sakaguchi Y, Suzuki T, Ueda H, Miyauchi K (2007) The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated. *Nat Struct Mol Biol* 14: 349-350.
32. Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, et al. (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313: 320-324.
33. Kirino Y, Mourelatos Z (2007) The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. *RNA* 13: 1397-1401.
34. Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, et al. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403: 901-906.

35. Batista PJ, Ruby JG, Claycomb JM, Chiang R, Fahlgren N, 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.
36. Das PP, Bagijn MP, Goldstein LD, Woolford JR, Lehrbach NJ, 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.
37. Wang G, Reinke V (2008) A *C. elegans* Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. *Current biology* : CB 18: 861-867.
38. Cox DN, Chao A, Baker J, Chang L, Qiao D, et al. (1998) A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes & development* 12: 3715-3727.
39. Ambros V, Lee RC, Lavanway A, Williams PT, Jewell D (2003) MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr Biol* 13: 807-818.
40. Han T, Manoharan AP, Harkins TT, Bouffard P, Fitzpatrick C, et al. (2009) 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* 106: 18674-18679.
41. Conine CC, Batista PJ, Gu W, Claycomb JM, Chaves DA, et al. (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* 107: 3588-3593.
42. Vasale JJ, Gu W, Thivierge C, Batista PJ, Claycomb JM, et al. (2010) Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. *Proc Natl Acad Sci U S A* 107: 3582-3587.
43. Yigit E, Batista PJ, Bei Y, Pang KM, Chen CC, et al. (2006) Analysis of the *C. elegans* Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* 127: 747-757.
44. Zhang C, Montgomery TA, Gabel HW, Fischer SE, Phillips CM, et al. (2011) mut-16 and other mutator class genes modulate 22G and 26G siRNA pathways in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America* 108: 1201-1208.
45. Gu W, Shirayama M, Conte D, Jr., Vasale J, Batista PJ, et al. (2009) Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. *Mol Cell* 36: 231-244.
46. Sijen T, Steiner FA, Thijssen KL, Plasterk RHA (2007) Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science* 315: 244-247.
47. Pak J, Fire A (2007) Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* 315: 241-244.
48. Okamura K, Chung WJ, Ruby JG, Guo H, Bartel DP, et al. (2008) The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* 453: 803-806.
49. Huang Y, Ji L, Huang Q, Vassilyev DG, Chen X, et al. (2009) Structural insights into mechanisms of the small RNA methyltransferase HEN1. *Nature* 461: 823-827.
50. Ameres SL, Hung J-H, Xu J, Weng Z, Zamore PD (2011) Target RNA-directed tailing and trimming purifies the sorting of endo-siRNAs between the two *Drosophila* Argonaute proteins. *RNA* 17: 54-63.

51. Yang Z, Vilkaitis G, Yu B, Klimasauskas S, Chen X (2007) Approaches for studying microRNA and small interfering RNA methylation in vitro and in vivo. *Meth Enzymol* 427: 139-154.
52. Frokjaer-Jensen C, Davis MW, Hopkins CE, Newman BJ, Thummel JM, et al. (2008) Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nat Genet* 40: 1375-1383.
53. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, et al. (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res* 33: e179.
54. Montgomery TA, Rim Y-S, Zhang C, Downen RH, Phillips CM, et al. (2012) PIWI associated siRNAs and piRNAs specifically require the *Caenorhabditis elegans* HEN1 ortholog henn-1. *PLoS Genetics*.
55. Duchaine TF, Wohlschlegel JA, Kennedy S, Bei Y, Conte D, Jr., et al. (2006) Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. *Cell* 124: 343-354.
56. Lee RC, Hammell CM, Ambros V (2006) Interacting endogenous and exogenous RNAi pathways in *Caenorhabditis elegans*. *RNA* 12: 589-597.
57. Tabara H, Hill RJ, Mello CC, Priess JR, Kohara Y (1999) pos-1 encodes a cytoplasmic zinc-finger protein essential for germline specification in *C. elegans*. *Development* 126: 1-11.
58. Reinke V, Gil IS, Ward S, Kazmer K (2004) Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. *Development* 131: 311-323.
59. Gent JI, Lamm AT, Pavelec DM, Maniar JM, Parameswaran P, et al. (2010) Distinct phases of siRNA synthesis in an endogenous RNAi pathway in *C. elegans* soma. *Molecular Cell* 37: 679-689.
60. Kawaoka S, Izumi N, Katsuma S, Tomari Y (2011) 3' end formation of PIWI-interacting RNAs in vitro. *Molecular Cell* 43: 1015-1022.
61. Forstemann K, Horwich MD, Wee L, Tomari Y, Zamore PD (2007) *Drosophila* microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* 130: 287-297.
62. Kamminga LM, van Wolfswinkel J, Luteijn MJ, Kaaij LJ, Bagjin MP, et al. (2012) Differential impact of the Hen1 homolog HENN-1 on 21U and 26G RNAs in the germline of *Caenorhabditis elegans*. *PLoS Genetics*.
63. Pelisson A, Sarot E, Payen-Groschene G, Bucheton A (2007) A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the *Drosophila* ovary. *Journal of Virology* 81: 1951-1960.
64. Cerutti L, Mian N, Bateman A (2000) Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. *Trends Biochem Sci* 25: 481.
65. Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, et al. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305: 1437-1441.
66. Song JJ, Smith SK, Hannon GJ, Joshua-Tor L (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305: 1434-1437.
67. Lingel A, Simon B, Izaurralde E, Sattler M (2004) Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat Struct Mol Biol* 11: 576-577.

68. Ma JB, Ye K, Patel DJ (2004) Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* 429: 318-322.
69. Tian Y, Simanshu DK, Ma JB, Patel DJ (2011) Structural basis for piRNA 2'-O-methylated 3'-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. *Proceedings of the National Academy of Sciences* 108: 903.
70. Boland A, Huntzinger E, Schmidt S, Izaurralde E, Weichenrieder O (2011) Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. *Proceedings of the National Academy of Sciences of the United States of America* 108: 10466-10471.
71. L'Hernault SW, Roberts TM (1995) Cell biology of nematode sperm. *Methods in cell biology* 48: 273-301.
72. Pall GS, Hamilton AJ (2008) Improved northern blot method for enhanced detection of small RNA. *Nature protocols* 3: 1077-1084.
73. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, et al. (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421: 231-237.
74. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, et al. (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Research* 33: e179.
75. Nolan T, Hands RE, Bustin SA (2006) Quantification of mRNA using real-time RT-PCR. *Nature protocols* 1: 1559-1582.
76. Allen MA, Hillier LW, Waterston RH, Blumenthal T (2011) A global analysis of *C. elegans* trans-splicing. *Genome research* 21: 255-264.
77. Crittenden S, Kimble J (2009) Preparation and immunolabeling of *Caenorhabditis elegans*. *Cold Spring Harbor protocols* 2009: pdb prot5216.
78. Girard LR, Fiedler TJ, Harris TW, Carvalho F, Antoshechkin I, et al. (2007) WormBook: the online review of *Caenorhabditis elegans* biology. *Nucleic Acids Research* 35: D472-475.
79. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology* 302: 205-217.
80. Poirot O, O'Toole E, Notredame C (2003) Tcoffee@igs: A web server for computing, evaluating and combining multiple sequence alignments. *Nucleic Acids Research* 31: 3503-3506.
81. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948.
82. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, et al. (2010) A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research* 38: W695-699.

CHAPTER THREE: A conserved upstream motif orchestrates autonomous, germline-enriched expression of *Caenorhabditis elegans* piRNAs

AUTHORS: Billi AC†, Freeberg MA†, Day AM, Chun SY, Khivansara V, Kim JK

† These authors contributed equally to this work

AUTHOR CONTRIBUTIONS: MAF, ACB, and JKK conceived and designed the experiments. AMD generated the high-copy and MosSCI ♂Tg1258, ♂Scram1258, ♂C>A1258, ♀Tg2502, ♀Scram2502, and ♀A>C2502 transgenic strains and performed the initial analyses of the high-copy transgenic strains. ACB generated the high-copy ♂21U>A1258, ♂21U>G1258, ♂Min1258, ♀21U>A2502, ♀21U>G2502, ♀Min2502, Tg1415-Tg2109, and Scram1415-Tg2109 transgenic strains and the MosSCI ♀Min2502 transgenic strain. ACB performed the experiments and analyses in Figures 3.3B-E,H,I, 3.4, 3.5B,C,F, 3.6, 3.S2C-E, 3.S4, and 3.S6. VK performed the experiments in Figures 3.3F,G and 3.S5. MAF performed the analyses in Figures 3.1, 3.2, 3.5D,E, 3.S1, 3.S2A,B,F,G, 3.S3, Tables 3.S1 and 3.S2, and Dataset 3.S1. SYC performed the analyses in Figure 3.S7. ACB and MAF wrote the manuscript. ACB, MAF, and JKK edited the manuscript.

CITATION: Billi AC* and Freeberg MA*, Day AM, Chun SY, Khivansara V, Kim JK. A conserved upstream motif drives autonomous, germline-specific expression of *C. elegans* piRNAs. *PLoS Genetics* 9, e1003392 (2013). * these authors contributed equally to this work.

ABSTRACT: Piwi-interacting RNAs (piRNAs) fulfill a critical, conserved role in defending the genome against foreign genetic elements. In many organisms, piRNAs

appear to be derived from processing of a long, polycistronic RNA precursor. Here, we establish that each *Caenorhabditis elegans* piRNA represents a tiny, autonomous transcriptional unit. Remarkably, the minimal *C. elegans* piRNA cassette requires only a 21 nucleotide (nt) piRNA sequence and an ~50 nt upstream motif with limited genomic context for expression. Combining computational analyses with a novel, in vivo transgenic system, we demonstrate that this upstream motif is necessary for independent expression of a germline-enriched, Piwi-dependent piRNA. We further show that a single nucleotide position within this motif directs differential germline enrichment. Accordingly, over 70% of *C. elegans* piRNAs are selectively expressed in male or female germline, and comparison of the genes they target suggests that these two populations have evolved independently. Together, our results indicate that *C. elegans* piRNA upstream motifs act as independent promoters to specify which sequences are expressed as piRNAs, how abundantly they are expressed, and in what germline. As the genome encodes well over 15,000 unique piRNA sequences, our study reveals that the number of transcriptional units encoding piRNAs rivals the number of mRNA coding genes in the *C. elegans* genome.

AUTHOR SUMMARY: Across the animal kingdom, Piwi-interacting small RNAs (piRNAs) protect genome integrity and promote fertility. While the functions of piRNAs are well-characterized, far less is known about how they are generated and how their expression is regulated. In the *Caenorhabditis elegans* genome, a conserved sequence motif lies upstream of many piRNA loci and appears to regulate their expression. We combined computational and experimental approaches to investigate the role of this motif in the expression of *C. elegans* piRNAs. We discovered that >70% of piRNAs are differentially enriched in male versus female germline, and these male and female piRNAs show different upstream motifs. Using a transgenic system for expressing synthetic piRNAs in vivo, we demonstrate that variation of a single nucleotide within this motif influences piRNA germline enrichment. We further show that the conserved motif is capable of driving piRNA expression in genomic isolation. Accordingly, the genomic distribution of these motifs determines which sequences are expressed as piRNAs in *C.*

elegans. Our results suggest that each *C. elegans* piRNA represents an independent transcript whose sequence, abundance, and germline enrichment are encoded by a variant upstream motif, defining a novel modality for expression of piRNAs.

INTRODUCTION

piRNAs and Piwi clade Argonautes arose in the primordial metazoan ancestor [1] and are generally restricted to the germline, where they act in an RNA-induced silencing complex (RISC) to silence foreign genetic elements. From protozoa to mammals, loss of Piwi proteins, and consequently piRNAs, results in abnormal fertility phenotypes or sterility, revealing their highly conserved and essential role in animal reproduction [2-8]. piRNAs are incredibly diverse, with tens of thousands of unique sequences expressed in any single organism. While piRNAs in many organisms map to large, broadly syntenic genomic clusters, the sequences are not conserved among even closely related species, and no unifying sequence features have been identified beyond a bias among primary piRNAs for a 5' uridine [9-15].

The mechanisms of de novo piRNA biogenesis remain elusive. In fly and mouse, primary piRNAs appear to be processed from long, single-stranded RNA precursors [9,10,14]. This long transcript is cleaved by the endoribonuclease Zucchini with little or no sequence specificity to generate candidate piRNA 5' ends [16,17], which are likely subsequently purified according to the binding preferences of the Piwi proteins that bind primary piRNAs [18]. Silkworm data suggest that the 3' ends of these piRNA precursors are then trimmed by a 3' to 5' exonuclease until the 3' end is sufficiently short for anchoring by Piwi to protect against further trimming [18]. The 3' end is then methylated to prevent degradation [19-23]. While recent studies have shed light on the biogenesis of primary piRNAs in many animal models, little is known in any organism about how primary piRNA expression is regulated or how specific sequences are designated as piRNAs.

21U RNAs, a class of germline-enriched small RNAs, represent the piRNAs of *Caenorhabditis elegans*. They are terminally methylated [24-26], show a 5' uridine bias [12], and are dependent upon and bound by the Piwi Argonaute PRG-1 [27,28], which is required for normal fertility [3]. Yet *C. elegans* piRNAs exhibit some unusual features. While the vast majority of 21U RNAs map to two large genomic clusters on chromosome IV, the loci do not exhibit prominent strand biases [12]. The 21U RNAs also do not appear to play a prominent role in silencing transposable elements, a main function of mouse and fly piRNAs, nor do they engage a ping-pong amplification mechanism [27,28]. Rather, PRG-1 and the 21U RNAs target aberrant and coding transcripts broadly via imperfect complementarity, triggering production of secondary endogenous siRNAs [27-31]. These 21U RNA-dependent 22G RNAs can induce chromatin changes to establish dominant, heritable target silencing [32-34]. 21U RNAs evolve rapidly, presumably constrained only by selection against sequences that silence mRNAs; thus, mismatch-tolerant 21U RNAs constitute an epigenetic memory of self versus non-self. Finally, a conserved motif lies upstream of 21U RNA genomic loci [12]. This stretch of sequence, which includes an eight-nucleotide (nt) core motif approximately 40 nt upstream of the 21U RNA locus, is conserved across divergent nematodes [12,13]. Recently, Cecere et al. found that this motif is bound by forkhead family transcription factors and that deletion of the core motif abrogates 21U RNA expression [35], but it is still unknown how 21U RNA sequences are defined and how their expression is regulated.

Here, we demonstrate that piRNAs are expressed autonomously in *C. elegans*. Combining computational and transgenic approaches, we find that the conserved core motif defines the piRNA transcriptional cassette, specifying expression of 21U RNAs from genomic thymidines situated at an optimal distance downstream to determine which genomic sequences are expressed as *C. elegans* piRNAs. Core motifs also encode information dictating germline-specific expression of 21U RNAs. We show that more than 70% of *C. elegans* piRNAs are preferentially enriched in male or female germline. Unexpectedly, this germline enrichment appears to be enforced by a single nucleotide position within the core motif. We demonstrate autonomous expression of

synthetic 21U RNAs from multiple minimal transgenic cassettes consisting only of the 8 nt core motif, the ~40 nt intervening genomic spacer, the 21U RNA sequence, and ~50-100 nt of flanking genomic context. Finally, we use single-copy transgenes integrated in genomic isolation to show that the clustered organization of endogenous piRNA loci is entirely dispensable for robust piRNA expression. Together, our results suggest that each 21U RNA locus encodes all of the information necessary for driving independent, autonomous transcription from more than 15,000 unique piRNA loci in *C. elegans*.

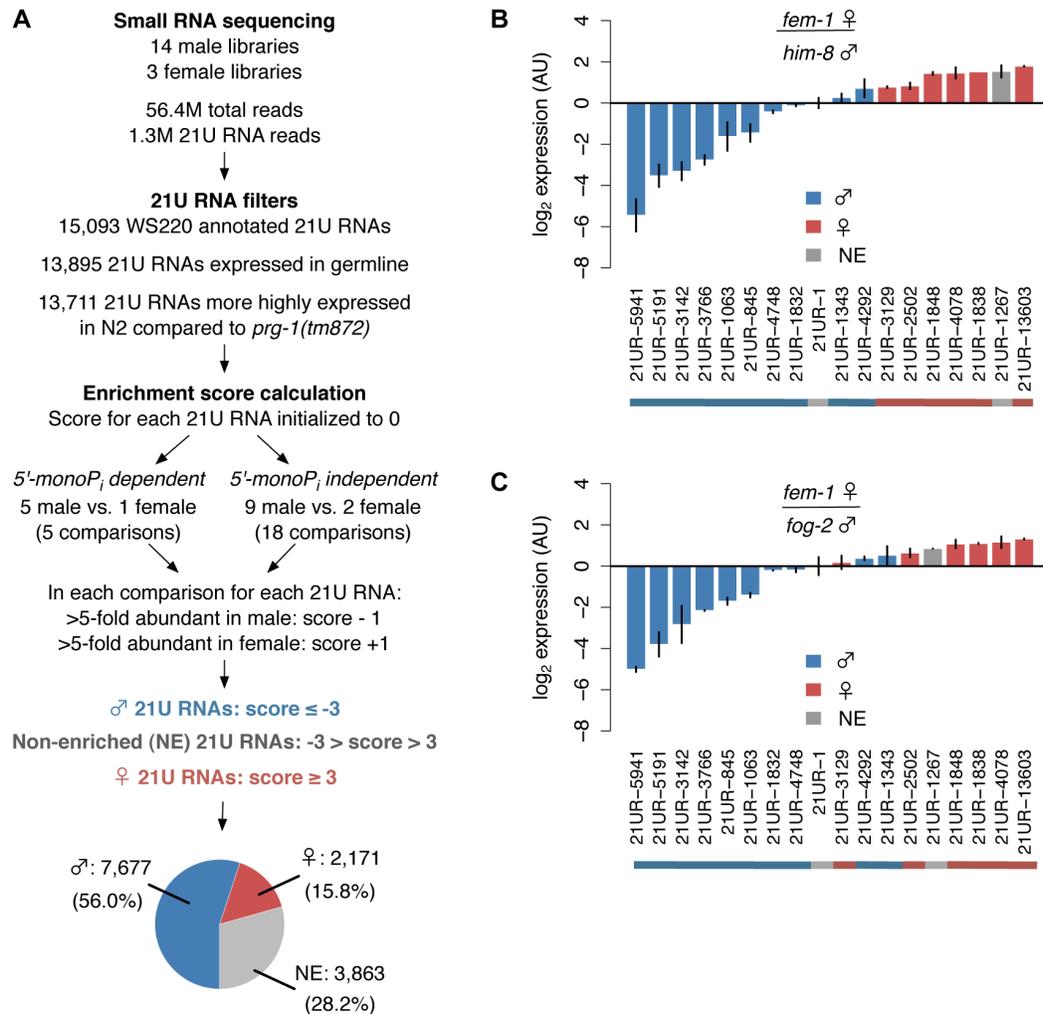
RESULTS

A majority of 21U RNAs are male or female germline-enriched

To investigate the mechanisms regulating piRNA expression, we first identified 21U RNA subclasses by performing a meta-analysis of over 50 million reads from published small RNA deep sequencing datasets [27,36-42] (Table 3.S1). Using the pipeline shown in Figure 3.1A, we determined that a majority of the 13,711 21U RNAs represented in our composite dataset show differential germline enrichment, distinguishing 7,677 (56.0%) unique male and 2,171 (15.8%) unique female germline-enriched 21U RNAs (hereafter, male and female 21U RNAs) (See Materials and Methods). The distribution of 21U RNA Enrichment scores is skewed toward the male (Figure 3.S1A), whereas randomly generated 21U RNA count data show no significant skewing (Binomial test, $p = 0.245$) and define a false discovery rate below 1% (Figure 3.S1B). To assess the reliability of the Enrichment score in classifying germline enrichment, we quantified the average relative abundance of every male 21U RNA between each pair of male and female libraries (Figure 3.S1C); the reciprocal calculation was performed for female 21U RNAs (Figure 3.S1D). On average, the abundance of male 21U RNAs is 6.8-fold higher in male libraries than female, whereas the abundance of female 21U RNAs is 2.4-fold higher in female libraries than male. Average abundance of 21U RNAs not classified as male or female (hereafter, non-enriched 21U RNAs) is approximately equal in male and female libraries (Figure 3.S1E). Taqman RT-qPCR of select 21U RNAs in *fem-1(hc17)*

adult female versus *him-8(e1489)* or *fog-2(q71)* adult male animals shows segregation of 21U RNAs according to germline enrichment classification (Figure 3.1B,C), endorsing our computational discovery of germline-enriched piRNA subclasses in *C. elegans*.

Figure 3.1: Over 70% of 21U RNAs show distinct germline enrichment. (A) Pipeline for computational identification of male and female 21U RNAs. A majority of 21U RNAs are classified as male or female germline-enriched. Pie chart depicts classification as proportion of 13,711 21U RNAs analyzed. (B,C) Male 21U RNAs are more highly expressed in male animals, and female 21U RNAs are more highly expressed in female animals. Relative expression of representative 21U RNAs was assayed by Taqman RT-qPCR in *him-8(e1489)* (B) and *fog-2(q71)* (C) male versus *fem-1(hc17)* female animals and normalized to non-enriched 21U RNA 21UR-1. Error bars: ± 1 standard deviation (SD) of two biological replicates. AU: arbitrary units.



Male and female 21U RNAs show different expression profiles in embryo

Our meta-analysis also revealed a subpopulation of 21U RNAs highly abundant in embryo. Comparison of the abundances of male and female 21U RNAs in mixed stage embryo sequencing libraries showed that female 21U RNAs were overrepresented in embryo relative to male. A higher proportion of unique female 21U RNAs were detected in embryo (χ^2 test, $p = 9.2e-45$) (Figure 3.S2A,B). Furthermore, unique female 21U RNAs were on average 4.4-fold more abundant in embryo than unique male species (Welch's t -test, $p = 3.4e-148$). The trend is corroborated by Taqman analysis showing depletion of male 21U RNAs and enrichment of female 21U RNAs in embryo (Figure 3.S2C-E). These data suggest that female piRNAs are preferentially inherited into *C. elegans* embryo, consistent with previous observations in fly [43-45]. Parallel classification and embryonic enrichment analysis of 26G RNAs, germline-enriched primary endo-siRNAs, recapitulated previously observed inheritance patterns [38] and validated the ability of our pipeline to identify germline-enriched small RNA subclasses (Figure 3.S2F,G).

Male 21U RNA targets reflect spermatogenic gonad restriction

21U RNAs target transcripts with imperfect complementarity of up to three mismatches to trigger production of antisense 22G RNAs proximal to the targeting site [29,30,32,33]. The lax complementarity requirement for piRNA-mediated silencing predicts widespread targeting capacity. Compartmentalization of piRNA expression to the male and female germline may help to confer specificity. To investigate the biological significance of germline-enriched 21U RNA subclasses, we first examined whether male and female 21U RNAs target distinct subsets of genes. We analyzed the overlap between their respective dependent 22G RNAs by identifying 22G RNAs that map antisense to within 40 nt of 21U RNA target sites [30] (See Materials and Methods). Ignoring 22G RNAs detected in *prg-1(n4357)* deep sequencing datasets, as these are likely not 21U RNA-dependent, we identified 11,377 (72.3%) unique 22G RNAs that are likely male 21U

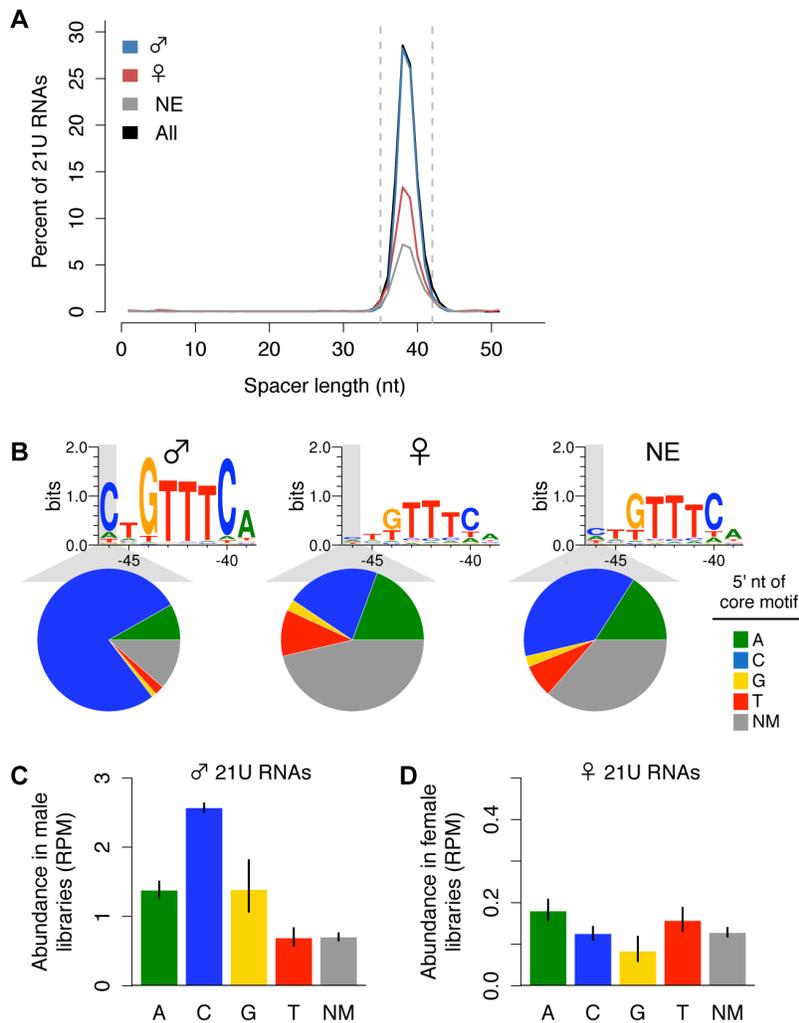
RNA-dependent and 3,855 (24.5%) unique 22G RNAs that are likely female 21U RNA-dependent (Figure 3.S3A). Only 494 (3.1%) unique 22G RNAs lie within 40 nt of both a male and female 21U RNA target site, precluding assignment to either category. This overlap is less than expected when 22G RNAs from random but similarly sized sets of 21U RNAs are compared (χ^2 test, $p = 0.012$). We then compared the 5,956 male and 1,387 female 21U RNA targets identified in young adult [29] and gravid [30] animals, respectively. Overlap between targets (149 overlapping targets) is significantly lower compared to random sets of genes (294 overlapping and 6,756 non-overlapping targets; χ^2 test, $p = 7.7e-13$) (Figure 3.S3B).

Because targets of 21U RNAs are subject to transgenerational silencing [32-34], 21U RNAs are unlikely to evolve to target transcripts required in the germline. Similarly, male 21U RNAs would not be expected to target transcripts required for spermatogenesis; however, temporal separation of the spermatogenic and oogenic gonads might permit evolution of male 21U RNAs capable of targeting transcripts required for oogenesis. We examined our data for evidence of this evolutionary signature. As comprehensive lists of genes required for spermatogenesis and oogenesis have yet to be assembled, we used as a proxy lists of transcripts identified by microarray studies as enriched during spermatogenesis (865 transcripts) or oogenesis (1,030) [46]. Comparing male 21U RNA targets to randomly generated gene lists, we found that male 21U RNA targets are indeed depleted of spermatogenesis transcripts (χ^2 test, $p = 0.044$), but neither enriched nor depleted for oogenesis transcripts (χ^2 test, $p = 0.76$) (Figure 3.S3C). Curiously, we do not observe the same signature for female 21U RNAs (Figure 3.S3D). Their targets are neither enriched nor depleted for spermatogenesis transcripts (χ^2 test, $p = 0.27$), as expected, but female 21U RNA targets are significantly enriched for oogenesis transcripts (χ^2 test, $p = 0.0017$). These differences between male and female 21U RNA targeting suggest that the evolutionary pressures acting on male and female 21U RNA sequences may differ (See Discussion).

Male and female 21U RNAs have distinct core upstream motifs

To investigate how 21U RNA germline enrichment information is genetically encoded, we analyzed the genomic loci of the 13,387 21U RNAs that map uniquely to the genome. Comparison of male and female 21U RNA sequences identified no differences in content; therefore, we evaluated the 21U RNA upstream region. The 8 nt core motif, with consensus sequence CTGTTTCA, is separated from the 21U RNA locus by an A/T-rich spacer of ~35 to 42 nt [12]. Scanning the 60 nt upstream of each 21U RNA for the best conserved central GTTTC of the core motif, we found that 6,615 of 7,677 (88%) male 21U RNAs show a canonical, GTTTC-containing core motif, compared to only 1,119 of 2,171 (54%) female 21U RNAs. While the length of the A/T-rich spacer does not differ between male and female 21U RNAs (Figure 3.2A), core motif sequence analysis revealed a striking difference: only the core motifs of male 21U RNAs are enriched for a 5' cytidine. 5,765 of 7,677 (77%) male 21U RNAs are located downstream of canonical, GTTTC-containing core motifs with a 5' cytidine, compared to only 443 of 2,171 (21%) female 21U RNAs (χ^2 test, $p = 7.9e-137$) (Figure 3.2B). To examine whether this 5' core motif position influences 21U RNA expression, we calculated the average abundance of male and female 21U RNAs grouped by 5' core motif nt. Male 21U RNAs with 5' cytidine core motifs are significantly more abundant than all other male 21U RNAs (Figure 3.2C, Welch's t -test p -values in Table 3.S2), consistent with the previous observation that 21U RNAs whose core motifs better match the consensus sequence are more highly expressed [12]. No other subgroup differs significantly in abundance from all others among the male, female, and non-enriched 21U RNAs (Figure 3.2C,D, Table 3.S2), suggesting that GTTTC-containing core motifs with a 5' cytidine are overrepresented among male 21U RNAs and may drive male germline expression.

Figure 3.2: Variation in the core upstream motif correlates with 21U RNA germline enrichment. (A) Spacer lengths follow expected distribution for all enrichment classifications. Dotted lines: canonical spacer length range (35-42 nt). (B) Male, but not female, 21U RNA loci show enrichment for core motifs with 5' cytidines. Significantly fewer female 21U RNAs exhibit a GTTTC-containing core motif than male. Top: Weblogo plots illustrate core motif differences. Bottom: Pie charts depict proportions of 21U RNAs with GTTTC-containing core motifs indicating the 5' nt (colors) or with no GTTTC-containing core motif (NM, no motif, dark grey). (C) Core motif variations correlate with male 21U RNA abundance in 5'-monophosphate-dependent libraries. Average 21U RNA abundance was calculated based on the 5' nt of the core motif. Error bars: ± 1 standard error of the mean (SEM). (D) Core motif variations do not correlate with female 21U RNA abundance in 5'-monophosphate-dependent libraries. Average 21U RNA abundance was calculated as in (C).



A transgenic synthetic 21U RNA recapitulates features of endogenous 21U RNAs

To explore the significance of variation at the 21U RNA upstream motif, we developed a transgenic system to express synthetic 21U RNAs from high-copy, integrated arrays in vivo (Figure 3.3A). 2-3 kilobase regions of genomic sequence from a chromosome IV piRNA cluster were cloned, and a central 21U RNA (male 21U RNA ♂21UR-1258 or female 21U RNA ♀21UR-2502) was mutated to a unique synthetic 21 nt sequence (21UR-synth) to distinguish transgenic from endogenous expression. The sequences were then further mutated to generate the panel of transgenes shown in Table 3.1. Transgenes are named for the endogenous 21U RNA replaced by 21UR-synth, with prefixes to indicate transgene type (e.g., ♀Tg2502 represents the otherwise wild-type transgene encoding 21UR-synth in place of ♀21UR-2502). These transgenes are carried by the vector pCFJ178 [47], which also expresses the *C. briggsae unc-119* gene (Figure 3.S4A), enabling gross normalization for variable array expression.

To validate our transgenic system, we examined whether 21UR-synth recapitulates all of the known features and genetic sensitivities of endogenous 21U RNAs. 21U RNAs are 2'-O-methylated at the 3' terminus by the conserved methyltransferase HENN-1 [24-26]. Northern blot for 21UR-synth in transgenic strains identified a 21 nt species that is terminally methylated in a *henn-1*-dependent manner (Figure 3.3B). Robust, specific detection of the 3' terminus by Taqman RT-qPCR [48] confirms that this species corresponds to 21UR-synth (Figure 3.3C). Levels of endogenous 21U RNAs ♂21UR-1258 and ♀21UR-2502 are largely unaffected by expression of the transgenes (Figure 3.S4B). Endogenous 21U RNAs are generated in the germline and require PRG-1 for accumulation [27,28]. Accordingly, 21UR-synth is highly depleted by loss of *prg-1* and in the *glp-4(bn2)* germline-deficient mutant (Figure 3.3C-E). 21UR-synth and endogenous 21U RNAs are also specifically detected in immunoprecipitated PRG-1 complexes, while a microRNA control is not (Figure 3.3F,G, Figure 3.S5). To rule out the unlikely possibility that transgenic products corresponding to the 21UR-synth sequence might be generated by an alternative, Dicer-dependent mechanism, we assayed 21UR-synth accumulation in a null mutant of *rde-4*. This gene encodes a dsRNA binding protein that

is a key cofactor of Dicer in siRNA biogenesis [49-51], but dispensable for 21U RNA production (Figure 3.S4B). Loss of *rde-4* does not impair 21UR-synth expression (Figure 3.3C-E), suggesting that 21UR-synth does not represent an siRNA generated from the high-copy transgenic array.

Finally, we examined whether the core motif is required for 21UR-synth expression. We scrambled the core motif to eliminate any resemblance to the consensus sequence (♂Scram1258 and ♀Scram2502 transgenes; Table 3.1). 21UR-synth levels in these strains are depleted by more than 100-fold after normalization for array expression (Figure 3.3H,I), consistent with previous findings that deletion of the core motif depletes 21U RNA expression [35]. Together, these data demonstrate that 21UR-synth represents a bona fide 21U RNA and support the use of this transgenic system for exploring 21U RNA biology in vivo.

Figure 3.3: A transgenic synthetic 21U RNA shows characteristics of endogenous 21U RNAs. (A) Diagram of Tg (dark grey) and Min (light grey) transgenes with core motif sequences shown. Asterisk indicates a 21U RNA whose core motif is disrupted by 21UR-synth and is therefore predicted not to express. (B) 21UR-synth is methylated by HENN-1. 21UR-synth is specifically detected in transgenic strains and is susceptible to β -elimination only in the *henn-1(tm4477)* background. Arrowhead represents migration of a 21 nt size marker. 21UR-synth blot was reprobed for miR-1. Endogenous ♀21UR-2502 is shown as a control. (C-E) 21UR-synth is a *prg-1*-dependent, germline-enriched 21U RNA. 21UR-synth detection by Taqman RT-qPCR (C) and northern blot (D,E) is greatly decreased in *prg-1(tm872)* and *glp-4(bn2)* germline-deficient mutant animals, but intact in *rde-4(ne301)* mutant animals. Error bars: ± 1 SD of three biological replicates. (F) anti-PRG-1 antibody immunopurifies PRG-1 complexes. CL: crude lysate, RIP: RNA immunoprecipitation. (G) 21UR-synth is bound by endogenous PRG-1. Error bars: ± 1 SD of two technical replicates; data are representative of two independent experiments. (H,I) Loss of the core motif dramatically decreases 21UR-synth expression by northern blot (H) and Taqman qRT-PCR (I). Error bars: ± 1 SD of three biological replicates.

Table 3.1: Descriptions of transgenic alleles and features of the transgenes.

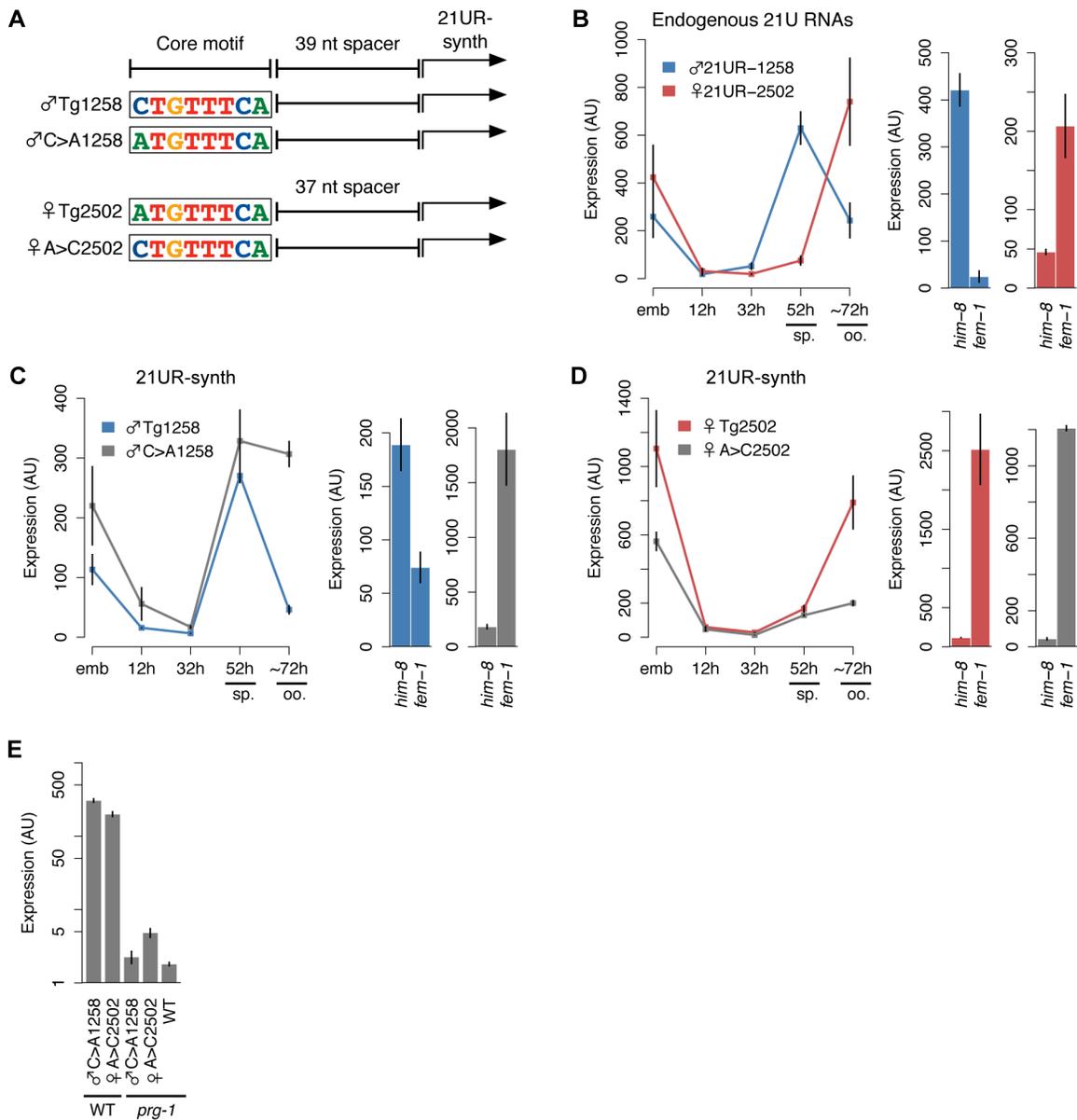
High-copy transgenes									
Transgene name	Transgene description	Size (kb)	Upstream motif	1 st nt	Chr	Allele	Strain	Backgrounds	
♂Tg1258	Cluster with synthetic 21U RNA replacing 21UR-1258	2.87	attaagc <u>CTGTTTC</u> caattttt	U	V	xkis1	QK7	WT, <i>prg-1, rde-4, glp-4, him-8, fem-1, heenn-1</i>	
♂Scram1258	♂Tg1258 with upstream motif scrambled	2.87	attatcagaccgctctattttt	U	N/D	xkis11	QK8	WT	
♂C>A1258	♂Tg1258 with upstream motif first nt mutated C>A	2.87	attaagc <u>ATGTTTC</u> caattttt	U	II	xkis6	QK9	WT, <i>prg-1, him-8, fem-1</i>	
♂21U>A1258	♂Tg1258 with synthetic 21U RNA first nt mutated U>A	2.87	attaagc <u>CTGTTTC</u> caattttt	A	N/D	xkis14	QK10	WT	
♂21U>G1258	♂Tg1258 with synthetic 21U RNA first nt mutated U>G	2.87	attaagc <u>CTGTTTC</u> caattttt	G	N/D	xkis15	QK11	WT	
♂Min1258	Minimal 21U RNA construct from ♂Tg1258	0.17	attaagc <u>CTGTTTC</u> caattttt	U	N/D	xkis16	QK12	WT, <i>prg-1, rde-4, glp-4</i>	
♀Tg2502	Cluster with synthetic 21U RNA replacing 21UR-2502	2.04	aaataaaa <u>TGTTTC</u> aactagtc	U	I	xkis5	QK13	WT, <i>prg-1, rde-4, glp-4, him-8, fem-1, heenn-1</i>	
♀Scram2502	♀Tg2502 with upstream motif scrambled	2.04	aaataaaggacaccttattttt	U	N/D	xkis12	QK14	WT	
♀A>C2502	♀Tg2502 with upstream motif first nt mutated C>A	2.04	aaataaaa <u>TGTTTC</u> aactagtc	U	X	xkis10	QK15	WT, <i>prg-1, him-8, fem-1</i>	
♀21U>A2502	♀Tg2502 with synthetic 21U RNA first nt mutated U>A	2.04	aaataaaa <u>TGTTTC</u> aactagtc	A	N/D	xkis17	QK16	WT	
♀21U>G2502	♀Tg2502 with synthetic 21U RNA first nt mutated U>G	2.04	aaataaaa <u>TGTTTC</u> aactagtc	G	N/D	xkis18	QK17	WT	
♀Min2502	Minimal 21U RNA construct from ♀Tg2502	0.23	aaataaaa <u>TGTTTC</u> aactagtc	U	N/D	xkis19	QK18	WT, <i>prg-1, rde-4, glp-4</i>	
♂Min1415	Minimal 21U RNA construct generated from 21UR-1415	0.22	ttttgcg <u>CTGTTTC</u> aggagtt	U	N/D	xkis20	QK19	WT	
♂MinC>A1415	♂Min1415 with upstream motif first nt mutated C>A	0.22	ttttgcg <u>ATGTTTC</u> aggagtt	U	N/D	xkis21	QK20	WT	
Tg1415-Tg2109	Minimal 21U RNA construct with 21UR-synth replacing 21UR-1415 and 21UR-synthB replacing 21UR-2109	0.26	ttttgcg <u>CTGTTTC</u> aggagtt taatctc <u>TGTTTC</u> caatatt	U	N/D	xkis22	QK21	WT	
Scram1415-Tg2109	Tg1415-Tg2109 with 21UR-1415 upstream motif scrambled	0.26	ttttgcg <u>taggaccctctagtt</u> taatctc <u>CTGTTTC</u> caatatt	U	N/D	xkis23	QK22	WT	
MosSCI transgenes									
Transgene name	Transgene description	Size (kb)	Upstream motif	1 st nt	Chr	Allele	Strain	Backgrounds	
♂Tg1258	Cluster with synthetic 21U RNA replacing 21UR-1258	2.87	attaagc <u>CTGTTTC</u> caattttt	U	IV	xkis3	QK23		
♂Scram1258	♂Tg1258 with upstream motif scrambled	2.87	attatcagaccgctctattttt	U	IV	xkis23	QK24		
♂C>A1258	♂Tg1258 with upstream motif first nt mutated C>A	2.87	attaagc <u>ATGTTTC</u> caattttt	U	IV	xkis17	QK25		
♀Tg2502	Cluster with synthetic 21U RNA replacing 21UR-2502	2.04	aaataaaa <u>TGTTTC</u> aactagtc	U	IV	xkis13	QK26		
♀Scram2502	♀Tg2502 with upstream motif scrambled	2.04	aaataaaggacaccttattttt	U	IV	xkis28	QK27		
♀A>C2502	♀Tg2502 with upstream motif first nt mutated C>A	2.04	aaataaaa <u>CTGTTTC</u> aactagtc	U	IV	xkis20	QK28		
♀Min2502	Minimal 21U RNA construct from ♀Tg2502	0.23	aaataaaa <u>TGTTTC</u> aactagtc	U	II	xkis30	QK29		

Both high-copy and MosSCI transgenes used in this study are listed with a short description, sequence characteristics, integration information, and strain notation. Full transgene data are listed in Materials and Methods. Bolded letters indicate mutated nucleotides. Eight nt core upstream motifs are capitalized while motif positions are underlined. N/D, not determined.

21U RNA core upstream motif variation influences germline enrichment

We then used our transgenic system to test whether variation at the core motif 5' position affects germline expression of 21UR-synth (Figure 3.4A). Endogenous male 21U RNA ♂21UR-1258, which lies downstream of a **CTGTTTCA** core motif, peaks in expression during spermatogenesis (52h time point) and is highly expressed in *him-8(e1489)* male adult; in contrast, expression of endogenous ♀21UR-2502, with an **ATGTTTCA** core motif, peaks after the spermatogenesis-to-oogenesis transition in adulthood (~72h) and is highly expressed in *fem-1(hc17)* female adult (Figure 3.4B). Accordingly, the ♂Tg1258 and ♀Tg2502 transgenes express 21UR-synth in similar male and female patterns, respectively (Figure 3.4C,D, colored lines/bars). Toggling the core motif from **CTGTTTCA** to **ATGTTTCA** (♂C>A1258 transgene) or **ATGTTTCA** to **CTGTTTCA** (♀A>C2502) disrupts these germline-specific expression patterns. Whereas 21UR-synth expression from ♂Tg1258 plummets after spermatogenesis, loss of the core motif 5' cytidine in the ♂C>A1258 transgenic strain results in sustained 21UR-synth expression through oogenesis; the ♂C>A1258 transgene also preferentially expresses 21UR-synth in *fem-1(hc17)* female (Figure 3.4C). Thus, mutating the 5' cytidine of a male 21U RNA core motif results in a failure to restrict 21U RNA expression to spermatogenesis. Similarly, introducing a 5' cytidine into a female 21U RNA core motif impairs restriction of expression to oogenesis: while ♀Tg2502 expression of 21UR-synth increases dramatically during the spermatogenesis-to-oogenesis transition, gain of the motif 5' cytidine in the ♀A>C2502 transgene dampens this increase (Figure 3.4D). These results suggest that this single nucleotide orchestrates the accurate switching of 21U RNA expression in the hermaphroditic germline. However, 21UR-synth expression from the ♀A>C2502 transgene is still high in *fem-1(hc17)* female, indicating that other elements contribute to female 21U RNA expression patterns. This is consistent with our finding that female 21U RNA core motifs show no bias at the 5' nucleotide, and indeed ~21% of female 21U RNA core motifs show a 5' cytidine (Figure 3.2B). As expected, 21UR-synth expression from the ♂C>A1258 and ♀A>C2502 transgenes is still dependent upon *prg-1* (Figure 3.4E).

Figure 3.4: A 5' cytidine in the core upstream motif promotes male germline expression pattern of 21UR-synth. (A) Schematic of transgenes with 5' nt of core motif mutated. (B) Left: Endogenous ♂21UR-1258 and ♀21UR-2502 peak during spermatogenesis (sp.) and oogenesis (oo.), respectively. Right: Germline enrichment patterns are recapitulated in *him-8(e1489)* male and *fem-1(hc17)* female animals. Error bars: ± 1 SD of three biological replicates. (C) The male expression pattern of 21UR-synth from ♂Tg1258 is disrupted by core motif mutation in ♂C>A1258. Error bars: ± 1 SD of three biological replicates. (D) The female expression pattern of 21UR-synth from ♀Tg2502 is disrupted by core motif mutation in ♀A>C2502, but expression in *fem-1(hc17)* female is not lost. Error bars: ± 1 SD of three biological replicates. (E) Mutating the 5' nt of the core motif does not affect 21UR-synth *prg-1* dependence.



A 5' thymidine is required for robust expression from the 21UR-synth locus

It is not yet known how individual genomic sequences are selected for expression as piRNAs. As the core motifs, but not the sequences, of 21U RNAs are conserved across *Caenorhabditis* species, it seemed possible that the core motifs themselves might determine what sequences are expressed as 21U RNAs by directing their expression from genomic thymidines located an optimal distance downstream. We explored this hypothesis by mutating the genomic thymidines encoding the first nucleotide of 21UR-synth to adenosine (21U>A transgenes) or guanosine (21U>G transgenes), such that the transgenes encode 21[U>A]R-synth or 21[U>G]R-synth, respectively (Figure 3.5A, Figure 3.S6A). These putative products emulate the 5' nucleotide identity of microRNAs (predominantly 5' uridine and adenosine) and endo-siRNAs (predominantly 5' guanosine). Small RNAs expressed from these transgenes and recognized by the 21UR-synth northern blot probe differ in size from and are less abundant than wild-type 21UR-synth (Figure 3.5B, Figure 3.S6B). By Taqman analysis, 21[U>A]R-synth and 21[U>G]R-synth are detected at levels more than 150-fold lower than 21UR-synth after normalization for array expression (Figure 3.5C, Figure 3.S6C), suggesting that 21[U>A]R-synth and 21[U>G]R-synth are poorly transcribed, stabilized, or both.

The genomic positioning of core motifs specifies 21U RNA sequences

We hypothesized that 21U RNA expression from a particular genomic thymidine may simply be a function of distance from a core motif (i.e., length of the intervening genomic spacer). Therefore, the presence of multiple thymidines within the optimal genomic window downstream of a core motif might result in expression of multiple, overlapping 21U RNAs. Indeed, many *C. elegans* piRNAs map to proximal genomic thymidines as members of “miniclusters” of overlapping 21U RNAs that appear to share an upstream core motif. To explore the relationship between core motif position and expression, we extracted read count information from deep sequencing of wild-type adult animals [27] for uniquely mapping 21U RNAs and analyzed their corresponding genomic loci. After

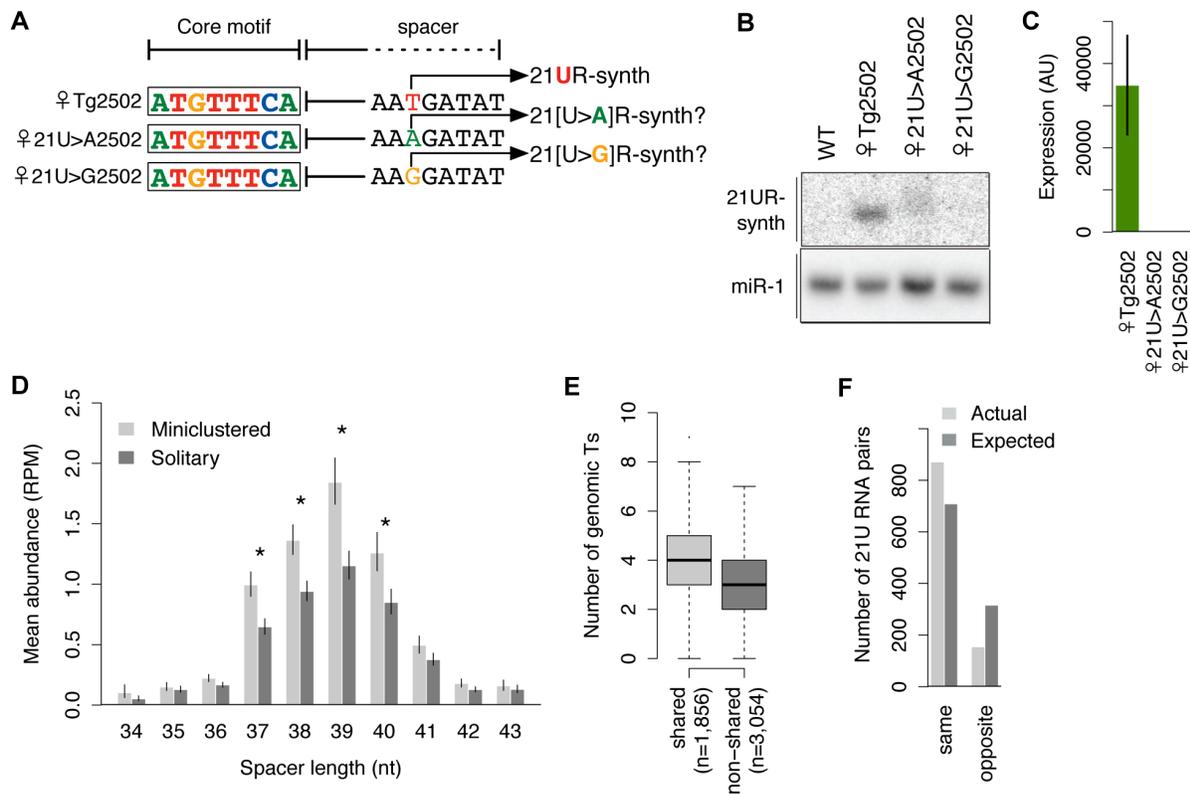
separating 21U RNAs into those that share a core motif with at least one other uniquely mapping 21U RNA (“miniclustered”; 4,550 21U RNAs) and those that do not (“solitary”; 8,837 21U RNAs), we grouped 21U RNAs by length of genomic spacer and examined their abundance. For both miniclustered and solitary 21U RNAs, the resulting distributions peak at a 39 nt spacer length and decrease as the spacer lengthens or shortens (Figure 3.5D). The evident correlation between spacer length and robustness of expression explains previous observations that miniclustered 21U RNAs routinely show great variation in abundance [39].

