Title

The RNA phosphatase PIR-1 regulates endogenous small RNA pathways in C. elegans

Journal

Molecular Cell, 81(3)

ISSN

1097-2765

Authors

Chaves, Daniel A
Dai, Hui
Li, Lichao
et al.

Publication Date

2021-02-01

DOI

10.1016/j.molcel.2020.12.004

Peer reviewed

Published in final edited form as:

Mol Cell. 2021 February 04; 81(3): 546–557.e5. doi:10.1016/j.molcel.2020.12.004.

The RNA phosphatase PIR-1 regulates endogenous small RNA pathways in *C. elegans*

Daniel A Chaves[1,6,*], **Hui Dai**[2,*], **Lichao Li**[2], **James J Moresco**[3], **Myung Eun Oh**[2], **Darryl Conte Jr**[4], **John R Yates III**[5], **Craig C Mello**[1,4], **Weifeng Gu**[2,7]

[1]Dept. of Molecular Medicine, University of Massachusetts Medical School

[2]Dept. Of Molecular, Cell and Systems Biology, University of California, Riverside

[3]Center for Genetics of Host Defense, UT Southwestern Medical Center

[4]RNA Therapeutics Institute, University of Massachusetts Medical School

[5]Department of Molecular Medicine, Scripps Research Institute

[6]Faculdade de Medicina, Universidade de Lisboa, Portugal

[7]Lead Contact: Weifeng Gu, Dept. of Molecular, Cell and Systems Biology, University of California, Riverside, 900 University Avenue, Riverside CA 92521

SUMMARY

Eukaryotic cells regulate 5’ triphosphorylated (ppp-) RNAs to promote cellular functions and prevent recognition by antiviral RNA sensors. For example, RNA capping enzymes possess triphosphatase domains that remove the γ phosphates of ppp-RNAs during RNA capping. Members of the closely related PIR1 family of RNA polyphosphatases remove both the β and γ phosphates from ppp-RNAs. Here we show that *C. elegans* PIR-1 dephosphorylates ppp-RNAs made by cellular RdRPs and is required for the maturation of 26G-RNAs, Dicer-dependent small RNAs that regulate thousands of genes during spermatogenesis and embryogenesis. PIR-1 also regulates the CSR-1 22G-RNA pathway and has critical functions in both somatic and germline development. Our findings suggest that PIR-1 modulates both Dicer-dependent and - independent Argonaute pathways, and provide insight into how cells and viruses use a conserved RNA phosphatase to regulate and respond to ppp-RNA species.

Graphical Abstract

weifeng.gu@ucr.edu, Tel: 951-827-3600.