We also observed that miniclustered 21U RNAs with 37-40 nt spacers are more abundant than solitary 21U RNAs at matched positions (Figure 3.5D, asterisks), suggesting that 21U RNA miniclusters may arise when expression is driven more robustly. To investigate this further, we compared the core motifs associated with miniclustered 21U RNAs (“shared” motifs) versus solitary 21U RNAs (“non-shared” motifs). We found that a significantly larger proportion of miniclustered 21U RNAs (3,580 of 4,550, 79%) than solitary 21U RNAs (5,667 of 8,837, 64%) are associated with canonical, GTTTC-containing core motifs (χ^2 test, $p = 1.4e-66$). Additionally, we observed significantly greater thymidine richness in the optimal genomic windows 35-42 nt downstream of shared GTTTC-containing motifs versus non-shared (Welch’s t -test, $p = 4.0e-95$) (Figure 3.5E). Therefore, particular sequences of 21U RNAs may not be specified intrinsically; rather, core motifs may simply direct expression of 21U RNAs from one or more downstream thymidines, depending on the strength of the motif and the number of optimally positioned thymidines.

To further confirm the association between the core motif and germline enrichment, we analyzed miniclusters consisting of two germline-enriched 21U RNAs (1,026 pairs). Random assortment of these 21U RNAs would predict 66% male:female, 4% female:female, and 31% male:female pairs; however, we observed 73% male:female, 12% female:female, and only 15% male:female pairs. Thus, 85% of pairs showed matching enrichment classification (Figure 3.5F), a significant departure from the 69% expected by random assortment (χ^2 test, $p = 9.6e-28$). We note that this paucity of

mixed male:female 21U RNA miniclusters likely contributes to the low number of 22G RNAs that can be attributed to both male and female 21U RNAs (Figure 3.S3A).

Figure 3.5: 21U RNA sequences are specified by the genomic positions of upstream core motifs. (A) Schematic of transgenes with 5' nt of 21U RNA mutated. (B-C) Mutation of the 5' genomic thymidine disrupts expression of 21UR-synth by northern blot (B) and Taqman assay (C). (D) 21U RNA abundances correlate with distances downstream of core motifs. Miniclustered 21U RNAs with 37-40 nt spacer lengths are more abundant than solitary 21U RNAs. Asterisks indicate Welch's *t*-tests, $p < 0.05$. Error bars: ± 1 SEM. (E) Optimal downstream windows are more thymidine-rich for shared core motifs than non-shared (Welch's *t*-test, $p = 2.5e-46$). The number of genomic thymidines located 35 - 42 nt downstream of each GTTTC-containing motif was counted. (F) 21U RNA miniclusters are significantly biased for being composed of 21U RNAs with the same, as opposed to opposite, germline enrichment than expected if the same 21U RNAs were randomly paired.



Each upstream motif and 21U RNA sequence constitutes a tiny, autonomous transcriptional unit

The absence of long, unidirectional 21U RNA clusters in the *C. elegans* genome and the presence of the conserved upstream motif have generated speculation that 21U RNAs represent autonomously transcribed units [12,27,28]. This is further suggested by our and others' findings that scrambling or deleting the core motif abrogates 21U RNA expression (Figure 3.3H,I and [35]). To test whether 21U RNAs express independently, we generated transgenes representing putative minimal 21U RNA transcriptional units. Each of these Min transgenes encodes only a single core motif, spacer, and 21U RNA, with limited 5' and 3' genomic context (Figure 3.3A). Strikingly, 21UR-synth expressed from this minimal context shows the same size, *prg-1* dependence, *rde-4* independence, and germline enrichment as endogenous 21U RNAs (Figure 3.6A,B), indicating that the sequence features conferring these 21U RNA characteristics are contained within a single 21U RNA transcriptional unit. To ensure that the 5' nucleotide of the core motif still influences germline enrichment within this minimal context, we also generated and tested an independent set of minimal 21UR-synth transgenes with core motif intact ($\text{\textcircled{M}}\text{in}1415$) or first nucleotide toggled ($\text{\textcircled{M}}\text{inC}>\text{A}1415$). These transgenes also showed impaired male germline enrichment upon toggling of the core motif 5' nucleotide (Figure 3.6C), reaffirming our conclusions that a core motif 5' cytidine helps to orchestrate 21U RNA male germline enrichment.

The 21U RNA transcriptional unit is autonomous

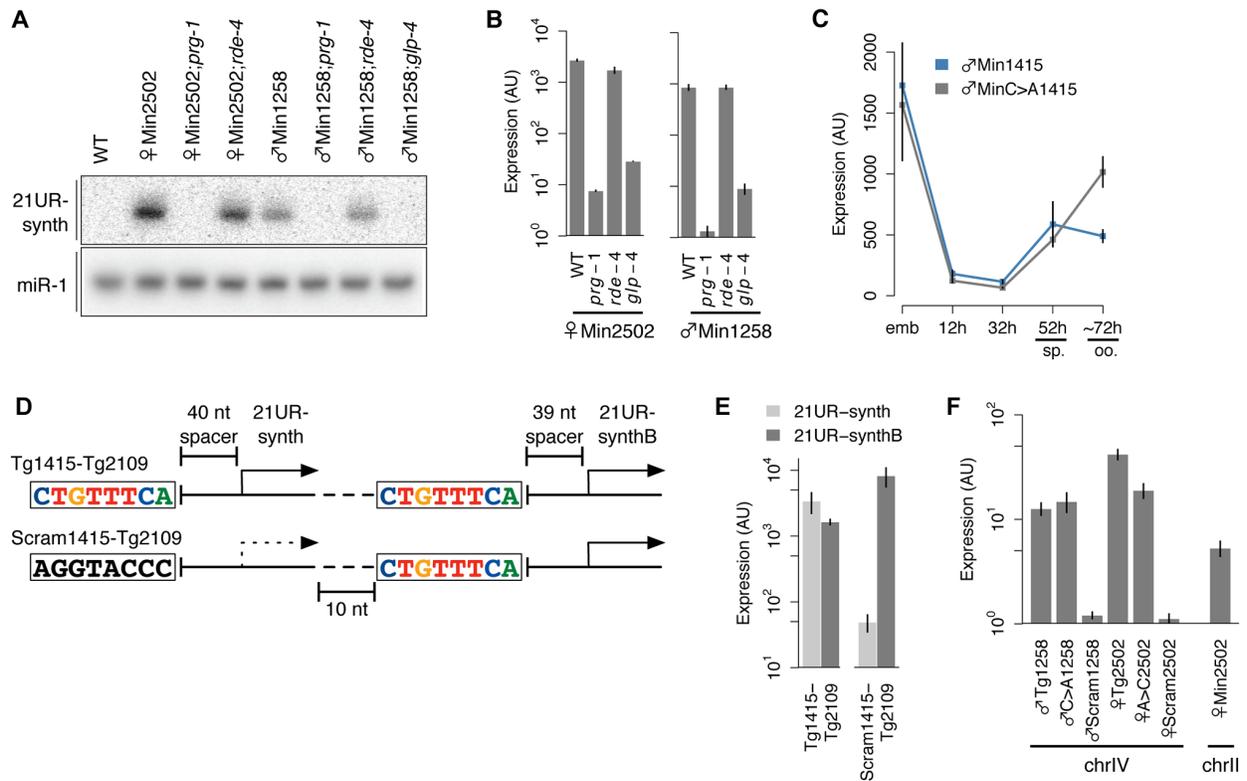
To explore the autonomy of the 21U RNA transcriptional unit further, we generated additional transgenes carrying <300 nt of genomic sequence encoding two adjacent 21U RNA transcriptional units on the same strand (Figure 3.6D). To create the "wild-type" Tg1415-Tg2109 transgene, the upstream 21U RNA locus, corresponding to 21UR-1415, was mutated to encode 21UR-synth, and the downstream locus, corresponding to 21UR-2109, was mutated to encode a different unique synthetic 21U

RNA (21UR-synthB). We then scrambled the core motif of the upstream 21U RNA locus to generate the Scram1415-Tg2109 transgene and measured relative expression of the two synthetic 21U RNAs from each transgene. Much as expression of 21UR-synth is vastly decreased by loss of the core motif in the ♂Scram1258 and ♀Scram2502 transgenes above (Figure 3.3H,I), 21UR-synth is expressed at far lower levels than 21UR-synthB from the Scram1415-Tg2109 transgene, whereas expression of the synthetic 21U RNAs from the Tg1415-Tg2109 transgene is comparable (Figure 3.6E). This experiment specifically pursues a recent finding by Cecere et al. that deletion of the core motif of one 21U RNA does not abrogate expression of neighboring 21U RNAs, although the species assessed were distant, separated by multiple 21U RNA loci, and encoded on both strands [35].

C. elegans 21U RNA loci, like the piRNA loci of mouse and fly [9,10,14], are genomically clustered. The overwhelming majority of 21U RNAs map to two large regions on chromosome IV, and GTTTC, the most highly conserved five nt of the core motif, occurs much more frequently on chromosome IV at these regions (4.0 occurrences per kilobase, occ/kb) than on chromosome IV outside these regions (0.4 occ/kb) or on other chromosomes (0.2 occ/kb). Furthermore, 21U RNAs encoded on chromosome IV are detected at much higher abundance (mean abundance: 148 RPM) than those encoded on other chromosomes (1 RPM) (Welch's *t*-test, $p = 2.4e-269$). These observations suggest the possibility of a positional requirement for expression of 21U RNA loci: a privileged genomic environment might contribute to the expression of 21U RNAs. To investigate the significance of 21U RNA genomic organization, we carried out rough mapping of the genomic insertion sites of several of the high-copy transgenic arrays. None of the integration loci mapped to chromosome IV (Table 3.1), indicating that these strains are not expressing 21UR-synth from the context of the 21U RNA genomic clusters. Yet the transgenic arrays themselves could represent 21U RNA-rich genomic microenvironments, much like the chromosome IV 21U RNA clusters. We therefore tested the true autonomy of the 21U RNA by using the MosSCI technique [47] to insert single-copy transgenes at a locus on chromosome IV not contained within the 21U RNA genomic clusters. Local 21U RNA concentration at the integration site is low,

and no 21U RNAs are annotated as mapping to the homology arms encoded on the pCFJ178 MosSCI plasmid. Unexpectedly, single-copy insertions of ♂Tg1258 and ♀Tg2502 transgenes express 21UR-synth at levels easily detectable, albeit tenfold lower than the high-copy arrays. As observed for the high-copy arrays, scrambling of the core motif severely diminishes expression of 21UR-synth from the single-copy transgenic insertions (Figure 3.6F). Finally, to exclude the remote possibility that chromosome IV origin itself is essential for 21U RNA expression, we used an alternative MosSCI plasmid to insert onto chromosome II a single copy of the ♀Min2502 transgene, which encodes no other 21U RNAs. Like the chromosome IV transgene insertions, ♀Min2502 expresses 21UR-synth robustly (Figure 3.6F), confirming that 21U RNAs can be autonomously transcribed.

Figure 3.6: 21U RNAs represent independent transcriptional units. (A-B) 21UR-synth expressed from a minimal transcriptional cassette shows *prg-1* dependence, *rde-4* independence, and germline enrichment by northern blot (A) and Taqman assay (B). **(C)** The male expression pattern of 21UR-synth from ♂Min1415 is disrupted by core motif mutation in ♂MinC>A1415. Error bars: ± 1 SD of three biological replicates. **(D)** Schematic of transgenes encoding two closely adjacent 21U RNAs. **(E)** Scrambling the core motif upstream of 21UR-synth abrogates 21UR-synth, but not 21UR-synthB, expression levels. **(F)** The ♂Tg1258, ♂C>A1258, ♀Tg2502, and ♀A>C2502 transgenes, but not the ♂Scram1258 or ♀Scram2502 transgenes, express from single copy insertions on chromosome IV. The ♀Min2502 transgene also expresses from a single-copy insertion on chromosome II.



DISCUSSION

piRNAs are transcribed as tiny, autonomous transcriptional units

Our data support a 21U RNA biogenesis mechanism wherein the upstream motif and 21U RNA sequence constitute a tiny, independent transcriptional unit that encodes regulated germline expression. The upstream motif as initially identified by Ruby et al. [12] is necessary for autonomous expression of a 21U RNA from one or more optimally situated downstream genomic thymidines. Importantly, this genomic thymidine may not represent a transcriptional requirement but rather reflect the binding preferences of the Argonaute PRG-1: a heterogeneous pool of candidate 21U RNA sequences may be transcribed and subsequently purified through preferential stabilization by PRG-1. Our transgenic studies showing greatly decreased expression when 21UR-synth is mutated to 21[U>A/G]R-synth cannot differentiate between a transcriptional or post-transcriptional requirement for a 5' uridine; however, findings in other organisms support the latter mechanism. In mouse and fly, the prevailing model posits that Zucchini generates candidate primary piRNA 5' ends with very little sequence specificity during the processing step, and then Piwi preferentially binds 5' uridine piRNAs during the loading step [16,17]. This is consistent with in vitro data showing that Siwi, the silkworm ortholog of PRG-1, preferentially incorporates ssRNAs bearing a 5' uridine [18].

On the evidence for transcription of 21U RNAs by RNA polymerase II

The upstream motif differences of male and female 21U RNAs suggest that germline enrichment could be achieved through selective transcription in male versus female germlines. Recently, Cecere et al. reported that 21U RNA upstream regions are depleted of nucleosomes [35]. They further observed that RNA polymerase II (Pol II) occupancy shows local peaks in this region, rising steadily over the interval of -300 nt to -50 nt from the genomic thymidine encoding the 5' uridine of the 21U RNA. Analyzing the same ChIP-seq dataset as Cecere et al., we noticed that the amplitude of the

changes in Pol II occupancy at 21U RNA loci is quite modest. Analyzing randomly generated intergenic windows from chromosome IV, we determined that the Pol II ChIP-seq background actually exceeds the “signal” at 21U RNA loci (Figure 3.S7A,B), indicating relative Pol II depletion. This overall depletion of Pol II occupancy at 21U RNA loci may indicate that transcription of 21U RNAs is a more transient process than transcription of genes with canonical promoter elements. Thus the ChIP-seq might capture only a small fraction of interactions between Pol II and DNA. However, the Pol II occupancy profiles for the loci encoding the top 25% and bottom 25% of 21U RNAs by abundance are virtually indistinguishable (Figure 3.S7C). Again, this is in stark contrast to mRNA coding loci, for which Pol II occupancy at the top 25% of mRNAs by abundance is much higher than at the bottom 25% (Figure 3.S7D). An alternative possibility is that the open chromatin of the nucleosome-depleted regions upstream of 21U RNA loci is more susceptible to incidental binding by Pol II, causing the modest increase in local occupancy observed by Cecere et al. Should this be the case, the products of Pol II transcription at these loci could be unrelated to 21U RNAs. Cecere et al. also identify a transcript whose 5' end extends 2 nt upstream of a 21U RNA locus and note that deep sequencing of 5' capped RNAs reveals many more such transcripts. While these transcripts may represent 21U RNA precursors, they may also represent the products of incidental transcription from 21U RNA loci exposed due to local nucleosome depletion. The levels of such long putative precursors were below the threshold of our detection, precluding further study. Nevertheless, the uncertain 5' nucleotide identity of the nascent 21U RNA transcript does not affect the interpretation of our results. Further studies, including identification of a cleavage mechanism for the 2 nt 5' overhang, are needed to confirm these capped transcripts as bona fide 21U RNA precursors. The Zucchini endoribonuclease, thought to generate piRNA 5' ends in mouse and fly [16,17], is not a likely candidate, as it has no obvious homolog in *C. elegans* and shows very little sequence specificity, nor is there any evidence in *C. elegans* for processing of a long 21U RNA precursor into multiple species.

How are the male and female subsets of 21U RNAs differentially expressed?

We show that the 5' nucleotide of the conserved core motif influences germline enrichment of the dependent 21U RNA species (Figures 3.4, 3.6). This differential expression of male and female 21U RNAs may be orchestrated by DNA-binding proteins that differ in germline expression patterns and/or binding affinity for 5' cytidine core motifs. Recently, Cecere et al. demonstrated that the forkhead transcription factors UNC-130, FKH-3, and FKH-5 specifically bind a CTGTTTCA-containing substrate dsDNA probe in vitro [35]. However, male and female 21U RNAs do not appear to be differentially sensitive to depletion of these forkhead proteins, nor do 21U RNAs with and without 5' cytidine motifs (data not shown and [35]). Cecere et al. propose that these forkhead proteins play a redundant role in transcription of 21U RNAs. While these are dispensable for viability and fertility, other forkhead proteins are required for development of the germline, precluding testing for a role in transcribing 21U RNAs; these additional forkhead proteins could indeed represent germline-specific or motif-specific transcription factors (Figure 3.S8).

Why are autonomous 21U RNA transcriptional units genomically clustered?

The autonomy of the *C. elegans* piRNA gene raises the questions of why 21U RNA loci exhibit genomic clustering on chromosome IV and why 21U RNAs encoded on chromosome IV are expressed more robustly. Perhaps the high density of 21U RNAs within these genomic clusters evolved as such: 21U RNA loci, defined by 21U RNA core motifs flanked by A/T richness, accumulated randomly on ancestral chromosome IV. Targeting of any overlapping genes resulted in silencing, subjecting the coding sequences of these genes to drift and eventual elimination. This would deplete the region of genes, reducing selection upon the genomic sequence and thereby permitting further accumulation of 21U RNA loci. The lack of selective pressure related to conservation of protein-coding genes might also explain why chromosome IV loci express 21U RNAs most robustly: the high density of coding and regulatory elements on other chromosomes likely constrains the evolution of features such as flanking A/T-richness that might enhance 21U RNA expression. It is also possible that different

transcriptional machineries or different chromatin configurations are required to transcribe 21U RNAs versus other elements.

Genomic clustering of piRNA loci has been proposed to provide a “trap” for mobile elements [14]. In organisms such as mouse and fly where these clusters are transcribed to generate long precursors from which piRNAs are processed [9,10,14], the trapping function of the genomic piRNA cluster is readily apparent. Although the 21U RNAs are independently transcribed, Bagijn et al. have identified a similar mechanism acting in *C. elegans*: the genome shows evidence of recent transposon integration downstream of the conserved upstream 21U RNA motif, sometimes generating 21U RNAs that are antisense to the transposon 3' end and capable of silencing it [29]. Each conserved upstream motif can therefore serve as an independent trap, with the result that increased accumulation of motifs enhances protection against mobile elements. While retroelements comprise over 40% of the human genome, they appear to have been strongly counterselected in *C. elegans*, where they constitute only 0.2% of the genome [52]. Perhaps the autonomous piRNA mechanism at play in *C. elegans* has rendered the animal less susceptible to this kind of mobile element over an evolutionary time scale. Intriguingly, however, *C. elegans* shows significantly higher rates of gene duplication than fly [53], and the *C. elegans* genome shows substantial expansions of gene families; for example, the *C. elegans* Argonaute family has expanded to over two dozen members, with the evolution of a worm-specific clade. As gene duplications, like mobile elements, may also be targeted by piRNAs, the preponderance of gene family expansions in *C. elegans* could suggest that this system confers enhanced protection against transposons at the expense of enhanced tolerance for gene duplications. Identification of additional organisms that use similar mechanisms for generating piRNAs will reveal whether this is a pattern or a peculiarity of *C. elegans*.

Note added in proof: Gu et al. recently identified global candidate RNA polymerase II transcription start sites by deep sequencing of capped RNAs [54]. For a large proportion of annotated 21U RNAs, the authors identified 5' capped, ~26 nt putative precursors with a 2 nt 5' overhang. Longer RNA reads (70-90 nt) were identified overlapping a very

small minority of 21U RNA loci. Abundance of these longer reads correlated poorly with 21U RNA abundance, while the abundance of the short, ~26 nt reads correlated well, suggesting they are likelier to represent 21U RNA precursors. The 5' cap structure of the putative 21U RNA precursor indeed suggests transcription by Pol II, although our analysis of Pol II occupancy data is inconclusive.

MATERIALS AND METHODS

Strains

C. elegans were maintained according to standard procedures. The Bristol strain N2 was used as the standard wild-type strain. The alleles used in this study, listed by chromosome, are: unmapped: *xkls11*[♂ *Scram1258* *cb-unc-119(+)*], *xkls12*[♀ *Scram2502* *cb-unc-119(+)*], *xkls14*[♂ 21U>A1258 *cb-unc-119(+)*], *xkls15*[♂ 21U>G1258 *cb-unc-119(+)*], *xkls16*[♂ *Min1258* *cb-unc-119(+)*], *xkls17*[♀ 21U>A2502 *cb-unc-119(+)*], *xkls18*[♀ 21U>G2502 *cb-unc-119(+)*], *xkls19*[♀ *Min2502* *cb-unc-119(+)*], *xkls20*[♂ *Min1415* *cb-unc-119(+)*], *xkls21*[♂ *MinC*>A1415 *cb-unc-119(+)*], *xkls22*[Tg1415-Tg2109 *cb-unc-119(+)*], *xkls23*[*Scram1415-Tg2109* *cb-unc-119(+)*]; LGX: *xkls10*[♀ A>C2502 *cb-unc-119(+)*]; LGI: *glp-4(bn2)*, *prg-1(tm872)*, *xkls5*[♀ Tg2502 *cb-unc-119(+)*]; LGII: *xkSi30* [♀ *Min2502* *cb-unc-119(+)*], *xkls6*[♂ C>A1258 *cb-unc-119(+)*]; LGIII: *rde-4(ne301)*, *henn-1(tm4477)*; LGIV: *xkSi3*[♂ Tg1258 *cb-unc-119(+)*], *xkSi13*[♀ Tg2502 *cb-unc-119(+)*], *xkSi17*[♂ C>A1258 *cb-unc-119(+)*], *xkSi20*[♀ A>C2502 *cb-unc-119(+)*], *xkSi23*[♂ *Scram1258* *cb-unc-119(+)*], *xkSi28*[♀ *Scram2502* *cb-unc-119(+)*], *fem-1(hc17)*, *him-8(e1489)*; LGV: *fog-2(q71)*, *xkls1*[♂ Tg1258 *cb-unc-119(+)*]. Transgenic allele details and corresponding strain names are shown in Table 3.1.

Sample collection and small RNA analysis

C. elegans samples were generated as previously described [24]. Samples for Taqman RT-qPCR validation of 21U RNA germline enrichment classification analysis were collected in biological duplicate. Samples collected for RNA-immunoprecipitation (RIP)

analysis were collected in biological duplicate and analyzed in independent experiments with technical duplicates. All other samples were collected in biological triplicate. All samples analyzed represent adult animals unless otherwise stated.

RNA isolation, beta-elimination, northern blot analysis, Taqman RT-qPCR, and mRNA quantitation were performed as previously described [24]. RIP analysis was performed as follows: A custom rabbit polyclonal anti-PRG-1 antibody was generated by Proteintech Group, Inc using an N-terminal peptide antigen (MASGSGRGRGRGSGSNNS (C)) conjugated to keyhole limpet hemocyanin (KLH) carrier protein. Antisera were affinity purified using Affi-Gel 10 gel (Bio-Rad). PRG-1 was purified from synchronized gravid animals using this anti-PRG-1 rabbit polyclonal antibody. For each IP, 10 µg of anti-PRG-1 antibody was cross-linked to Dynabeads Protein A (Invitrogen) and incubated with lysate prepared from 0.3 ml of frozen worms at 4°C for 1 hr. Beads were washed 4X with RIP wash buffer (50 mM Tris-HCL pH 7.5, 200 mM KCL and 0.05% NP-40). After final wash, beads were split into equal volumes for RNA extraction and western blot procedure. For western blot analysis: 30 ul of 1X Tris-glycine SDS sample buffer (Invitrogen) without DTT was added directly to beads and incubated at 50°C for 10 min. 0.1 M DTT was then added to samples and boiled for 5 min before loading on gel. Proteins immobilized on Immobilon-FL transfer membrane (Millipore) were probed with anti-PRG-1 rabbit polyclonal antibody or anti-gamma-tubulin rabbit polyclonal antibody (LL-17) (Sigma) (1:2,000). Peroxidase-AffiniPure goat anti-rabbit IgG secondary antibody was used at 1:10,000 (Jackson ImmunoResearch Laboratories) for detection using Pierce ECL Western Blotting Substrate (Thermo Scientific). For RNA extraction: 1 ml of TRI-Reagent (Ambion) was directly added to beads and incubated at room temperature for 5 min. RNAs were precipitated in isopropanol for 1 hr at -30°C followed by three washes with 70% ethanol.

Small RNA quantitation was performed as previously described [24]. All 21U RNA qPCR data from transgenic studies were normalized to miR-1 levels. As a result of this normalization, some small RNAs whose levels are not detectable (cycle number > 36) appear to be detected due to small variation in detection of miR-1. 21UR-synth is not

detectable in non-transgenic animals at any stage at which it was assessed. All *Cbr-unc-119* qPCR data were normalized to *act-1* mRNA levels. The sequence of 21UR-synth is 5' TGATATGCGATGTAGTAGACT 3'. The sequence of 21UR-synthB is 5' TTAGTCGTATGTGACGCTGCC 3'. Full small RNA sequences were submitted to Applied Biosystems for design of Taqman assays. Northern blot probe sequences used for this study: miR-1 5' TACATACTTCTTTACATTCCA /3StarFire/ 3'; ♀21UR-2502 5' CAGCAGTCTACTACAATTTCA /3StarFire/ 3'; 21UR-synth 5' AGTCTACTACATCGCATATCA /3StarFire/ 3'. RT-qPCR primer sequences used for this study are as follows: *act-1* F 5' CCAGGAATTGCTGATCGTATGCAGAA 3', R 5' TGGAGAGGGAAGCGAGGATAGA 3'; *Cbr-unc-119* F 5' AACGACGTTTTAGCACTTCCG 3', R 5' GGATTTGGAAGTGGTGAAGTTCG 3'.

C. *elegans* transgenesis

To generate the base of the 1258 transgene, sequence spanning genomic coordinates IV:14390835-14393692 was used; IV:14392513-14392673 was used for the ♂Min1258 transgene. To generate the base of the 2502 transgene, sequence spanning genomic coordinates IV:15395699-15397722 was used; IV:15396667-15396886 was used for the ♀Min2502 transgene. To generate the base of the Min1415 transgene, sequence spanning genomic coordinates IV:16564187-16564395 was used. To generate the base of the Tg1415-Tg2109 transgene, sequence spanning genomic coordinates IV:16564133-16564395 was used; the Tg1415-Tg2109 and Scram1415-Tg2109 transgenes carry a 13 nt deletion downstream of both 21U RNA loci. Coordinates were taken from the *C. elegans* genome WS220. The mutations described in Table 3.1 were introduced through site-directed mutagenesis or inverse PCR with phosphorylated primers. Transgenes were then subcloned into the pCFJ178 (IV) or pCFJ151 (II) vector. The chromosome IV transgene insertion site lies outside the larger 21U RNA genomic clusters, and the homology arms of chromosome IV MosSCI vector pCFJ178 do not encode any annotated 21U RNAs. Transgenes were confirmed by sequencing and injected into animals with pharyngeal and/or body wall muscle coinjection markers to distinguish transgenic animals. High-copy arrays were integrated through ultraviolet irradiation. MosSCI single-copy insertions were generated as previously described [47].

Small RNA sequencing data acquisition and linker removal

Raw data files from 24 small RNA sequencing experiments [27,36-42] were downloaded from NCBI Gene Expression Omnibus [55]. Artificial linker sequences were removed using an in-house linker removal pipeline. We first searched each sequence for a perfect match to the linker. If a perfect match was not found, we searched for an alignment to the linker with 1 mismatch. If not found, we searched for a perfect alignment between the last 5 nt of the sequence and the first 5 nt of the linker. If not found, we repeated this search allowing 1 mismatch. We continued this pattern to align 4 and 3 nt. Sequences with no linker alignment were discarded (~20% of reads).

Small RNA read alignment to genome and annotation to 21U RNAs

Reads were aligned to the reference *C. elegans* genome version WS220 using Bowtie [56] with the following parameters: -f -v 2 -k 50 --best --strata. Mapped read counts in each library were normalized to the number of total mapped reads in that library and to the number of mapped genomic loci. Sequence abundance is reported as reads per million mapped reads (RPM). To determine 21U RNA abundance, we first generated 21U RNA genomic coordinates by aligning 15,703 known 21U RNA sequences [27] to the *C. elegans* genome version WS220 using Bowtie. Perfect, full-length alignments for 15,093 of these sequences were considered valid 21U RNA coordinates. Reads mapping entirely within these coordinates were annotated to 21U RNAs.

Enrichment Score calculations

Germline enrichment classifications of 21U RNAs were generated based on read counts in 17 germline libraries: 14 male germline libraries prepared from isolated spermatogenic cells, isolated spermatids, or whole adult males; and 3 female germline libraries prepared from purified oocytes or whole adult hermaphrodites defective in sperm production (Table 3.S1). 1,198 21U RNAs had no read counts in any of these libraries and were removed from our analysis. 184 21U RNAs had higher read counts in a *prg-1(tm872)* young adult library compared to an N2 young adult library [27] and were removed from our analysis, leaving 13,711 21U RNAs for which we assessed germline enrichment. Libraries generated using a 5'-monophosphate-dependent (5 male, 1

female) versus -independent (9 male, 2 female) protocol were separated for calculation of the Enrichment Score as follows: For each 21U RNA, we calculated fold abundance difference between every male and female library, for a total of 23 comparisons. Each 21U RNA began with an Enrichment Score of 0. For every comparison, if the 21U RNA was more than 5-fold abundant in the male library, the Enrichment Score decreased by 1; if the 21U RNA was more than 5-fold abundant in the female library, the Enrichment Score increased by 1. Male 21U RNAs were defined as those with Enrichment Scores ≤ -3 , while female 21U RNAs were defined as those with Enrichment Scores ≥ 3 . Remaining 21U RNAs were classified as non-enriched. To validate enrichment classifications, the fold abundance differences for each 21U RNA were averaged across all 23 comparisons. Less than 1% of 21U RNAs classified as male or female do not show enrichment by average fold abundance in their respective libraries. These 21U RNAs were reclassified as non-enriched for subsequent analyses. 21U RNA Enrichment scores and germline enrichment classifications are in Dataset 3.S1.

Determination of false discovery rate

To approximate the number of 21U RNAs falsely classified as male or female germline-enriched by our method, we performed Enrichment Score calculations on randomly generated count data modeled from an N2 young adult library [27]. 11,458 21U RNAs are represented in this library. Because 17 germline libraries were used for the real analysis, we generated 17 control libraries as follows: For each 21U RNA, 17 random counts were generated from a Poisson distribution with $\lambda = \alpha$ (where α is set to the 21U RNA count in the N2 library) and assigned to one of 17 control libraries. After all counts were assigned, the 17 control libraries were randomly grouped to represent the number of male or female and 5'-monophosphate-dependent or -independent libraries used above. Enrichment Score calculations were then performed on these control libraries as described above, and the number of 21U RNAs classified as germline-enriched was calculated. This protocol was repeated 1,000 times. On the basis of this randomized data, we defined an Enrichment Score threshold of $-/+3$, inclusive, for classifying 21U RNAs as male or female germline-enriched, respectively. Application of this threshold to the randomized data resulted in classification of, on average, only 0.76% (101 of

11,458) of 21U RNAs as germline-enriched, corresponding to a false discovery rate below 1%. This value is consistent with the less than 1% of 21U RNAs classified as male or female that do not show enrichment by average fold abundance in their respective libraries.

Enrichment Score calculations performed on 26G RNAs

26G RNA annotations were taken from Han et al., 2009 [38]. The abundances of 4,002 26G RNAs were measured in 13 of the 17 libraries used for 21U RNA Enrichment Score calculations. Four male libraries (GSM465244, GSM503843, GSM459329, and GSM459331) were excluded because the animals used in preparation of the libraries carried mutations in genes required for 26G RNA expression [36,40,41]. Enrichment Score calculations were performed on the 13 remaining libraries as above, for a total of 16 male:female comparisons. We retained the Enrichment Score threshold for classifying 26G RNAs as male or female germline-enriched.

Analysis of 21U RNA-dependent 22G RNAs and 21U RNA targets

21U RNA target and 22G RNA information for young adult animals (N2 and *prg-1(n4357)*) was obtained from Bagijn et al. [29]; raw sequencing data files for gravid adult animals (N2 and *prg-1(n4357)*) were downloaded from GEO [30]. Raw sequences were processed as described above, and reads 22 nt long and starting with guanosine were annotated as 22G RNAs. 21U RNA targets were defined as transcripts with 0-3 mismatches to a 21U RNA sequence. 21U RNA-dependent 22G RNAs were defined as 22G RNAs that map antisense to transcripts within 40 nt of a 21U RNA target site. The number of 22G RNAs that map to both male and female 21U RNA target sites was compared to a control number of 22G RNAs that map to both a random set of male and a random set of female 21U RNA target sites. These random target sites were defined as the target sites of 7,677 randomly selected 21U RNAs representing “male” 21U RNAs and the target sites of 2,171 randomly selected (and not overlapping random male) 21U RNAs to represent “female” 21U RNAs. This random selection was repeated 1,000 times. A similar randomization process was repeated to compare with the number of genes targeted by both male and female 21U RNAs.

Core motif visualization

Core motifs of 21U RNAs were visualized using WebLogo and correcting for *C. elegans* genome nucleotide composition [57]. To account for variability in the location of core motifs relative to their 21U RNA loci, upstream regions were aligned by the central 3 Ts of the core motif. If no core motif was identified within 60 nt upstream of a 21U RNA, we aligned position -44 relative to the 21U RNA locus to the G of the core motif, corresponding to the previously identified most common position of the G [12]. Only 21U RNAs that map to a single locus in the genome (13,387 of 13,711 21U RNAs, 97.6%) were analyzed since 21U RNAs that map to more than 1 locus may have different upstream sequences.

Identification of genomic features for nucleosome and Pol II occupancy profiling

Nucleosome and Pol II occupancy profiling for 21U RNA loci was centered on the genomic thymidine encoding the 21U RNA 5' uridine. Profiling for transcripts was centered on transcription start sites (TSS) defined as the start of 5' UTRs annotated in the Ensembl66 database [58]. Intergenic regions were defined as regions absent of an annotated 5' UTR, exon, intron, 3' UTR or small RNA transcript that were partitioned into randomly distributed, non-overlapping 1,000 nt windows. Profiling for intergenic regions was centered on these 1,000 nt windows. Young adult TSS expression was calculated as fragments per kilobase per million mapped reads (FPKM) using biological replicates from a transcriptomic sequencing experiment [59]. Transcriptome sequence data were removed of linkers and aligned to the *C. elegans* genome version WS220 using TopHat [60]. Cufflinks [61] was used to calculate transcript isoform expression. Transcripts with an annotated 5' UTR were extracted from the Ensembl66 database. Average transcript FPKM across the two libraries was calculated, and the isoform with the highest expression was chosen for nucleosome and Pol II occupancy analyses. For isoforms with equivalent expression, a single isoform was randomly chosen.

Analysis of nucleosome and Pol II occupancy

Published nucleosome occupancy data [62] were downloaded from UCSC, and the genomic coordinates were lifted over from WS170 to WS220. Adjusted nucleosome

occupancy data centered on 21U RNAs, TSS, and an intergenic background control were averaged for each nucleotide. Pol II ChIP-seq data from young adult worms were downloaded from the modEncode repository [63]. Pol II signal to input ratios on chromosome IV were averaged for each nucleotide. TSS were further filtered to only include transcripts with at least 5 FPKM as calculated above. Pol II ChIP-seq data from young adult worms were downloaded from the modEncode repository. Pol II signal to input ratios were averaged for each nucleotide separately for male and female 21U RNAs on chromosome IV.

ACKNOWLEDGMENTS

We thank T. Han, J. Moran, S. Kalantry, E. Fearon, and S. Camper for discussions and comments on the manuscript and Germano Cecere and Alla Grishok for generously sharing methods. We also thank the *Caenorhabditis* Genetics Center for *C. elegans* for strains.

SUPPLEMENT

Figure 3.S1: Computational identification of male and female germline-enriched 21U RNAs. (A) Enrichment Score calculations performed on 17 small RNA sequencing libraries classify a majority of 21U RNAs as male (blue) or female (red) germline-enriched. Non-enriched (NE) 21U RNAs, grey. Numbers indicate percent of 13,711 21U RNAs analyzed. (B) Enrichment Score calculations performed on control data classify < 1% of 21U RNAs as male or female germline-enriched indicating a 1% false discovery rate. Numbers indicate percent of 11,458 21U RNAs analyzed. (C) Male 21U RNAs are more abundant in male libraries. Average relative abundance of each male 21U RNA was calculated between each of the 23 male:female library comparisons. (D) Female 21U RNAs are more abundant in female libraries. Average relative abundance of each female 21U RNA was calculated between each of the 23 female:male library comparisons. (E) Non-enriched 21U RNAs are equally abundant in male and female libraries.

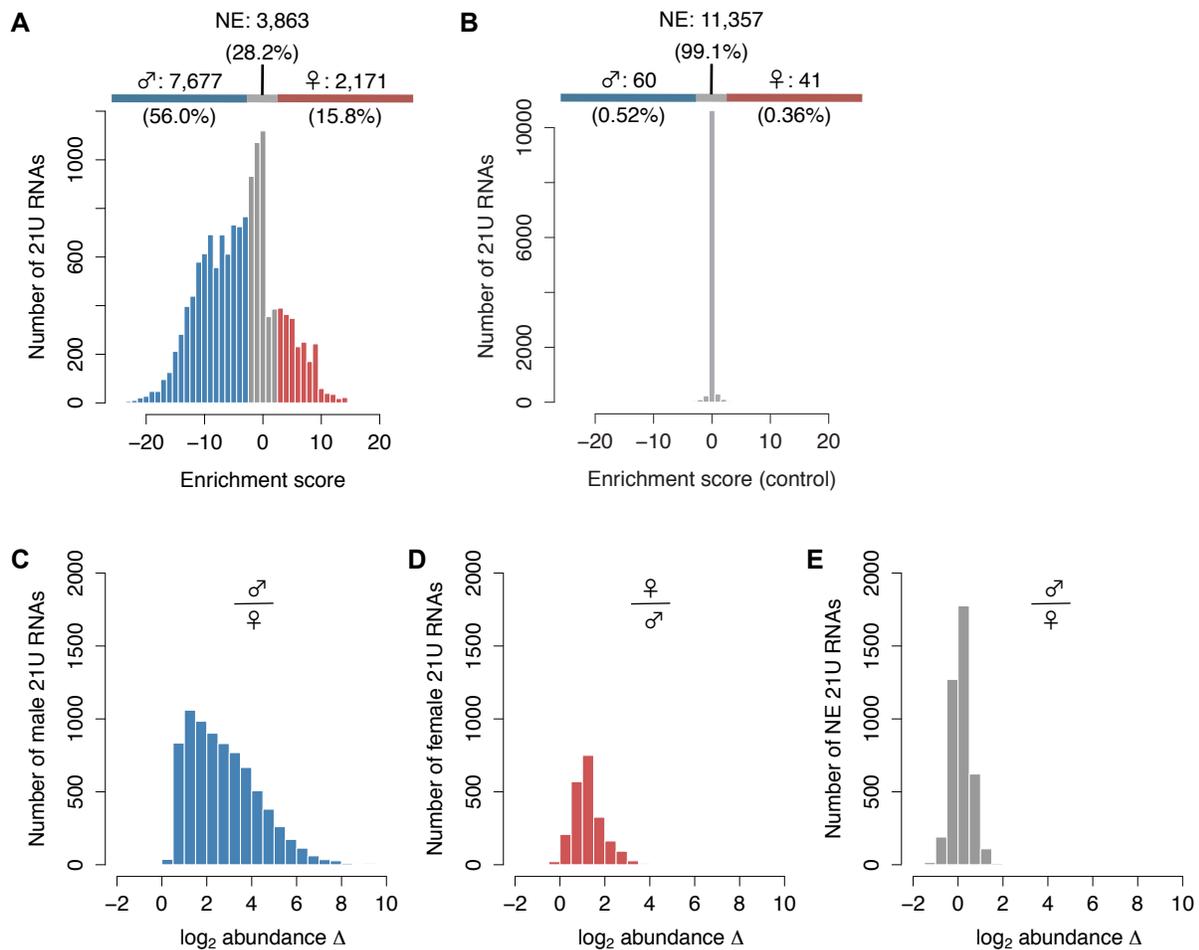


Figure 3.S2: Female 21U RNAs are preferentially abundant in embryo. (A) Relative male 21U RNA abundance is decreased in embryo. Average relative abundance of each male 21U RNA was calculated between each of 5 male and 4 mixed stage embryo libraries. Dotted line indicates equal male and embryo abundance. Pie chart depicts proportion of male 21U RNAs with reads in at least one embryo library (dark blue). (B) A population of female 21U RNAs shows increased abundance in embryo. Average relative abundance of each female 21U RNA was calculated between each of 1 female and 4 mixed stage embryo libraries. Pie chart depicts proportion of female 21U RNAs with reads in at least one embryo library (dark red). (C,D) Taqman RT-qPCR analysis corroborates male 21U RNA depletion in embryo. Expression of representative male 21U RNAs was assayed by Taqman in *him-8(e1489)* (C) and *fog-2(q71)* (D) male animals and N2 embryos. Error bars represent ± 1 SD from two biological replicates. (E) Taqman RT-qPCR analysis corroborates female 21U RNA enrichment in embryo. Expression of representative female 21U RNAs was assayed by Taqman in *fem-1(hc17)* female animals and N2 embryos. (F) Male germline-enriched 26G RNAs are generally absent in embryo. Average relative abundance of each male 26G RNA was calculated between each of 4 male and 4 mixed stage embryo libraries. (G) Female germline-enriched 26G RNAs are robustly expressed in embryo. Average relative abundance of each female 26G RNA was calculated between each of 1 female and 4 mixed stage embryo libraries.

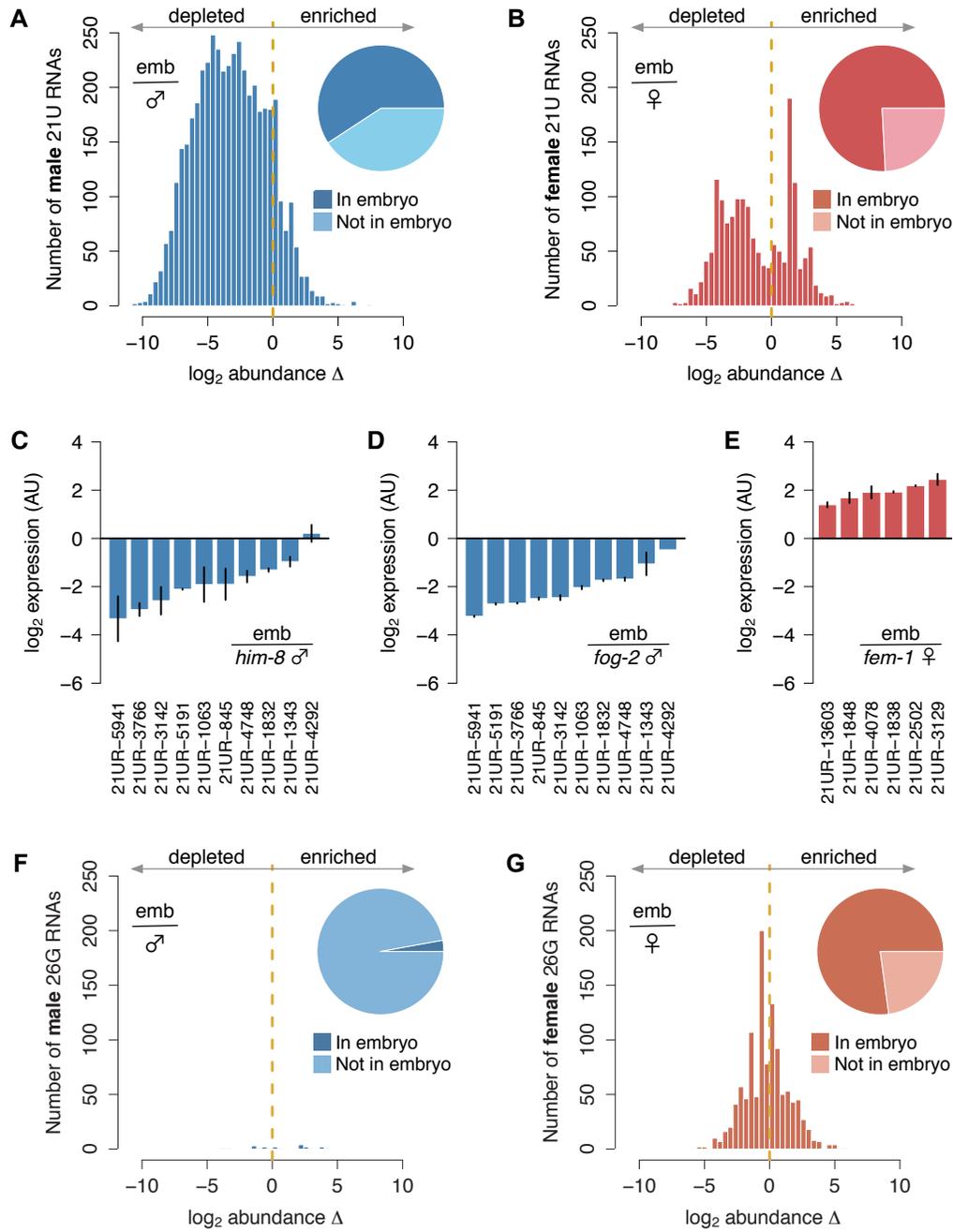


Figure 3.S3: 21U RNAs target significantly non-overlapping sets of genes. (A) 22G RNAs are almost exclusively derived from either male or female 21U RNAs, but not both. The number of unique 22G RNAs derived from both male and female 21U RNAs is significantly less than expected if 22G RNAs are selected at random (Fisher's exact test, $p=1.2e-02$). (B) Male and female 21U RNAs target significantly fewer overlapping genes compared to selecting random sets of genes (Fisher's exact test, $p=7.7e-13$). (C) 5,956 genes targeted by male 21U RNAs in young adult (YAd) animals are depleted of spermatogenesis genes compared to a random set of 5,956 genes. (D) 1,387 genes targeted by female 21U RNAs in gravid adult (GA) animals are enriched for oogenesis genes compared to a random set of 1,387 genes.

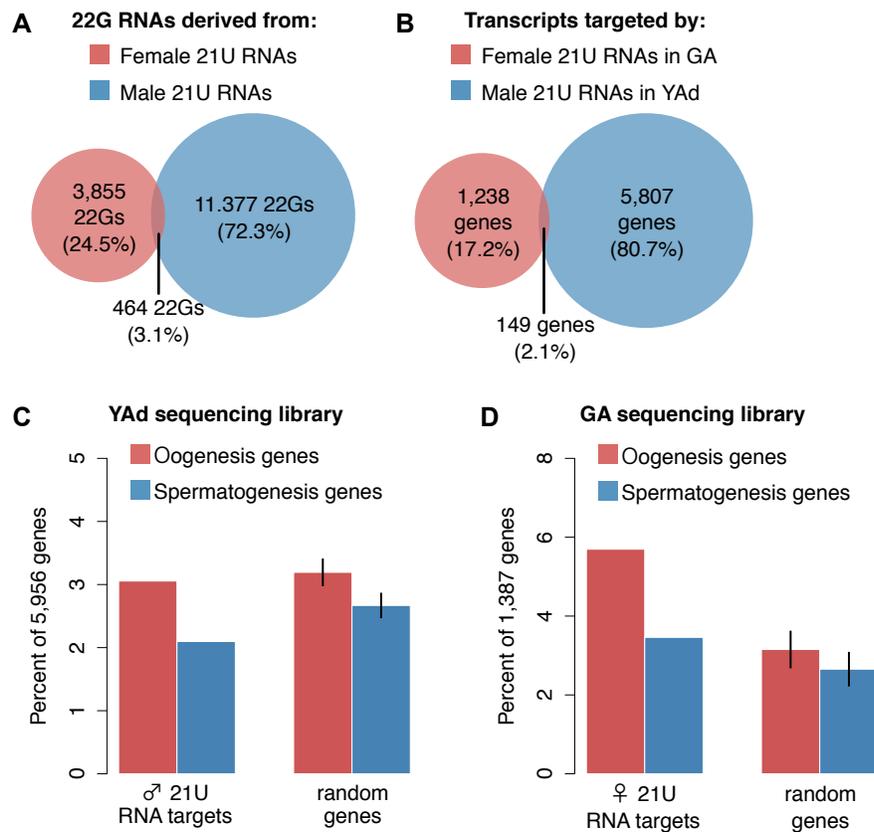


Figure 3.S4: Transgenic array expression varies across transgenes. (A) Levels of *Cbr-unc-119* mRNA in adult animals were assayed by RT-qPCR for all transgenes and normalized to *act-1* mRNA levels. (B) Expression of transgenic 21UR-synth does not affect expression of endogenous 21U RNA counterparts. Endogenous ♂21UR-1258 and ♀21UR-2502 levels were assayed by Taqman RT-qPCR and normalized to microRNA miR-1 levels in all samples.

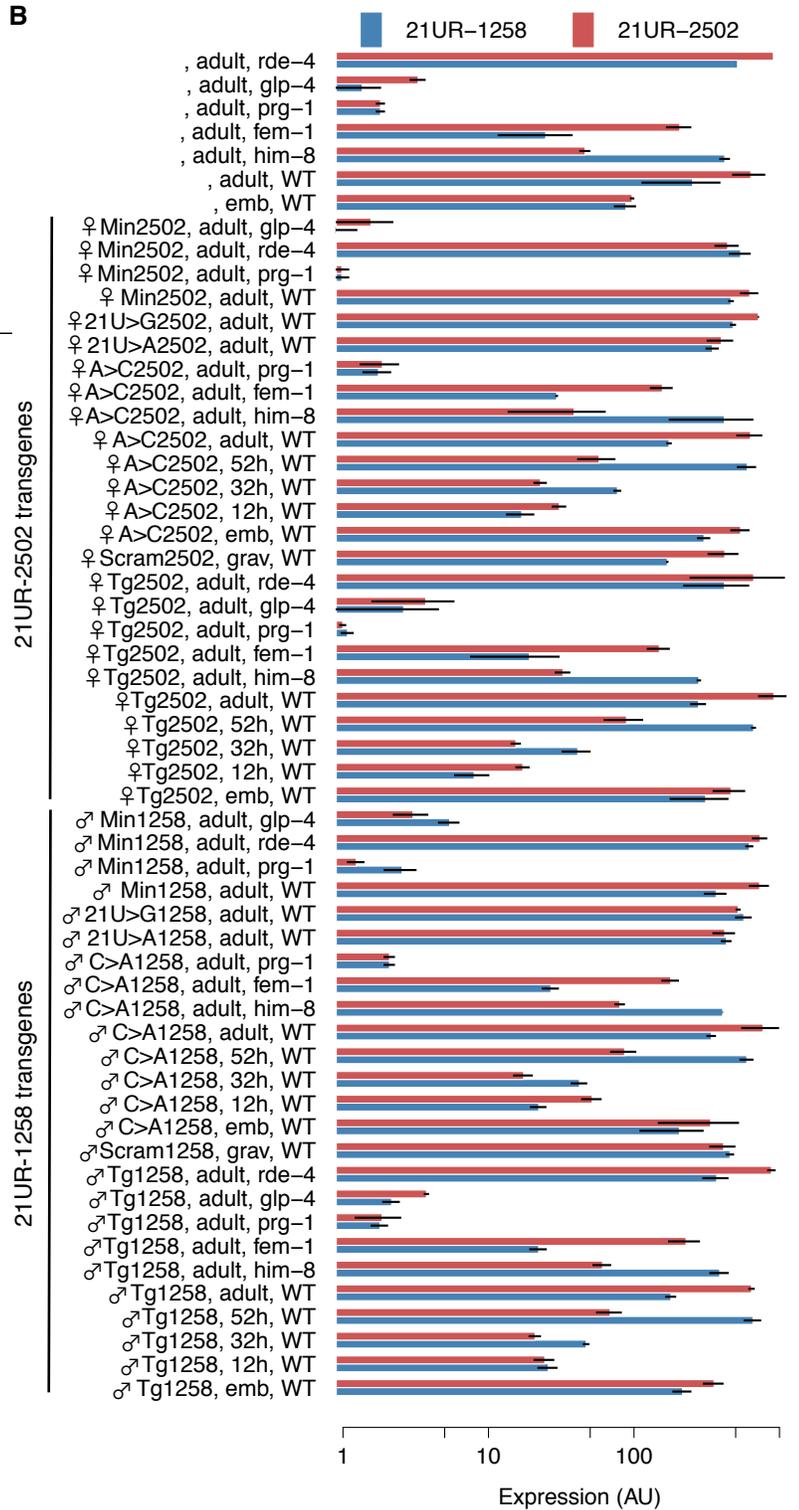
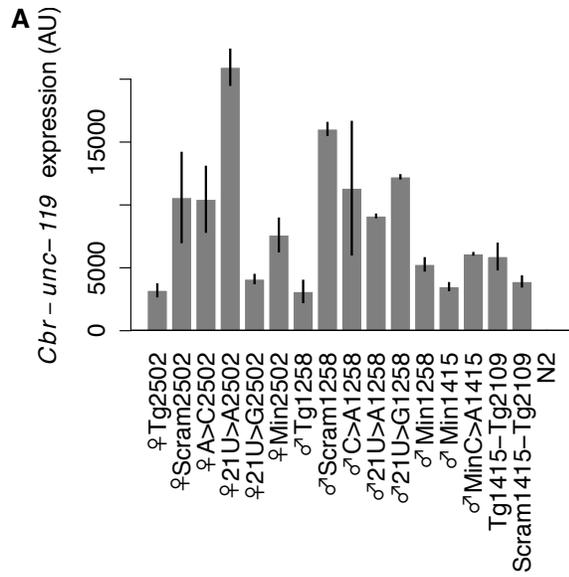


Figure 3.S5: 21U RNAs are specifically immunoprecipitated with PRG-1 complexes. (A) anti-PRG-1 antibody does not immunoprecipitate microRNA miR-1. (B) 21UR-synth expression does not interfere with association of endogenous 21U RNAs with PRG-1.

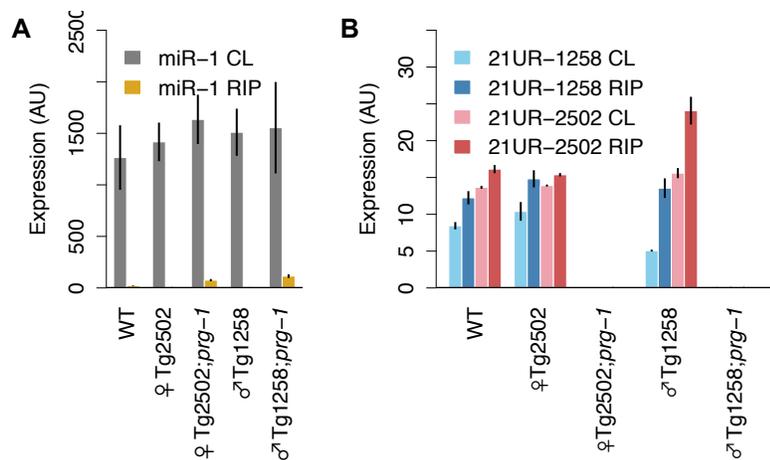


Figure 3.S6: 21U RNA expression requires a 5' genomic thymidine. (A) Schematic of transgenes encoding 21UR-synth with different 5' nt. (B-C) Mutation of the 5' genomic thymidine disrupts expression of 21UR-synth by northern blot (B) and Taqman assay (C). WT and ♀Tg2502 lanes in (B) are repeated from Figure 3.5B for clarity.

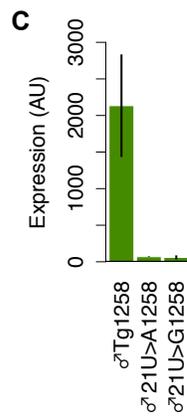
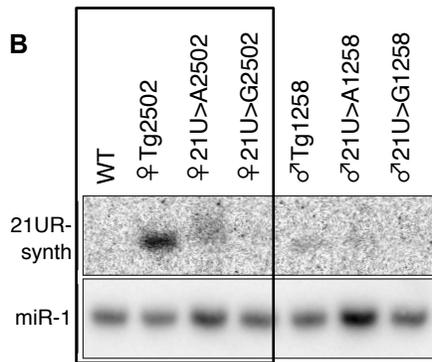
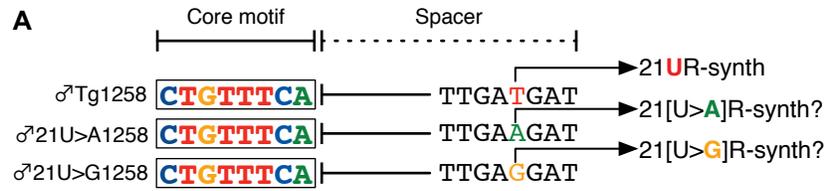


Figure 3.S7: RNA polymerase II occupancy at 21U RNA loci is below background level. (A) Average Pol II occupancy in a young adult library of 21U RNA loci expressing 21U RNAs with at least 5 RPM (red), transcriptional start sites (TSS) expressing transcripts with at least 5 FPKM (green), and randomized intergenic regions (yellow). Only regions on ChrIV were assayed (B) Pol II occupancy as described in (A) but independently scaled for each transcript type and plotted with average nucleosome occupancy (black line). Grey error bands: SEM. (C) Average Pol II occupancy of 21U RNA loci as (B) but showing the top 25% 21U RNAs by abundance (1st quartile) and the bottom 25% (4th quartile) separately. 21U RNAs on all chromosomes are shown. (D) Same as (E) but showing top and bottom 25% of TSS by transcript abundance.

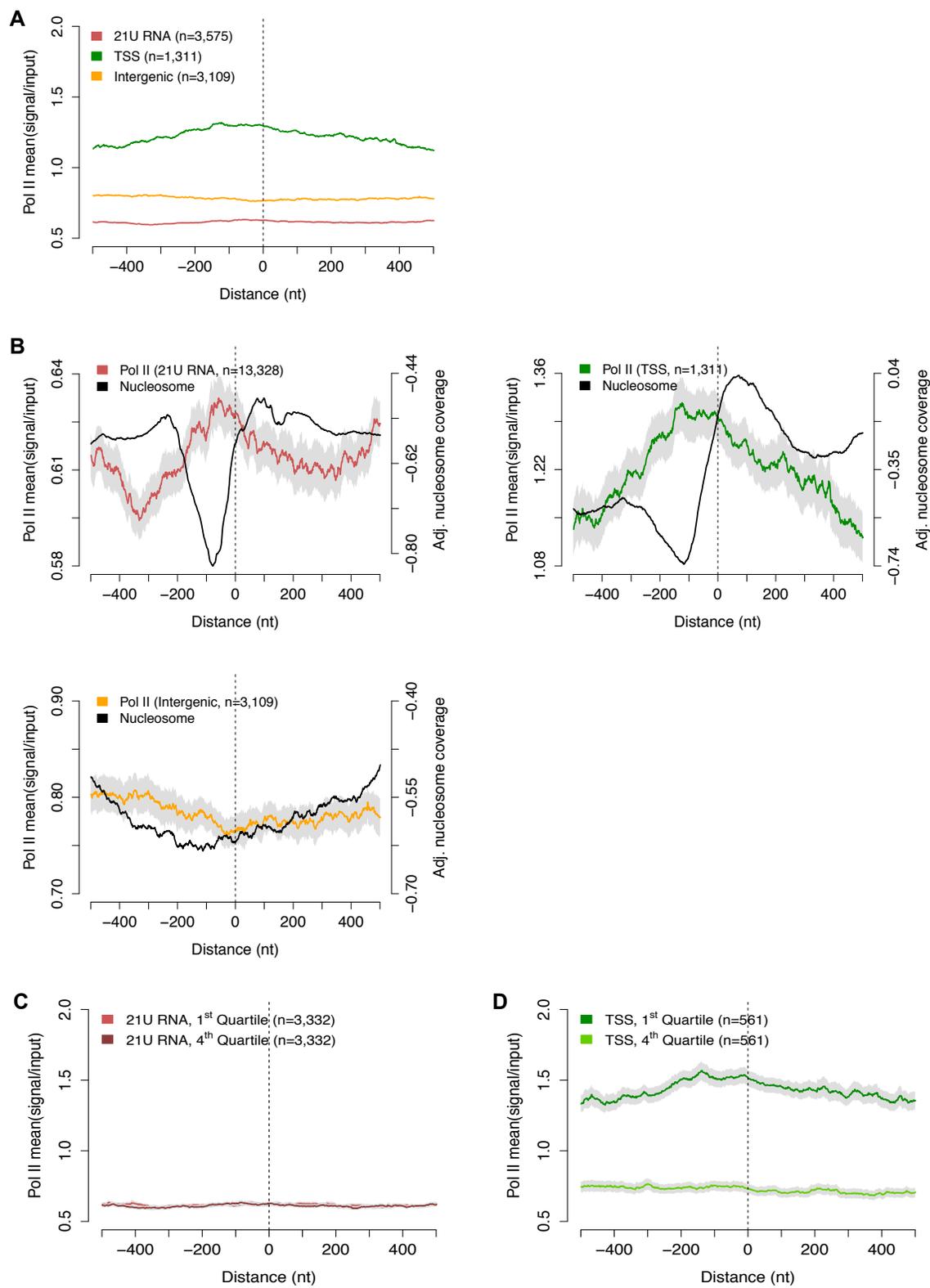


Figure 3.S8: Model of 21U RNA expression.

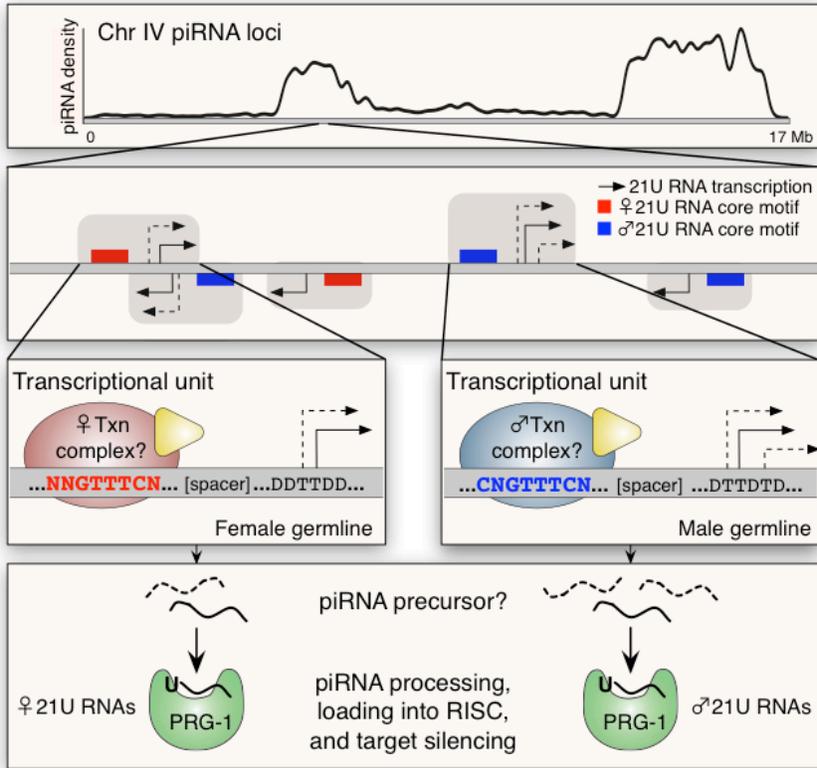


Table 3.S1: Descriptions of small RNA sequencing libraries used in this study. GEO Accessions for datasets and libraries used are listed. Libraries generated using 5'-monophosphate-dependent (Dep) or -independent (Indep) RNA extraction protocols are indicated along with how the library was used in this study ("Use" column).

Table S1. Descriptions of small RNA sequencing libraries used in this study.											
<i>GEO Accessions</i>											
Dataset	Library	Genotype	Developmental Stage	Extraction protocol	Sequencing platform	Raw reads	Mapped reads	%	21U RNA reads	%	Use
GSE20341	GSM510085	N2	mixed-stage embryos	Dep	Illumina	4,661,109	3,721,872	79.8	31,837	0.9	Embryo analysis
	GSM509932/ GSM510089	<i>him-8(e1489)</i>	isolated spermatogenic cells	Dep	Illumina/454	9,596,732	1,312,607	13.7	25,024	1.9	Germline enrichment
	GSM509933/ GSM510090	<i>fer-1(hc1)</i>	purified unfertilized oocytes	Dep	Illumina/454	6,488,731	2,165,341	33.4	47,731	2.2	Germline enrichment
GSE11738	GSM297742	N2	mixed-stage embryos	Dep	Illumina	2,730,450	2,382,829	87.3	52,072	2.2	Embryo analysis
	GSM297751	N2	young adult	Dep	Illumina	3,533,717	3,169,078	89.7	333,587	10.5	Random control, spacer analyses
	GSM297755 GSM297753	<i>prg-1(tm872)</i> <i>fog-2(q71)</i>	young adult young adult	Dep Dep	Illumina Illumina	3,588,293 3,387,268	3,303,711 2,960,986	92.1 87.4	2,577 297,715	0.1 10.1	21U RNA filtering Germline enrichment
GSE18215	GSM455395	<i>fem-1(hc17)</i>	purified oocytes	Indep	Illumina	8,496,639	7,575,752	89.2	53,848	0.7	Germline enrichment
GSE19414	GSM503834	<i>fem-1(hc17)</i>	adult	Indep	Illumina	389,636	369,130	94.7	224	0.1	Germline enrichment
	GSM503842	<i>mut-16(mg461); fem-3(q20)</i>	isolated spermatogenic cells	Dep	Illumina	425,438	399,905	94.0	94,863	23.7	Germline enrichment
	GSM503843	<i>rff-3(pk1426); fem-3(q20)</i>	isolated spermatogenic cells	Dep	Illumina	650,621	608,159	93.5	46,864	7.7	Germline enrichment
GSE17153	GSM427297	N2	mixed-stage embryos	Dep	Illumina	2,159,213	1,681,110	77.9	6,926	0.4	Embryo analysis
GSE13339	GSM336052	N2	mixed-stage embryos	Dep	Illumina	6,391,734	2,746,387	43.0	29,378	1.1	Embryo analysis
	GSM336086	<i>dpy-28(y1); him-8(e1489)</i>	young adult	Dep	Illumina	3,653,638	1,357,061	37.1	21,778	1.6	Germline enrichment
GSE18729	GSM465244	<i>alg-3(tm1155); alg-4(ok1041); fog-2(q71)</i>	adult	Indep	Illumina	3,216,031	3,003,318	93.4	237,635	7.9	Germline enrichment
	GSM465245	<i>fog-2(q71)</i>	adult	Indep	Illumina	821,513	757,771	92.2	45,592	6.0	Germline enrichment
	GSM465246	<i>fog-2(q71)</i>	adult	Indep	Illumina	2,740,511	2,562,914	93.5	157,991	6.2	Germline enrichment
	GSM465247	<i>fem-3(q20)</i>	isolated spermatids	Indep	Illumina	10,478,418	7,131,378	68.1	256,516	3.6	Germline enrichment
GSE18429	GSM459328	<i>fem-3(q20)</i>	isolated spermatogenic cells	Indep	Illumina	375,816	341,766	90.9	8,029	2.3	Germline enrichment
	GSM459329	<i>rff-3(pk1426); him-8(e1489)</i>	young adult	Indep	Illumina	1,756,561	1,673,756	95.3	15,702	0.9	Germline enrichment
	GSM459330	<i>him-8(e1489)</i>	young adult	Indep	Illumina	1,709,934	1,614,576	94.4	3,515	0.2	Germline enrichment
	GSM459331	<i>rff-3(pk1426); him-8(e1489)</i>	young adult	Indep	Illumina	1,492,360	1,366,804	91.6	9,238	0.7	Germline enrichment
	GSM459332	<i>him-8(e1489)</i>	young adult	Indep	Illumina	755,623	695,239	92.0	2,191	0.3	Germline enrichment

Table 3.S2: Welch's *t*-test p-values for all abundance comparisons between 21U RNAs with different core motifs. Highlighted are p-values <0.01. Identity of the 5' nt corresponding to higher 21U RNA abundance is indicated below each significant p-value. All *t*-tests are two-tailed. Comparisons of abundances in 5'-monophosphate-dependent and -independent libraries were performed separately.

Table S2. Welch's <i>t</i>-test p-values for all abundance comparisons between 21U RNAs with different core motifs.											
<i>Library type</i>	<i>Enrichment classification</i>	<i>A vs. C</i>	<i>A vs. G</i>	<i>A vs. T</i>	<i>A vs. N</i>	<i>C vs. G</i>	<i>C vs. T</i>	<i>C vs. N</i>	<i>G vs. T</i>	<i>G vs. N</i>	<i>T vs. N</i>
Male (5'-mP _i -Dep.)	Male	2.4E-13 C	3.5E-01	3.5E-03 A	2.0E-08 A	1.2E-03 C	6.0E-11 C	1.3E-56 C	3.6E-01	2.2E-01	7.7E-01
	Non-enriched	1.9E-01	4.2E-01	2.8E-01	3.3E-05 A	1.5E-01	3.0E-02	2.1E-12 C	9.0E-01	3.1E-01	5.3E-02
Female (5'-mP _i -Dep.)	Female	6.1E-03 A	2.4E-02	9.8E-01	1.4E-01	4.0E-01	1.9E-02	6.8E-02	3.1E-02	9.6E-02	2.2E-01
	Non-enriched	6.0E-03 A	5.0E-01	9.3E-01	4.6E-03 A	5.9E-01	3.6E-02	9.5E-01	4.9E-01	5.7E-01	3.2E-02
Male (5'-mP _i -indep.)	Male	4.06E-09 C	2.48E-01	6.40E-04 A	5.14E-07 A	1.09E-03 C	1.90E-11 C	8.21E-45 C	2.72E-01	3.28E-01	6.79E-01
	Non-enriched	4.04E-03 A	8.76E-01	9.41E-01	1.84E-04 A	2.59E-01	2.99E-02	2.73E-01	8.49E-01	1.33E-01	5.31E-03 T
Female (5'-mP _i -indep.)	Female	1.53E-05 A	2.62E-03 A	4.17E-01	4.44E-04 A	4.20E-01	5.98E-03 T	1.06E-01	1.72E-02	1.18E-01	6.69E-02
	Non-enriched	4.12E-06 A	8.19E-01	8.04E-01	6.90E-04 A	9.71E-02	1.99E-03 T	1.10E-01	9.42E-01	2.40E-01	2.63E-02

P-values colored gold are significant at p<0.01. Letters below p-values indicate which nt corresponds to higher abundance and are colored to match Weblogos. Boxed p-values/letters show that male 21U RNAs with 5'-cytidine motifs are more abundant than any other 5'-nt. mPi-indep: monophosphate independent; mPi-dep: monophosphate dependent.

Dataset 3.S1: List of 21U RNAs analyzed in this study, their Enrichment scores, their germline enrichment classifications based on Enrichment scores (Classification by score), and final germline enrichment classifications after removal of 21U RNAs whose average fold abundance in the enriched germline was not higher than the non-enriched germline (Adj. for avg. fold abundance).

<http://www.editorialmanager.com/pgenetics/download.aspx?id=308035&guid=57108f22-bdbb-44f2-b362-9a1858707257&scheme=1>

REFERENCES

1. Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, et al. (2008) Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* 455: 1193-1197.
2. Lin H, Spradling AC (1997) A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development* 124: 2463-2476.
3. Cox DN, Chao A, Baker J, Chang L, Qiao D, et al. (1998) A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev* 12: 3715-3727.
4. Schmidt A, Palumbo G, Bozzetti MP, Tritto P, Pimpinelli S, et al. (1999) Genetic and molecular characterization of sting, a gene involved in crystal formation and meiotic drive in the male germ line of *Drosophila melanogaster*. *Genetics* 151: 749-760.
5. Mochizuki K, Fine NA, Fujisawa T, Gorovsky MA (2002) Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell* 110: 689-699.
6. Carmell MA, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, et al. (2007) MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Developmental cell* 12: 503-514.
7. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, et al. (2008) DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes & Development* 22: 908-917.
8. Li C, Vagin VV, Lee S, Xu J, Ma S, et al. (2009) Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* 137: 509-521.
9. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, et al. (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 442: 203-207.
10. Girard A, Sachidanandam R, Hannon GJ, Carmell MA (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442: 199-202.
11. Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, et al. (2006) Characterization of the piRNA complex from rat testes. *Science* 313: 363-367.
12. Ruby JG, Jan C, Player C, Axtell MJ, Lee W, et al. (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* 127: 1193-1207.
13. de Wit E, Linsen SE, Cuppen E, Berezikov E (2009) Repertoire and evolution of miRNA genes in four divergent nematode species. *Genome Res* 19: 2064-2074.
14. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, et al. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128: 1089-1103.
15. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, et al. (2007) A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* 315: 1587-1590.

16. Ipsaro JJ, Haase AD, Knott SR, Joshua-Tor L, Hannon GJ (2012) The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature* 491: 279-283.
17. Nishimasu H, Ishizu H, Saito K, Fukuhara S, Kamatani MK, et al. (2012) Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature* 491: 284-287.
18. Kawaoka S, Izumi N, Katsuma S, Tomari Y (2011) 3' end formation of PIWI-interacting RNAs in vitro. *Mol Cell* 43: 1015-1022.
19. Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, et al. (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313: 320-324.
20. Kirino Y, Mourelatos Z (2007) Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. *Nature Structural & Molecular Biology* 14: 347-348.
21. Ohara T, Sakaguchi Y, Suzuki T, Ueda H, Miyauchi K (2007) The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated. *Nature Structural & Molecular Biology* 14: 349-350.
22. Kurth HM, Mochizuki K (2009) 2'-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in *Tetrahymena*. *RNA* 15: 675-685.
23. Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, et al. (2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* 129: 69-82.
24. Billi AC, Alessi AF, Khivansara V, Han T, Freeberg M, et al. (2012) The *Caenorhabditis elegans* HEN1 Ortholog, HENN-1, Methylates and Stabilizes Select Subclasses of Germline Small RNAs. *PLoS Genetics* 8: e1002617.
25. Montgomery TA, Rim Y-S, Zhang C, Downen RH, Phillips CM, et al. (2012) PIWI Associated siRNAs and piRNAs Specifically Require the *Caenorhabditis elegans* HEN1 Ortholog henn-1. *PLoS Genet* 8: e1002616.
26. Kamminga L, van Wolfswinkel J, Luteijn M, Kaaij L, Bagijn M, et al. (In Press) Differential impact of the Hen1 homolog HENN-1 on 21U and 26G RNAs in the germline of *Caenorhabditis elegans*. *PLoS Genet*.
27. Batista PJ, Ruby JG, Claycomb JM, Chiang R, Fahlgren N, et al. (2008) PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Molecular Cell* 31: 67-78.
28. Das PP, Bagijn MP, Goldstein LD, Woolford JR, Lehrbach NJ, et al. (2008) Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. *Molecular Cell* 31: 79-90.
29. Bagijn MP, Goldstein LD, Sapetschnig A, Weick EM, Bouasker S, et al. (2012) Function, Targets, and Evolution of *Caenorhabditis elegans* piRNAs. *Science*.
30. Lee HC, Gu W, Shirayama M, Youngman E, Conte D, Jr., et al. (2012) *C. elegans* piRNAs Mediate the Genome-wide Surveillance of Germline Transcripts. *Cell*.
31. Billi AC, Freeberg MA, Kim JK (2012) piRNAs and siRNAs collaborate in *Caenorhabditis elegans* genome defense. *Genome Biology* 13: 164.
32. Shirayama M, Seth M, Lee HC, Gu W, Ishidate T, et al. (2012) piRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. *Cell* 150: 65-77.

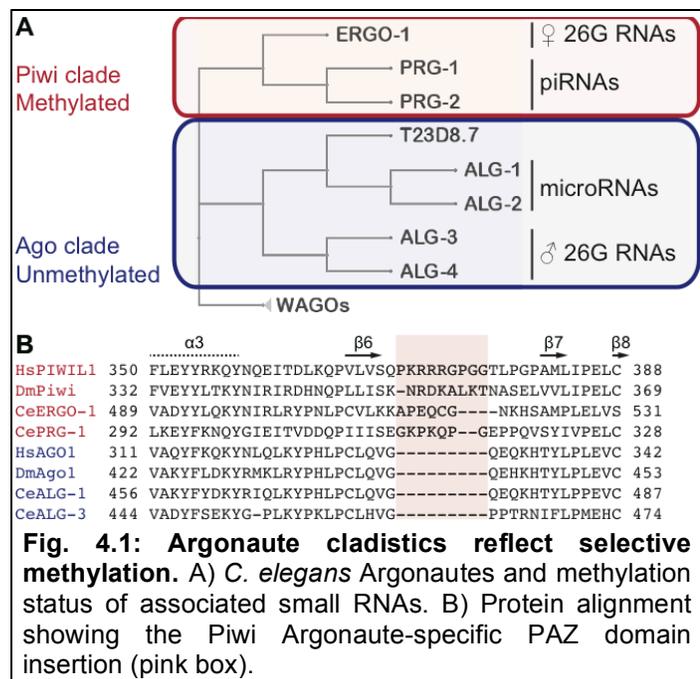
33. Ashe A, Sapetschnig A, Weick EM, Mitchell J, Bagijn MP, et al. (2012) piRNAs can trigger a multigenerational epigenetic memory in the germline of *C. elegans*. *Cell* 150: 88-99.
34. Luteijn MJ, van Bergeijk P, Kaaij LJ, Almeida MV, Roovers EF, et al. (2012) Extremely stable Piwi-induced gene silencing in *Caenorhabditis elegans*. *The EMBO journal* 31: 3422-3430.
35. Cecere G, Zheng GX, Mansisidor AR, Klymko KE, Grishok A (2012) Promoters recognized by forkhead proteins exist for individual 21U-RNAs. *Molecular Cell* 47: 734-745.
36. Gent JI, Schvarzstein M, Villeneuve AM, Gu SG, Jantsch V, et al. (2009) A *Caenorhabditis elegans* RNA-directed RNA polymerase in sperm development and endogenous RNA interference. *Genetics* 183: 1297-1314.
37. Gu W, Shirayama M, Conte D, Vasale J, Batista PJ, et al. (2009) Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. *Molecular Cell* 36: 231-244.
38. Han T, Manoharan AP, Harkins TT, Bouffard P, Fitzpatrick C, et al. (2009) 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in *Caenorhabditis elegans*. *Proc Natl Acad Sci USA* 106: 18674-18679.
39. Kato M, de Lencastre A, Pincus Z, Slack FJ (2009) Dynamic expression of small non-coding RNAs, including novel microRNAs and piRNAs/21U-RNAs, during *Caenorhabditis elegans* development. *Genome Biology* 10: R54.
40. Gent JI, Lamm AT, Pavelec DM, Maniar JM, Parameswaran P, et al. (2010) Distinct phases of siRNA synthesis in an endogenous RNAi pathway in *C. elegans* soma. *Molecular Cell* 37: 679-689.
41. Conine CC, Batista PJ, Gu W, Claycomb JM, Chaves DA, et al. (2010) Argonautes ALG-3 and ALG-4 are required for spermatogenesis-specific 26G-RNAs and thermotolerant sperm in *Caenorhabditis elegans*. *Proc Natl Acad Sci USA* 107: 3588-3593.
42. Stoeckius M, Maaskola J, Colombo T, Rahn HP, Friedlander MR, et al. (2009) Large-scale sorting of *C. elegans* embryos reveals the dynamics of small RNA expression. *Nature methods* 6: 745-751.
43. Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, et al. (2008) An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322: 1387-1392.
44. Chambeyron S, Popkova A, Payen-Groschene G, Brun C, Laouini D, et al. (2008) piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proceedings of the National Academy of Sciences of the United States of America* 105: 14964-14969.
45. Grentzinger T, Armenise C, Brun C, Mugat B, Serrano V, et al. (2012) piRNA-mediated transgenerational inheritance of an acquired trait. *Genome Research*.
46. Reinke V, Gil IS, Ward S, Kazmer K (2004) Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. *Development* 131: 311-323.

47. Frokjaer-Jensen C, Davis MW, Hopkins CE, Newman BJ, Thummel JM, et al. (2008) Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nature genetics* 40: 1375-1383.
48. Nolan T, Hands RE, Bustin SA (2006) Quantification of mRNA using real-time RT-PCR. *Nature protocols* 1: 1559-1582.
49. Tabara H, Yigit E, Siomi H, Mello CC (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.
50. Sijen T, Steiner FA, Thijssen KL, Plasterk RH (2007) Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science* 315: 244-247.
51. Vasale JJ, Gu W, Thivierge C, Batista PJ, Claycomb JM, et al. (2010) Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. *Proceedings of the National Academy of Sciences of the United States of America* 107: 3582-3587.
52. Bessereau JL (2006) Transposons in *C. elegans*. *WormBook : the online review of C elegans biology*: 1-13.
53. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290: 1151-1155.
54. Gu W, Lee HC, Chaves D, Youngman EM, Pazour GJ, et al. (2012) CapSeq and CIP-TAP Identify Pol II Start Sites and Reveal Capped Small RNAs as *C. elegans* piRNA Precursors. *Cell* 151: 1488-1500.
55. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, et al. (2011) NCBI GEO: archive for functional genomics data sets--10 years on. *Nucleic acids research* 39: D1005-1010.
56. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25.
57. Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Research* 14: 1188-1190.
58. Flicek P, Amode MR, Barrell D, Beal K, Brent S, et al. (2012) Ensembl 2012. *Nucleic acids research* 40: D84-90.
59. Hillier LW, Reinke V, Green P, Hirst M, Marra MA, et al. (2009) Massively parallel sequencing of the polyadenylated transcriptome of *C. elegans*. *Genome Research* 19: 657-666.
60. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25: 1105-1111.
61. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* 28: 511-515.
62. Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, et al. (2008) A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Research* 18: 1051-1063.
63. Celniker SE, Dillon LA, Gerstein MB, Gunsalus KC, Henikoff S, et al. (2009) Unlocking the secrets of the genome. *Nature* 459: 927-930.

CHAPTER FOUR: Future directions

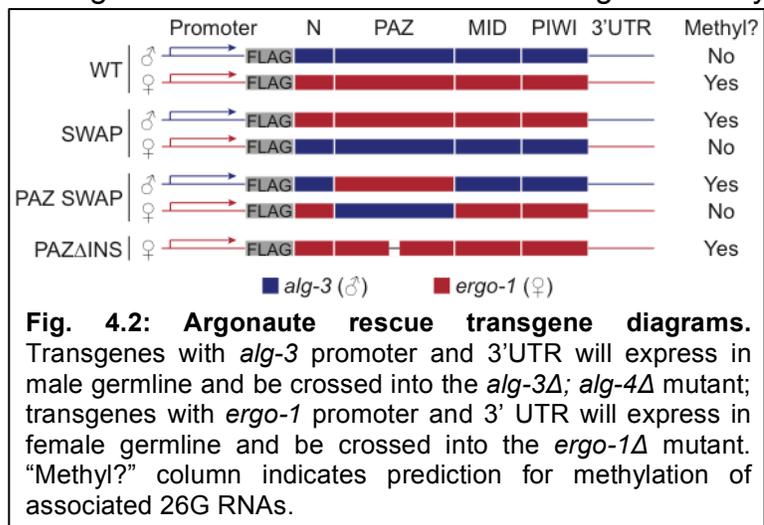
Investigating mechanisms and significance of selective endo-siRNA methylation

Germline small RNAs fulfill a conserved, essential role in promoting proper gamete formation and preserving genome integrity. Methylation by the conserved factor HEN1 represents a critical step in maturation of many germline small RNAs. This modification prevents degradation, promoting inheritance of parental small RNAs. While plant small RNAs are universally methylated, animal small RNA methylation is a regulated process: piRNAs are methylated, microRNAs are not, and methylation of siRNAs varies by class. However, the mechanisms governing selective methylation of animal small RNAs have not yet been definitively established. Recently, I and others determined that the *C. elegans* HEN1 ortholog, HENN-1 (hereafter, HEN1 for simplicity), methylates 26G RNAs generated in female, but not male, germline [1-3]. Intriguingly, 26G RNAs in male and female germlines are loaded onto unique, divergent Argonaute effector proteins whose cladistics reflect the methylation status of the associated small RNAs (Fig. 4.1A). The unmethylated male 26G RNAs are bound by redundant Ago clade Argonautes ALG-3 and ALG-4 in developing sperm, whereas the



methylated female 26G RNAs are bound by the ERGO-1 Argonaute in oocyte and developing embryo and require ERGO-1 for methylation [1]. These findings support a model wherein methylation status of an animal small RNA is dictated by the associated Argonaute. Comparison of the divergent PAZ domains of Piwi and Ago Argonautes reveals that the Piwi PAZ domain shows a small insertion that expands the binding pocket for the small RNA 3' end (**Fig. 4.1B**), where the stabilizing methyl group is added by HEN1 [4]. Using a panel of transgenic *C. elegans* strains that express wild-type or mutant 26G RNA Argonaute proteins (ALG-3 and ERGO-1) in their respective native germlines or ectopically in the non-native germlines, we will test the necessity and sufficiency of the Piwi Argonaute, its PAZ domain, and its PAZ insertion in directing methylation. We will then investigate the effects of aberrant methylation to gain insight into the functional significance of selective methylation and stabilization of germline small RNAs.

To test how Argonaute identity influences methylation of associated 26G RNAs, we are developing the transgenic rescue strains depicted in **Fig. 4.2** using a single-copy insertion method to evade transgene silencing in the germline [5]. These strains express N-terminal FLAG-tagged versions of wild-type or mutant 26G RNA Argonautes in the relevant null backgrounds and can be described in the following groups: 1) The WT strains express tagged wild-type ALG-3 or ERGO-1 under the native promoter in the relevant mutant background (*alg-3Δ*; *alg-4Δ* or *ergo-1Δ*, respectively); 2) The SWAP strains express ERGO-1 in the male germline or ALG-3 in the female germline by exchanging the coding sequences in the two transgenes; 3) The PAZ SWAP strains express mutated forms of ALG-3 and ERGO-1 in which the PAZ domains have been exchanged; 4) The PAZΔINS strain expresses a mutated form of ERGO-1 in which the Piwi



Argonaute-specific PAZ insertion has been deleted in frame. We will isolate single-copy transgene insertions as previously described [5] and confirm FLAG::Argonaute expression in the expected germline using anti-FLAG immunohistochemistry of extruded male and hermaphrodite gonads. After crossing the transgenes into the relevant null backgrounds, we will define the molecular consequences of rescue with

Table 4.1. Assays for 26G RNA biogenesis and function.		
Step tested	Assay	Ref
Methylation	Beta-elimination + northern blot	[6]
Accumulation	Taqman of 26G RNAs	[7]
Amplification	Taqman of dependent 22G RNAs	[8, 9]
Target silencing	RT-qPCR of published targets	[7]

the transgenic Argonaute, examining the effects on 26G RNA biogenesis and function as depicted in **Table 4.1**.

We hypothesize that the ALG-3 PAZ domain prohibits methylation of 26G RNAs and the ERGO-1 PAZ domain permits HEN1-mediated methylation, but only when its PAZ domain insertion is intact (see **Fig. 4.2**, “Methyl?” column). We therefore expect that SWAP and PAZ SWAP strains will show reversed methylation status for the native 26G RNAs by beta-elimination and northern blot. We further predict that the PAZΔINS strain will show loss of methylation for the relevant small RNA population. As loss of methylation decreases small RNA stability, *ergo-1Δ* animals rescued by predicted methylation-prohibitive Argonautes under the *ergo-1* promoter will likely show reduced female 26G RNA accumulation, with a corresponding decrease in dependent 22G RNAs. This will manifest most noticeably as a dramatic decrease in inheritance of female 26G RNAs, as in the *hen1Δ* mutant [1]. Female 26G RNA target levels may be largely unaffected, however, as the *hen1Δ* mutant shows little or no target upregulation [1]. The results in the male germline may reveal more. *alg-3Δ; alg-4Δ* mutants rescued by expression of putative methylation-permissive Argonautes under the *alg-3* promoter are expected to show increased male 26G RNA levels. This will likely correlate with increased production of dependent secondary 22G endo-siRNAs and enhanced silencing of target transcripts. The more abundant 22G RNAs, or possibly the stabilized male 26G RNAs themselves, may even show robust embryonic inheritance, as is observed for methylated female 26G RNAs and piRNAs. The significance of this is expanded upon below.

It is possible that the ERGO-1 PAZ domain may be required for 26G RNA methylation, but insufficient. In this case, our SWAP strains would show reversed methylation status, but no 26G RNAs would be methylated in the PAZ SWAP strains. Should we observe this, we would generate additional ERGO-1/ALG-3 hybrid transgenes to determine the minimal set of ERGO-1 domains required for 26G RNA methylation. Other domains identified as required might be involved in recruiting HEN1 to the Argonaute complex.

To understand the biological relevance of selective endo-siRNA methylation in the germline, we will complement these molecular studies with phenotypic analysis of the transgenic rescue strains, characterizing any defects in fertility and gametogenesis associated with 26G RNA dysregulation. To examine consequences of loss of female 26G RNA methylation, we will test for phenotypes observed in the *hen1Δ* mutant. We will assay for enhanced somatic sensitivity to exogenous dsRNA, which reflects the decreased competition with endo-siRNAs for limiting RNAi factors [7,10-12]. We will also perform a brood size assay, as the *hen1Δ* mutant exhibits a 25% decrease in brood size at elevated temperatures. Because the male 26G RNAs play an essential role in promoting sperm maturation at elevated temperatures [7,9], we will examine a broad range of possible phenotypes to determine the molecular consequences of aberrant male 26G RNA methylation. We will perform DAPI staining of the germline to quantitate sperm and assess their morphology. To test sperm maturation, we will conduct sperm activation assays as previously described [13]. We will assess sperm function by counting unfertilized oocytes and performing brood size assays. Any fertility defects will be further characterized by outcrossing transgenic males or females (*fem-1Δ* mutants that generate no sperm) to wild-type animals to explore the gametic origin of the defect.

We predict that transgenic Argonautes expected to prohibit female 26G RNA methylation will exhibit enhanced RNAi sensitivity, but not recapitulate the temperature-sensitive fertility defect observed in the *hen1Δ* mutant [1], as piRNA methylation will remain intact. It is far more difficult to predict the phenotypes of strains predicted to show aberrant methylation of male 26G RNAs. In these strains, we expect target

silencing to be enhanced and prolonged, which may affect gametogenesis. As male 26G RNAs target many spermatogenesis-enriched coding transcripts [7], enhanced silencing of these targets may disrupt sperm development, presenting phenotypically as decreased sperm number, increased unfertilized oocyte number, and complete rescue with outcrossing to wild-type males. The converse is possible as well: because silencing of male 26G RNA targets is required for normal sperm development, and particularly chromosome segregation [9], additional silencing may further increase production of functional sperm.

Aberrant methylation of male 26G RNAs may also affect zygotic viability or fitness. Zygotes formed from sperm carrying methylated male 26G RNAs or excess dependent secondary siRNAs may suffer erroneous silencing of transcripts common to zygote and sperm. Any inherited male 26G RNAs would also target imperfectly complementary transcripts in trans [3], triggering robust and most probably detrimental silencing by secondary siRNAs. As the *alg-3* promoter primarily drives sperm-specific expression, any somatic phenotypes observed in strains with aberrant male 26G RNA methylation would likely derive from erroneous inheritance of these endo-siRNAs from the paternal germline. This inherited effect could be conclusively demonstrated by examining offspring of heterozygous transgenic males outcrossed to wild-type hermaphrodites. Phenotypes observed in the nontransgenic cross progeny would represent heritable defects attributable to increased stability of male 26G RNAs in the parental germline. Such phenotypes would provide striking evidence in support of our hypothesis that methylation of only female 26G RNAs enforces selective inheritance of a maternal epigenetic program.

Female 26G RNAs are robustly inherited and therefore under selective pressure to avoid targeting essential zygotic transcripts. We expect aberrant methylation and inheritance of male 26G RNAs to prove deleterious to offspring, but only because male 26G RNAs have not evolved under the same constraint. It is intriguing to speculate on the biological advantage of inhibiting paternal inheritance of epigenetic signals. Additional mechanisms appear to exist to reinforce this bias: ALG-3 and PRG-1,

Argonautes that bind primary small RNAs in male germline, are largely excluded from mature sperm [9,14]. Accordingly, female piRNAs are dramatically overrepresented in the zygote [15] even though both male and female piRNAs are methylated [1]. Exclusion of ALG-3 and PRG-1 is not due to incapacity of the sperm to carry a significant Argonaute cargo, as secondary siRNA-binding Argonaute WAGO-1 is effectively packaged into the spermatid [9]. Several RNAi phenomena show interesting parent-of-origin effects that may be related to the maternal inheritance bias. Mutations in some genes required for transposon silencing cause complete, immediate transposon activation when maternally inherited, but only a gradual increase in desilencing over several generations when paternally inherited [16,17]. Transgenes that are heritably silenced at the chromatin level are capable of triggering immediate silencing in trans (cosuppression) of a homologous locus when maternally inherited, but cosuppression may show a generational lag or fail entirely when the silenced locus is paternally inherited [18-20].

Why should such a system of preferential maternal transmission evolve? The application of this seems somewhat unclear in *C. elegans*, as the vast majority of *C. elegans* zygotes arise from identical parental genomes through self-fertilization. However, the androdioecious (hermaphrodite and male) *C. elegans* species likely evolved from a relatively recent gonochoristic (male and female) ancestral state [21]. Perhaps inheritance of discordant epigenetic programs from two different parental genomes is detrimental to offspring, and selective inheritance of maternal small RNAs served until recently to prevent a clash of distinct maternal and paternal epigenetic programs. Other androdioecious and gonochoristic *Caenorhabditis* species also show distinct male and female 26G RNA classes [22]. Adding to the mystery, these species show target overlap between male 26G RNAs and 26G RNAs found in embryo, whereas male and embryo 26G RNA populations show virtually no target overlap in *C. elegans*. Notably, in these other three species, the *ergo-1* gene has undergone multiple duplications [22], and expression of one or more of these paralogs in male germline may confer methylation and thus inheritance of male 26G RNAs. While the proposed investigation of small RNA-mediated methylation in *C. elegans* will enhance our

understanding of structural features within the Argonaute family that dictate animal small RNA methylation, it will only begin to elucidate the biological function of selective germline endo-siRNA methylation in vivo. Examining Argonaute diversity and small RNA methylation in gametes of other *Caenorhabditis* species and additional organisms with diverse reproductive strategies may provide further clarity.

Identifying small RNA pathway cofactors through HEN1 proteomics

HEN1 is the only known factor required for small RNA methylation. In vitro, purified recombinant *C. elegans* HEN1 effectively methylates ssRNA substrates [2], suggesting that HEN1 may not require protein cofactors for methyltransferase activity; however, HEN1-mediated methylation appears to require Argonaute binding in vivo [1]. Proteins that associate with HEN1 may therefore be important for HEN1 recruitment or regulation; alternatively, they may represent Argonaute cofactors that mediate interaction with HEN1. While numerous members of the microRNA effector complex, miRISC, have been identified and characterized, the core Argonautes remain the only known components of the female 26G RNA and piRNA effector complexes. Previous mass spectrometry analysis of immunopurified ERGO-1 and PRG-1 complexes failed to distinguish protein cofactors essential for the female either pathway (data not shown). To identify HEN1 interactors that might contribute to methylation or effector function of these small RNAs, I introduced a rescuing HEN1::GFP transgene into the *hen1* Δ mutant [1] and used an α -GFP antibody to isolate HEN1 complexes from transgenic embryo and adult male extracts. Proteins were identified by mass spectrometry, and factors detected in nontransgenic control samples were excluded from subsequent analyses. I prioritized candidates by peptide coverage and putative function in small RNA pathways as indicated by emergence from a genome-wide transgene silencing RNAi screen performed in a sensitized genetic background (see below, the *eri-1; scm::gfp* screen). Our list of ten high-confidence candidates includes H27M09.1, a paralog of the known miRISC component CGH-1 [23], F43G9.1, an isocitrate dehydrogenase previously implicated in small RNA pathways [24], and TKT-1, the *C. elegans* ortholog of the

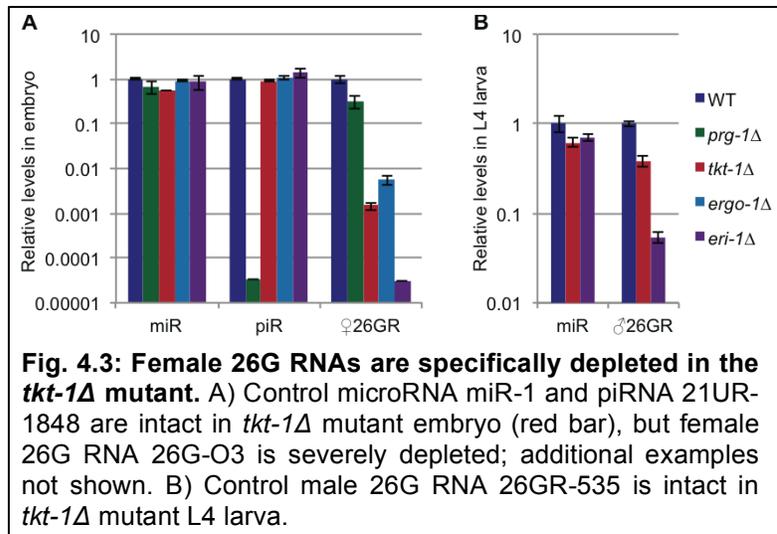
conserved metabolic enzyme transketolase. The last of these is discussed in detail in the next two sections.

To evaluate putative HEN1 interactors for a role in HEN1-mediated methylation, we will inactivate candidate genes by RNAi and assess progeny for loss of 26G RNA and piRNA methylation by beta-elimination and northern blot. For candidate genes with definitive roles in methylation, we will obtain mutants from the Caenorhabditis Genetics Center and subject them to the molecular and phenotypic assays described above to determine the candidate's relevance to germline small RNA dynamics. To validate interaction with HEN1, we will generate transgenic strains expressing FLAG-tagged candidate proteins to enable us to attempt reciprocal co-immunoprecipitations with HEN1. To determine if candidates mediate interaction between HEN1 and Argonautes, we will also attempt to reciprocally co-immunoprecipitate candidates and ERGO-1 and PRG-1 Argonautes using specific polyclonal antibodies we have previously generated. Candidates deemed to be Argonaute complex components will be similarly tested for interaction with Ago clade Argonautes to assess whether they represent factors unique to methylation-permissive silencing complexes. The results of these preliminary studies will guide further investigation of the roles of these candidates in small RNA methylation and/or effector function.

Establishing a role for HEN1 interactor TKT-1 in the female 26G RNA pathway

I have identified the highly conserved metabolic enzyme transketolase, encoded by *tkt-1*, as a HEN1 interactor. Our mass spectrometry analysis of immunopurified HEN1::GFP complexes recovered a total of four unique, non-overlapping TKT-1 peptides covering 11% of the protein. Although a role for TKT-1 in small RNA pathways has never before been demonstrated, TKT-1 also emerged from a proteomic analysis of complexes containing DCR-1 [11], the *C. elegans* ortholog of the conserved small RNA processing factor Dicer. In *C. elegans*, DCR-1 is required for cleavage of 26G RNAs, but not piRNAs [7,14,25]. Analysis of a mutant carrying a ~400 bp deletion in the coding

sequence of *tkt-1* reveals severe depletion (100- to 1000-fold) of female 26G RNAs in mutant embryo, whereas control microRNA and piRNA species show no significant change (**Fig. 4.3A**). Interestingly, a control male 26G RNA is intact in *tkt-1Δ* mutant young adult (**Fig.**



4.3B), indicating a specific role for *tkt-1* in accumulation of female 26G RNAs. These preliminary data have led us to hypothesize that TKT-1 fulfills a multipart role in 26G RNA biogenesis by promoting 26G RNA processing and assembly into a mature ERGO-1 Argonaute complex.

Interaction between TKT-1 and HEN1 suggests that they may collaborate to mediate methylation. To investigate the role of TKT-1 in germline small RNA methylation, we will assess the methylation status of female 26G RNAs and piRNAs in wild-type versus *tkt-1Δ* mutant embryo using the beta elimination assay as above. Based on the interaction of TKT-1 and HEN1 and the specific loss of female 26G RNAs in the *tkt-1Δ* mutant, we expect methylation of female 26G RNAs, but not piRNAs, to be compromised in the *tkt-1Δ* mutant. As methylation occurs following Argonaute loading, it is also possible that TKT-1 is required for loading of female 26G RNAs onto ERGO-1. We will test this by immunopurifying ERGO-1 complexes from wild-type, *eri-1Δ* mutant (which lacks 26G RNAs entirely), and *tkt-1Δ* mutant animals and quantifying associated 26G RNAs, normalizing the proportion of 26G RNAs recovered in the immunoprecipitate to levels in the crude lysate. Although female 26G RNA levels are very low in the *tkt-1Δ* mutant, we have previously demonstrated our ability to assess methylation status of highly depleted species by northern blot analysis of the *ergo-1Δ* mutant [1]. A direct or indirect role for TKT-1 in 26G RNA effector complex assembly will likely be shown by decreased relative detection of female 26G RNAs in ERGO-1 immunoprecipitates isolated from *tkt-1Δ* mutant, as this mutant shows female 26G RNA levels similar to those of the *ergo-1Δ*

mutant. If small RNA loading into the effector complex is inhibited in the *tkt-1Δ* mutant, our previous work suggesting that methylation occurs after Argonaute loading [1] would predict that the residual female 26G RNAs detected in the *tkt-1Δ* mutant will be unmethylated. Alternatively, if the residual female 26G RNAs in the *tkt-1Δ* mutant are still methylated, we would conclude that TKT-1 is dispensable for methylation, but functions in other steps of female 26G RNA biogenesis such as precursor processing, or possibly earlier.

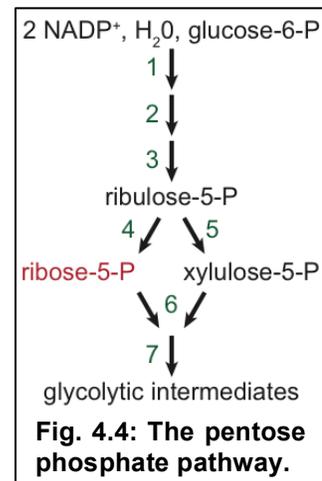
Because interaction was detected between TKT-1 and Dicer [11], we hypothesize that TKT-1 may also act in 26G RNA processing. The dsRNA dicing activity of Dicer complex can be faithfully recapitulated in vitro using embryo extract [26]. This dicing assay has been successfully adapted in our laboratory to investigate requirements for 26G RNA biogenesis in vitro. In extract from wild-type embryos supplemented with ATP and RNase inhibitor, radioactively labeled dsRNA precursor is processed to generate 26G RNAs, which can be visualized by polyacrylamide gel electrophoresis and autoradiography. This 26 nt dicing activity is absent in extract from *dcr-1(mg375)* mutant embryos, which lack Dicer helicase activity [27], as well as embryos lacking factors required for 26G RNA biogenesis (Ting Han, unpublished data). 26G RNA production is intact in the *ergo-1Δ* mutant, demonstrating that this assay reports on 26G RNA processing capacity rather than stabilization. To test if TKT-1 is required for 26G RNA processing, we will perform the dicing assay in embryonic extract from *tkt-1Δ* mutant embryos. We expect to see decreased accumulation of the 26 nt species in *tkt-1Δ* mutant extract relative to wild-type, indicating that the 26 nt dicing activity is compromised in *tkt-1* mutant extract.

Testing TKT-1 involvement at each stage will enable definitive identification of steps in the female 26G RNA pathway that are defective in the *tkt-1Δ* mutant; however, the separation of these steps may be somewhat artificial. Argonaute loading may be coupled to 26G RNA transcription and processing, and our previous results indicate that methylation is dependent upon Argonaute loading [1]. TKT-1 may represent a linchpin, regulating at multiple steps, perhaps acting to coordinate female 26G RNA biogenesis, Argonaute loading, and methylation. A similar role was recently shown to be played by a

human prion protein that binds Dicer and the microRNA Argonaute AGO, facilitating formation of a mature effector complex [28]. One experiment that may help us distinguish whether TKT-1 acts in transcription and processing, loading and methylation, or both would be to assess female 26G RNA levels in a *tkt-1Δ*; *ergo-1Δ* double mutant. If TKT-1 acts only in Argonaute loading and methylation, female 26G RNA levels should not decrease further; if, however, TKT-1 contributes to biogenesis, the defect in female 26G RNA levels should be enhanced relative to the *ergo-1Δ* single mutant. The results obtained from these studies will not only provide clarity as to how TKT-1 promotes accumulation of female germline endo-siRNAs but also guide the following investigation addressing the most important question: Why is the conserved metabolic enzyme transketolase acting in the female 26G RNA pathway? The function of female 26G RNAs is still not yet definitively established, and the involvement of transketolase may provide a critical clue.

Exploring a novel connection between nutrition status and maternal endo-siRNAs

The transketolase enzyme is broadly conserved in animals, plants, and bacteria, but plays no known role in small RNA pathways. Transketolase acts in the pentose phosphate pathway (PPP), a metabolic shunt with two distinct phases (**Fig. 4.4**). In the first phase, irreversible oxidative reactions generate reducing equivalents (NADPH) and a five-carbon sugar that can be isomerized to ribose-5-phosphate, the basic sugar component of the nucleic acid. Transketolase acts in the second phase, during which reversible non-oxidative reactions transform these five-carbon sugars into glycolytic intermediates. When nucleotide biosynthesis is in high demand, as in a highly proliferative cell, ribose-5-phosphate can be generated through both branches of the PPP [29]. Our preliminary data indicate a specific role for the *C. elegans* transketolase ortholog, TKT-1, in the female 26G RNA pathway, raising the possibility that one or both branches of the PPP may have evolved to fulfill a novel



function in germline endo-siRNA dynamics. Indeed our proteomic studies suggest such an undiscovered role: HEN1 and the female 26G RNA Argonaute, ERGO-1, associate specifically with other enzymes in the PPP (**Table 4.2**). These interactions suggest that the PPP may provide an unprecedented link between germline endo-siRNAs and maternal nutritional status through nucleotide metabolism.

#	Enzyme	Ce ortholog	Evidence for small RNA role
1	glucose-6-P dehydrogenase	GSPD-1	
2	6-Pgluconolactonase	Y57G11C.3	3 peptides, ERGO-1 mass spec (17% coverage)
3	6-Pgluconate dehydrogenase	T25B9.9	
4	ribulose-5-P isomerase	RPIA-1	1 peptide, HEN1 mass spec (8% coverage)
5	ribulose-5-P epimerase	F08F8.7	
6	transketolase	TKT-1	3 peptides, HEN1 mass spec (11% coverage) Dicer mass spec (13% coverage) [11]
7	transaldolase	Y24D9A.8	

Guided by the results of the TKT-1 studies described above, we propose to test other members of the PPP for roles in female 26G RNA biogenesis or stability to demonstrate a role for this metabolic pathway in regulating maternal endo-siRNA accumulation and inheritance. We will deplete the expression of each gene of the PPP by RNAi using clones from a *C. elegans* genome RNAi library [30] and assay for phenocopy of the selective female 26G RNA depletion observed in the *tkt-1* Δ mutant. Based on the HEN1 protein interaction evidence, we expect phenocopy of *tkt-1* Δ mutant defects upon RNAi-mediated knockdown of members of the non-oxidative branch of the pathway; however, if the primary activity of this pathway is routing energy from glycolytic intermediates into nucleotides, the oxidative branch may be dispensable to the female 26G RNA pathway, failing to produce phenotypes upon its loss. If indeed the PPP promotes female 26G RNA accumulation through provision of ribose-5-phosphate, supplementing the *E. coli* food source with excess ribose-5-phosphate may rescue female 26G RNA levels in animals lacking TKT-1 or other PPP enzymes. Supplementation of other intermediates in the PPP may further help to define how molecular energy flows through the reversible or non-reversible branches of this pathway into female 26G RNA synthesis.

The most intriguing facet of this investigation is examining the hypothesized connection between nutrient availability and female germline endo-siRNA accumulation. To explore this, we will compare female 26G RNA levels in late larval, young adult, and early gravid wild-type animals that have been collected from nutrient-rich environments or transferred to starvation conditions for four hours prior to collection. As a control, we will also assay male 26G RNA levels in the late larval starved and fed animals. To normalize for potential differences in germline development or proliferation due to starvation, we will assess levels of germline-enriched and oocyte-specific transcripts between starved and fed animals for normalization of germline small RNA levels. If we find that female but not male 26G RNA levels are sensitive to nutrient availability, the results would provide evidence for a novel connection between maternal nutritional status and germline endo-siRNA production. In the germline of a normal adult hermaphrodite, the PPP could route energy stores toward nucleotide synthesis and storage in the form of heritable endo-siRNAs. Under starvation conditions, this pathway could resume conversion of ribose-5-phosphate to glycolytic intermediates, depleting nucleotide stores and decreasing female 26G RNA accumulation.

The hypothesis we posit here would define an unprecedented function of the transketolase in funneling excess nucleotide components into endo-siRNA production in the female germline during periods of high nutrient availability. However, a connection between energy status and endo-siRNAs does not require that the PPP as a whole be involved. If we find that TKT-1 acts independently of the PPP to promote female 26G RNA accumulation, TKT-1 may still be sensitive to the broader nutritional status of the animal and be sequestered by the PPP when that shunt is most active, such as during oxidative stress when cells require additional NADPH. Thus, oxidative stress could reduce female 26G RNA levels. Developing *C. elegans* subjected to crowding or other unfavorable environmental conditions enter an alternative larval stage called dauer. This fate decision to become the hardier, longer-lived dauer form occurs early in larval development prior to germline expansion. Once in a favorable environment, the dauer larva will resume development into adulthood, with full germline expansion. Animals that have recovered from a dauer stage have been found to show changes in gene

expression and produce larger brood sizes [31]. A recent study has shown that recovered dauer adults also show altered small RNA expression levels, and small RNA pathway factors are required for the brood size increase [32]. This indicates a connection between fertility and a history of stress prior to germline expansion. Once the germline has expanded, a flow of nutrients through transketolase and the PPP may regulate maternal endo-siRNA accumulation to influence inheritance.

The completion of this investigation will either reveal a non-canonical function for the conserved enzyme transketolase in female 26G RNA biogenesis that is independent of its role in the PPP or establish an unprecedented link between these two pathways. The proposed studies may also provide evidence for a novel connection between metabolism and maternal germline endo-siRNAs. This discovery would lead to an exciting new line of inquiry: Can oocyte endo-siRNAs serve as a transgenerational energy source? Beyond their critical role in transmission of genetic information, nucleotides represent a rich source of cellular energy. In *C. elegans*, proteins and lipids derived from the yolk are mobilized to support rapid embryonic development. Polymerized nucleotides in the form of oocyte endo-siRNAs may represent an analogous reservoir of nucleotides for fueling the maternal to zygotic transmission.

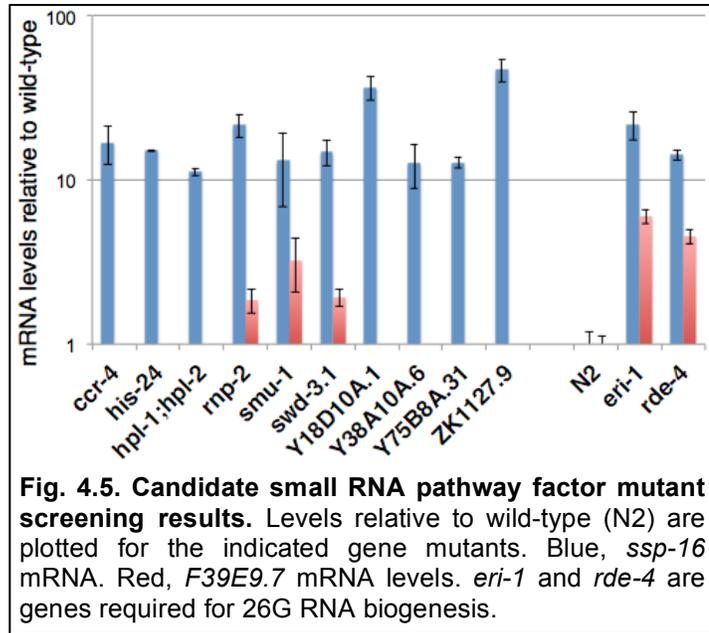
Mining genome-wide RNAi screen data for novel endo-siRNA pathway factors

Across metazoans, germline small noncoding RNAs, including piRNAs and endogenous siRNAs (endo-siRNAs), fulfill a conserved function in promoting germline immortality. These small RNAs act as an immune system for the genome, passing from parent to offspring and silencing genetic elements that pose a threat to genome integrity such as transposons. In higher animals, compromise of the piRNA or endo-siRNA pathway results in sterility [33-37], revealing their essential roles in reproduction. Similarly, compromise of 26G RNAs and piRNAs in *C. elegans* causes sterility at elevated temperatures [7,14,25,38]. Upon target recognition, these primary small RNAs trigger the production of 22G secondary siRNAs that amplify their repressive signal and effect silencing [8,9,25,39-41]. The secondary siRNAs associate with the expansive, worm-

specific clade of WAGO Argonautes and are accordingly called WAGO 22G RNAs. It was recently determined that WAGO 22G RNAs triggered by exo-siRNAs, endo-siRNAs, and piRNAs are capable of mediating not only post-transcriptional silencing but also heritable co-transcriptional silencing in the nuclei of germ cells; this germline nuclear RNAi pathway promotes ongoing expression of secondary siRNAs across generations to mediate H3K9 methylation and transgenerational silencing [18-20,41-43]. NRDE-2, NRDE-1, NRDE-4, and nuclear WAGO Argonaute HRDE-1 represent the core germline nuclear RNAi machinery, but other chromatin-associated factors required for transgenerational silencing have been identified [18-20,42,44-46]. At elevated and, in some cases, even permissive temperatures, loss of the core germline nuclear RNAi factors results in progressively worsening defects in gametogenesis that lead to sterility [42]. Thus, *C. elegans* endo-siRNA and piRNA pathways converge on an endo-siRNA amplification system with both cytoplasmic and nuclear branches for target silencing, the latter of which is critical to germline immortality.

Systematic RNAi screening of the *C. elegans* genome has proven extremely valuable for identifying genes involved in a common pathway. Somewhat surprisingly, the utility of this approach extends even to discovery of genes involved in the RNAi pathway itself [24]. Although the core factors of many small RNA pathways have been characterized, the frequent identification of novel RNAi factors suggests we are far from the point of saturation for discovery of such genes. An additional genome-wide RNAi screen with enhanced sensitivity to detect putative small RNA pathway factors was therefore devised and executed [47]. Briefly, this screen tests for desilencing of a high-copy array expressing GFP in a population of somatic cells (*scm::gfp*) in the enhanced RNAi background of the *eri-1Δ* mutant; this sensitizing background lacks 26G RNAs, thereby liberating downstream WAGO 22G RNA machinery to increase the potency of exo-RNAi. The *scm::gfp* transgene is fully silenced in the *eri-1Δ* mutant background. The genome was screened as previously described [24], and all genes whose depletion by RNAi resulted in any transgene desilencing were retested in triplicate. This approach distinguished ~900 candidate small RNA pathway factors, including many that had been previously discovered and validated.

To identify optimal candidates for further screening, I selected those candidates receiving a desilencing score ≥ 1.5 out of 4 for which viable, probable loss-of-function mutants are available from the *Caenorhabditis Genetics Center* or through our collaborator at the *National Bioresource Project for the Experimental Animal "Nematode C. elegans"* at Tokyo Women's Medical University School of



Medicine. I filtered this list further by identifying candidates whose protein structure or loss-of-function phenotypes suggested possible involvement in small RNA pathways. My final refined list contained over 60 candidate mutant strains. These were subjected to RT-qPCR analysis of levels of male 26G RNA target mRNA *ssp-16*, female 26G RNA target mRNA *F39E9.7*, piRNA target *Tc3* transposase mRNA, piRNA non-target *Tc1* transposase mRNA, and control mRNA *act-1* in early gravid adult biological duplicate RNA samples. No candidate mutant showed selective upregulation of *Tc3* but not *Tc1* transposase mRNA, which would suggest specific compromise of the piRNA pathway [14,25]. Candidates that showed possible 26G RNA target mRNA upregulation were confirmed by RT-qPCR analysis of a new set of RNA samples, defining 10 mutants representing 11 genes as positive in the RT-qPCR screening (**Fig. 4.5**). Three of our positive genes, *rmp-2*, *Y38A10A.6*, and *ZK1127.9*, were previously identified in the first genome-wide RNAi screen for factors involved in RNAi [24], and *swd-3.1* was previously found to be required for full sensitivity to exogenous RNAi [48]. Subsequent to this RT-qPCR screening, *hpl-1/2* were found to be required for transgenerational silencing [18,20], and *ZK1127.9* was found to be required for silencing of a transgene by a WAGO 22G RNA that engages the somatic nuclear RNAi pathway [3].

The next step in pursuing these candidates will be to determine at which stage in small RNA pathways their respective proteins function. As the *eri-1Δ; scm::gfp* screen was conducted in a mutant background lacking 26G RNAs, genes that function only in 26G RNA biogenesis or effector function would be unlikely to emerge from the screen. These mutants are thus not predicted to show decreased levels of 26G RNAs. As none of these mutants shows a specific defect in silencing of *Tc3* transposase mRNA (data not shown), levels of piRNAs are also not expected to be affected. Rather, silencing of the somatic high-copy *scm::gfp* array is likely mediated by a combination of 21U RNAs recognizing non-self elements of the array and exo-siRNA-like siRNAs generated by incidental bidirectional transcription within the array. The latter small RNAs may direct array transcripts for deadenylation and degradation via the recently identified RDE-10/RDE-11 complex [49,50]; however, neither *rde-10* nor *rde-11* was identified in the *eri-1Δ; scm::gfp* screen [47], suggesting that silencing of the sensor transgene may be independent of RDE-10/RDE-11. The transgenic system used for this screen is therefore likely to report on the function of WAGO 22G RNA-mediated cytoplasmic, post-transcriptional silencing of array transcripts and/or transcriptional silencing through engagement of the somatic nuclear RNAi pathway [44-46,49-51]. Consistent with this interpretation, many factors involved in these mechanisms emerge from the *eri-1Δ; scm::gfp* screen.

For all candidate factors, we will apply the same approach. Accumulation of WAGO 22G RNA species will be quantified by Taqman RT-qPCR to detect a defect in WAGO 22G RNA biogenesis or Argonaute loading and stabilization. Each gene will also be inactivated by RNAi in a GFP::NRDE-3 transgenic background; failure of the somatic nuclear RNAi Argonaute, NRDE-3, to localize to the nucleus will indicate a defect in accumulation or Argonaute loading of WAGO 22G RNAs that engage somatic nuclear RNAi pathway [44]. We will then test the integrity of the exo-RNAi pathway in germline and soma by subjecting the mutants to RNAi-mediated inactivation of *pos-1* and *dpy-13*, respectively. The integrity of the nuclear RNAi pathway will be specifically assessed by assaying mutant sensitivity to RNAi targeting the *lir-1* polycistronic pre-mRNA in the nucleus. Defects observed in the nuclear RNAi pathway will be further pursued using

assays for transcriptional gene silencing, target H3K9me3 levels, single-generation and transgenerational RNAi inheritance, and mortal germline phenotyping as previously described [42,45,46,51].

CCR-4, RNP-2, and SMU-1 have established roles in RNA metabolism or regulation, and Y38A10A.6 encodes a probable ATP-dependent RNA helicase. It is therefore possible that these factors will contribute to cytoplasmic small RNA pathway processes such as WAGO 22G RNA biogenesis or post-transcriptional target silencing, although they may also contribute to nuclear RNAi by interacting with nascent transcripts. In contrast, HIS-24, HPL-1/2, SWD-3.1, Y75B8A.31, and ZK1127.9 putatively act at the level of chromatin: HIS-24 is the only of eight *C. elegans* H1 linker histone variants that promotes germline development [52]. HIS-24 is required for germline silencing of extrachromosomal arrays and functions synergistically with the SIR-2.1 deacetylase to maintain germline heterochromatin; simultaneous loss of these two proteins results in a mortal germline phenotype [52,53]. HPL-1 and HPL-2 represent the two *C. elegans* heterochromatin protein 1 homologs; HPL-2 interacts with H3K9me2/3, and its loss results in desilencing of germline extrachromosomal arrays and sterility at elevated temperatures [53,54]. SWD-3.1 is a member of the conserved H3K4 trimethylation complex, which regulates lifespan in a germline-dependent manner and enables inheritance of longevity [55,56]. Y75B8A.31 contains an Arb2 domain, found in Hda1, a histone deacetylase factor, and Arb2; in fission yeast, Arb1 and Arb2 constitute the Argonaute siRNA chaperone complex and are required for siRNA biogenesis, H3K9 methylation, and heterochromatin assembly [57]. ZK1127.9 is a transcription elongation factor that also regulates lifespan in a germline-dependent manner [58]. Thus, many of our candidates may contribute to the chromatin-level effector function of WAGO 22G RNAs. At present, very little is known about the mechanisms required for endo-siRNA-mediated chromatin modification in *C. elegans*. Definitively linking the nuclear RNAi pathway to candidate RNAi factors with such diverse protein domains will help guide future investigation of these emerging mechanisms.

REFERENCES

1. Billi AC, Alessi AF, Khivansara V, Han T, Freeberg M, et al. (2012) The *Caenorhabditis elegans* HEN1 ortholog, HENN-1, methylates and stabilizes select subclasses of germline small RNAs. *PLoS genetics* 8: e1002617.
2. Kamminga LM, van Wolfswinkel JC, Luteijn MJ, Kaaij LJ, Bagijn MP, et al. (2012) Differential Impact of the HEN1 Homolog HENN-1 on 21U and 26G RNAs in the Germline of *Caenorhabditis elegans*. *PLoS genetics* 8: e1002702.
3. Montgomery TA, Rim YS, Zhang C, Downen RH, Phillips CM, et al. (2012) PIWI associated siRNAs and piRNAs specifically require the *Caenorhabditis elegans* HEN1 ortholog henn-1. *PLoS genetics* 8: e1002616.
4. Tian Y, Simanshu DK, Ma JB, Patel DJ (2011) Structural basis for piRNA 2'-O-methylated 3'-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. *Proceedings of the National Academy of Sciences* 108: 903.
5. Frokjaer-Jensen C, Davis MW, Hopkins CE, Newman BJ, Thummel JM, et al. (2008) Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nat Genet* 40: 1375-1383.
6. Yang Z, Vilkaitis G, Yu B, Klimasauskas S, Chen X (2007) Approaches for studying microRNA and small interfering RNA methylation in vitro and in vivo. *Meth Enzymol* 427: 139-154.
7. Han T, Manoharan AP, Harkins TT, Bouffard P, Fitzpatrick C, et al. (2009) 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* 106: 18674-18679.
8. Vasale JJ, Gu W, Thivierge C, Batista PJ, Claycomb JM, et al. (2010) Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. *Proc Natl Acad Sci U S A* 107: 3582-3587.
9. Conine CC, Batista PJ, Gu W, Claycomb JM, Chaves DA, et al. (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* 107: 3588-3593.
10. Yigit E, Batista PJ, Bei Y, Pang KM, Chen CC, et al. (2006) Analysis of the *C. elegans* Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* 127: 747-757.
11. Duchaine TF, Wohlschlegel JA, Kennedy S, Bei Y, Conte D, Jr., et al. (2006) Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. *Cell* 124: 343-354.
12. Lee RC, Hammell CM, Ambros V (2006) Interacting endogenous and exogenous RNAi pathways in *Caenorhabditis elegans*. *RNA* 12: 589-597.
13. Shakes DC, Ward S (1989) Initiation of spermiogenesis in *C. elegans*: a pharmacological and genetic analysis. *Developmental biology* 134: 189-200.
14. Batista PJ, Ruby JG, Claycomb JM, Chiang R, Fahlgren N, 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.

15. Billi AC, Freeberg MA, Day AM, Chun SY, Khivansara V, et al. (2013) A Conserved Upstream Motif Orchestrates Autonomous, Germline-Enriched Expression of *Caenorhabditis elegans* piRNAs. *PLoS Genet* 9: e1003392.
16. Ketting RF, Haverkamp TH, van Luenen HG, Plasterk RH (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.
17. Mori I, Moerman DG, Waterston RH (1990) Interstrain crosses enhance excision of Tc1 transposable elements in *Caenorhabditis elegans*. *Molecular & general genetics* : MGG 220: 251-255.
18. Ashe A, Sapetschnig A, Weick EM, Mitchell J, Bagijn MP, et al. (2012) piRNAs Can Trigger a Multigenerational Epigenetic Memory in the Germline of *C. elegans*. *Cell* 150: 88-99.
19. Luteijn MJ, van Bergeijk P, Kaaij LJ, Almeida MV, Roovers EF, et al. (2012) Extremely stable Piwi-induced gene silencing in *Caenorhabditis elegans*. *The EMBO journal* 31: 3422-3430.
20. Shirayama M, Seth M, Lee HC, Gu W, Ishidate T, et al. (2012) piRNAs Initiate an Epigenetic Memory of Nonself RNA in the *C. elegans* Germline. *Cell* 150: 65-77.
21. Cutter AD, Wasmuth JD, Washington NL (2008) Patterns of molecular evolution in *Caenorhabditis* preclude ancient origins of selfing. *Genetics* 178: 2093-2104.
22. Shi Z, Montgomery TA, Qi Y, Ruvkun G (2013) High-throughput sequencing reveals extraordinary fluidity of miRNA, piRNA, and siRNA pathways in nematodes. *Genome research*.
23. Hammell CM, Lubin I, Boag PR, Blackwell TK, Ambros V (2009) *nhl-2* Modulates microRNA activity in *Caenorhabditis elegans*. *Cell* 136: 926-938.
24. Kim JK, Gabel HW, Kamath RS, Tewari M, Pasquinelli A, et al. (2005) Functional genomic analysis of RNA interference in *C. elegans*. *Science* 308: 1164-1167.
25. Das PP, Bagijn MP, Goldstein LD, Woolford JR, Lehrbach NJ, 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.
26. Ketting RF, Fischer SE, Bernstein E, Sijen T, Hannon GJ, et al. (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes & development* 15: 2654-2659.
27. Pavelec DM, Lachowiec J, Duchaine TF, Smith HE, Kennedy S (2009) Requirement for the ERI/DICER complex in endogenous RNA interference and sperm development in *Caenorhabditis elegans*. *Genetics* 183: 1283-1295.
28. Gibbings D, Leblanc P, Jay F, Pontier D, Michel F, et al. (2012) Human prion protein binds Argonaute and promotes accumulation of microRNA effector complexes. *Nature structural & molecular biology* 19: 517-524, S511.
29. Gottlieb E (2011) p53 guards the metabolic pathway less travelled. *Nature cell biology* 13: 195-197.
30. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, et al. (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421: 231-237.
31. Hall SE, Beverly M, Russ C, Nusbaum C, Sengupta P (2010) A cellular memory of developmental history generates phenotypic diversity in *C. elegans*. *Current biology* : CB 20: 149-155.

32. Hall SE, Chirn GW, Lau NC, Sengupta P (2013) RNAi pathways contribute to developmental history-dependent phenotypic plasticity in *C. elegans*. *RNA* 19: 306-319.
33. Reuter M, Berninger P, Chuma S, Shah H, Hosokawa M, et al. (2011) Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature* 480: 264-267.
34. Carmell MA, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, et al. (2007) MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Developmental cell* 12: 503-514.
35. Kuramochi-Miyagawa S, Kimura T, Ijiri TW, Isobe T, Asada N, et al. (2004) Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* 131: 839-849.
36. Murchison EP, Stein P, Xuan Z, Pan H, Zhang MQ, et al. (2007) Critical roles for Dicer in the female germline. *Genes & development* 21: 682-693.
37. Suh N, Baehner L, Moltzahn F, Melton C, Shenoy A, et al. (2010) MicroRNA function is globally suppressed in mouse oocytes and early embryos. *Current biology* : CB 20: 271-277.
38. Wang G, Reinke V (2008) A *C. elegans* Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. *Current biology* : CB 18: 861-867.
39. Gu W, Shirayama M, Conte D, Jr., Vasale J, Batista PJ, et al. (2009) Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. *Mol Cell* 36: 231-244.
40. Bagijn MP, Goldstein LD, Sapetschnig A, Weick EM, Bouasker S, et al. (2012) Function, targets, and evolution of *Caenorhabditis elegans* piRNAs. *Science* 337: 574-578.
41. Lee HC, Gu W, Shirayama M, Youngman E, Conte D, Jr., et al. (2012) *C. elegans* piRNAs Mediate the Genome-wide Surveillance of Germline Transcripts. *Cell* 150: 78-87.
42. Buckley BA, Burkhart KB, Gu SG, Spracklin G, Kershner A, et al. (2012) A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. *Nature*.
43. Gu SG, Pak J, Guang S, Maniar JM, Kennedy S, et al. (2012) Amplification of siRNA in *Caenorhabditis elegans* generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. *Nature genetics* 44: 157-164.
44. Guang S, Bochner AF, Pavelec DM, Burkhart KB, Harding S, et al. (2008) An Argonaute transports siRNAs from the cytoplasm to the nucleus. *Science* 321: 537-541.
45. Guang S, Bochner AF, Burkhart KB, Burton N, Pavelec DM, et al. (2010) Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. *Nature* 465: 1097-1101.
46. Burkhart KB, Guang S, Buckley BA, Wong L, Bochner AF, et al. (2011) A pre-mRNA-associating factor links endogenous siRNAs to chromatin regulation. *PLoS genetics* 7: e1002249.
47. Tabach Y, Billi AC, Hayes GD, Newman MA, Zuk O, et al. (2013) Identification of small RNA pathway genes using patterns of phylogenetic conservation and divergence. *Nature* 493: 694-698.

48. Simonet T, Dulermo R, Schott S, Palladino F (2007) Antagonistic functions of SET-2/SET1 and HPL/HP1 proteins in *C. elegans* development. *Developmental biology* 312: 367-383.
49. Zhang C, Montgomery TA, Fischer SE, Garcia SM, Riedel CG, et al. (2012) The *Caenorhabditis elegans* RDE-10/RDE-11 complex regulates RNAi by promoting secondary siRNA amplification. *Current biology* : CB 22: 881-890.
50. Yang H, Zhang Y, Vallandingham J, Li H, Florens L, et al. (2012) The RDE-10/RDE-11 complex triggers RNAi-induced mRNA degradation by association with target mRNA in *C. elegans*. *Genes & development* 26: 846-856.
51. Burton NO, Burkhart KB, Kennedy S (2011) Nuclear RNAi maintains heritable gene silencing in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America* 108: 19683-19688.
52. Jedrusik MA, Schulze E (2001) A single histone H1 isoform (H1.1) is essential for chromatin silencing and germline development in *Caenorhabditis elegans*. *Development* 128: 1069-1080.
53. Wirth M, Paap F, Fischle W, Wenzel D, Agafonov DE, et al. (2009) HIS-24 linker histone and SIR-2.1 deacetylase induce H3K27me3 in the *Caenorhabditis elegans* germ line. *Molecular and cellular biology* 29: 3700-3709.
54. Couteau F, Guerry F, Muller F, Palladino F (2002) A heterochromatin protein 1 homologue in *Caenorhabditis elegans* acts in germline and vulval development. *EMBO reports* 3: 235-241.
55. Greer EL, Maures TJ, Hauswirth AG, Green EM, Leeman DS, et al. (2010) Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in *C. elegans*. *Nature* 466: 383-387.
56. Greer EL, Maures TJ, Ucar D, Hauswirth AG, Mancini E, et al. (2011) Transgenerational epigenetic inheritance of longevity in *Caenorhabditis elegans*. *Nature* 479: 365-371.
57. Buker SM, Iida T, Buhler M, Villen J, Gygi SP, et al. (2007) Two different Argonaute complexes are required for siRNA generation and heterochromatin assembly in fission yeast. *Nature structural & molecular biology* 14: 200-207.
58. Ghazi A, Henis-Korenblit S, Kenyon C (2009) A transcription elongation factor that links signals from the reproductive system to lifespan extension in *Caenorhabditis elegans*. *PLoS genetics* 5: e1000639.

APPENDIX A: piRNAs and siRNAs collaborate in *Caenorhabditis elegans* genome defense

AUTHORS: Billi AC, Freeberg MA, Kim JK

AUTHOR CONTRIBUTIONS: ACB wrote the paper. ACB, MAF, and JKK edited the paper.

CITATION: Billi AC, Freeberg MA, Kim JK. piRNAs and siRNAs collaborate in *C. elegans* genome defense. *Genome Biology* 13: 164 (2012).

ABSTRACT: *Caenorhabditis elegans* piRNAs promote genome surveillance by triggering siRNA-mediated silencing of nonself DNA in competition with licensing programs that support endogenous gene expression.

MAIN TEXT

Piwi-interacting (piRNAs) are a conserved class of small RNAs that defend against selfish genetic elements in the animal germline. The Piwi Argonautes associate with piRNAs to recognize and silence complementary transcripts. piRNA sequence diversity is immense, enabling targeting of various transposons and repetitive sequences, but mechanisms of target selection are incompletely resolved. Here, we discuss four recent publications [1-4] that provide important insights into piRNA-mediated genome surveillance mechanisms in *C. elegans*.

The canonical piRNA pathway promotes genome integrity

In the germline, piRNAs act to silence mobile elements that can be deleterious to the genome. In flies and mammals, these elements can become “trapped” after integration into genomic piRNA clusters; this chance event induces biogenesis of piRNAs from that element, enabling silencing in *trans* of copies located elsewhere in the genome. Additionally, through an amplification loop termed the ping-pong cycle, the piRNA genome surveillance system can be tuned to recognize and selectively repress actively transcribed mobile elements (Reviewed in [5]). The cycle is primed by primary piRNAs generated through largely unknown mechanisms from sense or antisense transcripts corresponding to target elements. Loaded into a Piwi Argonaute, primary piRNAs direct the cleavage of complementary transcripts of the opposite sense; these cleavage products are in turn incorporated into distinct Piwi Argonautes as secondary piRNAs to direct the generation of still more piRNAs. Thus, active mobile elements provide substrate transcripts for this amplification loop. In fly, the ping-pong cycle not only amplifies piRNA silencing, but also provides a mechanism for epigenetic transmission of silencing to progeny. Maternally inherited Piwi-piRNA complexes are required for continued genome surveillance in developing progeny and may indeed serve as primary piRNAs to trigger silencing in the filial germline.

C. *elegans* piRNAs act through secondary siRNAs

The *C. elegans* genome encodes two highly homologous Piwi Argonautes, PRG-1 and PRG-2; the latter is dispensable for the piRNA pathway and may represent a pseudogene [6, 7]. PRG-1 binds and is required for the production of 21U RNAs, a population of 21 nucleotide small RNAs with a 5' uridine that exhibit the high sequence diversity, genomic clustering, germline enrichment, and terminal methylation characteristic of piRNAs [6, 7]. Unlike canonical piRNAs, however, the mechanism of action of 21U RNAs is poorly understood: their targets and functions are largely unknown, and they exhibit no evidence of a ping-pong amplification cycle. Rather, they

were previously shown to act upstream of an endogenous siRNA pathway [7], but the specifics of the targeting mechanism and the nature of the secondary siRNAs were not reported.

Recent studies by Bagijn *et al.* [1] and Lee *et al.* [2] provide new clarity to these pathways through deep sequencing of small RNAs in *prg-1* mutant strains and transgenic sensor strains engineered to express complementary 21U RNA target sites. Independently, these two groups show that PRG-1 and the 21U RNA pathway trigger the biogenesis of secondary 22G RNAs of the WAGO pathway in order to effect target silencing. The WAGO 22G RNAs, which associate with the worm-specific WAGO clade of Argonautes, represent a point of convergence for multiple *C. elegans* small RNA pathways, including both the primary endogenous siRNA (26G RNA) pathway and exogenous RNA interference (RNAi).

Through activity that is independent of PRG-1 “slicer” endonuclease function, 21U RNAs guide PRG-1 to target transcripts with up to three or four mismatches, promoting the association of factors involved in WAGO 22G RNA biogenesis to mount a localized silencing response [1, 2]. Whereas earlier reports identified only a single transposon silenced by the *C. elegans* piRNA pathway [6, 7], Bagijn *et al.* and Lee *et al.* identify numerous additional *C. elegans* piRNA pathway targets that include multiple transposable elements and pseudogenes [1, 2], strengthening the previously tenuous connection between 21U RNAs and transposon defense. They further show that many factors required for WAGO 22G RNA biogenesis are also necessary for silencing of 21U RNA genomic targets, indicating that 22G RNAs mediate the silencing effects of piRNAs. Such laxity in piRNA targeting requirements raises the question of how selectivity is achieved for 21U RNA-triggered repression. The answer may lie with the CSR-1 22G RNAs, another class of endogenous siRNAs required for chromosome segregation. The CSR-1 22G RNAs are primarily antisense to germline-expressed genes and may recruit the Argonaute CSR-1 to protein-coding genomic loci to promote proper chromatin organization through embryonic mitotic divisions [8]. Intriguingly, Bagijn *et al.* observe that transcripts silenced by the 21U RNA pathway are significantly depleted of protein-

coding genes [1]. Similarly, Lee *et al.* show that 21U RNAs that map to CSR-1 22G RNA targets trigger a less robust secondary siRNA response compared to those mapping to WAGO 22G RNA targets [2]. Consistent with this finding, a *gfp::histone* fusion transgene shows inconsistent targeting by 21U RNA-dependent secondary siRNAs: the exogenous sequence encoding the GFP moiety is robustly targeted by 22G RNAs, whereas the endogenous sequence encoding the histone moiety largely evades 21U RNA targeting [2]. This leads the authors to suggest that CSR-1-dependent “licensing” may protect endogenous protein-coding sequences (self) from piRNA-mediated silencing. Unlicensed, nonself genes, in contrast, are silenced by PRG-1-dependent WAGO 22G RNAs. Once such a piRNA immune response is mounted against nonself genes, this silencing is heritable and no longer requires 21U RNAs [2-4].

siRNAs enforce an epigenetic identity of self versus nonself

The mechanism of 21U RNA-dependent, epigenetic memory is further explored by Ashe *et al.* [3] and Shirayama *et al.* [4]. In related studies, these two groups describe a phenomenon of heritable transcriptional and post-transcriptional silencing initiated by piRNAs. This pathway is also triggered by exogenous RNAi, which likewise engages the WAGO 22G RNA pathway [3]. Trans-generational silencing is observed under diverse circumstances. Shirayama *et al.* show that single-copy transgenes that include lengthy foreign sequences show permanent, PRG-1-dependent silencing that they call RNA-induced epigenetic silencing (RNAe) [4]. Ashe *et al.* demonstrate heritable epigenetic silencing of a single-copy 21U RNA target transgene as well as a transgene targeted by canonical, exogenous RNAi [3]. Trans-generational silencing requires an intact WAGO 22G RNA response, including genes involved in nuclear RNAi, and correlates with *de novo* production of filial 22G RNAs. Chromatin factors are also necessary for trans-generational silencing, solidifying a connection between chromatin modification and epigenetic inheritance of silencing [3, 4]. Heritably silenced single-copy alleles show enrichment for H3K9me3, a histone mark associated with silenced chromatin [4]. This finding correlates with recent work showing that exogenous RNAi of diverse

endogenous loci results in trans-generational silencing that requires the WAGO 22G RNA pathway and results in H3K9me3 accumulation [9, 10].

Study of transgenic *C. elegans* is greatly confounded by germline silencing. While the advent of single-copy transgene insertion has revolutionized *C. elegans* transgenesis, even these non-repetitive transgenes can be subject to piRNA-dependent silencing [4]. How therefore does silencing machinery recognize self versus nonself? The studies discussed here propose a memory of self genomically encoded in mismatch-tolerant piRNAs. Presumably, piRNA sequences, constrained only by selection against sequences that silence mRNAs, evolve rapidly, enabling targeting of diverse foreign genetic material [1]. An unpaired DNA silencing response may also aid in recognition of foreign sequences, as stable silencing of a piRNA sensor transgene was achieved when present in a heterozygous state over multiple generations [3].

There is, however, competition between the silencing and licensing signals. Endogenous genes targeted by exogenous RNAi generally recover expression after several generations, and even low-copy transgenes containing foreign DNA may become resistant to permanent silencing after propagation for years [4]. Continued propagation of foreign DNA seems to confer self-identity and thus unchecked expression; this process is likely enhanced by experimental selection by investigators for transgenic animals that maintain robust transgene expression. Intriguingly, silenced single-copy *gfp* transgenes can be activated in *trans* by the presence of ancient, licensed transgenes carrying *gfp* [4]. The agents responsible for this antagonism may well be CSR-1 22G RNAs. Thus, *C. elegans* may distinguish self from nonself through piRNA-mediated surveillance and the activities of competing, complementary siRNA pathways.

Questions remain regarding the function of *C. elegans* piRNAs. What are the consequences of loss of 21U RNAs over many generations? How is genome integrity affected by desilencing of the suite of elements regulated by PRG-1, only a minority of which are transposons? What other factors influence the outcome when silencing and

licensing programs clash? Importantly, are analogous epigenetic programs somehow enacted in higher organisms, which lack the arsenal of Argonaute proteins and RNA-dependent RNA polymerases that generate *C. elegans* siRNAs? And finally, whole areas of piRNA biology remain largely uncharted – most notably, primary piRNA biogenesis – and await further study in *C. elegans* and higher organisms.

ACKNOWLEDGMENTS

The authors thank S. Fischer and T. Han for comments. The authors apologize for omission of many relevant references in the piRNA field due to space limitations. ACB was supported by NIH Genetics Training Grant graduate fellowship T32GM007544-34. MAF was supported by the National Science Foundation Open Data IGERT Grant 0903629. JKK was supported by grants from the National Institute of General Medical Sciences (NIGMS) R01GM088565 and the Pew Charitable Trusts.

REFERENCES

1. Bagijn MP, Goldstein LD, Sapetschnig A, Weick EM, Bouasker S, Lehrbach NJ, Simard MJ, Miska EA: Function, Targets, and Evolution of *Caenorhabditis elegans* piRNAs. *Science* 2012.
2. Lee HC, Gu W, Shirayama M, Youngman E, Conte D, Jr., Mello CC: *C. elegans* piRNAs Mediate the Genome-wide Surveillance of Germline Transcripts. *Cell* 2012.
3. Ashe A, Sapetschnig A, Weick EM, Mitchell J, Bagijn MP, Cording AC, Doebley AL, Goldstein LD, Lehrbach NJ, Le Pen J, et al: piRNAs Can Trigger a Multigenerational Epigenetic Memory in the Germline of *C. elegans*. *Cell* 2012.
4. Shirayama M, Seth M, Lee HC, Gu W, Ishidate T, Conte D, Jr., Mello CC: piRNAs Initiate an Epigenetic Memory of Nonself RNA in the *C. elegans* Germline. *Cell* 2012.
5. Siomi MC, Sato K, Pezic D, Aravin AA: PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* 2011, 12:246-258.
6. Batista PJ, Ruby JG, Claycomb JM, Chiang R, Fahlgren N, Kasschau KD, Chaves DA, Gu W, Vasale JJ, Duan S, et al: PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Molecular Cell* 2008, 31:67-78.
7. Das PP, Bagijn MP, Goldstein LD, Woolford JR, Lehrbach NJ, Sapetschnig A, Buhecha HR, Gilchrist MJ, Howe KL, Stark R, et al: Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. *Molecular Cell* 2008, 31:79-90.
8. Claycomb JM, Batista PJ, Pang KM, Gu W, Vasale JJ, van Wolfswinkel JC, Chaves DA, Shirayama M, Mitani S, Ketting RF, et al: The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. *Cell* 2009, 139:123-134.
9. Gu SG, Pak J, Guang S, Maniar JM, Kennedy S, Fire A: Amplification of siRNA in *Caenorhabditis elegans* generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. *Nature genetics* 2012, 44:157-164.
10. Burton NO, Burkhart KB, Kennedy S: Nuclear RNAi maintains heritable gene silencing in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America* 2011, 108:19683-19688.

APPENDIX B: Identification of small RNA pathway genes using patterns of phylogenetic conservation and divergence

AUTHORS: Tabach Y, Billi AC*, Hayes GD*, Newman MA, Zuk O, Gabel H, Kamath R, Yacoby K, Chapman B, Garcia SM, Borowsky M, Kim JK, Ruvkun G

* These authors contributed equally to this work

AUTHOR CONTRIBUTIONS: YT, JKK, and GR designed experiments; YT developed analytical tools and analysed data; and YT, ACB, GDH, MAN, SMG, HG, RK, and JKK designed and carried out experiments. OZ gave statistical support and conceptual advice. YT, KY, BC, and MB wrote code. YT, ACB, JKK and GR wrote the paper. GR and JKK supervised the project.

CITATION: Tabach Y, Billi AC*, Hayes GD*, Newman MA, Zuk O, Gabel H, Kamath R, Yacoby K, Chapman B, Garcia SM, Borowsky M, Kim JK, Ruvkun G. Identification of small RNA pathway genes using patterns of phylogenetic conservation and divergence. *Nature* 493: 694-698 (2013). * these authors contributed equally to this work.

ABSTRACT: Genetic and biochemical analyses of RNA interference (RNAi) and microRNA (miRNA) pathways have revealed proteins such as Argonaute and Dicer as essential cofactors that process and present small RNAs to their targets. Well-validated small RNA pathway cofactors such as these show distinctive patterns of conservation or divergence in particular animal, plant, fungal and protist species. We compared 86 divergent eukaryotic genome sequences to discern sets of proteins that show similar phylogenetic profiles with known small RNA cofactors. A large set of additional candidate small RNA cofactors have emerged from functional genomic screens for defects in miRNA- or short interfering RNA (siRNA)-mediated repression in

Caenorhabditis elegans and *Drosophila melanogaster*^{1,2}, and from proteomic analyses of proteins co-purifying with validated small RNA pathway proteins^{3,4}. The phylogenetic profiles of many of these candidate small RNA pathway proteins are similar to those of known small RNA cofactor proteins. We used a Bayesian approach to integrate the phylogenetic profile analysis with predictions from diverse transcriptional coregulation and proteome interaction data sets to assign a probability for each protein for a role in a small RNA pathway. Testing high-confidence candidates from this analysis for defects in RNAi silencing, we found that about one-half of the predicted small RNA cofactors are required for RNAi silencing. Many of the newly identified small RNA pathway proteins are orthologues of proteins implicated in RNA splicing. In support of a deep connection between the mechanism of RNA splicing and small-RNA-mediated gene silencing, the presence of the Argonaute proteins and other small RNA components in the many species analysed strongly correlates with the number of introns in those species.

MAIN TEXT

Proteins with similar patterns of conservation or divergence across different organisms are more likely to act in the same pathways⁵. To identify proteins that share an evolutionary history with validated small RNA pathway proteins, we determined the phylogenetic profiles of approximately 20,000 *C. elegans* proteins in 85 genomes, representing diverse taxa of the eukaryotic tree of life: 33 animals, 6 land plants, 1 alga, 31 Ascomycota fungi, 3 Basidiomycota fungi and 12 protists. Of the ~20,000 *C. elegans* proteins, 10,054 show homologues in non-nematode eukaryotic genomes (Supplementary Table 1). Following correlation and clustering, this analysis sorts genes into clades of conservation and relative divergence or loss in the various organisms as suites of genes are maintained from common ancestors or diverge in particular lineages⁶. Protein divergence or loss in particular taxonomic clades is not random; entire suites of proteins can diverge or be lost as particular taxa specialize and no longer require ancestral functions. The correlated loss of proteins has been used to

assign roles for nuclear-encoded mitochondrial proteins⁷ and eukaryotic cilia-associated proteins⁸.

We developed a non-binary method of phylogenetic profiling to cluster all protein sequences encoded by *C. elegans* genes. BLAST scores were normalized to the length of the query sequence and for relative phylogenetic distance between *C. elegans* and the queried organism⁹. The matrix of 864,644 conservation scores for the 10,054 *C. elegans* proteins in the 86 genomes was queried either with a single protein to generate a ranking of other *C. elegans* proteins with the most similar pattern of conservation values or using a more global hierarchical clustering method (Fig. 1a). Proteins of the same families exhibit similar patterns of phylogenetic conservation and therefore tend to group together in the hierarchical clustering. However, many phylogenetic clusters include proteins with no sequence similarity; only their conservation or divergence in genomes is correlated. The ability of this non-binary method of phylogenetic profiling to cluster proteins based on function is exemplified by the clustering of proteins known to act as members of complexes. For example, the known protein components of the sensory cilium have highly correlated phylogenetic profiles characterized by loss in particular vertebrates and all fungi and plants and retention in particular protists, whereas the extraordinarily high and universal conservation of ribosomal and translation factor proteins clusters many of these translation components (Supplementary Fig. 1a, b).

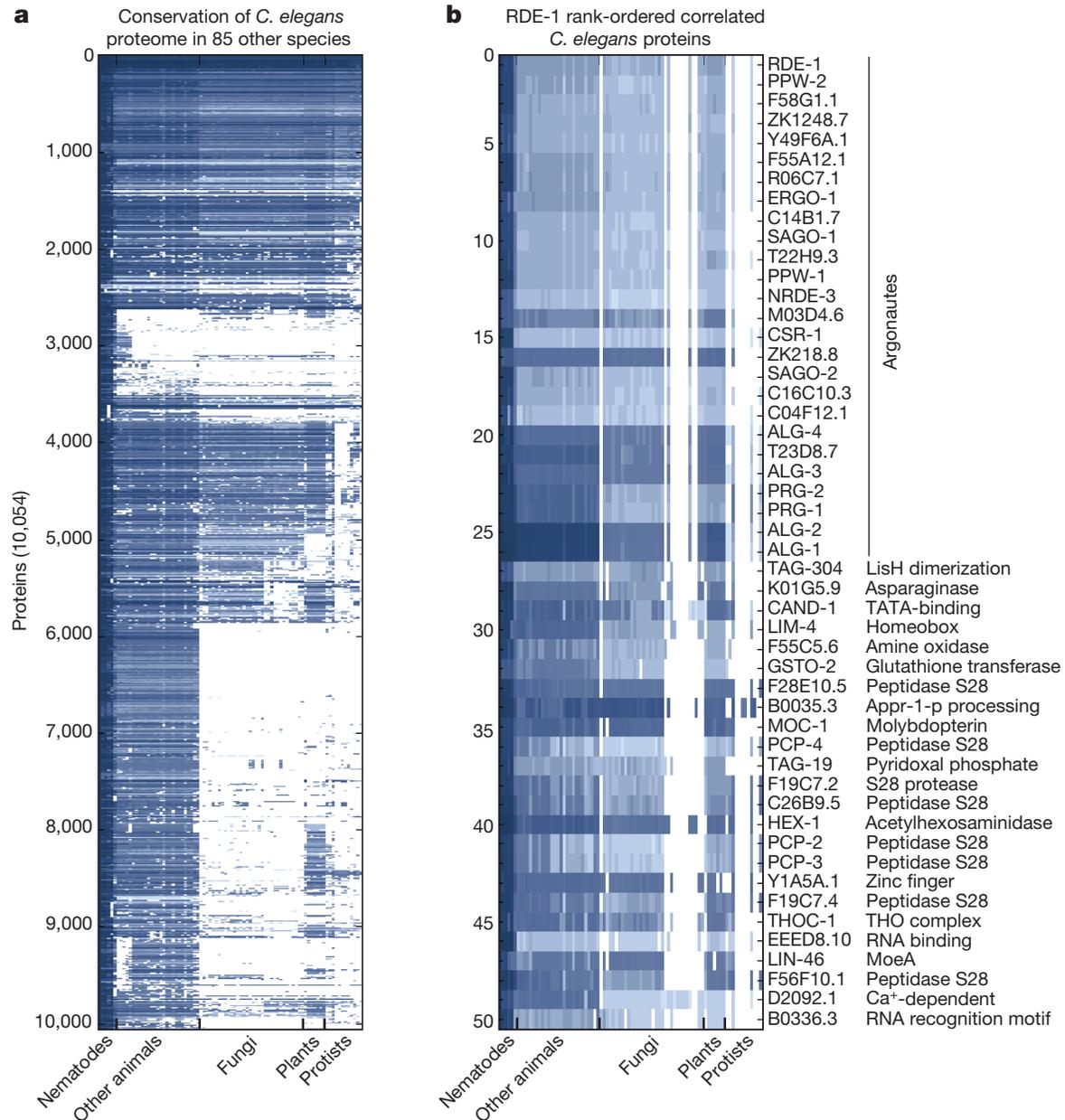
With a simple query of one of the central proteins in RNAi, the Argonaute RDE-1, we generated a rank-ordered list of proteins with phylogenetic profiles most similar to that of RDE-1 (Fig. 1b). The 26 other *C. elegans* Argonautes represent the top correlated proteins, a trivial consequence of protein sequence similarity within the Argonaute family. The signature phylogenetic profile of the Argonaute proteins is that they are absent in 9 out of 31 Ascomycota species, 1 out of 3 Basidiomycota species, and 6 out of 14 protist species, but have not been lost in any of the 33 animal or 6 land plant species compared. The retention of Argonaute proteins correlates with the ability to inactivate genes by RNAi¹⁰, and the loss of RNAi in about one-half of the sequenced

Ascomycota fungi is correlated with the 'killer' RNA virus¹¹. Additional *C. elegans* proteins that cluster with the Argonautes but show no sequence similarity include an asparaginase encoded by *K01G5.9*, the CAND-1 elongation factor and another elongation factor, the THO complex protein THOC-1. THO complex members have emerged from genetic screens for defective transgene and RNAi silencing in *Arabidopsis thaliana*¹².

Another validated *C. elegans* RNAi protein is MUT-2, a polyA polymerase implicated in a step downstream of the production of primary siRNAs by Dicer¹³. Out of the 50 *C. elegans* proteins with phylogenetic profiles most closely correlated with MUT-2 (Supplementary Fig. 1c), 10 are Argonautes, which bear no sequence similarity to MUT-2, demonstrating the efficacy of this approach to detect validated small RNA pathway proteins. The splicing components MAG-1, RSP-8, RNP-4, RSP-5 and DDB-1 and the translation factors EIF-3.D and EIF-3.E, many of which score in the validation tests below, also have similar phylogenetic profiles. In addition, out of the proteins most correlated with the *C. elegans* orthologue of Dicer (DCR-1), a nuclease that processes siRNAs and miRNAs, 3 Argonaute proteins emerge among the top 50 correlated phylogenetic profiles (Supplementary Fig. 1d and Supplementary Table 2).

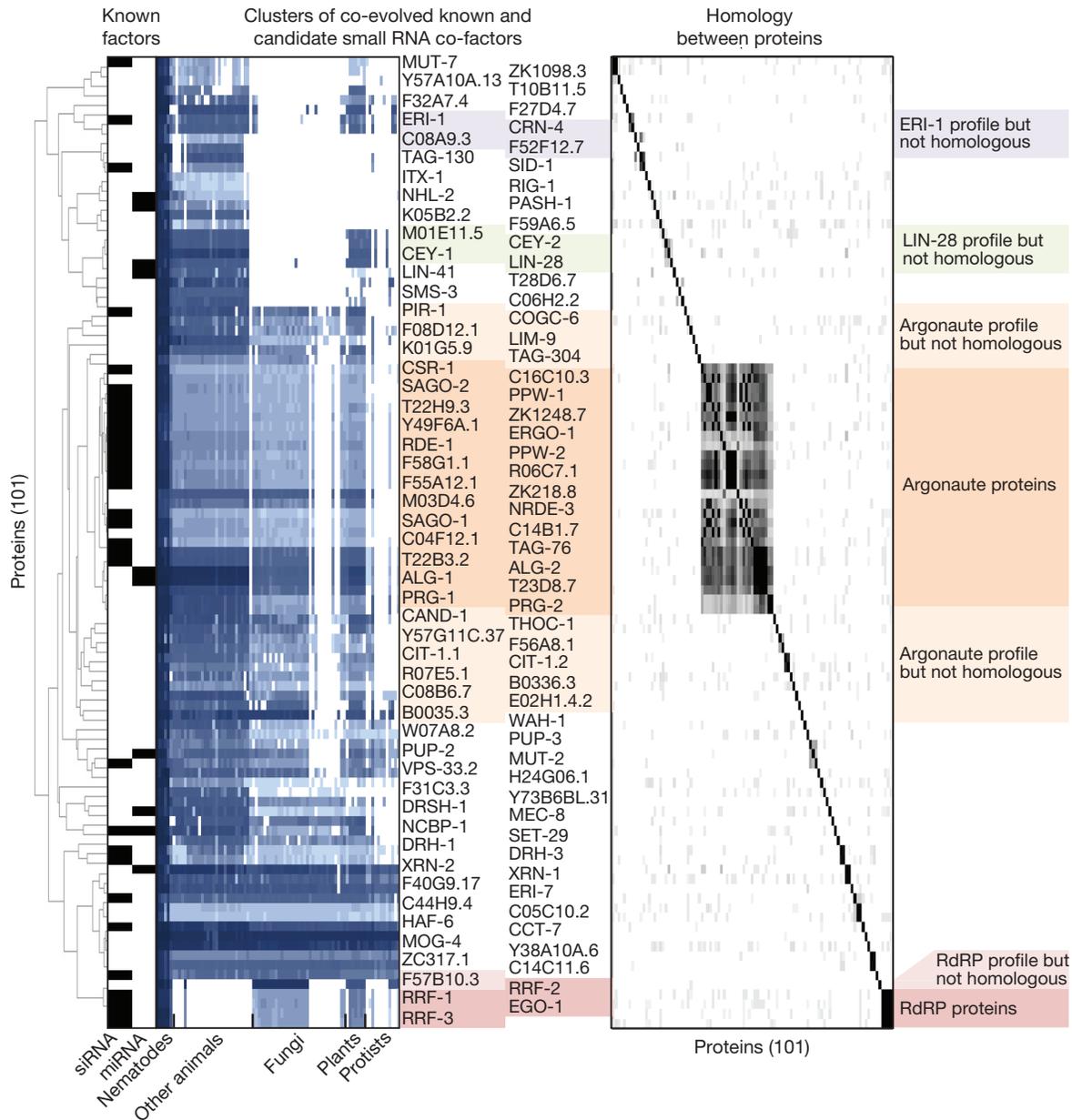
The RNA-dependent RNA polymerases¹⁴, siRNA-amplifying cofactors, are present in only 5 out of 27 animals (all the nematode species and, surprisingly, the tick), in all of the land plants, in 2 out of 4 Basidiomycota fungi, in 18 out of 27 Ascomycota fungi and in 4 out of 14 protists, but are not present in green algae. A query of the RNA-dependent RNA polymerase RRF-3 (Supplementary Fig. 1e) revealed the cofactor-independent phosphoglycerate mutase F57B10.3 as a dramatically correlated non-homologous protein ($R = 0.93$). Inactivation of this phosphoglycerate mutase gene causes defects in the endogenous siRNA response as well as transgene silencing, validating its role in RNA silencing (Supplementary Table 2). It is possible that either the biochemical substrate or product of this glycolysis pathway protein, or its enzymatic activity as a phosphatase, couples it to small RNA pathways.

Figure B.1 | Phylogenetic profiling analysis shows correlated conservation patterns of *C. elegans* proteins. **a**, Phylogenetic profiles of 10,054 conserved *C. elegans* proteins across 85 other eukaryotic genomes. For each *C. elegans* query protein, the normalized ratio of the BLAST score for the top-scoring protein sequence similarity is indicated in the column corresponding to each genome. Values range from 0 (white, no similarity) to 1 (blue, 100% similarity). **b**, Phylogenetic profiles of validated RNAi factor RDE-1 and the 49 most correlated proteins in rank order.



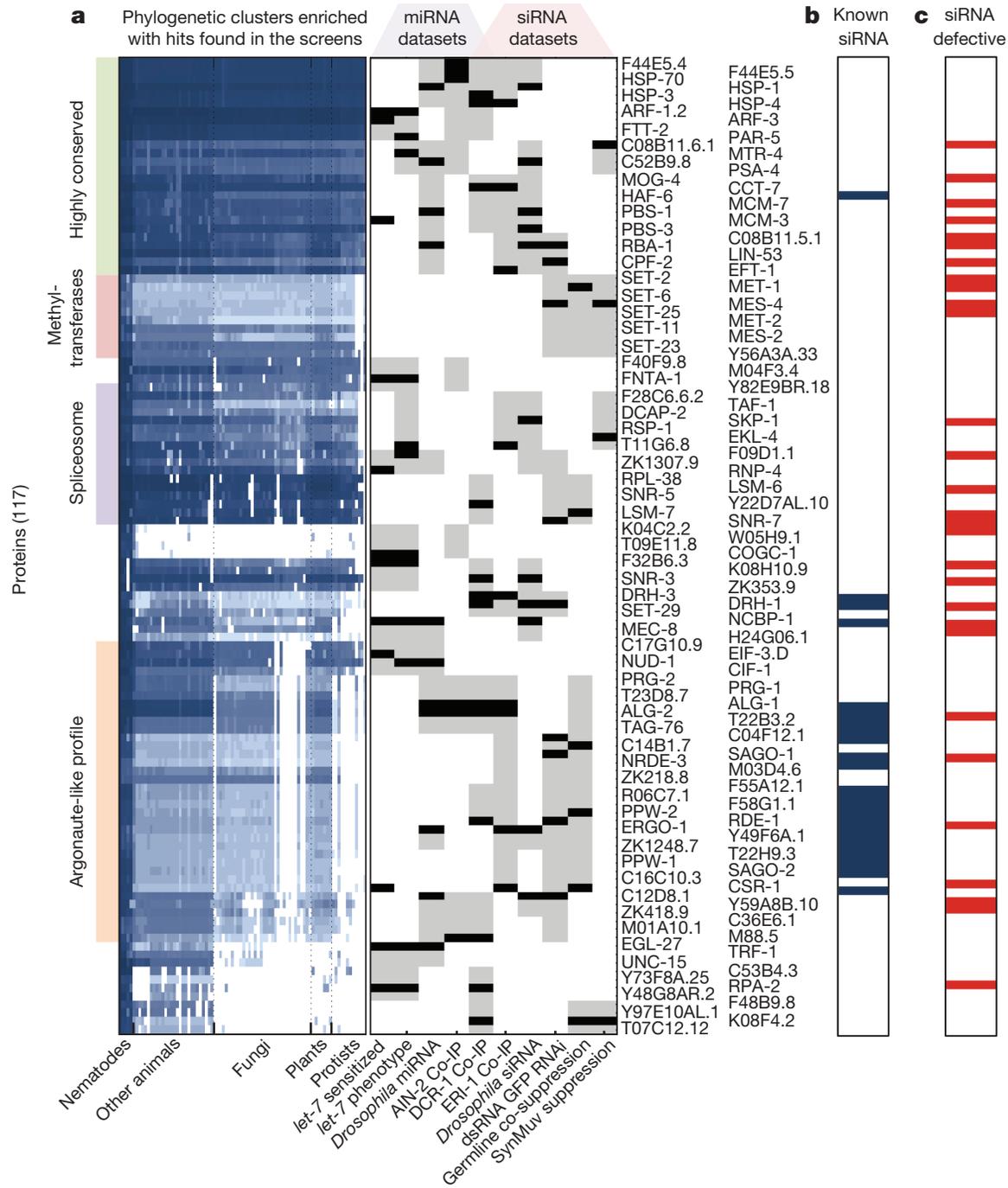
To identify candidate small RNA pathway proteins more comprehensively, we globally ranked proteins based on phylogenetic-profile correlation with multiple validated siRNA and miRNA cofactors. After assigning all conserved *C. elegans* proteins to hierarchical clusters, we gave each protein a score to reflect its phylogenetic clustering with the validated set of small RNA proteins (Supplementary Fig. 2). This analysis identified 60 proteins not previously implicated in small RNA pathways whose phylogenetic profiles correlate highly with those of validated siRNA and miRNA pathway proteins (Fig. 2).

Figure B.2 | Phylogenetic clusters of candidate small RNA pathway proteins. Validated miRNA and siRNA pathway proteins map non-randomly on the phylogenetic profile; proteins that map to the same clusters are likely to function in small RNA pathways. Left panel, clusters enriched for validated miRNA and siRNA pathway proteins (black boxes). Darker blue, higher protein-sequence similarity. Right panel, pairwise local protein-sequence alignment of all pairs of proteins in the cluster. White, no similarity; black, significant similarity.



The validated siRNA and miRNA protein cofactors identified so far probably constitute a small fraction of the total number of proteins that mediate small RNA function. Full-genome RNAi screens for defects in siRNA or miRNA pathway function have identified hundreds of additional candidate small RNA pathway proteins. We integrated ten genome-scale studies into the phylogenetic cluster analysis: five *C. elegans* gene-inactivation screens for defects in RNAi or miRNA function^{1,15,16}, *C. elegans* orthologues of *Drosophila* genes identified in two full genome RNAi screens for impaired siRNA or miRNA response² and three proteomic studies of complexes containing the known RNAi proteins DCR-1 (ref. 4), ERI-1 (ref. 17) and AIN-2 (ref. 18). Candidate genes identified in these studies show little overlap (Supplementary Table 3 and Supplementary Fig. 3a, b). However, the candidates from the different studies have similar phylogenetic profiles to each other and to validated small RNA cofactors (Fig. 3, Supplementary Fig. 3c, d and Supplementary Table 4).

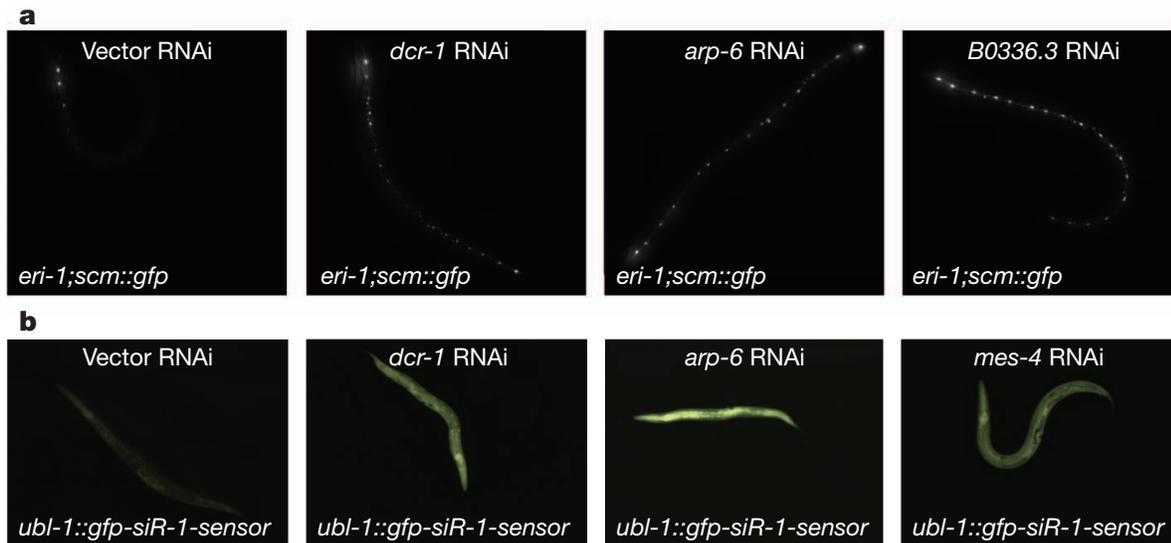
Figure B.3 | Select phylogenetic clusters enriched with hits from proteomic and functional genomic small RNA screens. **a**, The phylogenetic profile matrix was clustered and a Max Ratio score (MRS) was calculated for every protein in each screen; 117 proteins scored significantly in miRNA (56 proteins) or siRNA (75 proteins) functional genomic screens, or both (14 proteins). Middle panel, black tick, hit in screens; grey tick, significant MRS. **b**, Blue boxes, the 23 known small RNA pathway proteins identified. **c**, From the 117 proteins predicted by the phylogenetic profile, 28 proteins (red boxes) show defects in siRNA silencing ($P < 3 \times 10^{15}$).



We used a naive Bayesian classifier to assign predictive values to six genome-scale studies of RNAi cofactors and five miRNA cofactors (see Supplementary Methods)^{19,20}. To the phylogenetic profiles, we added a score for each *C. elegans* gene that is co-expressed on microarrays²¹ or whose encoded gene product interacts with validated small RNA pathway proteins²². The top 105 genes identified by this analysis are enriched with 41 well-validated siRNA pathway genes (Supplementary Fig. 7 and Supplementary Table 2). The other genes on this list are excellent candidates to mediate siRNA or related small RNA functions. More than 20 of these genes encode RNA recognition motifs including RNP ($P < 0.00001$) and helicase ($P < 0.00001$), an approximately 20-fold enrichment relative to the entire data set. Nine proteins from this list constitute components of the spliceosome (Supplementary Fig. 3).

From the proteins best correlated with validated small RNA pathway cofactors by phylogenetic profile or in the naive Bayesian analysis (Figs 1–3), we tested 87 representative candidates using two different tests for defects in RNAi. Transgene silencing in the somatic cells of the enhanced RNAi mutant *eri-1(mg366)* is mediated by an RNAi mechanism¹. We tested a set of 87 predicted small RNA pathway genes using this strain, and 43 scored as significantly RNAi-defective (Supplementary Table 2, and Fig. 4a). We also tested candidates using a green fluorescent protein (GFP)-based sensor for the abundant *C. elegans* endogenous siRNA 22G siR-1 (ref. 23) to monitor whether any of the gene inactivations affect the production or response to this endogenous siRNA. Thirty-three out of 87 genes tested scored in this assay (Supplementary Table 2 and Fig. 4b). Eight of the nine predicted splicing components scored strongly in these validation screens.

Figure B.4 | Inactivation of genes implicated in RNAi pathways re-animates transgenes that are silenced by RNAi. a, Expression of *scm::gfp* in the seam cells of an *eri-1(mg366)* mutant, where it is normally silenced by RNAi. Animals shown were treated with control, *dcr-1*, *arp-6* or *B0336.3* RNAi. **b,** GFP expression from the *ubl-1::gfp-siR-1-sensor* transgene, which is normally silenced by the siR-1 endogenous siRNA. Animals shown were treated with control, *dcr-1*, *arp-6* or *mes-4* RNAi.



The enrichment for RNA splicing components (Supplementary Fig. 4) points to a close mechanistic connection between splicing and small RNA regulation. Among the Ascomycota and protist species that have lost the Argonaute proteins, most show an extreme loss of introns, from 10^4 - 10^5 introns in species with Argonautes to 10^2 or fewer introns in most species without Argonautes (Supplementary Fig. 5). We screened for defects in RNAi a cherry-picked gene inactivation sublibrary of *C. elegans* orthologues of known splicing factors that have emerged from biochemical and genetic screens for splicing components from other systems. From a set of 46 *C. elegans* genes annotated in KEGG (Kyoto Encyclopedia of Genes and Genomes) to encode the orthologues of known splicing proteins that could be tested for roles in RNAi in our assays, 16 and 22 of these splicing-factor genes scored strongly in the *eri-1* transgene desilencing assay and the endogenous 22G siR-1 sensor assay. Many of the splicing components that scored strongly in these screens show a phylogenetic profile similar to the Argonaute proteins (Supplementary Fig. 6 and Supplementary Table 6). However, a subset of splicing factors that are well conserved across phylogeny also scored strongly in these assays.

We used the *eri-1* transgene desilencing system to conduct a full genome screen for gene inactivations that disable transgene silencing and identified 855 genes required for transgene silencing, with more than 200 scoring above 3 on a scale of 0 to 4 for desilencing (Supplementary Table 7). Among gene inactivations that caused the greatest desilencing, 11% correspond to the highest ranked predictions from the siRNA naive Bayesian analysis, a 30-fold enrichment ($P = 4.7 \times 10^{-13}$ using a hypergeometric test) for positives. Out of the 84 splicing factors that have been assigned to specific splicing steps, 49 scored in the full genome screen as required for transgene silencing, and 32 showed phylogenetic profiles clustering with known small RNA factors. The splicing factors that couple to small RNA pathways were not isolated to any particular step of RNA splicing. Splicing factor mutations in *Schizosaccharomyces pombe* disrupt the RNAi-based centromeric silencing²⁴. Both splicing proteins and siRNA and miRNA pathway proteins co-localize to cytoplasmic processing bodies (P-bodies) and nuclear

Cajal bodies²⁵, further supporting the possibility of functional crosstalk between splicing and RNAi.

Early genome sequence comparisons of *S. pombe*, *Saccharomyces cerevisiae* and a small set of eukaryotes suggested that loss of introns and splicing components is highly correlated with loss of Argonaute proteins²⁶. One interpretation was that the loss of RNAi in *S. cerevisiae* enabled viral invasion and a subsequent loss of introns through reverse transcription of genes by the invading viral replication enzymes. However, such a scenario would not predict that inactivation of splicing components in a species bearing the RNAi apparatus would cause an RNAi-defective phenotype. One model is that splicing could regulate RNAi indirectly by modulating spliced isoforms of key RNAi factors. However, the observations that only a subset of splicing cofactors are required for RNAi and the co-immunoprecipitation of splicing factors and DCR-1, ERI-1 and AIN-2 disfavours this indirect model. A mechanistic coupling between RNAi and RNA splicing explains these new data better. RNAi factors also affect splicing: Dicer is required for efficient spliceosomal RNA maturation in *Candida albicans*²⁷. If RNAi engages introns intimately by, for example, engaging nascent transcripts through the Argonaute NRDE-3 before splicing²⁸, then the selective advantage of introns may fade once the RNAi pathway is lost.

Our data suggest that a large subset of the proteins that mediate steps in the maturation of mRNAs bearing introns are also required for RNAi, and that those genomes that have lost most of their introns no longer require the RNAi pathway. Superimposed on the mRNA splicing pathway is an RNA surveillance system that eliminates aberrantly processed or mutant pre-mRNAs and mRNAs. It is possible that RNAi constitutes another level of mRNA surveillance that acts in parallel to—and using many of the same components as—the splicing quality control surveillance pathways.

METHODS SUMMARY

Informatics

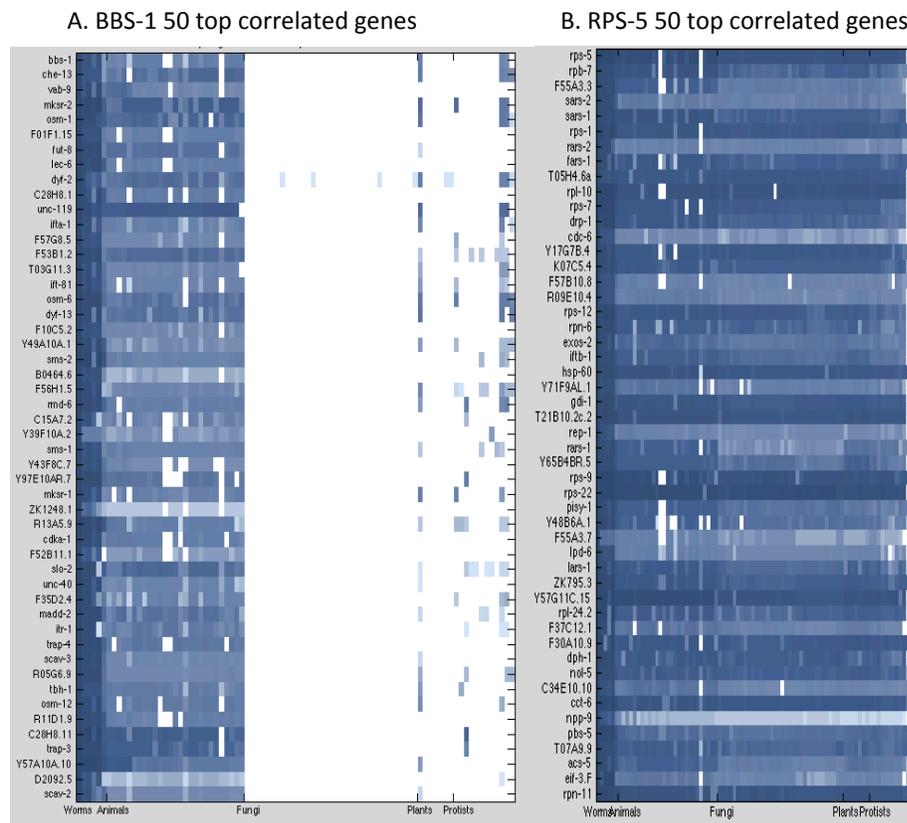
The Normalized Phylogenetic Profile (NPP) data matrix was clustered through MATLAB statistical toolbox using the average linkage method and Pearson correlation coefficient as a similarity measure. Clustering was performed on the rows of the matrix. To identify *C. elegans* proteins with phylogenetic profiles similar to published small RNA co-factors (Supplementary Table 9), the fraction of the validated proteins in each phylogenetic cluster was calculated and optimized to define a Max Ratio Score (MRS) (Supplementary Fig. 2).

ACKNOWLEDGMENTS

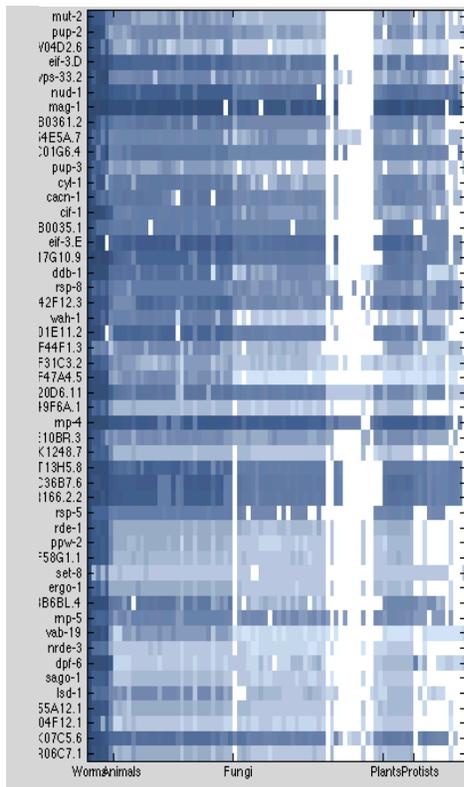
We thank T. Duchaine for access to his ERI-1 proteomic data before it was published and to S. Fischer, C. Zhang and T. Montgomery for helpful discussions. The work was supported by NIH GM088565 and the Pew Charitable Trusts (J.K.K.) and NIH GM44619 and GM098647 (G.R.).

SUPPLEMENT

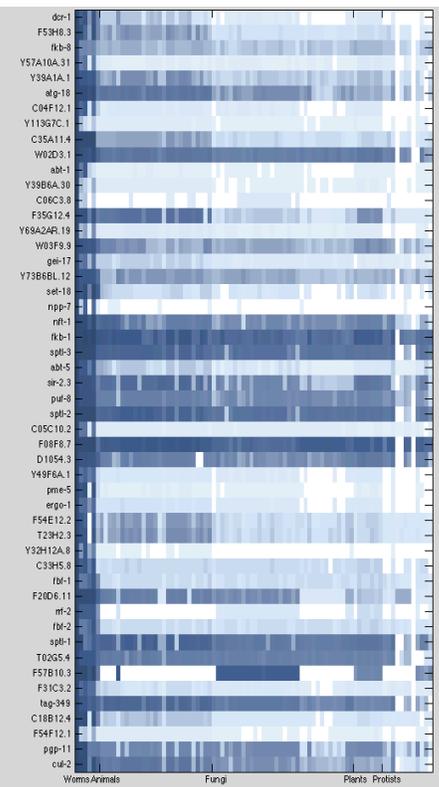
Figure B.S1: Phylogenetic profiles of the 50 proteins mostly correlated with: A. BBS-1, B. RPS-5, C. MUT-2, D. RRF-3, E. DCR-1. Correlation coefficients were calculated using the normalized phylogenetic profile matrix (NPP) and genes were rank ordered. Each row represents a gene; dark blue corresponds to high conservation of the *C. elegans* gene in that organism; white denotes no similarity. **A.** A query of the ribosomal S5 protein RPS-5 identifies in the top 30 proteins most correlated in phylogenetic profile 7 other ribosomal proteins with no similarity to RPS-5 as well as 6 tRNA synthetases also involved in translation. The ribosome is one of the most conserved components of the cell; strong conservation across nearly the entire phylogeny correlates the profiles of these proteins. **B.** A query of the ciliated sensory ending component BBS-1 detects the known ciliated ending components CHE-13, MKSR-2, OSM-1, IFTA1, IFT-81, DYF-2, OSM-6, and DYF-13 in the top 20 proteins with a correlated phylogenetic profile, BBS-1 shows no protein sequence similarity to any of these phylogenetically correlated *C. elegans* factors²⁹. The driving pattern of this phylogenetic profile correlation is strong conservation in all animals and particular protists, but no homologue in any of the fungi or plants tested.



C. MUT-2 50 top correlated genes



D. DCR-1 50 top correlated genes



E. RRF-3 50 top correlated genes

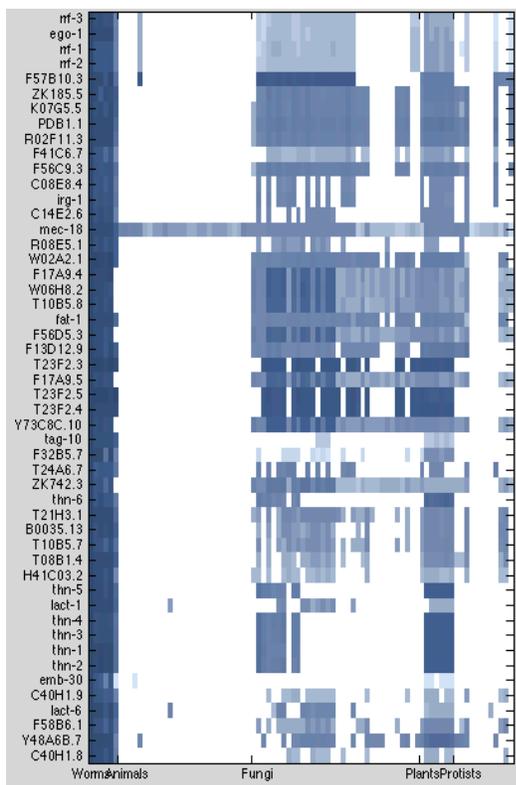
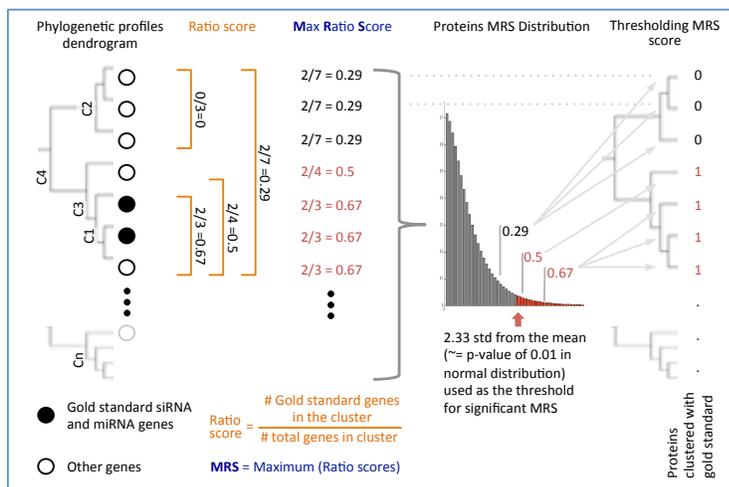


Figure B.S2: Identification of proteins that cluster phylogenetically with known small RNA co-factors or with hits from a set of small RNA genetic and biochemical screens. A. Hierarchical clustering of the NPP was used to cluster the proteins such that each could be assigned to several clusters, ranging from small, tight clusters (i.e. c1, c2) to clusters that contain more members (c3 or the even looser c4). The ratio of the number of validated RNAi pathway proteins to the total number of proteins in each cluster was calculated (termed the ratio score). Because each protein can have several ratio scores, depending on the number of clusters it belongs to, the highest ratio score per protein was used (termed the Maximum Ratio Score (MRS)). To identify those proteins with a significant MRS, we applied a filter, retaining only proteins with $MRS \geq 2.33$ Standard Deviations (SD) from the mean (p -value < 0.01). **B.** MRS calculation and thresholding was applied to each protein in the six datasets used to identify siRNA cofactors (see Supplementary Methods). Proteins that passed the threshold of 2.33 in at least three of the six datasets were considered positives and reported in Figure 3 (similar analysis was done to identify candidate miRNA pathway proteins).

A. Max Ratio Score (MRS) calculation



B. Using MRS to identify Phylogenetic profile clusters enriched with hits from a range of screens (used to generate Figure 3).

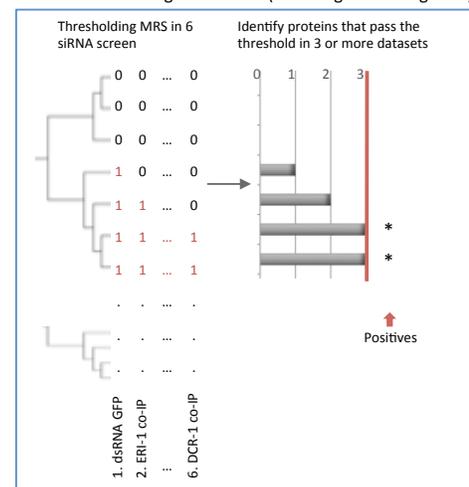


Figure B.S3: Overlap of genes and known small RNA factors between different screens without (A-B) and with (C-D) taking into account the phylogenetic clustering. **A.** Histogram of proteins that emerged from the siRNA (white) or miRNA (black) screens that were hits in 1, 2, 3, or 4 screens to identify siRNA or miRNA factors (see Methods). Absolute numbers are given above the bars. **B.** The number of previously validated siRNA (white) or miRNA (black) pathway proteins identified as hits in 0 to 4 screens. **C.** Histogram of the ratio of proteins (among those that emerged from the siRNA (white) or miRNA (black) screens) that passed the Max Ratio Score (MRS) threshold (Supplementary Figure 2) in the analysis of 1, 2, 3, or 4 screens. Absolute numbers are given above the bars. **D.** The number of previously validated small RNA pathway proteins that obtained a significant score in the MRS analysis of 0 to 5 screens.

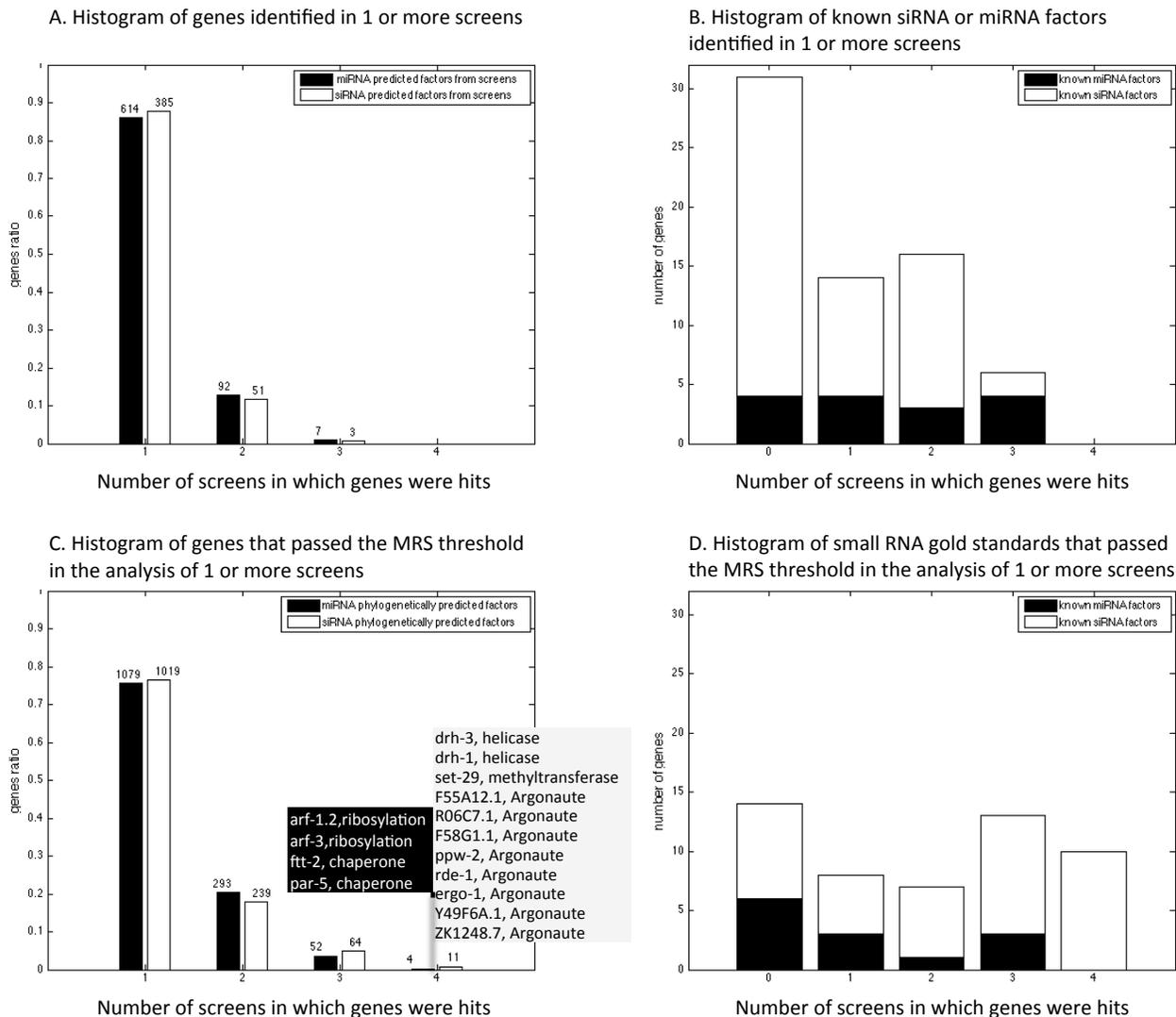


Figure B.S4: Proteins assigned to the spliceosome by KEGG pathway analysis. Cyan boxes represent the proteins that received the high scores in the Bayesian classification for siRNA co-factors (Supplementary Table 4), proteins that mapped to the same phylogenetic profile clusters as known small RNA-related factors (Figure 2), or proteins found in clusters enriched with hits from a range of proteomic and functional genomic small RNA screens (Figure 3).

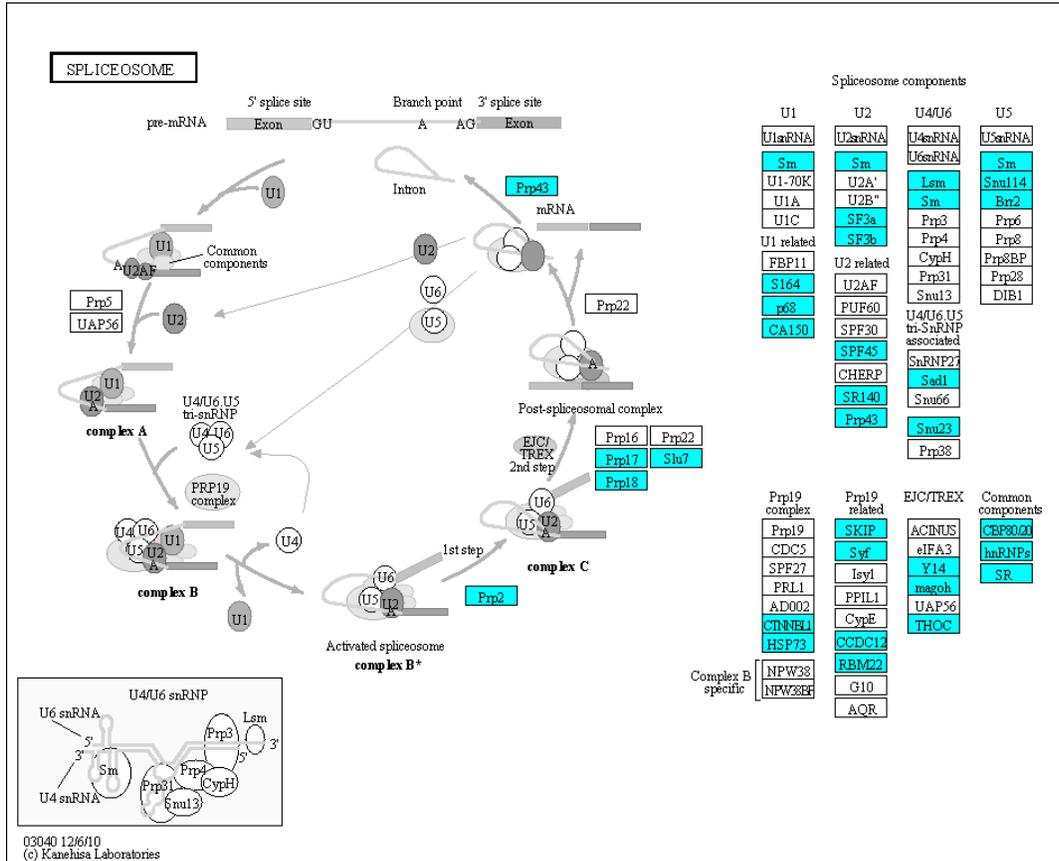


Figure B.S5: Relation between the average number of introns per gene in each species and the presence or absence of Argonaute proteins. Average number of introns per gene were taken from Kaplunovsky et al.³⁰ and Koralewski et al.³¹ (x-axis). Protein similarity to the *C. elegans* ALG-1 was calculated using blastp (see Supplementary Methods). Since all metazoans have many introns and all have conserved Argonaute proteins, we present one representative metazoan, *Homo sapiens*. There is a clear general trend for organisms with more introns to have significant Argonaute homologues. But there are outliers that are not explained by our model: for example, the fungus *Magnaporthe grisea* and the chromalveolate *Theileria parva* have no Argonaute but retain significant intron numbers.

taxonomic group	Organism	Introns/ gene	Homology to ALG-1
Amoebozoa	Dictyostelium discoideum	1.2850	0.08651801
Chromalveolata	Theileria parva	2.5804	0
Chromalveolata	Hemiselmis andersenii	0.0000	0
Chromalveolata	Guillardia theta	0.0253	0
Chromalveolata	Plasmodium falciparum	1.4038	0
Chromalveolata	Cryptosporidium parvum	0.0130	0
Protista	Leishmania braziliensis	0.0128	0
Protista	Leishmania infantum	0.0095	0
Protista	Trypanosoma brucei	0.0002	0
Fungi	Cryptococcus neoformans	5.2729	0.182631579
Fungi	Aspergillus fumigatus	1.9345	0.16
Fungi	Ustilago maydis	0.7510	0
Fungi	Eremothecium gossypii	0.0479	0
Fungi	Pichia stipitis (Scheffersomyces stipitis)	0.4414	0
Fungi	Encephalitozoon cuniculi	0.0075	0
Fungi	Magnaporthe grisea	1.6766	0
Fungi	Aspergillus niger	2.5760	0.1568326
Fungi	Gibberella zeae	2.2235	0.1458903
Fungi	Neurospora crassa	1.7423	0.1481193
Fungi	Schizosaccharomyces pombe	0.9367	0.2172181
Fungi	Yarrowia lipolytica	0.1114	0
Fungi	Saccharomyces cerevisiae	0.0553	0
Fungi	Debaryomyces hansenii	0.0526	0
Fungi	Kluyveromyces lactis	0.0244	0
Fungi	Candida glabrata	0.0161	0
Fungi	Malassezia globosa	0.5000	0
Plantae	Micromonas sp. rec299	0.5160	0.183684211
Plantae	Ostreococcus lucimarinus	0.2790	0
Plantae	Vitis vinifera	4.7500	0.3458917
Plantae	Arabidopsis thaliana	4.1479	0.3568341
Plantae	Oryza sativa	3.7902	0.3517682
Mammalia	Homo sapiens	7.7219	0.6281633

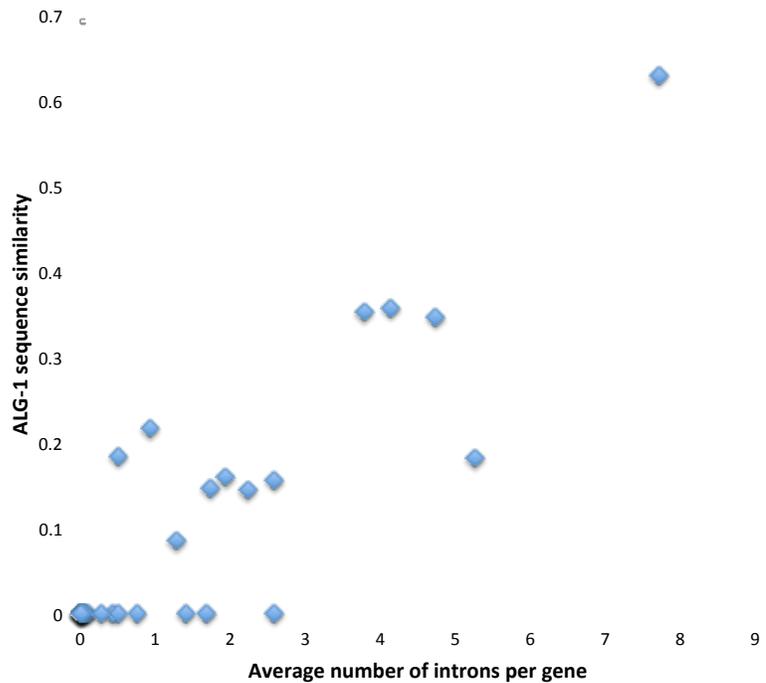


Figure B.S6: Inactivation of splicing factors that are implicated in the RNAi pathways reanimates transgenes targeted by RNAi. A. Expression of *scm::gfp* in the seam cells of an *eri-1(mg366)* mutant, where it is normally silenced by an RNAi pathway. Animals shown were treated with control, *rde-4*, *rnp-6*, *cdtl-7*, *rsp-3*, *prp-17*, *C0749.2*, *hrp-1*, *mnp-3*, or *ncbp-1* RNAi. **B.** GFP expression from the *ubl-1::gfp-siR-1* sensor transgene, which is normally silenced by the siR-1 endogenous siRNA. Animals shown were treated with control, *sap-1*, *lsm-6*, *mtr-4* or *rnp-6* RNAi.

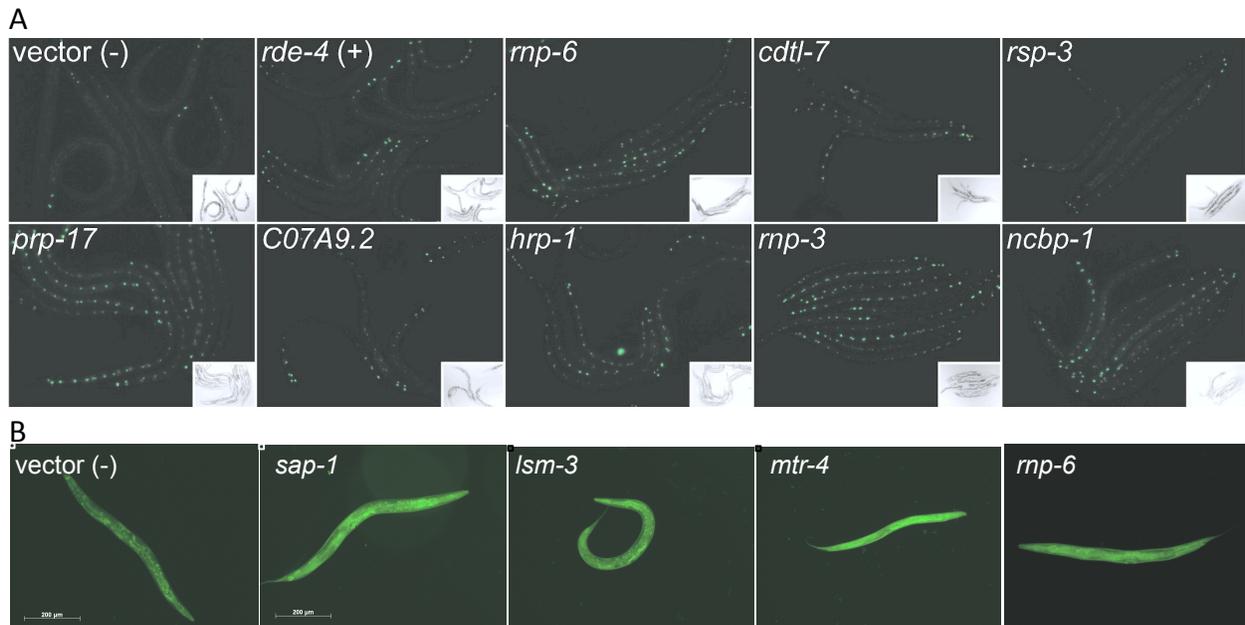


Figure B.S7: Receiver operating characteristic (ROC) analysis. Graphical representation of the Naïve Bayesian Classifier performance (see methods) in discovery of known siRNA factors (in red) compared to single datasets (in blue). For each dataset, a Likelihood Ratio score was calculated and the sensitivity as function of the specificity (Or number of known RNAi factors compared to other genes) was plotted.

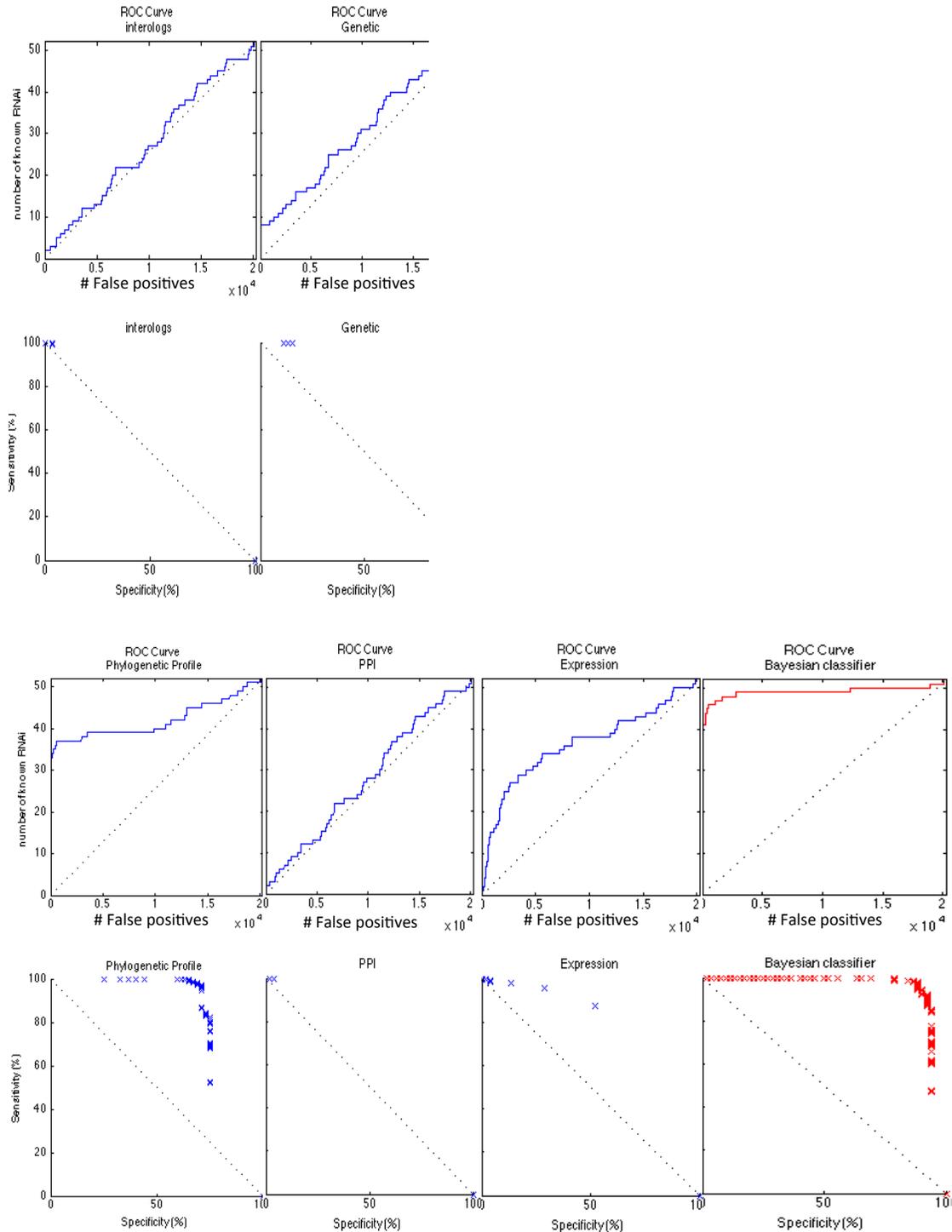


Table B.S1: The *C. elegans* phylogenetic profile database. Each row is a blastp bit-score between a single *C. elegans* protein and the top blast hit in each of the 85 other genomes. Among the ~20,000 *C. elegans* proteins, 10,054 are conserved proteins that have homologues (bearing significant protein domain sequence similarity) or orthologues (reciprocal top blast hit in each species) in other eukaryotic genomes. The result is a table of 10,054 proteins X 86 species. The table continues the gene list from Figures 1-3.

<http://www.nature.com/nature/journal/v493/n7434/extref/nature11779-s2.xlsx>

Table B.S2: Top siRNA pathway candidates and experimental tests of informatic predictions. Column A-M: Likelihood Ratio score indicating the contribution of being a positive in each of 10 different screens, gene coexpression, or protein-protein interaction to the probability of being a small RNA cofactor relative to baseline (see Supplementary Methods). **Column J-M:** The 87 genes were chosen for further validation based on: high Naïve Bayesian Classifier score, similar phylogenetic profile to RDE-1, similar phylogenetic profile to other known siRNA genes, or high CR score (Figure 1-3). **Column R,S:** score in the *eri-1* transgene desilencing screens. **Column T:** score in the 22G-siR-1 siRNA sensor screen.

<http://www.nature.com/nature/journal/v493/n7434/extref/nature11779-s3.xlsx>

Table B.S3: Overlap between positives in each of the functional genomic and proteomic screens and with the lists of known siRNA and miRNA pathway proteins. The table presents the percent of the known siRNA and miRNA proteins that were hits in each screen, the number of hits identified in each screen (the gray diagonal), the number of proteins that were also hits in other screens, (upper triangle) and the hyper-geometric p-value for such an overlap (lower triangle).

Studies	% siRNA (52)	% MicroRNA (15)	let-7 sensitized	let-7 phenotype	Drosophila miRNA	AIN-2 Co-IP	DCR-1 Co-IP	ERI-1 Co-IP	Drosophila siRNA	ds GFP RNAi	Germline Suppression defect	SynMuv suppression	Suppression of transgene silencing in eri-1
let-7 sensitized	7.7%	40.0%	319	78	7	3	3	5	11	7	5	1	68
let-7 phenotype	11.5%	33.3%	0	296	7	1	4	6	12	13	5	2	63
Drosophila miRNA	5.8%	33.3%	4E-04	3E-04	71	2	2	4	63	4	0	2	13
AIN-2 Co-IP	1.9%	20.0%	0.034	0.492	0.011	38	6	3	0	0	0	0	1
DCR-1 Co-IP	19.2%	20.0%	0.243	0.078	0.057	5E-08	95	22	3	3	1	1	10
ERI-1 Co-IP	19.2%	20.0%	0.026	0.005	6E-04	0.001	0	89	5	4	1	0	9
Drosophila siRNA	7.7%	20.0%	2E-05	1E-06	0	1	0.028	4E-04	120	6	0	3	23
ds GFP RNAi	23.1%	13.3%	0.001	5E-09	5E-04	1	0.012	0.001	4E-05	90	3	6	66
Germline suppression defect	11.5%	6.7%	0.01	0.008	1	1	0.317	0.312	1	0.006	71	1	11
SynMuv suppression	0.0%	0.0%	0.444	0.104	0.008	1	0.155	1	0.001	1E-08	0.122	31	17
Suppression of transgene silencing in eri-1	26.9%	46.7%	0	0	3E-05	0.846	0.012	0.027	1E-08	0	5E-04	5E-15	829

Table B.S4: Phylogenetic Clustering of hits from small RNA functional genomic screens.

Positives from the RNAi screens for factors in miRNA or siRNA pathway proteins tend to aggregate into phylogenetic profile clusters (column B), with an average of 80% conserved proteins (defined as top hit blastp scores >50 in more than 8 organisms). The Phylogenetic Coherence (PC) score (column C) was calculated for the conserved proteins in each screen to measure similarity among the phylogenetic profiles in a group of proteins (column D) (see Supplementary Methods).

screens	% Conserved genes	PC score	p-value
let-7 sensitized	82.4%	0.069	0.01181
let-7 phenotype	84.5%	0.055	0.25922
Drosophila miRNA	100.0%	0.154	<0.00001
AIN-2 Co-IP	94.7%	0.084	0.07598
DCR-1 Co-IP	78.9%	0.068	0.12217
ERI-1 Co-IP	79.8%	0.096	0.00658
Drosophila siRNA	98.3%	0.128	<0.00001
ds GFP RNAi	78.9%	0.103	0.00249
Germline Supression defect	62.0%	0.044	0.55185
SynMuv supression	87.1%	0.097	0.05415
Suppression of transgene silencing in eri-1	72.0%	0.064	0.00526
Known RNAi factors	76.6%	0.164	0.00004

Table B.S5: Top miRNA pathway candidates by Bayesian analysis. To estimate the likelihood of protein *a* being part of the siRNA pathway, we examined its score relative to the scores of the highly validated miRNA proteins in the relevant datasets. This was performed in two stages: First, we computed the likelihood ratio of protein *a* being associated with the miRNA pathway given the evidence from a single dataset (columns F-P). Next, we combined all likelihoods from the individual datasets into one predictive score (column Q).

<http://www.nature.com/nature/journal/v493/n7434/extref/nature11779-s4.xlsx>

Table B.S6: RNA interference defects after gene inactivations of *C. elegans* orthologues of known splicing factors. Columns F-H: The scores of gene inactivations of splicing factors with a transgene silenced by one particular endogenous siRNA or a transgene that is desilenced if RNAi is defective. Columns I-R: scores for these same gene inactivations from 10 other full genome screens for small RNA defects. Columns S: if the gene maps to a phylogenetic cluster of a known small RNA factor and the rank of the correlation. Above (in gray) are the p-values that were calculated for over-representation of splicing factors in each of the genome-wide small RNA studies. Gene inactivation of 33 out of 89 splicing factors from the KEGG dataset caused embryonic or early larval arrest that interfered with these tests, so only 46 of the 89 gene inactivations could be tested.

<http://www.nature.com/nature/journal/v493/n7434/extref/nature11779-s5.xlsx>

Table B.S7: Genome-wide transgene desilencing screen positives. Roughly 800 RNAi inactivations caused transgene desilencing. Of these, 448 were strong hits (scoring 2 or more). Genes targeted by positive clones are listed with average score as determined by screening process described in Supplementary Methods and Kim et al. (Kim et al., 2005).

<http://www.nature.com/nature/journal/v493/n7434/extref/nature11779-s6.xlsx>

Table B.S8: The validated siRNA and the miRNA pathway proteins. These genes and their encoded proteins were included in the validated list if the factor has been genetically or biochemically found to be a component of small RNA pathways. Using these criteria, we assembled a list of 52 factors that act in the siRNA pathway and a list of 15 factors that act in the miRNA pathway.

miRNA factors	siRNA Factors			
ain-1	C04F12.1	eri-9	rde-10	tsn-1
ain-2	cgh-1	haf-6	rde-2	vig-1
alg-1	cid-1	mut-14	rde-4	wago-2
alg-2	csr-1	mut-15	rrf-1	wago-4
dcr-1	dcr-1	mut-16	rrf-2	Y49F6A.1
drsh-1	drh-1	mut-2	rrf-3	ZK1248.7
lin-28	drh-3	mut-7	rsd-2	ZK757.2
lin-41	ego-1	ncbp-1	rsd-3	
ncbp-1	ekl-1	ncbp-2	rsd-6	
ncbp-2	ergo-1	nrde-3	sago-1	
nhl-2	eri-1	pir-1	sago-2	
pash-1	eri-3	ppw-1	sid-1	
pup-2	eri-5	ppw-2	sid-2	
xpo-1	eri-6	R06C7.1	T22B3.2	
xrn-2	eri-7	rde-1	T22H9.3	

SUPPLEMENTARY METHODS

Almost 50% of *C. elegans* genes encode proteins that are nematode-specific and excluded from this phylogenetic analysis. While the expected trend for conservation of most *C. elegans* proteins correlates with phylogenetic distance, with higher conservation in animals, less conservation in fungi and plants, and even less in protists. However, there are numerous dramatic examples of much higher divergence or even disappearance of homologues in particular clades; we focus on one such example, the small RNA cofactors.

The validated small RNA pathway factors are broadly conserved among RNAi-competent organisms. Furthermore, candidates identified by RNAi screens for small RNA pathway factors are highly enriched for conserved proteins (proteins that have homologous protein outside nematode), with an average of 80% conserved (Supplementary Table 2), and tend to aggregate into phylogenetic profile clusters (as measured using Phylogenetic Coherence score; see below). This suggests that the analysis captures much of the small RNA pathway despite the exclusion of nematode-specific proteins.

Phylogenetic profile generation

Protein sequences for *C. elegans* were downloaded using BioMart version 0.7 from the Ensembl project (release 60). When different splice variants existed for a gene, the longest variant was used. The resulting 20,242 protein sequences of *C. elegans* were compared using blastp of all open reading frames (ORFs) of 85 additional organisms. From the existing genomes available in the Ensembl database (release 60), we filtered a set of 53 fully sequenced eukaryotic genomes with no more than one genome per genus (except *Caenorhabditis*). Because Ensembl includes only a limited number of fungi and protists, 33 additional high quality genomes from the NCBI genome database were added to supplement the analysis. The blastp comparison generates a matrix P of

size 20,242 x 86 where each entry P_{ab} is the best blastp bit score between a *C. elegans* protein sequence 'a' and the top result in organism 'b'. The blastp scores provide a continuous phylogenetic profile, indicating homology level at each species. This approach is more sensitive than traditional binary phylogenetic profiles, which are based only on a comparison of the presence or absence pattern of suites of factors in particular clades of organisms^{32,33}.

Preprocessing and clustering the phylogenetic profiles

Preprocessing and normalization were applied to the profile matrix P prior to clustering. We used a preprocessing approach similar to that described by Enault et al.³⁴, related to the original binary phylogenetic profile preprocessing³².

Our method included several steps that were performed on the phylogenetic profile matrix P:

1. Thresholding low blastp bit scores: To reduce the influence of random matches in the phylogenetic profiles, low blastp bit scores (<50) were assigned a value of 1 (if $P_{ab} < 50$ then we set $P_{ab} = 1$).
2. Excluding poorly conserved proteins from the phylogenetic analysis: We have excluded proteins with less than five orthologues in the 81 non-nematode organisms from further phylogenetic analysis, since calculating the correlation between poorly conserved proteins is mainly governed by the zeros (no homologue found) across the phylogenetic matrix, and therefore such correlation measurement is likely less reliable. From a total of 20,242 worm proteins, only 10,054 passed this filter and were used for the subsequent phylogenetic profiling analysis.
3. Normalizing the blastp bit scores for protein length: Since the blastp score depends linearly on the length of protein 'a', long alignments would tend to have higher scores

independently of whether the aligned segments show sequence similarity, resulting a bias towards longer proteins. We therefore next normalized the phylogenetic profile matrix values to remove biases resulting from variations in protein lengths. In addition to P_{ab} , the best blastp bit score between a *C. elegans* protein *a* and all ORFs of a eukaryote genome '*b*', we computed P_{aa} , defined as the self-similarity score of the *C. elegans* protein '*a*' when blasted against itself. In LenNPP, the normalized phylogenetic profile matrix, each entry in the row corresponding to protein '*a*' is computed as: $\text{LenNPP}_{ab} = \log_2(P_{ab}/P_{aa})$. The normalized blastp score represents the (log)-ratio of the observed blastp score and the best possible blastp score of the same length (the self-similarity score), thus eliminating dependence on alignment length³⁴.

4. Normalizing for organisms with different evolutionary distance: A second normalization procedure was applied in order to compensate for the different protein similarity (i.e. score) expected when *C. elegans* proteins are compared to proteins from eukaryotes of highly variable evolutionary distance. For this purpose we normalized the values in each column *b* (i.e. each organism) by subtracting their average μ_b and dividing by their standard deviation σ_b , yielding:

$$\text{NPP}_{ab} = (\text{LenNPP}_{ab} - \mu_b) / \sigma_b$$

The normalized matrix NPP was used for subsequent clustering analysis.

For the more global clustering of proteins, a phylogenetic profile correlation (R) was calculated for each pair of the 10,054 proteins in the dataset. These R-values were used to cluster the proteins by average linkage, yielding groups of proteins with similar phylogenetic profiles.

Phylogenetic Coherence score

To measure if a particular set of proteins tends to have a more similar phylogenetic

profile than a random set of genes, we have developed the Phylogenetic Coherence (PC) score. The PC score measures how close on average are the phylogenetic profiles of proteins within a set compared to within a random set of proteins. A high PC score indicates that proteins within a set show similar phylogenetic profiles, a characteristic known to be associated with similar function^{32,33,35}. The PC score is a variation of the Expression Coherence (EC) score, which was originally developed to measure how similar a set of proteins is with regard to their expression profiles across different conditions^{36,37}.

To calculate the PC score for a given set A of K genes, the Pearson correlation between the normalized phylogenetic profiles (the NPP matrix) of each of the $K \times (K - 1) / 2$ pairs of proteins in A was calculated. The phylogenetic coherence score is simply defined as the fraction of pairs whose score exceeds a threshold, $PC(A) = p(A,S) / (K(K - 1) / 2)$, where $p(A,S)$ is the number of gene pairs in set A whose phylogenetic similarity is better than a threshold similarity S. We determined the value of the threshold S as follows: We calculated the Phylogenetic correlation between all 10,057 conserved *C. elegans* protein sequence pairs ($10,054 \times 10,053 / 2 = 50,536,431$) and then defined S as the 95th percentile of the distribution of these similarities (such that a random set of K sequences should get, on average, a PC score of ~0.05). If the sequences in our set K tend to have more similar phylogenetic profiles than a random set, their PC score should be > 0.05 .

To assign a p-value for the PC score of a list of sequences A of size K, the process was repeated 10,000 times for random sets of sequences of the same size K. PC scores were calculated for the random sets and used to rank of the true set's score PC (A) among the 10,000 randomized scores, yielding an empirical p-value for the PC score of the true set A. Finally, to test the robustness of the method to the threshold choice, alternative thresholds (S) were tested. These yielded similar p-values and identified similar factors as significant (data not shown).

The validated siRNA and the miRNA pathway factors

To identify new proteins that are part of small RNA pathways, we compiled two gold standard lists of factors with evidence in the literature for a role in either the siRNA or miRNA pathway. A factor was included in a gold standard list if the factor has been genetically or biochemically found to be a component of the small RNA pathways. Using these criteria, we assembled a gold standard list of 52 factors that are part of the siRNA pathway and a list of 15 factors that are part of the miRNA pathway (see lists in Supplemental table 8). Three factors, DCR-1, NCBP-1, and NCBP-2, are in both lists. The average linkage method produces a hierarchical clustering (dendrogram), and distinct clusters were obtained by ‘cutting’ the dendrogram at various thresholds, producing different numbers of clusters.

The Cluster Ratio and Max Ratio Scores

Given a pre-defined set of proteins of interest G_{interest} (for example, proteins with shared biological function such as the siRNA pathway factors, or factors obtained as results of a certain biological assay such as an RNAi screen), we wanted to identify which other *C. elegans* proteins might be related to this set based on similarity in their phylogenetic profiles. For this purpose, we have clustered the NPP and used the obtained dendrogram to score proteins for phylogenetic similarity with the list of validated factors. The dendrogram was thresholded to obtain N distinct clusters using the MATLAB ‘cluster’ function, for different clustering resolutions N . Next, we looked for each factor a at the overlap between the cluster to which it was assigned $G_{\text{cluster}(a,N)}$ and the list of factors of interest G_{interest} . To quantify this overlap, we have calculated for each factor a the Cluster Ratio (CR) score $CR_{a,N}$, which is the fraction of factors from the cluster $G_{\text{cluster}(a)}$ that belong to the list of interest (G_{interest}).

$$CR_{a,N} = |G_{\text{interest}} \cap G_{\text{cluster}(a,N)}| / |G_{\text{cluster}(a,N)}|$$

Where here $|A|$ denotes the number of factors in a set A. (see supplementary fig 2 showing the cluster ratio score)

The CR score captures the tendency of factors to appear together with the list of interest based on the clustering dictated by our dendrogram, with factors having a high CR score showing similar phylogenetic profile to one or several factors in our list of interest. Such genes represent candidate factors predicted to have similar function with our list of validated miRNA and siRNA pathway factors.

For the phylogenetic profile of each factor a , we have tested the similarity to the profiles of factors on the list of interest at various similarity levels by modifying the clustering resolution. This was achieved by altering the number of clusters N obtained from the dendrogram, with N values chosen to be $N = 10, 50, 100, 200, \dots, 9000, 10000$. This resulted in 102 different Cluster Ratio ($CR_{a,N}$) scores for each factor a . Finally, for each factor we chose the clustering resolution maximizing the cluster ratio, giving us the gene's Max Ratio Score: $MRSa = \max (CR_{a,10}, CR_{a,50}, \dots, CR_{a,10000})$; when cluster is defined as a group of 3 or more proteins with most similar profile to each other. The MRS for each factor a represents the optimized phylogenetic clustering resolution achieving the highest enrichment for factors of interest in a cluster containing gene a .

Integration of genome-scale data sets

Sixteen recently published studies and genome-wide databases were integrated using a Naïve Bayesian Classifier (see below) to predict new factors that are part of the siRNA or miRNA pathways. From the 16 datasets described below, 12 were used to predict new factors in the siRNA pathway and 11 were used to predict new factors in the miRNA pathway, as indicated below:

***let-7* sensitized background screen (miRNA):** The *let-7* miRNA is conserved in other organisms^{38,39}. A sensitized background of a weak *let-7* allele, *mg279*, was used to

identify miRNA pathway factors by genome-wide RNAi screening for enhancement of the *let-7(mg279)* vulval rupture phenotype⁴⁰. Screen positives were divided into three categories: weak, medium, and strong. From the total of 332 hits in the screen, 105 were not repeated in a secondary screen and considered as weak hits (we scored them 1), 169 genes retested positive in triplicate, considered as a medium hits (scored 2), and 45 were validated by genetic tests and declared strong hits (scored 3). Three genes didn't match our gene database, and all the other genes in the database were scored 0.

Vulval bursting phenotype screen (miRNA): The *let-7* miRNA controls the L4-to-adult transition. *let-7* mutants fail to execute this transition and die by bursting through the vulva³⁹. This vulval bursting phenotype can therefore indicate defects in miRNA pathway function. We have downloaded from WormBase (WS220) a list of 296 genes with the exploded through vulva phenotype in RNAi experiments. These genes were scored 1 to indicate a vulval bursting phenotype, and all other genes were scored 0.

***D. melanogaster* miRNA type (imperfect duplex) 3' UTR reporter screen (miRNA):** A genome-wide RNAi screen was performed in *D. melanogaster* S2 cells to identify factors that impact miRNA pathway function⁴¹. *C. elegans* orthologues of tested protein sequences were scored 1 if positive, 0 if not. *C. elegans* proteins whose orthologues were not tested were assigned a null score.

AIN-2 Co-immunoprecipitation (miRNA): AIN-2 interacts with miRNA-specific Argonaute proteins and regulates the expression of miRNA targets. To identify proteins interacting with AIN-2, which could represent miRNA pathway factors, a mass spectrometry-based proteomics approach was applied⁴². The 38 identified AIN-2-interacting factors were scored 1, and all others were scored 0.

DCR-1 Co-immunoprecipitation (siRNA and miRNA): A mass spectrometry-based proteomics approach was used to identify DCR-1-interacting proteins⁴³. The purification process was performed in duplicate under native conditions in embryos and gravid adults⁴³. We scored as follows: Proteins identified in mass spectrometry of DCR-1

complexes in both embryonic and adult purifications received a score of 2. Proteins identified in two repeats of a single purification (embryonic or adult) received a score of 1. Otherwise, proteins were scored according to the peptide coverage ratio, which was always less than one (i.e. for peptide coverage of 26%, the gene score is 0.26).

ERI-1 Co-immunoprecipitation (siRNA): A mass spectrometry-based proteomics approach was used to identify ERI-1-interacting proteins. A tagged ERI-1 protein was purified using standard protein biochemistry under native conditions, washed extensively, and interacting proteins were identified by mass spectroscopy. The ERI-1-interacting factors were scored 1, and all others were scored 0.

***D. melanogaster* siRNA type (perfect duplex) 3' UTR reporter screen (siRNA):** A genome-wide RNAi screen was performed in *D. melanogaster* S2 cells to identify genes that impact siRNA pathway function⁴¹. *C. elegans* genes orthologous to tested genes were scored 1 if positive, 0 if not. *C. elegans* genes whose orthologues were not tested were assigned a null score.

Transgene RNAi screen (siRNA): A genome-wide RNAi screen was performed in an engineered RNAi sensor strain of *C. elegans* to identify genes required for RNAi. Genes corresponding to the RNAi clones were scored on a GFP intensity and penetrance scale of 0 (no GFP expression) to 4 (highly penetrant, strong GFP expression), and those that scored an average of 2 or greater were designated candidate RNAi genes⁴⁴. We used numerical scores as reported in the paper.

Germline cosuppression defect screen (siRNA): During silencing of repetitive transgenes, a trans effect (“cosuppression”) occurs that results in silencing of cognate endogenous genes. A genome-wide RNAi screen was performed in an engineered germline cosuppression sensor strain of *C. elegans* to identify factors required for cosuppression in the germline⁴⁵. Positives were scored 1, and all others were scored 0.

Suppression of synMuvB and synMuvA synthetic multivulva (Muv) phenotype

screen (siRNA): SynMuv B genes are involved in multiple cellular functions during development including RNA interference⁴⁶. A genome-wide RNAi screen was performed in the *lin-15AB(n765)* background to identify suppressors of the Muv phenotype⁴⁶. SynMuv suppressor genes were scored 1, and all others were scored 0.

Phylogenetic profiling analysis (siRNA and miRNA): We have generated phylogenetic profiles for the entire worm proteome by blastp, searching all ~20,000 worm proteins across all 86 genomes (see Methods, above). Proteins were clustered based on phylogenetic profile similarity, and the score used for each is the Max Ratio score (MR) (see Methods, The Cluster Ratio and Max Ratio scores).

Co-expression analysis (siRNA and miRNA): For each gene in the gold standard groups (siRNA or miRNA) we identified, using the SPELL engine (Serial Pattern of Expression Levels Locator)⁴⁷, the 100 genes that correlate best in 72 different gene expression data sets. The results are 100 x 51 (for the siRNA) and 100 x 14 (for the miRNA) tables of the most correlated genes for each of the gold standard genes. For each gene, independent siRNA and miRNA co-expression scores were calculated as the number of time the gene is found in each of the tables (e.g. *inx-22* was among the top 100 co-expressed genes of 15 of the siRNA and 2 of the miRNA gold standard factors; hence, its scores are 15 for siRNA and 2 for miRNA).

Protein-protein interactions: A genome-scale protein-protein interaction map generated from yeast two-hybrid data was downloaded from the Worm Interactome version 8⁴⁸. We scored each gene by calculating the ratio of its number of interactions with the siRNA or miRNA gold standard factors to its total number of interactions.

Interologs: protein-protein interactions of orthologues of *C. elegans* protein coding genes: Predicted pairs of *C. elegans* interactors whose respective orthologues were experimentally shown to interact in another organism were downloaded from Worm Interactome version 8⁴⁸. We scored each factor by calculating the ratio of its number of interactions with the siRNA or the miRNA gold standard factors to its total number of

interactions.

Predicted genetic interactions from text mining: WormBase provides a list of genetic interactions that are text processed and manually curated⁴⁸. We scored each gene by calculating the ratio of its number of interactions with the siRNA or the miRNA gold standard factors to its total number of interactions.

Gross phenotypic signatures: A list of genes pairs that share phenotypic similarity were download from the Worm Interactome version 8⁴⁸. We scored each gene by calculating the ratio of its number of pairings with the siRNA or the miRNA gold standard factors to its total number of pairings.

For each of the two pathways, the entire dataset was represented by one data matrix D , where D_{ab} represents the value obtained for factor 'a' in dataset b . Values were either binary (e.g. for the vulval bursting phenotype screen), or quantitative (e.g. for the protein-protein interaction dataset). In all datasets, higher values suggest a higher probability of a factor belonging to the siRNA or miRNA pathways.

For brevity, we describe here the analysis for the siRNA pathway. The miRNA pathway analysis is identical, except for a different gold standard set and data matrix D used. To estimate the likelihood of factor 'a' being part of the siRNA pathway, we examined its score relative to the scores of the gold standard genes in all datasets. This was performed in two stages: First, we computed the likelihood of factor 'a' being associated with the siRNA pathway given the evidence from a single dataset. Next, we combined all likelihoods from the individual datasets into one predictive score. For the true status of factor 'a' is marked by a binary variable Y_a , which is equal to one if the factor is part of the siRNA pathway. Since we don't know if factor a is part of the siRNA pathway, Y_a is unknown, and our goal is to predict it as accurately as possible, given the dataset D . Methods are defined in the following sections.

A screen that is useful for our analysis is indicated by scores for the gold standard

factors that are higher than expected by chance. Therefore, a factor getting a high score is more likely to function in the siRNA pathway. We utilized this information to define a likelihood ratio score as follows: For each factor a in each dataset b , we defined a threshold score t_{ab} , such that all factors with scores in the dataset greater or equal to this threshold are considered positives, and other factors are considered negatives. For binary traits, the threshold t_{ab} was simply chosen to be $t_{ab} = D_{ab}$, such that positives are either all factors with score '1' (in case factor 'a' got a '1' score, giving evidence for it being part of the siRNA pathway) or all factors (in case factor 'a' got a '0', offering no evidence for pathway membership). For quantitative datasets, threshold selection was slightly more complex. The use of D_{ab} as a threshold might be sub-optimal and even misleading - this is particularly true in cases when D_{ab} is very high and none of the gold standard factors passed D_{ab} . We therefore examined all thresholds $t \leq D_{ab}$ and calculated the likelihood ratio $LR_{ab+}(t)$ for each possible threshold (as described below). We then set the threshold t_{ab} as the one maximizing the obtained likelihood ratio, and took

$$LR_{ab+} = \text{MAX}_t\{LR_{ab+}(t)\}.$$

Once a threshold has been set, we have computed a Likelihood Ratio score LR_{ab+} , a measure of a test power indicating how the knowledge of a specific score changes the likelihood of a factor being part of the siRNA pathways from baseline. More precisely, the likelihood ratio score is defined as $LR_{ab+} = \Pr(Y_a = 1 \mid D_{ab}) / \Pr(Y_a = 0 \mid D_{ab})$; i.e. $LR_{ab+}(t)$ is the ratio of probabilities of a factor a being part of the siRNA pathway versus not being part of this pathway given the evidence provided by dataset b . For each dataset we set LR_{ab+} as $LR_{ab+}(t)$ using the threshold t chosen as above. In practice, it is computed by comparing the proportion of gold standard factors among the positives (genes which scored above the threshold t_{ab}) and negatives (factors scoring below the threshold), as detailed below.

The value $\Pr(Y_a = 1 \mid D_{ab})$ is also often termed sensitivity, and the value $\Pr(Y_a = 0 \mid D_{ab})$ is known as one minus the specificity. The sensitivity and specificity values for a given

score D_{ab} are defined as:

where:

(i) TP_{ab} denotes True Positives, the number of gold standard factors with scores equal to or higher than the score threshold t_{ab}

(ii) TN_{ab} denotes True Negatives, the number of non-gold standard factors with scores lower than the score threshold t_{ab} .

(iii) FP_{ab} denotes False Positives, the number of non-gold standard factors with scores equal to or higher than the score threshold t_{ab} .

(iv) FN_{ab} denotes False Negatives, the number of gold standard factors with scores lower than the score threshold t_{ab} .

The likelihood ratio, computed via sensitivity and specificity is then given by:

$$\text{specificity} = \frac{TN_{ab}}{TN_{ab} + FP_{ab}} = \frac{\text{TrueNegatives}}{\text{TrueNegatives} + \text{FalsePositives}}$$
$$\text{sensitivity} = \frac{TP_{ab}}{TP_{ab} + FN_{ab}} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalseNegatives}}$$

Finally, we used a Naïve Bayesian Classifier to merge the LR_{ab} scores from the different datasets and assign a final score. Naïve Bayesian Classifiers provide a simple, standard, and scalable method for utilizing the power of different data sources and types for prediction by assuming conditional independence of the various predictors given the outcome. It has been used successfully in various genomics applications⁴⁹⁻⁵¹ and was used here to predict likelihood of membership in the siRNA pathway for a given factor.

We define the final score for a factor 'a' (S_a) as the log likelihood ratio of the probability of factor a being in the siRNA pathway to the probability of factor 'a' not being in the pathway given evidence collected from all 12 datasets used for the siRNA classifier:

$$S_a = \log(\Pr(Y_a = 1 \mid D_{a1}, D_{a2}, \dots, D_{a12}) / \Pr(Y_a = 0 \mid D_{ab}, D_{a2}, \dots, D_{a12}))$$

An underlying assumption of the Naïve Bayesian procedure is that the individual data sets are independent of each other. As such, we can compute S_a by simply summing the log-likelihood ratios:

$$S_a = \sum_b (LR_{ab+})$$

where LR_{ab+} is the likelihood ratio score of factor 'a' in data set b .

The independence assumption is rarely strictly satisfied in practice⁴⁹ Hence treating the dataset as independence may be sub-optimal. Nevertheless, we used the Naïve Bayesian model for two reasons: first, our goal in this work was to show that combining different data sources in a simple manner enables us to reliably predict new siRNA pathway factors; and second, reliably estimating and exploiting the dependencies in our databases is difficult, and often requires larger amounts of data. Better modeling of the dependencies between the different data sources will likely lead to even better classifiers and thus more accurate prediction of gene membership in the pathway.

Validation screens

A transgene that expresses GFP in the hypodermal cells in wild type is silenced in an *eri-1(mg366)* mutant, but RNAi targeting of genes encoding validated small RNA pathway cofactors such as *rde-1*, *rde-4*, or *dcr-1* causes transgene desilencing. *wls54(scm:gfp)* in *eri-1(mg366)* is silenced in seam cells⁴⁴. Desilencing of the *wls54(scm:gfp)* transgene in the *eri-1(mg366)* mutant and desilencing of the *ubl-*

1::GFP::siR-1 endo siRNA sensor transgene was tested in two samples of each of 87 gene inactivations and scored 4 for most desilencing to 0 for least. For the 87 top ranked genes from the Bayesian analysis tested, the sequences of the gene inactivating dsRNAs were verified. In the full genome screen with the *wls54* in *eri-1(mg366)*, every gene knockdown that caused in any degree of desilencing (score > 0) in the primary screen was subjected to secondary screening in triplicate, scoring 4 for the most desilencing down to 0 for no desilencing. Due to the large number of positives emerging from the full genome screen, plasmids for RNAi clones were not re-sequenced.

Images

Images were captured using a Zeiss Axioplan microscope equipped with a Hamamatsu digital camera and Zeiss Axiovision software. Images compared to each other were captured using the same exposure settings and processed identically. Control RNAi bacteria expressed double-stranded RNA homologous to no worm gene.

REFERENCES

1. Kim, J. K. et al. Functional genomic analysis of RNA interference in *C. elegans*. *Science* 308, 1164–1167 (2005).
2. Zhou, R. et al. Comparative analysis of argonaute-dependent small RNA pathways in *Drosophila*. *Mol. Cell* 32, 592–599 (2008).
3. Meister, G. et al. Identification of novel argonaute-associated proteins. *Curr. Biol.* 15, 2149–2155 (2005).
4. Duchaine, T. F. et al. Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. *Cell* 124, 343–354 (2006).
5. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl Acad. Sci. USA* 96, 4285–4288 (1999).
6. Gabaldón, T. Evolution of proteins and proteomes: a phylogenetics approach. *Evol. Bioinform. Online* 1, 51–61 (2005).
7. Pagliarini, D. J. et al. A mitochondrial protein compendium elucidates complex I disease biology. *Cell* 134, 112–123 (2008).
8. Avidor-Reiss, T. et al. Decoding cilia function: defining specialized genes required for compartmentalized cilia biogenesis. *Cell* 117, 527–539 (2004).
9. Enault, F., Suhre, K., Abergel, C., Poirot, O. & Claverie, J. M. Annotation of bacterial genomes using improved phylogenomic profiles. *Bioinformatics* 19 (Suppl. 1), i105–i107 (2003).
10. Drinnenberg, I. A. et al. RNAi in budding yeast. *Science* 326, 544–550 (2009).
11. Drinnenberg, I.A., Fink, G. R. & Bartel, D. P. Compatibility with killer explains the rise of RNAi-deficient fungi. *Science* 333, 1592 (2011).
12. Yelina, N. E. et al. Putative Arabidopsis THO/TREX mRNA export complex is involved in transgene and endogenous siRNA biosynthesis. *Proc. Natl Acad. Sci. USA* 107, 13948–13953 (2010).
13. Ketting, R. F. & Plasterk, R. H. A genetic link between co-suppression and RNA interference in *C. elegans*. *Nature* 404, 296–298 (2000).
14. Simmer, F. et al. Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. *Curr. Biol.* 12, 1317–1319 (2002).
15. Cui, M., Kim, E. B. & Han, M. Diverse chromatin remodeling genes antagonize the Rb-involved SynMuv pathways in *C. elegans*. *PLoS Genet.* 2, e74 (2006).
16. Parry, D. H., Xu, J. & Ruvkun, G. A whole-genome RNAi Screen for *C. elegans* miRNA pathway genes. *Curr. Biol.* 17, 2013–2022 (2007).
17. Thivierge, C. et al. Tudor domain ERI-5 tethers an RNA-dependent RNA polymerase to DCR-1 to potentiate endo-RNAi. *Nature Struct. Mol. Biol.* 19, 90–97 (2012).
18. Zhang, L. et al. Systematic identification of *C. elegans* miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2. *Mol. Cell* 28, 598–613 (2007).
19. Calvo, S. et al. Systematic identification of human mitochondrial disease genes through integrative genomics. *Nature Genet.* 38, 576–582 (2006).

20. Jansen, R. et al. A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science* 302, 449–453 (2003).
21. Hibbs, M. A. et al. Exploring the functional landscape of gene expression: directed search of large microarray compendia. *Bioinformatics* 23, 2692–2699 (2007).
22. Simonis, N. et al. Empirically controlled mapping of the *Caenorhabditis elegans* protein-protein interactome network. *Nature Methods* 6, 47–54 (2009).
23. Montgomery, T. A. et al. PIWI associated siRNAs and piRNAs specifically require the *Caenorhabditis elegans* HEN1 ortholog henn-1. *PLoS Genet.* 8, e1002616 (2012).
24. Bayne, E. H. et al. Splicing factors facilitate RNAi-directed silencing in fission yeast. *Science* 322, 602–606 (2008).
25. Pontes, O. & Pikaard, C. S. siRNA and miRNA processing: new functions for Cajal bodies. *Curr. Opin. Genet. Dev.* 18, 197–203 (2008).
26. Aravind, L., Watanabe, H., Lipman, D. J. & Koonin, E. V. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc. Natl Acad. Sci. USA* 97, 11319–11324 (2000).
27. Bernstein, D. A. et al. *Candida albicans* Dicer (CaDcr1) is required for efficient ribosomal and spliceosomal RNA maturation. *Proc. Natl Acad. Sci. USA* 109, 523–528 (2012).
28. Guang, S. et al. An Argonaute transports siRNAs from the cytoplasm to the nucleus. *Science* 321, 537–541 (2008).
29. Inglis, P. N., Blacque, O. E. & Leroux, M. R. Functional genomics of intraflagellar transport-associated proteins in *C. elegans*. *Methods Cell Biol* 93, 267–304, doi:S0091-679X(08)93014-4 [pii] 10.1016/S0091-679X(08)93014-4 (2009).
30. Kaplunovsky, A., Ivashchenko, A. & Bolshoy, A. Statistical analysis of exon lengths in various eukaryotes. *Open Access Bioinformatics* 2011:3, 1-15 doi:http://dx.doi.org/10.2147/OAB.S14448 (2011).
31. Koralewski, T. E. & Krutovsky, K. V. Evolution of exon-intron structure and alternative splicing. *PLoS One* 6, e18055, doi:10.1371/journal.pone.0018055 (2011).
32. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A* 96, 4285–4288 (1999).
33. Avidor-Reiss, T. et al. Decoding cilia function: defining specialized genes required for compartmentalized cilia biogenesis. *Cell* 117, 527–539, doi:S009286740400412X [pii] (2004).
34. Enault, F., Suhre, K., Abergel, C., Poirot, O. & Claverie, J. M. Annotation of bacterial genomes using improved phylogenomic profiles. *Bioinformatics* 19 Suppl 1, i105–107 (2003).
35. Pagliarini, D. J. et al. A mitochondrial protein compendium elucidates complex I disease biology. *Cell* 134, 112–123, doi:S0092-8674(08)00768-X [pii] 10.1016/j.cell.2008.06.016 (2008).
36. Pilpel, Y., Sudarsanam, P. & Church, G. M. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat Genet* 29, 153–159, doi:10.1038/ng724 ng724 [pii] (2001).
37. Tabach, Y. et al. The promoters of human cell cycle genes integrate signals from

- two tumor suppressive pathways during cellular transformation. *Mol Syst Biol* **1**, 2005 0022, doi:msb4100030 [pii] 10.1038/msb4100030 (2005).
38. Pasquinelli, A. E. *et al.* Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**, 86-89, doi:10.1038/35040556 (2000).
 39. Reinhart, B. J. *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901-906, doi:10.1038/35002607 (2000).
 40. Parry, D. H., Xu, J. & Ruvkun, G. A whole-genome RNAi Screen for *C. elegans* miRNA pathway genes. *Curr Biol* **17**, 2013-2022, doi:S0960-9822(07)02155-0 [pii] 10.1016/j.cub.2007.10.058 (2007).
 41. Zhou, R. *et al.* Comparative analysis of argonaute-dependent small RNA pathways in *Drosophila*. *Mol Cell* **32**, 592-599, doi:S1097-2765(08)00734-X [pii] 10.1016/j.molcel.2008.10.018 (2008).
 42. Zhang, L. *et al.* Systematic identification of *C. elegans* miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2. *Mol Cell* **28**, 598-613, doi:S1097-2765(07)00626-0 [pii] 10.1016/j.molcel.2007.09.014 (2007).
 43. Duchaine, T. F. *et al.* Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. *Cell* **124**, 343-354, doi:S0092-8674(05)01394-2 [pii] 10.1016/j.cell.2005.11.036 (2006).
 44. Kim, J. K. *et al.* Functional genomic analysis of RNA interference in *C. elegans*. *Science* **308**, 1164-1167, doi:1109267 [pii] 10.1126/science.1109267 (2005).
 45. Robert, V. J., Sijen, T., van Wolfswinkel, J. & Plasterk, R. H. Chromatin and RNAi factors protect the *C. elegans* germline against repetitive sequences. *Genes Dev* **19**, 782-787, doi:gad.332305 [pii] 10.1101/gad.332305 (2005).
 46. Cui, M., Kim, E. B. & Han, M. Diverse chromatin remodeling genes antagonize the Rb-involved SynMuv pathways in *C. elegans*. *PLoS Genet* **2**, e74, doi:10.1371/journal.pgen.0020074 (2006).
 47. Hibbs, M. A. *et al.* Exploring the functional landscape of gene expression: directed search of large microarray compendia. *Bioinformatics* **23**, 2692-2699, doi:btm403 [pii] 10.1093/bioinformatics/btm403 (2007).
 48. Simonis, N. *et al.* Empirically controlled mapping of the *Caenorhabditis elegans* protein-protein interactome network. *Nat Methods* **6**, 47-54 (2009).
 49. Calvo, S. *et al.* Systematic identification of human mitochondrial disease genes through integrative genomics. *Nat Genet* **38**, 576-582, doi:ng1776 [pii] 10.1038/ng1776 (2006).
 50. Jansen, R. *et al.* A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science* **302**, 449-453, doi:10.1126/science.1087361 302/5644/449 [pii] (2003).
 51. Grossman, S. R. *et al.* A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* **327**, 883-886, doi:science.1183863 [pii] 10.1126/science.1183863 (2010).

APPENDIX C: MORC family ATPases required for heterochromatin condensation and gene silencing

AUTHORS: Moissiard G, Cokus SJ, Cary J, Feng S, Billi AC, Stroud H, Husmann D, Zhan Y, Lajoie BR, McCord RP, Hale CJ, Feng W, Michaels SD, Frand AR, Pellegrini M, Dekker J, Kim JK, Jacobsen S

AUTHOR CONTRIBUTIONS: JKK conceived the experiment depicted in Figure 3D. ACB carried out the experiment.

CITATION: Moissiard G, Cokus SJ, Cary J, Feng S, **Billi AC**, Stroud H, Husmann D, Zhan Y, Lajoie BR, McCord RP, Hale CJ, Feng W, Michaels SD, Frand AR, Pellegrini M, Dekker J, Kim JK, Jacobsen S. MORC Family ATPases Required for Heterochromatin Condensation and Gene Silencing. *Science* 336: 1448-51 (2012).

ABSTRACT: Transposable elements (TEs) and DNA repeats are commonly targeted by DNA and histone methylation to achieve epigenetic gene silencing. We isolated mutations in two genes, *CRT1* and *CRH6*, which cause de-repression of DNA-methylated genes and TEs, but no losses of DNA or histone methylation. *CRT1* and *CRH6* are members of the conserved *Microrchidia* (*MORC*) ATPase family, predicted to catalyze alterations in chromosome superstructure. The *crt1* and *crh6* mutants show decondensation of pericentromeric heterochromatin, increased interaction of pericentromeric regions the rest of the genome, and transcriptional defects that are largely restricted to loci residing in pericentromeric regions. Knockdown of the single MORC homolog in *Caenorhabditis elegans* also impairs transgene silencing. We propose that the MORC ATPases are conserved regulators of gene silencing in eukaryotes.

MAIN TEXT

Gene silencing in the Arabidopsis genome is highly correlated with DNA methylation, which is found in three different cytosine contexts. Methylation of symmetric CG and CHG sites (in which H = A, T, or C) are mediated by DNA METHYLTRANSFERASE1 (MET1) and CHROMOMETHYLASE3 (CMT3), respectively, whereas CHH methylation is mainly catalyzed by DOMAINS REARRANGED METHYLTRANSFERASE2 (DRM2) (1). Silent loci are also enriched in the repressive histone H3 lysine 9 dimethylation mark (H3K9me2) (2, 3).

Suppressor of *drm2 cmt3* (*SDC*) is a gene whose repression in most tissues depends on the redundant activities of DRM2 and CMT3 (4, 5). Hence, a loss of *SDC* silencing is observed in the *drm2 cmt3* double mutant but not in *drm2* or *cmt3* single mutants. The *SDC* promoter carries seven tandem repeats, which recruit the DNA methylation machinery and cause transcriptional gene silencing. We engineered a green fluorescent protein (GFP)–based sensor construct controlled by the *SDC* promoter (fig. S1A). The *SDC::GFP* transgene behaves similarly to endogenous *SDC*, and GFP fluorescence is not detectable in wild-type, *drm2*, or *cmt3* plants but is highly expressed in *drm2 cmt3* double mutant (Fig. 1A).

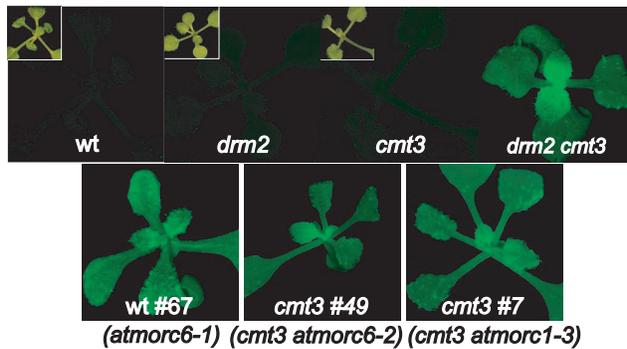
We carried out ethyl methanesulfonate (EMS) mutagenesis screens in wild-type (wt) or *cmt3* backgrounds for mutants showing *SDC::GFP* overexpression and identified the wt #67, *cmt3* #7, and *cmt3* #49 mutants (Fig. 1, A and B). Mapping experiments using bulk segregant analysis coupled to deep genome resequencing indicated that *cmt3* #7 contained a mutation in *At4g36290* (*AtMORC1*), previously also named *COMPROMISED RECOGNITION OF TCV-1* (*CRT1*) (6, 7), whereas wt #67 and *cmt3* #49 both contained mutations in *At1g19100* (*AtMORC6*) (7) (figs. S1B, S2, and S3A). An *atmorc1* allele was previously reported to show reduced resistance to the turnip crinkle virus (TCV) (6, 7), suggesting that *AtMORC1* is involved in viral resistance in addition to its role in gene silencing described in this study, whereas mutations in *AtMORC6* have not been described. To ensure that *atmorc1* and *atmorc6* mutations

were those responsible for the loss of *SDC* silencing, we isolated knock-out transferred DNA (T-DNA) insertion lines *atmorc1-4* and *atmorc6-3* and confirmed *SDC* overexpression in these two mutant alleles (fig. S3, B to D). Genetic complementation crosses between the recessive EMS and T-DNA mutants confirmed *AtMORC1* and *AtMORC6* as the mutated genes responsible for *SDC::GFP* activation in the three EMS lines (fig. S3E). Therefore, #7, #67, and #49 were renamed *atmorc1-3*, *atmorc6-1*, and *atmorc6-2*, respectively.

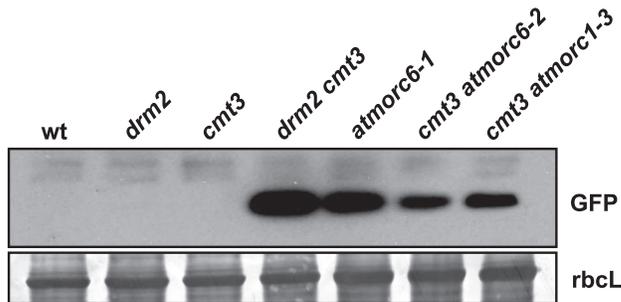
By using RNA sequencing (RNA-seq) (8), we found that the majority of RNAs significantly affected in the *atmorc1* and *atmorc6* mutants showed up-regulation, and many of these were transposable elements (TEs) belonging to various transposon superfamilies, including, among others, the LTR/Gypsy, LTR/Copia, DNA/MuDR, and DNA/Harbinger families (Fig. 1, C and D; fig. S4A; table S1). The expression defects in the *atmorc1* and *atmorc6* mutants were very similar, with all but two of the transposons up-regulated in *atmorc1* also up-regulated in *atmorc6* (fig. S4B). Protein-coding genes overexpressed in the *atmorc1* and *atmorc6* EMS and T-DNA mutants included endogenous *SDC* (table S2). There was a high degree of overlap between the genes up-regulated in *atmorc1* and *atmorc6* (fig. S4C), most of them corresponding to DNA-methylated and silenced loci (fig. S4, D and E). We also performed RNA-seq in the *atmorc1 atmorc6* double mutant and found a very similar set of genes and transposons up-regulated, with only a few genes up-regulated in the double mutant that were not up-regulated in each of the single mutants (table S3), suggesting that *AtMORC1* and *AtMORC6* may act together to enforce gene silencing.

Fig. C.1. Mutations of two MORC homologs induce *SDC::GFP* and TE overexpression. **(A)** wt, *drm2* mutant, and *cmt3* mutant plants carrying *SDC::GFP* showed no GFP fluorescence under ultraviolet (UV) light (insets show each plant under white light), and *drm2 cmt3* double mutant and EMS-mutagenized lines wt #67, *cmt3* #49, and *cmt3* #7 plants showed strong GFP fluorescence. **(B)** Western blot using antibody against GFP (anti-GFP) confirms *SDC::GFP* overexpression in the EMS mutants. Coomassie staining of the large Rubisco subunit (*rbcl*) is used as loading control. **(C)** Number of TEs overexpressed in *atmorc1* and *atmorc6* mutants and classified by superfamily. For each mutant, only TEs with at least a fourfold increase in both the EMS and T-DNA alleles over wt and with a $P \leq 0.05$ are represented. **(D)** Relative fold increase of four TE transcripts in *atmorc1-4* and *atmorc6-3* over wt assayed by real-time quantitative polymerase chain reaction (RT-qPCR) and normalized to *ACTIN7*. Errors bars indicate standard deviation based on three independent biological replicates.

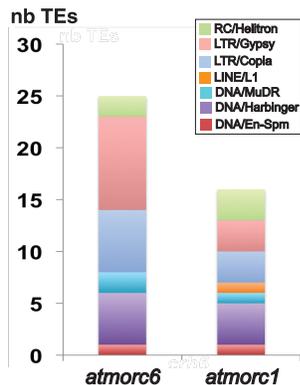
A



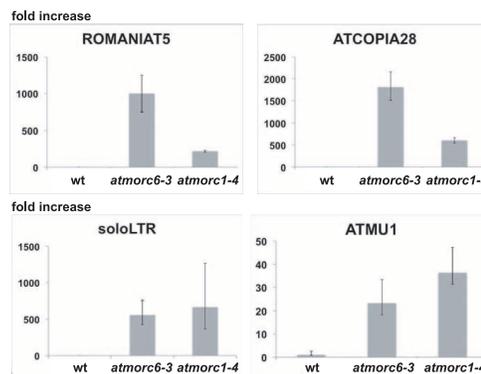
B



C



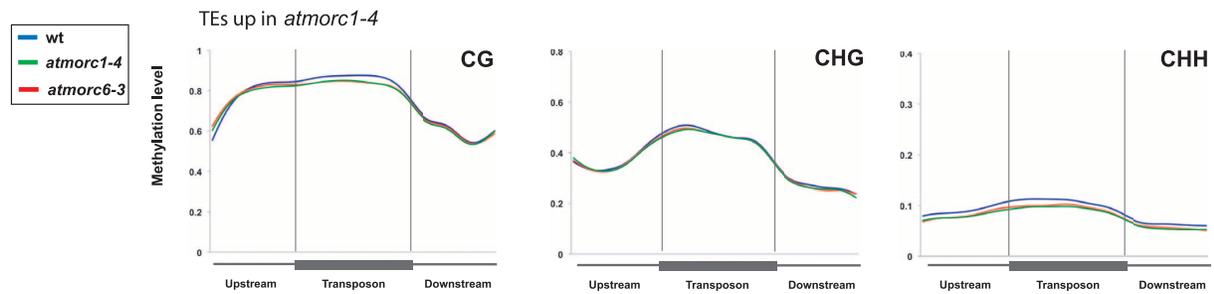
D



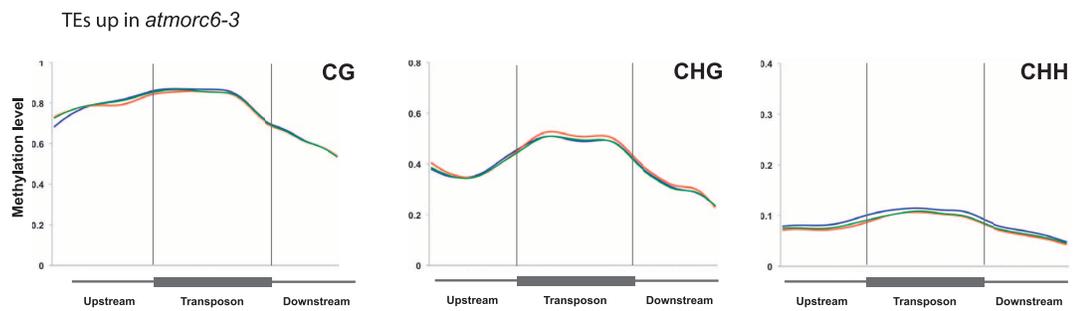
Whole-genome bisulfite sequencing (BS-seq) (9) revealed that DNA methylation levels in all sequence contexts were unaltered in *atmorc1* or *atmorc6* relative to wild type at TEs up-regulated in *atmorc1* or *atmorc6* (Fig. 2, A and B), nor were there any bulk alterations in protein-coding genes or TEs in the genome (fig. S5, A and B). In addition, analyses at the pericentromeric satellite CEN180 repeats and five loci up-regulated in *atmorc1* and *atmorc6* showed that the DNA methylation patterns in *atmorc1-4* and *atmorc6-3* were similar to those of wild type (Fig. 2, C and D). Chromatin immunoprecipitation sequencing (ChIP-seq) analyses of H3K9me2 also did not reveal any changes in the *atmorc1* or *atmorc6* mutants at *SDC* or other up-regulated locations (fig. S6, A and B). Lastly, small RNA sequencing analyses showed that elements up-regulated in *atmorc1* and *atmorc6* mutants were enriched in small interfering RNAs (siRNAs), but these siRNA levels did not change in the mutants (fig. S7). Thus, *AtMORC1* and *AtMORC6* are not required to maintain DNA methylation, H3K9me2, or siRNAs, suggesting that AtMORC1 and AtMORC6 are likely to either act downstream of DNA methylation or enforce silencing by a novel mechanism.

Fig. C.2. DNA methylation is not impaired in *atmorc1* and *atmorc6* mutants. **(A and B)** Metaplot analyses show DNA methylation level in *atmorc1-4*, *atmorc6-3*, and wt for the set of TEs up-regulated in *atmorc1-4* (A) and *atmorc6-3* (B). The gray vertical lines mark the boundaries between 1 kilobase upstream and downstream regions of TEs. **(C)** Southern blot analyses assayed CG methylation level at CEN180 repeats by using *Hpa*II-treated genomic DNAs. m, methylated; u, unmethylated. *met1-3* genomic DNA is used as positive control for loss of CG methylation (23). **(D)** Percent DNA methylation at SDC and four TEs overexpressed in *atmorc1-4* and *atmorc6-3* mutants assayed by bisulfite sequencing. Twenty-four clones were analyzed for each individual analysis.

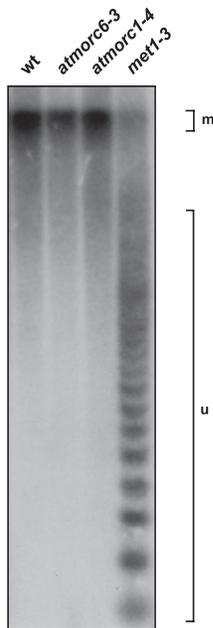
A



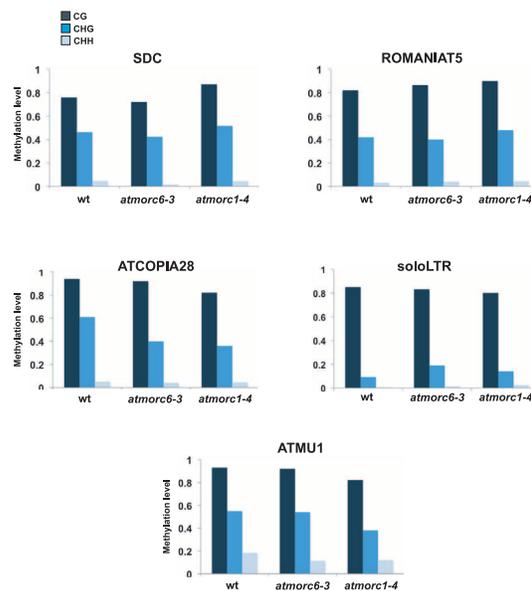
B



C



D

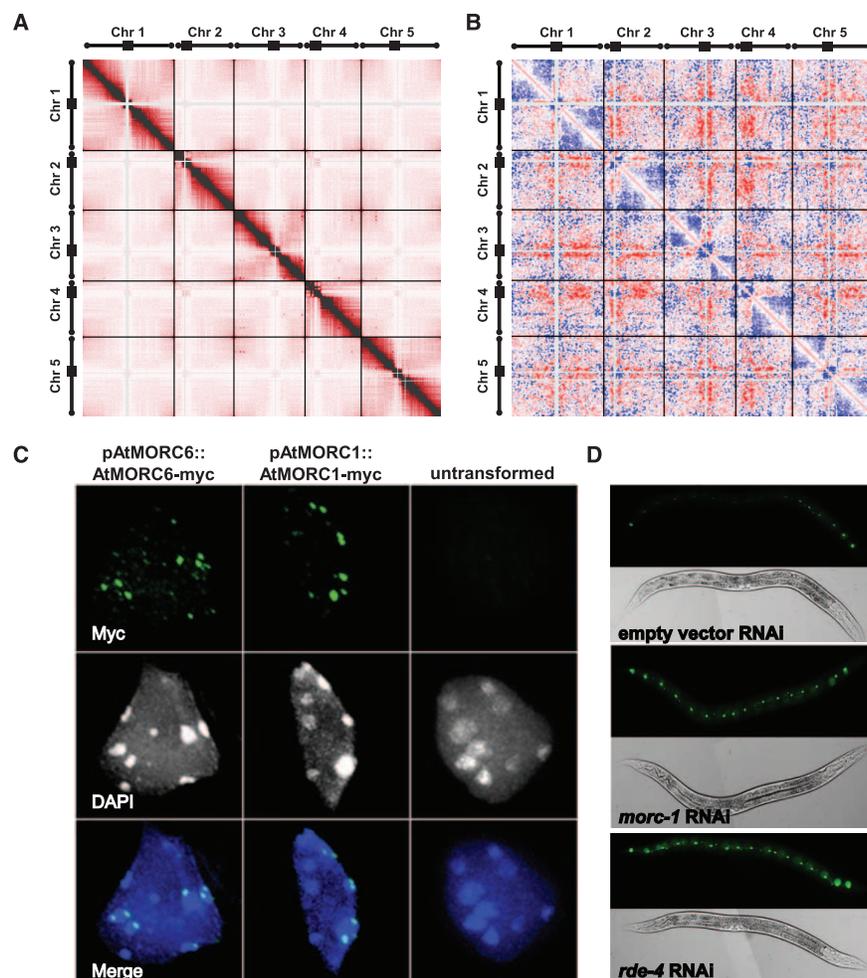


AtMORC1 and *AtMORC6* are homologs of mouse *Microrchidia1* (*MORC1*) (10, 11) and contain gyrase, Hsp90, histidine kinase, and MutL (GHKL) and S5 domains, together comprising an adenosine triphosphatase (ATPase) module (6) in addition to a putative C-terminal coiled-coil domain (fig. S1B). The EMS mutations found in *atmorc1-3*, *atmorc6-1*, and *atmorc6-2* alleles all introduced premature stop codons within the GHKL domain (fig. S1B). Because of the similarity of *AtMORC1* and *AtMORC6* to ATPases involved in manipulating chromatin superstructure (12), these proteins may affect gene silencing through higher-order compaction of methylated and silent chromatin. In wild-type nuclei, pericentromeric heterochromatin forms densely staining nuclear bodies called chromocenters that localize to the nuclear periphery (13). We observed decondensation of chromocenters in the *atmorc1* and *atmorc6* mutants (as well as in *atmorc1 atmorc6* double mutant) (figs. S8 to S11) and found that loci transcriptionally derepressed in the mutants mostly localized to pericentromeric heterochromatin (fig. S12 and tables S1 and S3). To directly examine whole-genome chromatin interactions, we performed Hi-C analyses in wild type and *atmorc6-1* (14). Consistent with previous cytological studies (13), the wild-type genome showed interactions between telomeres as well as between euchromatic regions on the same chromosome arm (Fig. 3A). In contrast, pericentromeric heterochromatin regions interacted very weakly with the rest of the genome, consistent with their compaction in chromocenters (Fig. 3A). Although *atmorc6-1* showed a roughly similar chromatin architecture (fig. S13), plotting the differences between mutant and wild type showed that *atmorc6-1* shows an increase in interactions between the pericentromeric regions of all chromosomes with the euchromatic arms of all chromosomes and a corresponding depletion of interactions of euchromatic arms with themselves. Because the analysis reports relative changes with the sum of differences set to zero, the most likely interpretation of these findings is that pericentromeric regions interact more strongly with the euchromatic arms in *atmorc6-1*, although we cannot exclude that the mutant also has effects on the euchromatic arms (Fig. 3B). This interpretation is consistent with the cytological observations showing that chromocenters expand out into a larger area of the nucleus in the mutants (fig. S8). We also found, by using complementing myc-tagged transgenes, that *AtMORC1* and *AtMORC6* proteins formed small nuclear bodies that were usually adjacent to but not

within chromocenters (Fig. 3C and figs. S14 and S15). These results are all consistent with a model in which AtMORC1 and AtMORC6 enforce compaction and gene silencing of pericentromeric heterochromatin, although it is also possible that changes in chromatin and gene expression in the mutant secondarily lead to the observed changes in chromatin compaction. Mutation of the plant-specific *MOM1* gene has also been shown to affect gene silencing but not DNA methylation in *Arabidopsis*; however, *mom1* mutants do not show chromocenter decondensation and therefore are likely to act via a different mechanism (15, 16).

A single MORC homolog, *morc-1*, is present in the worm *Caenorhabditis elegans*, which is devoid of DNA methylation (17). To test whether the *C. elegans morc-1* (ZC155.3) is involved in gene silencing, we performed RNA interference (RNAi)-mediated knockdown of *morc-1* in the *eri-1* sensitized background, in which a GFP transgene is silenced in most of the worm seam cells (Fig. 3D) (18). *morc-1*-depleted worms showed GFP reactivation similar to worms depleted of *rde-4*, an essential component of gene silencing in *C. elegans* (Fig. 3D) (19). These results suggest that MORCs may play an ancient and conserved role in gene silencing. In addition, the observation that *morc-1* is required for gene silencing in *C. elegans* reinforces our view that MORCs in *Arabidopsis* are enforcing silencing by a mechanism that may not be directly linked with DNA methylation. It is interesting to note that the phenotype of the *Morc1*-knockout mouse resembles *Miwi2*- and *Dnmt3L*-knockout mouse phenotypes, showing male-specific meiotic defects during spermatogenesis (10, 20–22). *Miwi2* and *Dnmt3L* are both required for TE silencing, and it is possible that *Morc1* might be involved in transposon silencing in mammals as well. We propose that MORC family ATPases act to regulate chromatin architecture and gene silencing in a wide variety of eukaryotes.

Fig. C.3. AtMORC1 and AtMORC6 are required for maintenance of chromatin architecture and form nuclear bodies near chromocenters, and *morc-1* is involved in gene silencing in *C. elegans*. **(A)** Interaction matrix of the wt *Arabidopsis* genome from Hi-C analysis. Positions along the five chromosomes are shown from left to right and top to bottom, and each pixel represents interactions from uniquely mapping paired end reads in 200-kilobase bins. Black bars and circles mark the positions of the pericentromeric and telomeric regions, respectively. Light gray regions represent areas masked out because of problematic mapping. Black bars show separation between chromosomes. **(B)** Difference plot shows enrichment of Hi-C interactions in *atmorc6-1* in red and interactions depleted in *atmorc6-1* in blue. **(C)** Anti-Myc immunostaining showing localization of pAtMORC6::AtMORC6-Myc and pAtMORC1::AtMORC1-Myc in nuclear bodies adjacent to chromocenters. AtMORC1 and AtMORC6 showed 2.0 T 1.0 (average T standard deviation) and 2.5 T 1.2 bodies per chromocenter, respectively. DAPI (4',6-diamidino-2-phenylindole) staining shows chromocenter location. Bottom images are merges. **(D)** A silenced seam cell-specific GFP transgene in the *eri-1(mg366)* sensitized background is overexpressed in worms fed with bacteria expressing double-stranded RNA targeting *morc-1* or *rde-4* but not in worms fed with bacteria expressing a control empty vector. Results are representative of five independent replicates.



ACKNOWLEDGMENTS

We thank M. Akhavan for sequencing, L. Goddard and L. Iruela-Arispe for assistance with confocal microscopy, and P. Fransz and I. Schubert for helpful discussions. S.F. is a Special Fellow of the Leukemia and Lymphoma Society. J.C. is supported by the Ruth L. Kirschstein National Research Service Award GM007185. R.P.M. is supported by the National Institute of General Medical Sciences (grant F32GM100617), and J.D. is supported by a W. M. Keck Foundation Distinguished Young Scholar in Medical Research Award. Research in the Jacobsen, Kim, Michaels, and Dekker laboratories was supported by NIH grants GM60398, GM088565, GM075060, and HG003143, respectively. Sequencing files have been deposited at Gene Expression Omnibus (GEO) (accession code GSE37644). The authors declare no competing financial interests. S.E.J. is an investigator of the Howard Hughes Medical Institute. Correspondence and requests for materials should be addressed to S.E.J.

MATERIALS AND METHODS

Plant material and growing conditions

All mutants are in the Columbia (Col) ecotype. *atmorc6-3* (GK_599B06) and *atmorc1-4* (SAIL_1239_C08) T-DNA lines were obtained from GABI-Kat (24) at University of Bielefeld, Germany and ABRC at Ohio State University, respectively. *drm2-2* (SALK_150863.37.35) and *cmt3-11* (SALK_148381) were previously described (4). T-DNA insertions were confirmed by PCR-based genotyping. Primer sequences are described in Table S4. *Arabidopsis* plants were grown under continuous light.

Cloning of *SDC::GFP*

NLS-GFP-35S terminator was PCR amplified and cloned into pCambia3300. The *SDC*

promoter corresponding to a region of ~2.4 kb upstream of *SDC* transcriptional start site was PCR amplified from wild type genomic DNA and cloned into pCR2.1 TOPO vector (Invitrogen). Quick change site-directed mutagenesis (Stratagene) was performed to create a polymorphism (NlaIII -> BamHI) within the *SDC* promoter, which was subsequently mobilized into pCambia3300 upstream of NLS-GFP sequence. *drm2 cmt3* double mutant plants were transformed with the *SDC::GFP* construct using the *Agrobacterium-mediated* floral dip method (25). Transgenic plants showing strong GFP fluorescence were backcrossed with a wild type plant to ensure proper silencing of *SDC::GFP* in the F1 generation. F1 plants were self-crossed and their progenies (F2) were screened for GFP fluorescence and PCR-based genotyped to obtain the following genetic backgrounds: *SDC::GFP* wt, *SDC::GFP drm2*, *SDC::GFP cmt3* and *SDC::GFP drm2 cmt3*. Primer sequences used for *SDC::GFP* cloning are described in Table S4.

Cloning of *pAtMORC1::AtMORC1-Myc* and *pAtMORC6::AtMORC6-Myc*

AtMORC1 and *AtMORC6* genomic regions were PCR amplified and the Myc epitope was added to the C-terminus of each protein as previously described (26). In both cases, the amplified region includes a ~1Kb promoter sequence upstream of the respective transcriptional start site. Primer sequences are described in Table S4.

EMS mutagenesis, GFP screening and mapping analyses

Two thousand seeds from *SDC::GFP* wt and *SDC::GFP cmt3* lines were mutagenized in 0.3% EMS solution for 13 hours with rotation. Seeds were subsequently washed with water and planted onto soil. For each background, approximately one thousand M2 populations were collected and subsequently screened for GFP fluorescence under UV light using a Leica MZ16F Fluorescence Stereomicroscope coupled with the GFP Plus fluorescence filter. Pictures were taken using the DFC300 FX digital camera kit.

Mapping and identification of the three EMS mutations responsible for the phenotypes were performed by bulk segregant analysis coupled with deep genome re-sequencing as previously described (27), using single nucleotide polymorphisms (SNPs) between the Landsberg (Ler) and Col ecotypes derived *de novo* from data from a large number of mapping crosses.

Western Blotting

Western blots against GFP were performed using the GFP-specific antibody (Invitrogen, AA1122). Western blots against Myc were performed as previously described (26).

RNA analyses

Total RNAs were extracted from two-week-old seedlings using Trizol (Ambion RNA technology). Two μg of total RNAs were subsequently used to generate libraries for High Throughput RNA sequencing (TruSeq RNA, Illumina) per manufacturer instruction. For RNA-seq analyses, sequencing reads were mapped with Bowtie (28) allowing up to 2 mismatches. Gene and transposon expression was measured by calculating reads per kilobase per million mapped reads (RPKM) (29). *p-values* were calculated using Fisher's exact test and Benjamini corrected for multiple testing (30). Differentially expressed elements in wild type and mutants were defined by applying $\log_2(\text{mutant}/\text{wild type}) > 2$ and $P < 0.05$ cutoffs. For quantitative PCR analysis, total RNAs were converted into cDNA using SuperScript III Reverse Transcriptase (Invitrogen) per manufacturer instructions. Quantitative PCR was carried out using SyBr Green PCR mastermix (Roche) and gene- or transposon-specific primers (see Table S4) on a Stratagene Mx real-time thermocycler.

DNA methylation analyses

Whole genome BS-seq libraries were performed as previously described (9), except directly with pre-methylated final adapters. BS-seq data was mapped with BS seeker as previously described (31). For traditional bisulfite sequencing, genomic DNA extracted from two-week-old seedlings was bisulfite converted using MethylEasy (Human Genetic Signature) and processed as previously described (5). Primer sequences used for bisulfite sequencing are described in Table S4. Southern blot was performed as previously described (2).

H3K9me2 ChIP-seq analyses

Two grams of 3-week-old seedlings were crosslinked with formaldehyde and chromatin-IP experiments were performed as previously described (32). A mouse monoclonal antibody was used for H3K9me2 immunoprecipitation (Abcam ab1220). ChIP-seq library was generated per manufacturer instructions (Illumina). Demultiplexed (by exact match to canonical 6-mers) single end 50-mer HiSeq PF-passing reads were aligned to the TAIR8 reference genome using Bowtie 1, keeping all hits with at most two or fewer mismatches in the first 28 cycles and with total sum of Phred quality scores at mismatches up to 100, further filtered to only keep reads with a unique hit of fewest total mismatches, retaining only that unique hit. Reads were extended downstream to total length 220 nucleotides to reflect nominal library fragment lengths and single-stranded per-base pair coverage tallied, normalized by total nuclear chromosome coverage to account for variation in sequencing depth.

siRNA analyses

Total RNAs were extracted from flowers using Trizol (Ambion RNA technology) and siRNAs were purified as previously described (33) with the following modifications. Polyacrylamide gel-excised siRNAs were eluted in 0.3M NaCl overnight at 4°C. Gel debris were filtered using 5µm Filter tubes (IST Engineering Inc) and, ethanol-

precipitated siRNAs were resuspended in 5ul of nuclease-free H₂O to subsequently generate libraries for High Throughput small RNA sequencing (TruSeq small RNA, Illumina) per manufacturer instruction. Small RNA libraries were Illumina sequenced to 50bp length, and resulting reads were trimmed for adapter sequences and then aligned to the TAIR8 genome using Bowtie (28).

Hi-C analyses

Two grams of 3-week-old seedling leaves were crosslinked with formaldehyde as previously described (32). Hi-C experiments were performed as previously described (34), with the exception that plant nuclei were prepared following a previously published *Arabidopsis* ChIP protocol (32). Hi-C libraries were sequenced on a HiSeq 2000 sequencer (Illumina) obtaining paired end 50+50 nucleotide reads. Sequencing reads were mapped to the TAIR8 *A. thaliana* reference genome using Bowtie 1 to obtain all zero-mismatch hits of ends independently and keeping only paired end read pairs with each end having exactly one hit, obtaining 21,379,391 wild type and 14,815,038 *atmorc6-1* pairs. Paired reads with ends aligning to the same HindIII fragment were discarded. Hi-C interaction counts were summed within disjoint symmetric 2-D bins 200 kilobase pairs tiling the genome. HindIII fragments which overlapped regions of poor reference genome quality were excluded from the analysis. Genomic 1-D bins (rows and columns) in which > 50% of the sequence length was excluded by these filters were treated as missing data, excluded from further analyses, and appear in figures as empty regions. The “raw” coverage of each genomic 2-D bin was taken as the number of paired end reads lying in that bin. Whole genomic 1-D bins with a total coverage more than 3 standard deviations greater than or less than the mean were excluded and then the matrix of interactions was corrected for 1-D bin coverage variation as previously described (34) using 50 iterations of that procedure. The comparison between the wild type and *atmorc6-1* mutant was expressed as the difference divided by the mean within each bin with smoothing plus or minus one bin.

Immunofluorescence

Immunofluorescence experiments examining chromocenter condensation were performed as previously described (35) with the following modifications. Leaves from three-week-old plants were fixed in 4% paraformaldehyde in TRIS buffer (10mM TRIS pH 7.5, 10mM EDTA, 100mM NaCl) for 20 minutes and washed twice in TRIS buffer. Leaves were chopped in 400 microliters lysis buffer (15mM TRIS pH 7.5, 2mM EDTA, 0.5mM spermine, 80mM KCl, 20mM NaCl, 0.1% Triton X-100) and filtered through a 35 micron cell strainer. Five microliters of nuclei suspension was added to sorting buffer (100mM TRIS pH 7.5, 50mM KCl, 2mM MgCl₂, .05% Tween-20, 20.5% sucrose) and air dried on microscope slides for two hours and then post-fixed in 4% paraformaldehyde in PBS for 20 minutes. Slides were washed three times in PBS and incubated in blocking buffer (3% BSA, 10% horse serum in PBS) for 30 minutes at 37°C. Nuclei were incubated at 4°C overnight in mouse monoclonal antibody against H3K9me₂ (Abcam ab1220; 1:200). Slides were washed in PBS and incubated with goat anti-mouse FITC antibody (Abcam ab7064; 1:200) for 90 minutes at room temperature. Following PBS washes, nuclei were counterstained and mounted in Vectashield mounting media with DAPI (Vector H-1200). Nuclei were analyzed with a Zeiss Axio Imager Z1 microscope at 100X magnification and images were captured with a Hamamatsu ORCA-ER Camera. For detection of Myc epitope tagged proteins, nuclei were isolated as above. Following preparation of nuclei suspension, nuclei were spun down for 2 minutes at 2,000rpm and resuspended in PBS. Blocking and antibody incubations were performed in suspension, followed by pellet washing with PBS. Myc epitope was detected with mouse monoclonal antibody (Abcam 9E10; 1:200) and goat anti-mouse FITC (abcam ab7064; 1:200). Two microliters of prepared nuclei were mounted in Vectashield media. Nuclei were analyzed with the Applied Precision DeltaVision DV Live Cell Imaging System using Olympus IX-71 Customized Inverted Microscope and Photometrics CoolSNAP HQ2 CCD Camera (Figure 3). Additional images of nuclei were taken using a Zeiss LSM 510 META confocal microscope with Axiocam camera (Figure S15).

RNA interference in *C. elegans*

RNAi experiments were carried out as reported previously (36) using the *eri-1(mg366); [wls54(scm::gfp)]* strain, which shows increased sensitivity to RNAi. Briefly, bacterial strains carrying plasmids expressing double-stranded RNA targeting *morc-1* or *rde-4* were obtained from the Ahringer RNAi library (37). Hatched L1 *eri-1(mg366); [wls54(scm::gfp)]* larvae were cultured on empty vector (L4440), *morc-1*, or *rde-4* RNAi bacteria for two generations at 22.5°C. Images of F1 L4 larvae were captured on an Olympus BX61 epifluorescence compound microscope with a Hamamatsu ORCA ER camera using Slidebook 4.0.1 digital microscopy software (Intelligent Imaging Innovations) and processed using ImageJ.

SUPPLEMENT

Fig. C.S1. (A) Schematic representation of the *SDC::GFP* construct. The *SDC* promoter carries seven tandem repeats (black arrows) targeted by DNA methylation. The red bar corresponds to the *Simian Virus 40* (SV40) Nuclear Localization Signal fused to GFP. **(B)** Schematic of AtMORC1 and AtMORC6 proteins showing the GHKL and S5 ATPase domains together with putative Coiled-Coil (CC) domains. The location of nonsense mutations within the protein sequences is shown for each EMS allele. aa; amino acid. Based on contextual associations of prokaryotic and eukaryotic MORC family members with associated domains, as well as with genes in operons, eukaryotic MORCs were predicted to have originated from bacterial restriction modification systems, and in eukaryotes have been proposed to remodel chromatin superstructure in response to epigenetic signals such as histone and DNA methylation (12). For instance, topoisomerases and MutL (a factor involved in methylated DNA directed mismatch repair) proteins use ATP hydrolysis to mediate large movements and looping of DNA. Similar to other GHKL ATPases (38), MORC3 in mouse was shown to act as a molecular clamp, interacting with itself constitutively through its coiled-coil domain, and also interacting via its ATPase domain in an ATP-dependent manner (39). In this way, MORC protein architectures are also reminiscent of the structural maintenance of chromosomes (SMC) family of proteins, which control chromosome condensation and cohesion (40). In addition, a very distant class of ATPase homologs contain the GHKL domain fused with the hinge and coiled coil domains of SMC-like ATPases (12) one of which, SmcHD1, is required for maintenance of X chromosome inactivation in mouse (41). Interestingly, in *Arabidopsis*, the protein DMS3/IDN1, which contains an SMC-like Hinge domain, has also been involved in gene silencing (42, 43) and GMI1, which is a GHKL protein carrying a Hinge domain has recently been involved in DNA repair (44).

A



B

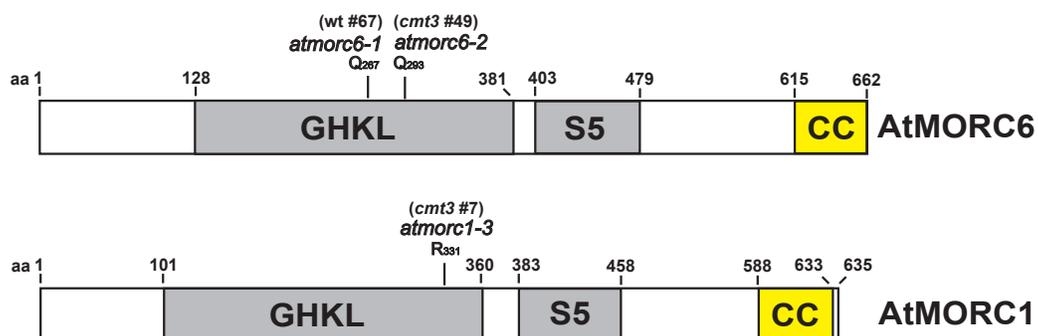


Fig. C.S2. Mapping of *cmt3* #7, wt #67 and *cmt3* #49 mutations by bulk segregant analysis coupled with whole genome re-sequencing (27). **(A to C)** Depletion in percentage of Landsberg (Ler) single nucleotide polymorphisms (SNPs), defining the linkage interval for the population *cmt3* #7 (chromosome 4 at ~17-18 megabases) (A), and defining the linkage intervals for the populations wt #67 (B) and *cmt3* #49 (C) (chromosome 1 at ~6-7 megabases). Red arrows mark the linkage intervals.

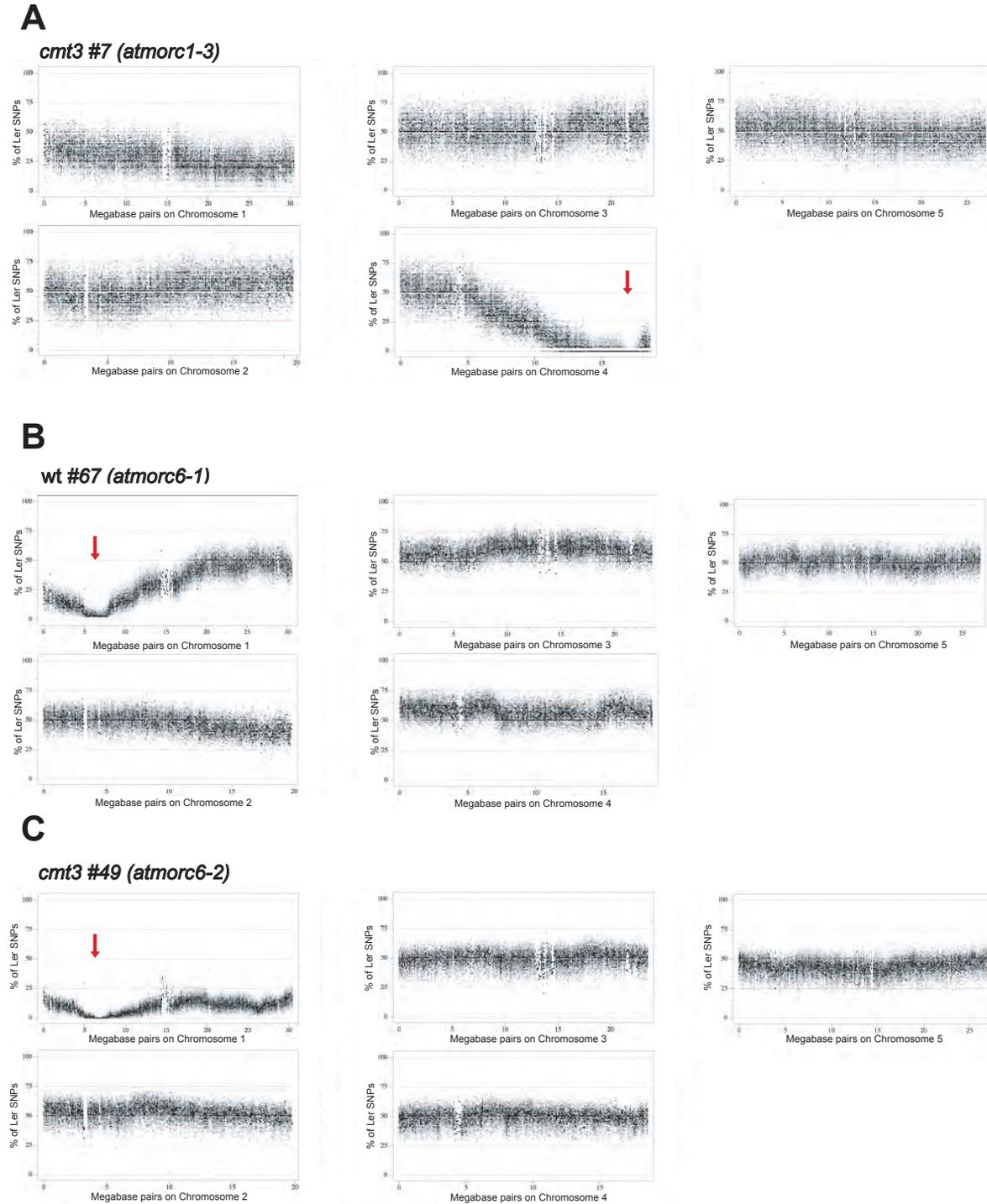


Fig. C.S3. Characterization of *atmorc1* and *atmorc6* EMS and T-DNA mutant alleles. (A) Location of EMS point mutations in *atmorc1-3*, *atmorc6-1*, and *atmorc6-2* alleles confirmed by traditional DNA sequencing. *AtMORC1* and *AtMORC6* reference sequences are shown for comparison. bp, base pair. (B) Gene structures of *AtMORC1* and *AtMORC6* showing exons (E, dark gray boxes) and introns (gray lanes). Light gray boxes correspond to 5' and 3' untranslated regions. The locations of EMS point mutations and T-DNA insertions are described above the gene structures. (C and D) Relative *AtMORC1* and *AtMORC6* RNA levels (C), and *SDC* RNA level (D) in *atmorc1-4* and *atmorc6-3* over wild type assayed by RT-qPCR and normalized to *ACTIN7*. Errors bars indicate standard deviation based on three independent biological replicates. Red arrows shown in (B) correspond to the primer locations within *AtMORC1* and *AtMORC6* mRNAs used in (C). (E) Genetic complementation tests and backcrosses showing that wt #67, *cmt3* #49, and *atmorc6-3* are three recessive allelic mutations in *AtMORC6* (top panels), while *cmt3* #7 and *atmorc1-4* are two recessive allelic mutations in *AtMORC1* (two bottom left panels). The cross between wt #67 and *cmt3* #7 confirms that these two mutations are non-allelic (bottom right panel). Pictures represent leaves from F1 plants observed under UV light for GFP fluorescence.

Fig. C.S4. Similar sets of TEs and protein-coding genes are upregulated in *atmorc1* and *atmorc6*. **(A)** Relative fold increase of *SDC* and four TE transcripts in *atmorc6-1*, *cmt3-11* and *cmt3-11 atmorc1-3* over wild type assayed by RT-qPCR and normalized to *ACTIN7*. Errors bars indicate standard deviation based on three independent biological replicates. **(B-D)** Venn diagrams showing overlap of upregulated TEs (B), upregulated protein-coding genes (C), and upregulated protein-coding genes associated with DNA methylation (D) in *atmorc1* and *atmorc6*. These analyses include the genes that were 4-fold up-regulated in RNA-seq experiments in both the EMS and T-DNA alleles. **(E)** Gene ontology (GO) analyses of genes upregulated in *atmorc1* and *atmorc6* showing no significant over-representation of any GO category in both *MORC* mutants. GO *p*-values are shown in parenthesis.

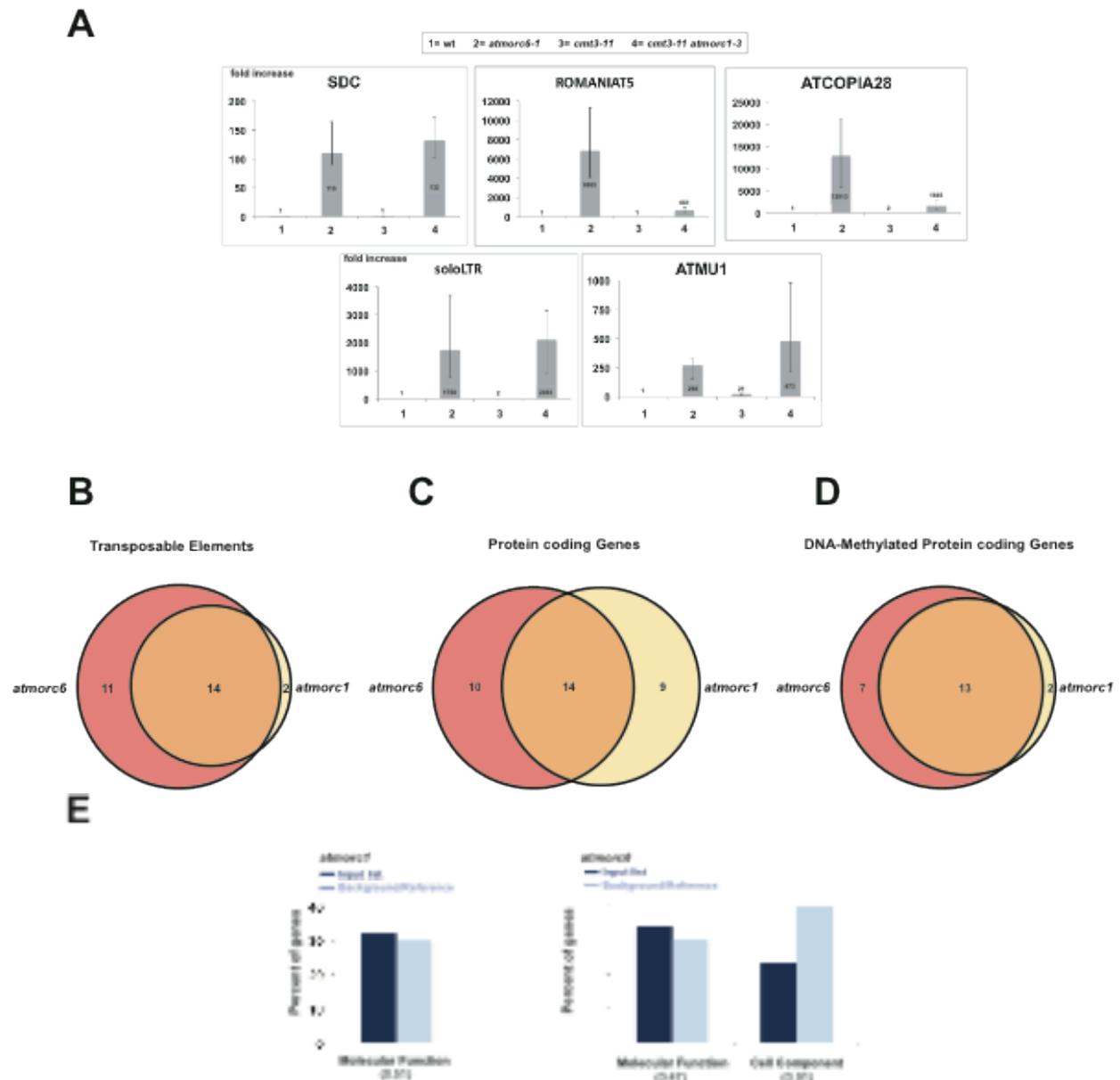
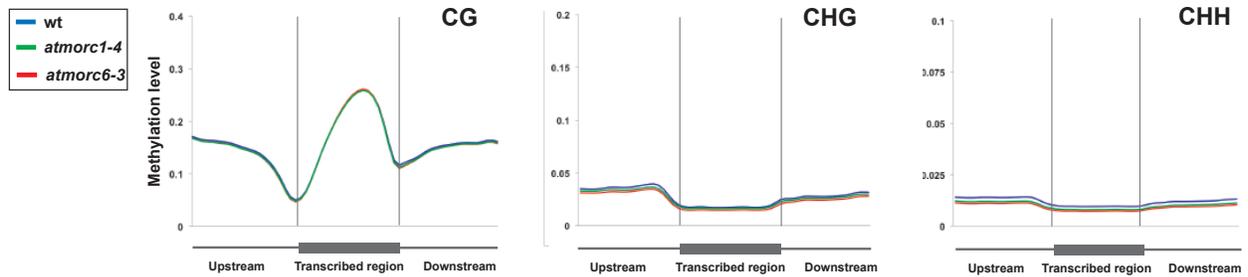


Fig. C.S5. DNA methylation at all protein-coding genes and transposons is not altered in *atmorc1* and *atmorc6*. (**A** and **B**) Metaplot analyses showing DNA methylation percent within protein-coding genes (A) and transposons (B) in *atmorc1-4*, *atmorc6-3* and wild type plants. The gray vertical lines mark the boundaries between 1 kilobase (Kb) upstream regions and gene bodies or transposons (left), and between gene bodies or transposons and 1Kb downstream regions (right).

A



B

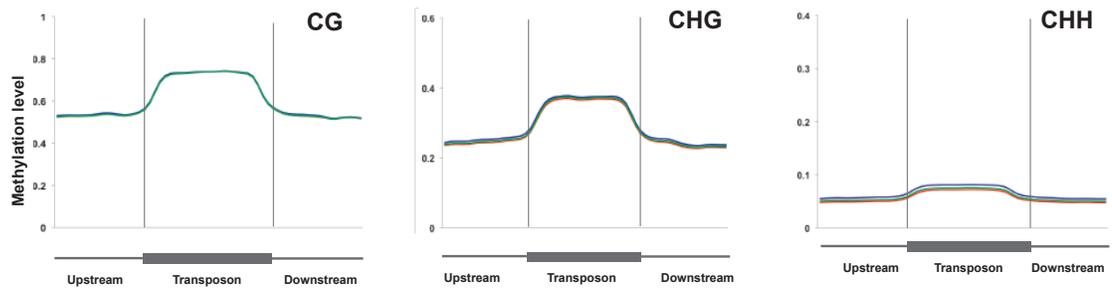


Fig. C.S6. H3K9me2 is not altered in *atmorc1* and *atmorc6*. **(A)** log₂ ratios of H3K9me2/H3 in *atmorc1-4* and *atmorc6-3* over wild type do not show H3K9me2 changes in *atmorc6* or *atmorc1* mutants at the *SDC* promoter region. The H3K9me2/H3 wild type ratio is shown over the *SDC* promoter region in which H3K9me2 is enriched. RNA-seq reads in the different genetic backgrounds are shown to define the *SDC* transcribed region. **(B)** Metaplot analyses showing log₂ ratios of H3K9me2/H3 in *atmorc1-4* and *atmorc6-3* over wild type at the set of transposable elements and DNA-methylated genes that are upregulated in *atmorc1-4* and *atmorc6-3*. The gray vertical lines mark the boundaries between 1 kilobase upstream regions and transposons/PCGs (left) and between transposons/PCGs and 1 kilobase downstream regions (right). PCGs, Protein-coding genes.

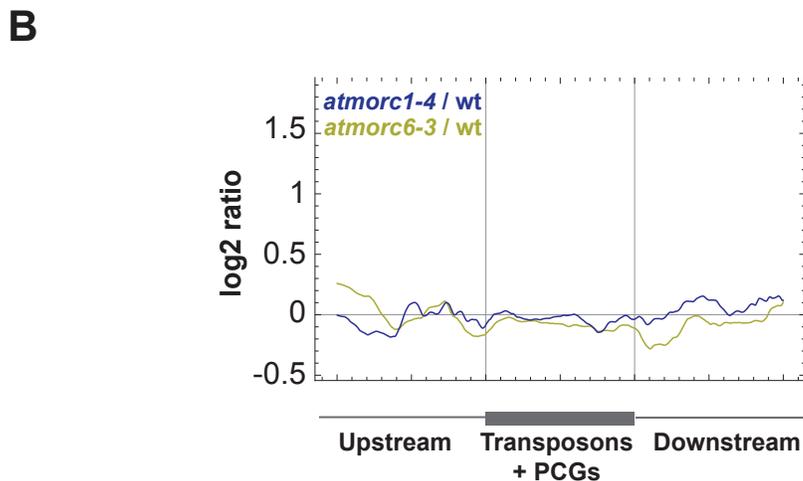
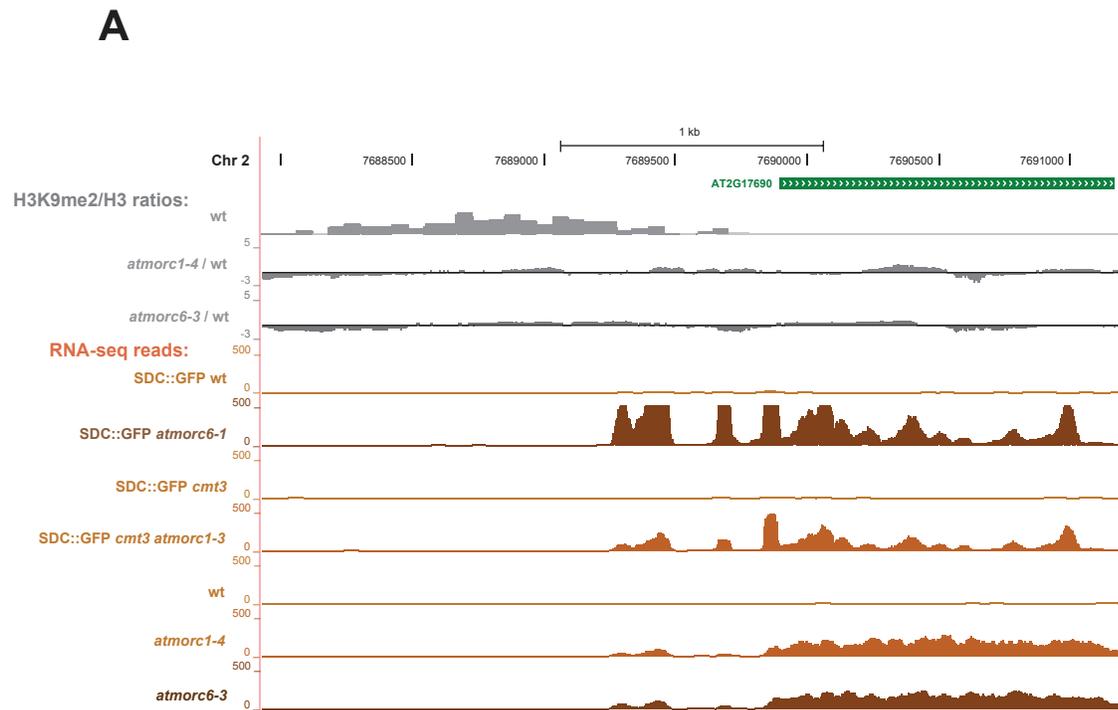


Fig. C.S7. Small RNA accumulation is unaltered in *atmorc1* and *atmorc6*. Metaplot analyses showing no difference in the level of siRNAs in *atmorc1-4* and *atmorc6-3* compared to wild type at the set of transposable elements that are upregulated in *atmorc1-4* and *atmorc6-3*. The gray vertical lines mark the boundaries between 2 kilobases upstream regions and transposons (left) and between transposons and 2 kilobases downstream regions (right). Data is illustrated for different size classes of small RNAs.

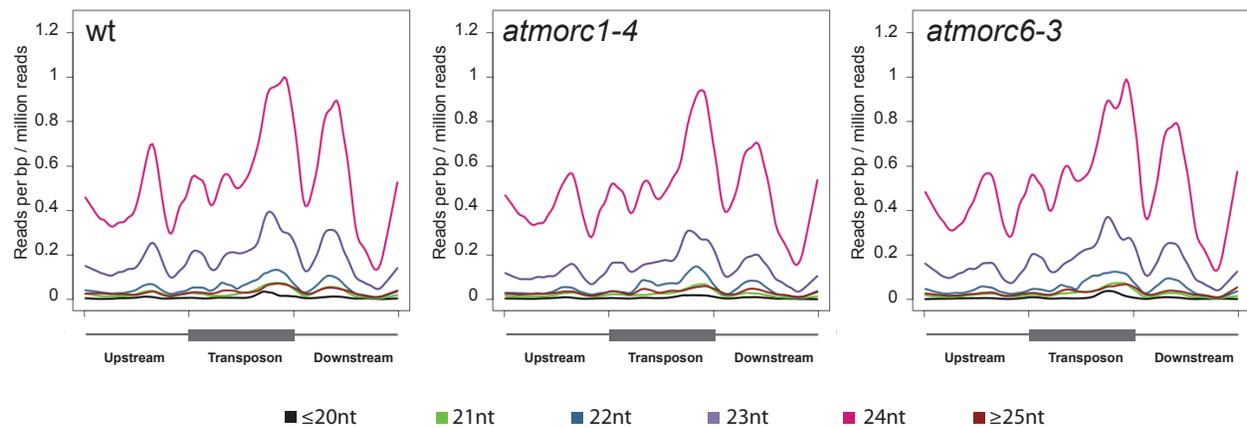


Fig. C.S8. AtMORC1 and AtMORC6 are required for chromocenter condensation. Percentage of nuclei showing decondensed, partially decondensed (intermediate) and wild type chromocenters in *atmorc6-1*, *cmt3-11 atmorc1-3* and in *atmorc1-3 atmorc6-1* double mutants in comparison to control backgrounds after immunostaining of nuclei using an antibody against H3K9me2. Pictures on top panels show examples of the three different patterns of chromocenters.

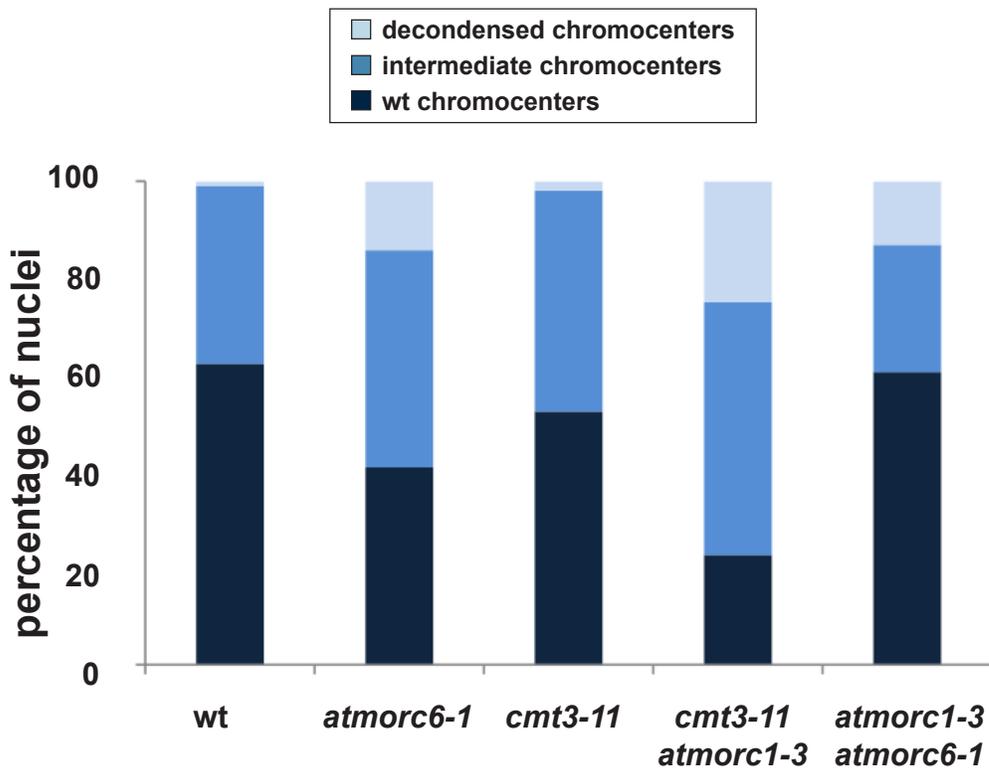
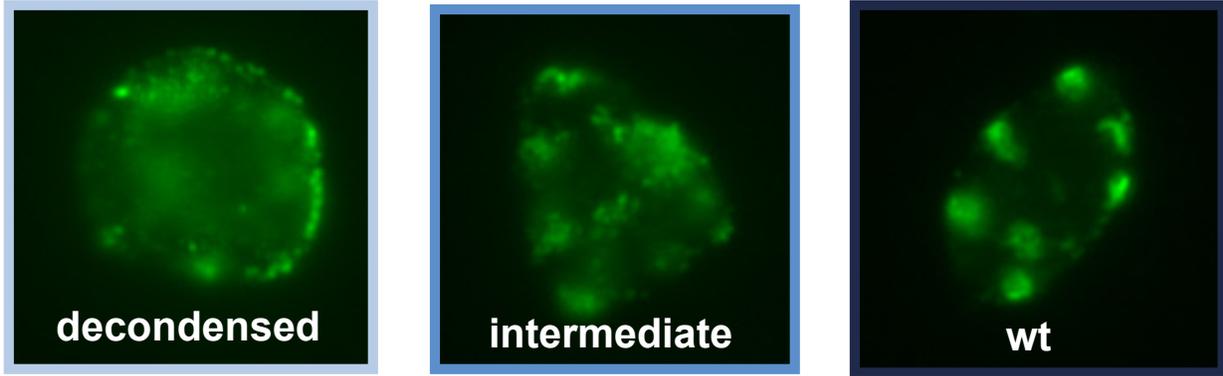


Fig. C.S9. H3K9me2 and DAPI staining show similar chromocenter condensation patterns in nuclei defined in Fig. S8 as nuclei showing wild type, intermediate or decondensed chromocenters in *cmt3-11 atmorc1-3*.

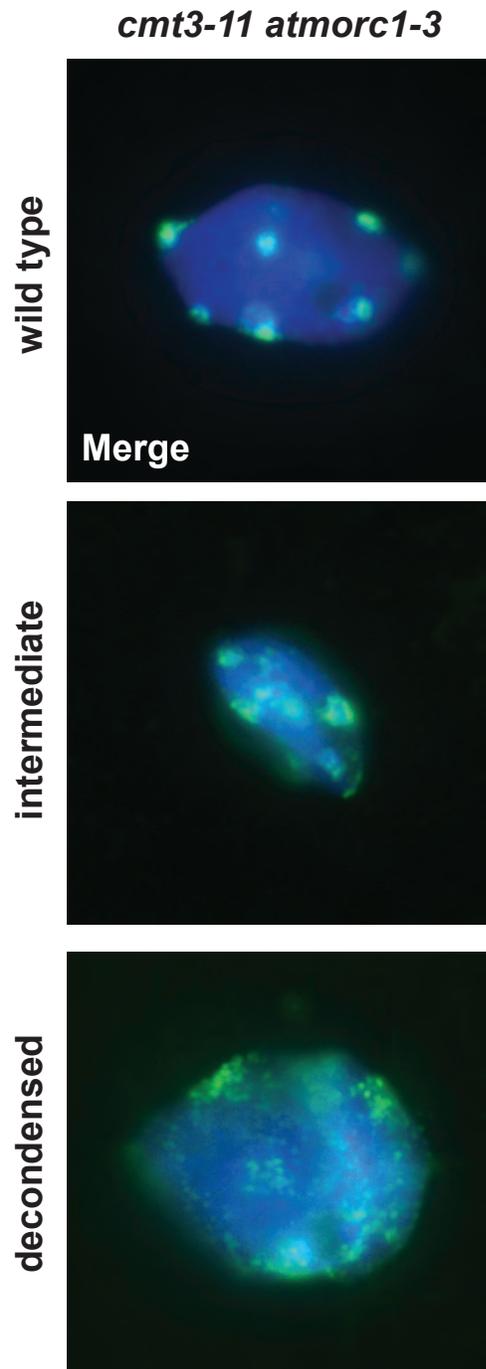


Fig. C.S10. H3K9me2 and DAPI staining show similar chromocenter condensation patterns in nuclei defined in Fig. S8 as nuclei showing wild type, intermediate or decondensed chromocenters in *atmorc6-1*.

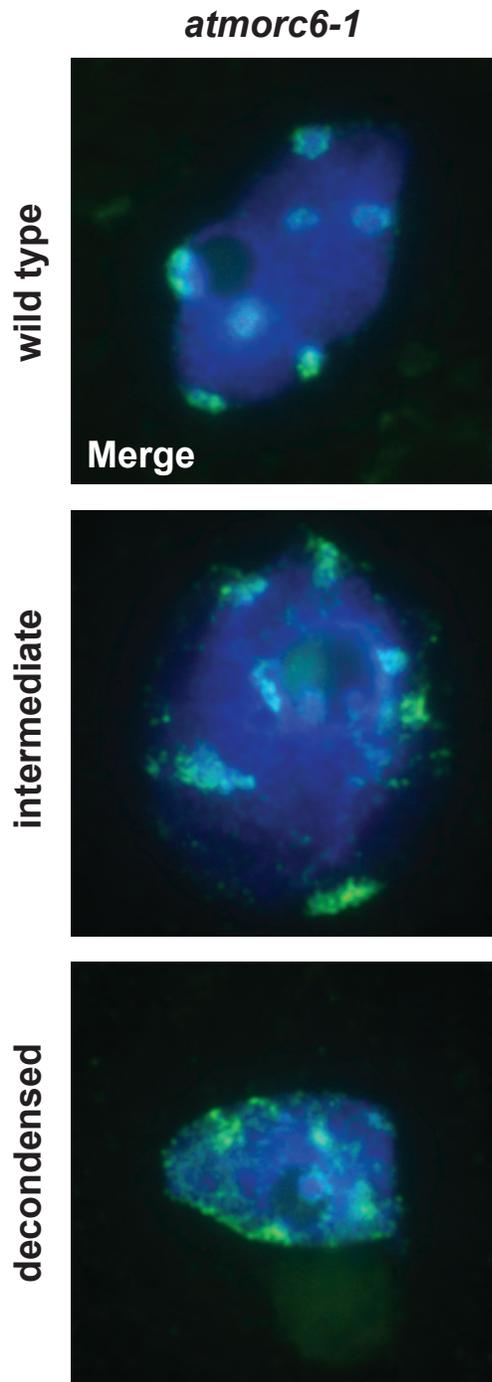


Fig. C.S11. H3K9me2 and DAPI staining show similar chromocenter condensation patterns in nuclei defined in Fig. S8 as nuclei showing wild type, intermediate or decondensed chromocenters in *atmorc1-3 atmorc6-1*.

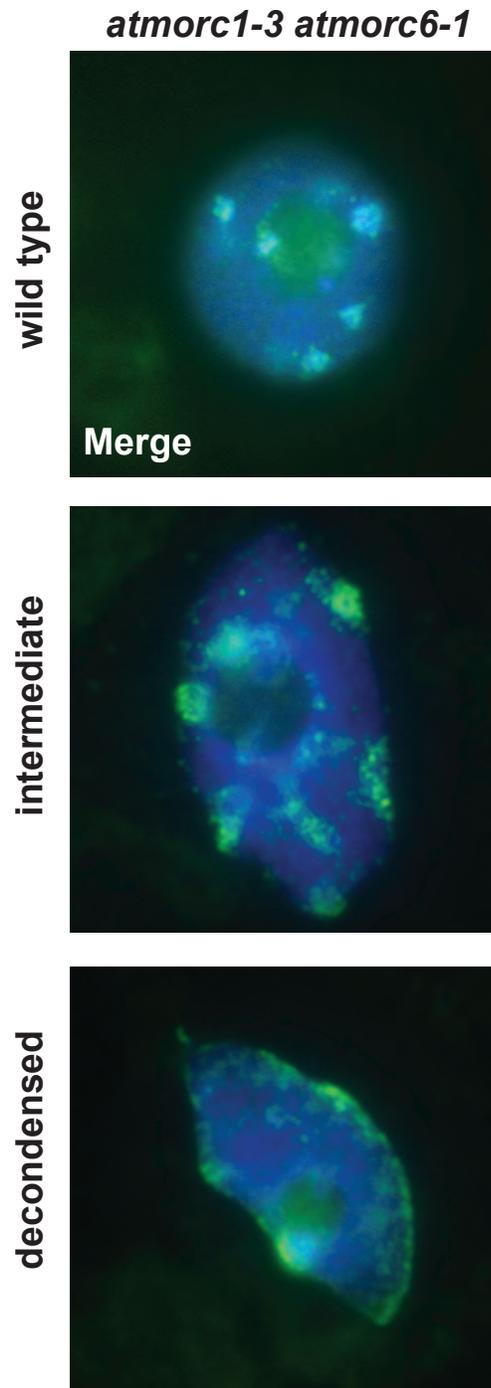


Fig. C.S12. Loci upregulated in *atmorc1* and *atmorc6* mostly localize to pericentromeric regions. Chromosomal views showing the log₂ ratios (*atmorc6-3/wild type* and *atmorc1-4/wild type*) of RNA sequencing reads in 100 Kb bins. The two red bars on each chromosome delimit the pericentromeric region with white circles representing the centromeres.

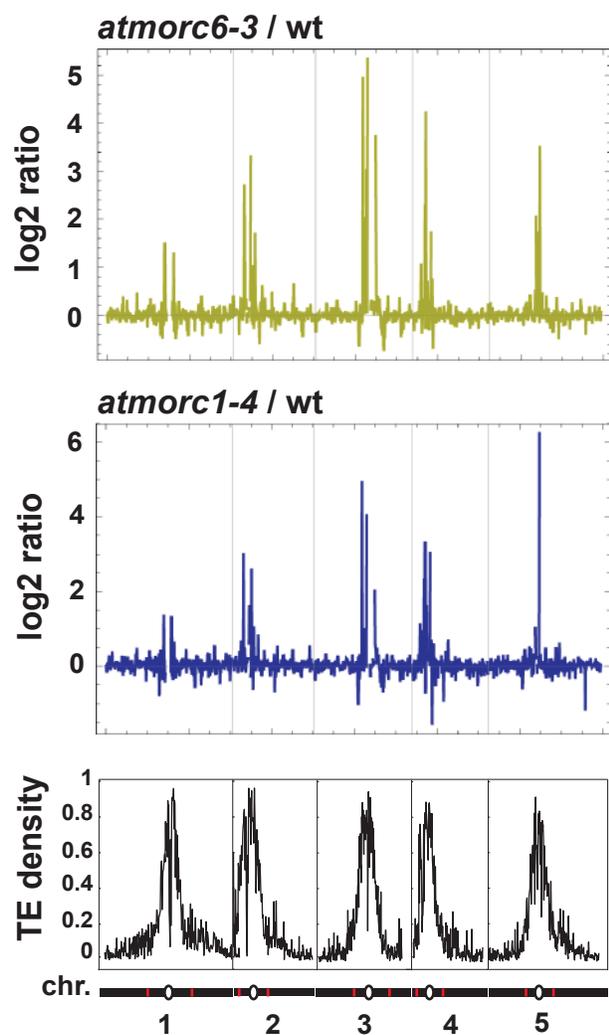


Fig. C.S13. Interaction matrix of the *atmorc6-1* genome from Hi-C analysis. Positions along the 5 chromosomes are shown from left to right and top to bottom, and each pixel represents interactions from uniquely mapping paired end reads in 200 kilobase bins. Black bars and circles mark the positions of the pericentromeric and telomeric regions, respectively. Light grey regions represent areas masked out due to problematic mapping. Black bars show separation between chromosomes.

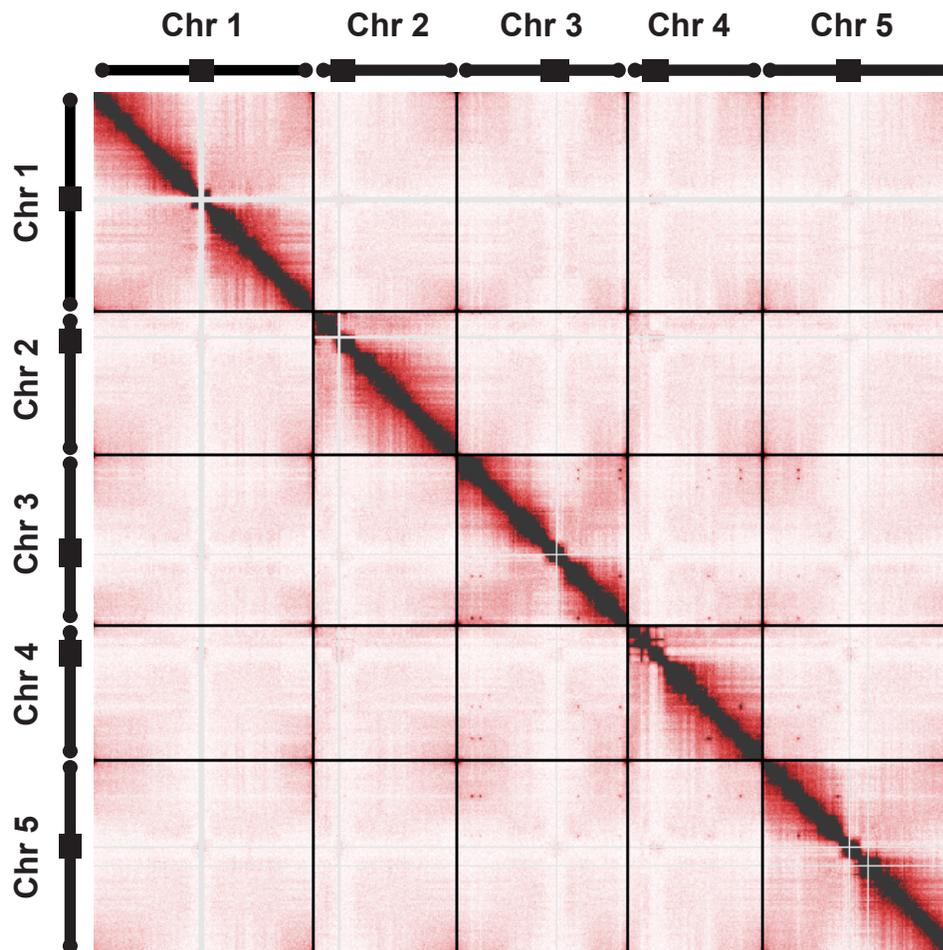


Fig. C.S14. Myc tagged AtMORC1 and AtMORC6 complement the EMS mutant lines. **(A)** Western blot using an antibody against Myc confirms that pAtMORC1::AtMORC1-Myc and pAtMORC6::AtMORC6-Myc are expressed in transformed plants but not in untransformed control. **(B)** Similar analyses using an antibody against GFP show that lines expressing AtMORC1-Myc and AtMORC6-Myc do not express the *SDC::GFP* transgene, confirming that pAtMORC1::AtMORC1-Myc and pAtMORC6::AtMORC6-Myc proteins are functional. *cmt3-11 atmorc1-3* and *atmorc6-1* mutants are used as GFP positive controls for AtMORC1-Myc and AtMORC6-Myc complementation, respectively. Coomassie staining of the large Rubisco subunit (*rbcL*) is used as loading control.

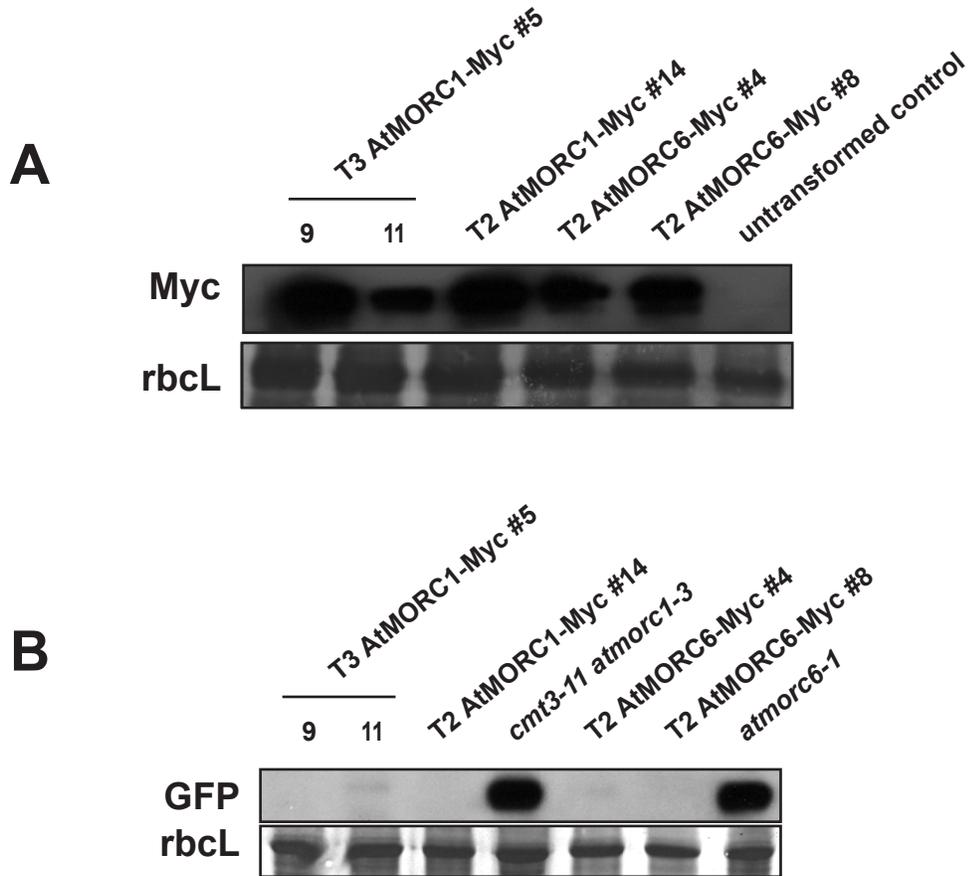
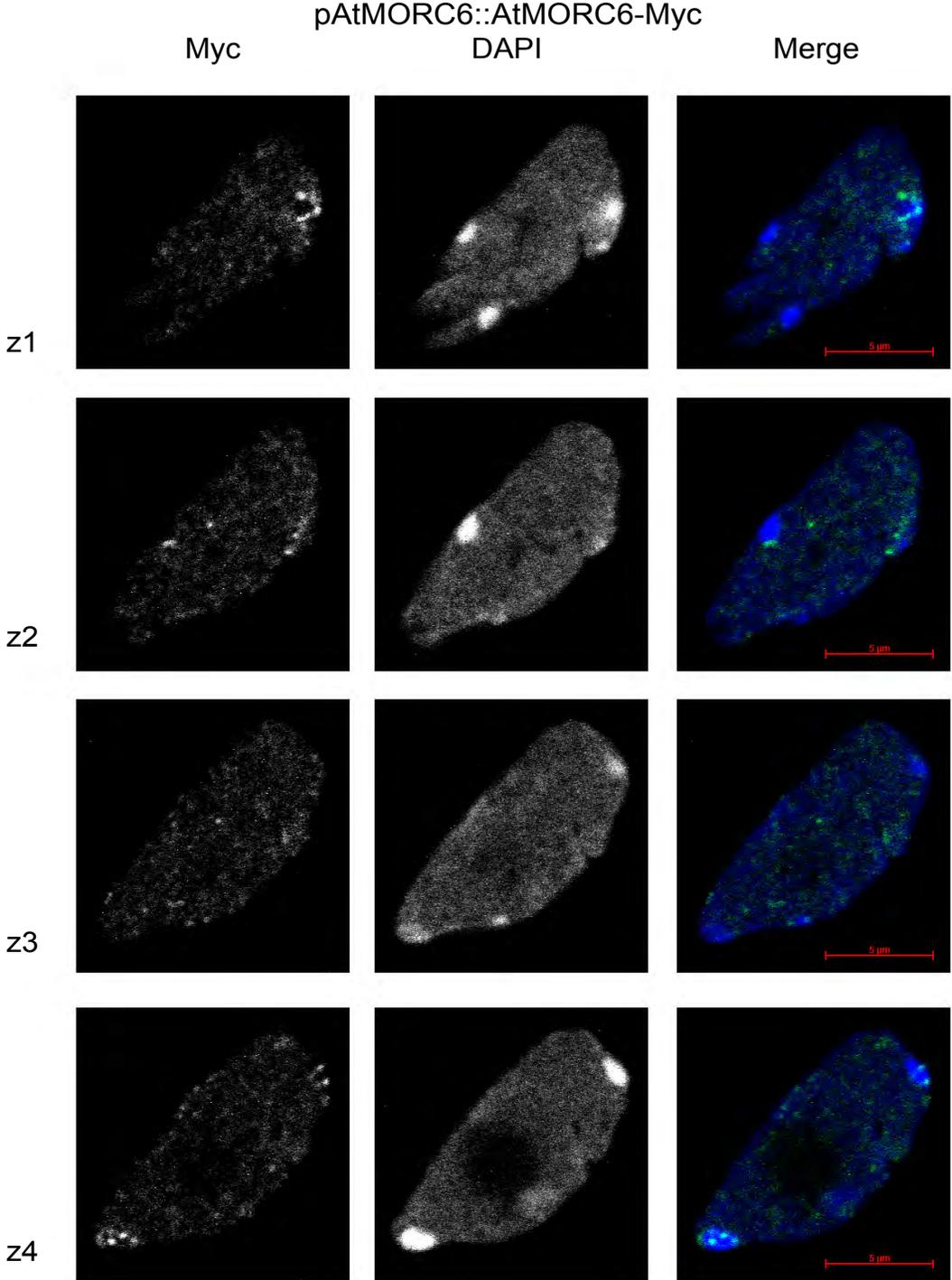


Fig. C.S15. AtMORC6 bodies are adjacent to chromocenters. Additional images of nuclei expressing AtMORC6-Myc were obtained using a Zeiss LSM 510 META confocal microscope. Eight images are displayed, taken at depth intervals of 1 micron. Multiple chromocenters are bordered by AtMORC6 bodies.



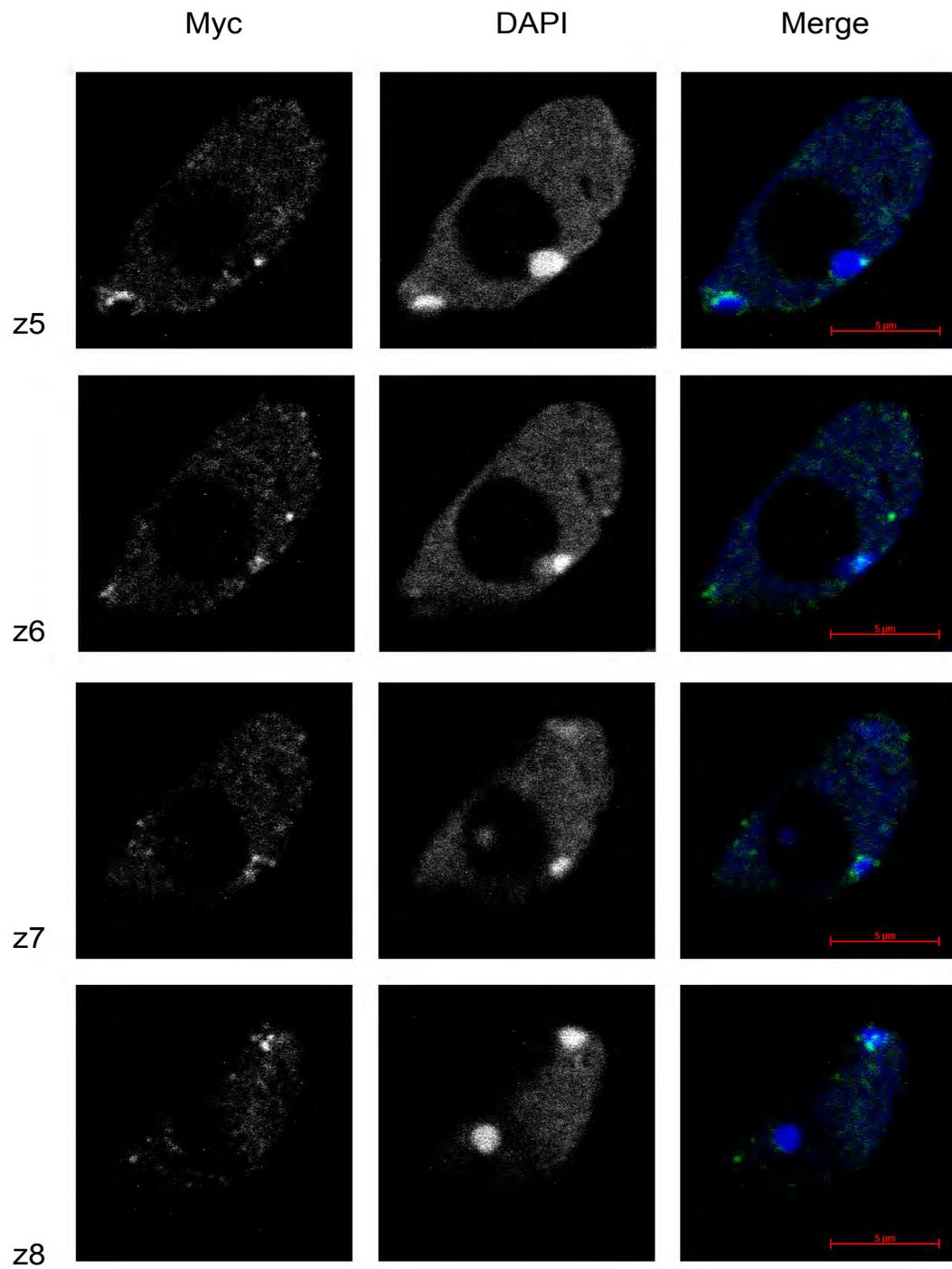


Table C.S1.

Table S1: Lists of Transposable elements (TEs) 4-fold upregulated in both EMS and T-DNA <i>atmorc6</i> and <i>atmorc1</i> mutant alleles. The TE soloLTR was identified by RT-qPCR. Therefore, it does not appear in Venn diagram shown in Fig. S4b.			
TEs upregulated in <i>atmorc6</i>			
NAME	FAMILY NAME	SUPERFAMILY NAME	CHROMOSOMAL LOCATION
AT1TE43225	ROMANIAT5	LTR/Copia	chr1:13,232,205-13,236,935
AT1TE45510	ATENSPM6	DNA/En-Spm	chr1:13,872,370-13,872,595
AT1TE51360	ROMANIAT5	LTR/Copia	chr1:15,613,068-15,617,844
AT2TE07145	ATCOPIA95	LTR/Copia	chr2:1,547,658-1,552,030
AT2TE13060	ATCOPIA32	LTR/Copia	chr2:2,995,748-2,996,367
AT2TE16220	HELITRON2	RC/Helitron	chr2:3,749,179-3,756,447
AT2TE18240	ATIS112A	DNA/Harbinger	chr2:4,343,480-4,344,832
AT2TE28020	ATMU1	DNA/MuDR	chr2:6,888,343-6,891,692
AT2TE41170	ATIS112A	DNA/Harbinger	chr2:9,660,885-9,661,799
AT3TE51895	ROMANIAT5	LTR/Copia	chr3:12,605,745-12,609,307
AT3TE51900	ATCOPIA28	LTR/Copia	chr3:12,609,702-12,610,774
AT3TE51910	VANDAL18NA	DNA/MuDR	chr3:12,612,304-12,613,201
AT3TE51930	ATGP1	LTR/Gypsy	chr3:12,615,787-12,623,418
AT3TE60425	ATHILA3	LTR/Gypsy	chr3:14,799,085-14,800,505
AT3TE60460	ATHILA6A	LTR/Gypsy	chr3:14,806,306-14,815,940
AT3TE63540	ATHILA2	LTR/Gypsy	chr3:15,722,463-15,733,680
AT4TE04415	ATIS112A	DNA/Harbinger	chr4:860,921-864,078
AT4TE09845	ATIS112A	DNA/Harbinger	chr4:2,076,980-2,077,351
AT4TE15005	ATHILA3	LTR/Gypsy	chr4:3,266,370-3,288,501
AT4TE15025	ATHILA0_I	LTR/Gypsy	chr4:3,273,946-3,285,539
AT4TE15030	ATHILA6A	LTR/Gypsy	chr4:3,274,431-3,285,232
soloLTR	soloLTR	retroelement soloLTR	chr5:9,871,837-9,872,408
AT5TE39630	ATIS112A	DNA/Harbinger	chr5:10,901,855-10,903,021
AT5TE47200	ATHILA2	LTR/Gypsy	chr5:13,334,474-13,345,770
AT5TE48715	ATREP15	RC/Helitron	chr5:13,705,744-13,706,653
AT5TE48720	HELITRONY1E	RC/Helitron	chr5:13,706,654-13,706,934
TEs upregulated in <i>atmorc1</i>			
NAME	FAMILY NAME	SUPERFAMILY NAME	CHROMOSOMAL LOCATION
AT1TE43225	ROMANIAT5	LTR/Copia	chr1:13,232,205-13,236,935
AT1TE45510	ATENSPM6	DNA/En-Spm	chr1:13,872,370-13,872,595
AT1TE51360	ROMANIAT5	LTR/Copia	chr1:15,613,068-15,617,844
AT2TE13060	ATCOPIA32	LTR/Copia	chr2:2,995,748-2,996,367
AT2TE18240	ATIS112A	DNA/Harbinger	chr2:4,343,480-4,344,832
AT2TE28020	ATMU1	DNA/MuDR	chr2:6,888,343-6,891,692
AT2TE28025	ATGP10	LTR/Gypsy	chr2:6,891,693-6,892,236
AT3TE51895	ROMANIAT5	LTR/Copia	chr3:12,605,745-12,609,307
AT3TE51900	ATCOPIA28	LTR/Copia	chr3:12,609,702-12,610,774
AT3TE60460	ATHILA6A	LTR/Gypsy	chr3:14,806,306-14,815,940
AT4TE04415	ATIS112A	DNA/Harbinger	chr4:860,921-864,078
AT4TE09845	ATIS112A	DNA/Harbinger	chr4:2,076,980-2,077,351
soloLTR	soloLTR	retroelement soloLTR	chr5:9,871,837-9,872,408
AT5TE39630	ATIS112A	DNA/Harbinger	chr5:10,901,855-10,903,021
AT5TE48715	ATREP15	RC/Helitron	chr5:13,705,744-13,706,653
AT5TE48720	HELITRONY1E	RC/Helitron	chr5:13,706,654-13,706,934
AT5TE48740	ATLINE1_4	LINE/L1	chr5:13,712,007-13,715,840

Table C.S2.

Table S2: Lists of Protein coding genes 4-fold upregulated in both EMS and T-DNA *atmorc6* and *atmorc1* mutant alleles. With the exception of genes in italics, all genes in these lists are located in pericentromeric regions and show DNA methylation or H3K9me2 silencing marks.

Protein coding genes upregulated in <i>atmorc6</i>	
NAME	GENE ANNOTATION
AT1G33840	unknown protein
AT1G35730	APUM9 (ARABIDOPSIS PUMILIO 9); RNA binding
AT1G36675	glycine-rich protein
AT1G53480	unknown protein
AT1G59930	unknown protein
AT1G60110	jacalin lectin family protein
AT1G67105	non coding RNA
AT2G07215	unknown protein
AT2G07240	Ulp1 protease family protein
AT2G10975	unknown protein
AT2G13770	unknown protein
AT2G17690	SDC, F-box family protein
<i>AT3G20340</i>	<i>unknown protein</i>
AT3G29639	unknown protein
AT3G30842	ATPDR10/PDR10 (PLEIOTROPIC DRUG RESISTANCE 10)
AT3G33528	unknown protein
AT3G42850	galactokinase, putative
<i>AT4G12490</i>	<i>protease inhibitor/seed storage/lipid transfer protein (LTP) family protein</i>
AT5G07550	GRP19 (Glycine rich protein 19)
AT5G07560	GRP20 (Glycine rich protein 20); nutrient reservoir
<i>AT5G15500</i>	<i>ankyrin repeat family protein</i>
AT5G35480	unknown protein
AT5G35490	unknown protein
<i>AT5G36910</i>	<i>THI2.2 (THIONIN 2.2); toxin receptor binding (THI2.2)</i>
Protein coding genes upregulated in <i>atmorc1</i>	
NAME	GENE ANNOTATION
<i>AT1G19610</i>	<i>LCR78/PDF1.4 (Low-molecular-weight cysteine-rich 78)</i>
<i>AT1G26390</i>	<i>FAD-binding domain-containing protein</i>
<i>AT1G26410</i>	<i>FAD-binding domain-containing protein</i>
AT1G33840	unknown protein
AT1G35730	APUM9 (ARABIDOPSIS PUMILIO 9)
AT1G36675	glycine-rich protein (AT1G36675)
<i>AT1G47890</i>	<i>disease resistance family protein</i>
AT2G07215	unknown protein
AT2G10975	unknown protein
AT2G13770	unknown protein
AT2G17690	SDC, F-box family protein
<i>AT2G17740</i>	<i>DC1 domain-containing protein</i>
<i>AT2G45220</i>	<i>pectinesterase family protein</i>
AT3G29639	unknown protein
AT3G30842	ATPDR10/PDR10 (PLEIOTROPIC DRUG RESISTANCE 10)
AT3G33528	unknown protein
AT3G55970	oxidoreductase, 2OG-Fe(II) oxygenase family protein
<i>AT4G12490</i>	<i>protease inhibitor/seed storage/lipid transfer protein (LTP) family protein</i>
<i>AT4G18170</i>	<i>WRKY28 (WRKY DNA-binding protein 28); transcription factor</i>
AT5G07550	GRP19 (Glycine rich protein 19)
AT5G35480	unknown protein
AT5G35490	unknown protein
AT5G35510	unknown protein

Table C.S3.

Table S3: Lists of Transposable elements (TEs) and protein-coding genes (PCGs) that were 4-fold upregulated in the <i>atmorc1-3 atmorc6-1</i> double mutant. * defines loci found to be 4-fold upregulated only in the <i>atmorc1-3 atmorc6-1</i> double and not in the <i>atmorc1</i> or <i>atmorc6</i> single mutants. All the other loci have been found to be upregulated in either EMS or T-DNA <i>atmorc1</i> and/or <i>atmorc6</i> alleles at least in one RNA sequencing dataset.			
TEs upregulated in the <i>atmorc1-3 atmorc6-1</i> double mutant			
NAME	FAMILY NAME	SUPERFAMILY NAME	CHROMOSOMAL LOCATION
AT1TE30845	HELITRON1	RC/Helitron	chr1:9,574,349-9,575,358
AT1TE43225	ROMANIAT5	LTR/Copia	chr1:13,232,205-13,236,935
AT1TE45135	ATHILA2	LTR/Gypsy	chr1:13,872,370-13,872,595
AT1TE45510	ATENSPM6	DNA/En-Spm	chr1:13,872,370-13,872,595
AT1TE53070	ATCOPIA87	LTR/Copia	chr1:16,118,258-16,118,383
AT1TE59745	ATCOPIA49	LTR/Copia	chr1:18,005,845-18,011,038
AT2TE07145	ATCOPIA95	LTR/Copia	chr2:1,547,658-1,552,030
AT2TE08840	TA11	LINE/L1	chr2:1,921,517-1,925,279
AT2TE13060	ATCOPIA32	LTR/Copia	chr2:2,995,748-2,996,367
AT2TE15415	ATGP10	LTR/Gypsy	chr2:3,533,343-3,539,331
AT2TE15880	ATHILA6A	LTR/Gypsy	chr2:3,643,453-3,652,409
AT2TE16220	HELITRON2	RC/Helitron	chr2:3,749,179-3,756,447
AT2TE18240	ATIS112A	DNA/Harbinger	chr2:4,343,480-4,344,832
AT2TE19615	VANDAL21	DNA/MuDR	chr2:4,736,486-4,744,627
AT2TE19625	ATHILA4C	LTR/Gypsy	chr2:4,738,315-4,739,542
AT2TE29450	ATCOPIA70	LTR/Copia	chr2:7,231,166-7,236,206
AT2TE38900	ATCOPIA76	LTR/Copia	chr2:9,194,572-9,198,735
AT2TE38905	ATGP2N	LTR/Gypsy	chr2:9,198,736-9,200,805
AT2TE66360	ATCOPIA50	LTR/Copia	chr2:14,938,850-14,939,092
AT3TE50570	VANDAL3	DNA/MuDR	chr3:12,177,936-12,189,847
AT3TE50595	ATHILA2	LTR/Gypsy	chr3:12,189,865-12,193,865
AT3TE51895	ROMANIAT5	LTR/Copia	chr3:12,605,745-12,609,307
AT3TE51900	ATCOPIA28	LTR/Copia	chr3:12,609,702-12,610,774
AT3TE51910	VANDAL18NA	DNA/MuDR	chr3:12,612,304-12,613,201
AT3TE51930	ATGP1	LTR/Gypsy	chr3:12,615,787-12,623,418
AT3TE60425	ATHILA3	LTR/Gypsy	chr3:14,799,08- 14,800,505
AT3TE60460	ATHILA6A	LTR/Gypsy	chr3:14,806,306-14,815,940
AT3TE63540	ATHILA2	LTR/Gypsy	chr3:15,722,463-15,733,680
AT4TE09845	ATIS112A	DNA/Harbinger	chr4:2,076,980-2,077,351
AT4TE10335	ATCOPIA58	LTR/Copia	chr4:2,198,163-2,203,991
AT4TE15005	ATHILA3	LTR/Gypsy	chr4:3,266,370-3,288,501
AT4TE15025	ATHILA0_I	LTR/Gypsy	chr4:3,273,946-3,285,539
AT4TE15030	ATHILA6A	LTR/Gypsy	chr4:3,274,431-3,285,232
AT4TE16900	ATHILA	LTR/Gypsy	chr4:3,838,635-3,843,992
AT4TE17090	ATCOPIA41	LTR/Copia	chr4:3,909,249-3,910,990
AT4TE17115	ATHILA6A	LTR/Gypsy	chr4:3,915,207-3,916,819
AT4TE20120	ATENSPM5	DNA/En-Spm	chr4:4,830,277-4,836,762
AT4TE25590	ATCOPIA49	LTR/Copia	chr4:6,067,394-6,072,632
AT5TE35950	HELITRONY1D	RC/Helitron	chr5:9,871,833-9,872,483
AT5TE39170	ATGP1	LTR/Gypsy	chr5:10,764,247-10,772,507
AT5TE39630	ATIS112A	DNA/Harbinger	chr5:10,901,855-10,903,021

AT5TE43315	ATHILA	LTR/Gypsy	chr5:12,182,325-12,187,588
AT5TE46210	ATHILA8B	LTR/Gypsy	chr5:13,008,325-13,013,667
AT5TE47200	ATHILA2	LTR/Gypsy	chr5:13,334,474-13,345,770
PCGs upregulated in <i>in the atmorc1-3 atmorc6-1 double mutant</i>			
NAME	GENE ANNOTATION		
AT3G20710	F-box protein-related		
AT4G05370	unknown protein		
AT3G13220	ABC transporter family protein		
AT3G15440	unknown protein		
AT3G22860	TIF3C2 (eukaryotic translation initiation factor 3 subunit C2)		
AT5G29560	Ca ²⁺ -binding EF hand family protein		
AT1G45063	copper ion binding / electron carrier		
AT3G44460	DPBF2 (BASIC LEUCINE ZIPPER TRANSCRIPTION FACTOR 67)		
AT3G30842	ATPDR10/PDR10 (PLEIOTROPIC DRUG RESISTANCE 10)		
AT1G27570	phosphatidylinositol 3- and 4-kinase family protein		
AT2G17690	F-box family protein		
AT3G29639	unknown protein		
AT2G04050	MATE efflux family protein		
AT5G60260	unknown protein		
AT4G05380	AAA-type ATPase family protein		
AT1G33840	unknown protein		
AT2G10975	unknown protein		
AT5G18840	sugar transporter, putative		
AT1G36675	glycine-rich protein		
AT1G35730	APUM9 (ARABIDOPSIS PUMILIO 9); RNA binding (APUM9)		
AT3G01345	unknown protein		
AT1G15150	MATE efflux family protein		
AT4G36850	PQ-loop repeat family protein / transmembrane family protein		
AT1G27565	unknown protein		
AT2G07213	other_rna		
AT5G23020	MAM-L (METHYLTHIOALKYLMALATE SYNTHASE-LIKE); 2-isopropylmalate synthase		
AT5G45095	unknown protein		
AT4G20820*	FAD-binding domain-containing protein		
AT2G07215	unknown protein		
AT3G14380	integral membrane family protein		
AT3G47340	ASN1 (DARK INDUCIBLE 6)		
AT5G07700*	MYB76 (myb domain protein 76); DNA binding / transcription factor		
AT5G41080	glycerophosphoryl diester phosphodiesterase family protein		
AT3G33528	unknown protein		
AT3G09450	unknown protein		
AT5G38386	F-box family protein		
AT2G18193	AAA-type ATPase family protein		
AT5G36910	THI2.2 (THIONIN 2.2); toxin receptor binding (THI2.2)		
AT3G44990*	XTR8 (xyloglucan:xyloglucosyl transferase 8); hydrolase, acting on glycosyl bonds (XTR8)		
AT3G01600	ANAC044 (Arabidopsis NAC domain containing protein 44); transcription factor		
AT1G07450*	tropinone reductase, putative / tropine dehydrogenase, putative		
AT4G15680	glutaredoxin family protein		

Table C.S4.

Table S4: Sequences of primers used in this study		
T-DNA genotyping		
JP9339	atmorc6-3 LP	GGAAAGCTGGAAGCTATAATGATG
JP9340	atmorc6-3 RP	GATGACATCTGCCCAAGTCTC
JP7707	GABI-KAT LB O8409	ATATTGACCATCATACTCATTGC
JP8938	atmorc1-4 LP	CGTATCTCAGCCGCTAACTTG
JP8939	atmorc1-4 RP	AAGCAGCTGCAGTGGATTATG
JP2207	LB3 SAIL T-DNA	TAGCATCTGAATTTATAACCAATCTCGATACAC
JP8509	drm2-2 LP	AGATCGCTTCCAGAGTTAGCC
JP8510	drm2-2 RP	TTGTCGCAAAAAGCAAAAAGAG
JP2922	cmt3-11 LP	TAACGGAAGGATGCCAGATT
JP2923	cmt3-11 RP	CAAGAAATGGGCTGTTGACAT
JP2410	LBA1 SALK T-DNA	TGGTTCACGTAGTGGGCCATCG
RT-qPCR		
JP2452	ACTIN 7 LP	TCGTGGTGGTGAATTTGTTAC
JP2453	ACTIN 7 RP	CAGCATCATCACAAGCATCC
JP3395	SDC LP	AATGTAAGTTGTAAACCATTGTAACGTGACC
JP3396	SDC RP	CAGGCATCCGTAGAACTCATGAGC
JP9642	ROMANIAT5 LP	GTATCCTTTGGCCCGGTATT
JP9643	ROMANIAT5 RP	GCCTCTTCGAAATGCCATAA
JP9055	ATCOPIA28 LP	AGTCCTTTTGGTTGCTGAACA
JP9056	ATCOPIA28 RP	CCGGATGTAGCAACATTCCT
JP9640	soloLTR LP	AACTAACGTCATTACATACACATCTTG
JP9641	soloLTR RP	AATTAGGATCTTGTGGCCAGCTA
JP9057	ATMU1 LP	TAATTTGGCTGACGGAATCAC
JP9058	ATMU1 RP	ATTTGGGGGAAAACAAATGAG
JP9646	atmorc6-3 LP	CATGTGCACCCTATGTTCCCT
JP9647	atmorc6-3 RP	ATCCCTTGGATTTGTGGTTTT
JP9680	atmorc1-4 LP	ATCAAGGAGGCCCTAAACTT
JP9681	atmorc1-4 RP	TGTGACAGTGATTTGCCAGT
BS-sequencing		
JP6349	SDC BSPCR LP	GAAAAAGTTGGAATGGGTTTGGAGAGTTTAA
JP6350	SDC BSPCR RP	CAACAAACCCTAATATATTTTATATATAAAC
JP9775	ROMANIAT5 BSPCR LP	GTAAGTGGATTAGTTATATAAAGAGAGTT
JP9776	ROMANIAT5 BSPCR RP	ATAAATAAACATCATCTACATCTTATAA
JP9120	ATCOPIA28 BSPCR LP	TATTTATTTYGTTTCAATTTGGATTAGTTTT
JP9121	ATCOPIA28 BSPCR RP	ACRATATCAAATAATTATCATCATCTTAA
JP9377	soloLTR BSPCR LP	GATATAAAGGAATGGTTAGATAATATGYGATT
JP9378	soloLTR BSPCR RP	CRATATAACTCAAATTTATATTACTCTTAA
JP9177	ATMU1 BSPCR LP	TTATGAATTAGTTAGGTTATAGTTTGTATT
JP9178	ATMU1 BSPCR RP	ATTCCTCRTCTTCTACAACATCATTTAA
Cloning		
JP5275	PstI_SDCpro_fwd	GAACTGCAGTGATGCTCTAACAATCTTCCACAAGACC
JP5276	PstI_SDCpro_rev	GAAACTGCAGTTCTCTCCCTGTTTTGCTACTATTG
JP5235	HindIII-NLS-eGFP_FWD	TGGCAAGCTTATGGCTCCAAGAAGAAGAAAGGTCATG
JP5236	HindIII_35Sterminator_REV	TAGCAAGCTTCTCTCAACACATGAGCGAAACCTATAAG
JP8644	ATMORC1_Ctertag_LP	CACCGTTGATTTGGTTTTGTCTGGTC
JP8645	ATMORC1_Ctertag_RP	AACTGTTGCATCTCCTTCTTC
JP8646	ATMORC6_Ctertag_LP	CACCAGTATGATGTGAGGTTAGTGAG
JP8647	ATMORC6_Ctertag_RP	CGTATTACATTTCTTCTGTGC

REFERENCES

1. J. A. Law, S. E. Jacobsen, Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204 (2010). doi:10.1038/nrg2719 Medline
2. J. P. Jackson, A. M. Lindroth, X. Cao, S. E. Jacobsen, Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* 416, 556 (2002). doi:10.1038/nature731 Medline
3. F. Malagnac, L. Bartee, J. Bender, An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. *EMBO J.* 21, 6842 (2002). doi:10.1093/emboj/cdf687 Medline
4. X. Zhang *et al.*, Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* 126, 1189 (2006). doi:10.1016/j.cell.2006.08.003 Medline
5. I. R. Henderson, S. E. Jacobsen, Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Genes Dev.* 22, 1597 (2008). doi:10.1101/gad.1667808 Medline
6. H. G. Kang, J. C. Kuhl, P. Kachroo, D. F. Klessig, CRT1, an Arabidopsis ATPase that interacts with diverse resistance proteins and modulates disease resistance to turnip crinkle virus. *Cell Host Microbe* 3, 48 (2008). doi:10.1016/j.chom.2007.11.006 Medline
7. H. G. Kang *et al.*, Endosome-associated CRT1 functions early in resistance gene-mediated defense signaling in Arabidopsis and tobacco. *Plant Cell* 22, 918 (2010). doi:10.1105/tpc.109.071662 Medline
8. Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57 (2009). doi:10.1038/nrg2484 Medline
9. S. J. Cokus *et al.*, Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452, 215 (2008). doi:10.1038/nature06745 Medline
10. M. L. Watson *et al.*, Identification of morc (microorchidia), a mutation that results in arrest of spermatogenesis at an early meiotic stage in the mouse. *Proc. Natl. Acad. Sci. U.S.A.* 95, 14361 (1998). doi:10.1073/pnas.95.24.14361 Medline
11. N. Inoue *et al.*, New gene family defined by MORC, a nuclear protein required for mouse spermatogenesis. *Hum. Mol. Genet.* 8, 1201 (1999). doi:10.1093/hmg/8.7.1201 Medline
12. L. M. Iyer, S. Abhiman, L. Aravind, MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. *Biol. Direct* 3, 8 (2008). doi:10.1186/1745-6150-3-8 Medline
13. P. Fransz, J. H. De Jong, M. Lysak, M. R. Castiglione, I. Schubert, Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl. Acad. Sci. U.S.A.* 99, 14584 (2002). doi:10.1073/pnas.212325299 Medline
14. E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289 (2009). doi:10.1126/science.1181369 Medline
15. P. Amedeo, Y. Habu, K. Afsar, O. Mittelsten Scheid, J. Paszkowski, Disruption of

- the plant gene MOM releases transcriptional silencing of methylated genes. *Nature* 405, 203 (2000). doi:10.1038/35012108 Medline
16. A. V. Probst, P. F. Fransz, J. Paszkowski, O. Mittelsten Scheid, Two means of transcriptional reactivation within heterochromatin. *Plant J.* 33, 743 (2003). doi:10.1046/j.1365-313X.2003.01667.x Medline
 17. V. J. Simpson, T. E. Johnson, R. F. Hammen, *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Res.* 14, 6711 (1986). doi:10.1093/nar/14.16.6711 Medline
 18. S. Kennedy, D. Wang, G. Ruvkun, A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* 427, 645 (2004). doi:10.1038/nature02302 Medline
 19. H. Tabara *et al.*, The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* 99, 123 (1999). doi:10.1016/S0092-8674(00)81644-X Medline
 20. M. A. Carmell *et al.*, MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell* 12, 503 (2007). doi:10.1016/j.devcel.2007.03.001 Medline
 21. D. Bourc'his, T. H. Bestor, Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 431, 96 (2004). doi:10.1038/nature02886 Medline
 22. A. A. Aravin *et al.*, A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol. Cell* 31, 785 (2008). doi:10.1016/j.molcel.2008.09.003 Medline
 23. M. W. Kankel *et al.*, Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics* 163, 1109 (2003). Medline
 24. N. Kleinboelting, G. Huet, A. Kloetgen, P. Viehoveer, B. Weisshaar, GABI-Kat SimpleSearch: new features of the Arabidopsis thaliana T-DNA mutant database. *Nucleic Acids Res.* 40, (Database issue), D1211 (2012). doi:10.1093/nar/gkr1047 Medline
 25. S. J. Clough, A. F. Bent, Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J.* 16, 735 (1998). doi:10.1046/j.1365-313x.1998.00343.x Medline
 26. J. A. Law, A. A. Vashisht, J. A. Wohlschlegel, S. E. Jacobsen, SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. *PLoS Genet.* 7, e1002195 (2011). doi:10.1371/journal.pgen.1002195 Medline
 27. M. V. Greenberg *et al.*, Identification of genes required for de novo DNA methylation in Arabidopsis. *Epigenetics* 6, 344 (2011). doi:10.4161/epi.6.3.14242 Medline
 28. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25 (2009). doi:10.1186/gb-2009-10-3-r25 Medline
 29. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621 (2008). doi:10.1038/nmeth.1226 Medline
 30. Y. Benjamini, Y. Hochberg, *J. R. Stat. Soc., B* 57, 12 (1995).
 31. P. Y. Chen, S. J. Cokus, M. Pellegrini, BS Seeker: precise mapping for bisulfite sequencing. *BMC Bioinformatics* 11, 203 (2010). doi:10.1186/1471-2105-11-203

Medline

32. J. Lee *et al.*, Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *Plant Cell* 19, 731 (2007). doi:10.1105/tpc.106.047688 Medline
33. C. Lu, B. C. Meyers, P. J. Green, Construction of small RNA cDNA libraries for deep sequencing. *Methods* 43, 110 (2007). doi:10.1016/j.ymeth.2007.05.002 Medline
34. Y. Zhang *et al.*, Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* 148, 908 (2012). doi:10.1016/j.cell.2012.02.002 Medline
35. W. J. Soppe *et al.*, DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. *EMBO J.* 21, 6549 (2002). doi:10.1093/emboj/cdf657 Medline
36. J. K. Kim *et al.*, Functional genomic analysis of RNA interference in *C. elegans*. *Science* 308, 1164 (2005). doi:10.1126/science.1109267 Medline
37. R. S. Kamath *et al.*, Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421, 231 (2003). doi:10.1038/nature01278 Medline
38. Y. Y. Polosina, C. G. Cupples, Wot the 'L-Does MutL do? *Mutat. Res.* 705, 228 (2010). doi:10.1016/j.mrrev.2010.07.002 Medline
39. Y. Mimura, K. Takahashi, K. Kawata, T. Akazawa, N. Inoue, Two-step colocalization of MORC3 with PML nuclear bodies. *J. Cell Sci.* 123, 2014 (2010). doi:10.1242/jcs.063586 Medline
40. T. Hirano, The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion, and repair. *Genes Dev.* 16, 399 (2002). doi:10.1101/gad.955102 Medline
41. M. E. Blewitt *et al.*, SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat. Genet.* 40, 663 (2008). doi:10.1038/ng.142 Medline
42. I. Ausin, T. C. Mockler, J. Chory, S. E. Jacobsen, IDN1 and IDN2 are required for de novo DNA methylation in *Arabidopsis thaliana*. *Nat. Struct. Mol. Biol.* 16, 1325 (2009). doi:10.1038/nsmb.1690 Medline
43. T. Kanno *et al.*, A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nat. Genet.* 40, 670 (2008). doi:10.1038/ng.119 Medline
44. G. Böhmendorfer *et al.*, GMI1, a structural-maintenance-of-chromosomes-hinge domain-containing protein, is involved in somatic homologous recombination in *Arabidopsis*. *Plant J.* 67, 420 (2011). doi:10.1111/j.1365-313X.2011.04604.x Medline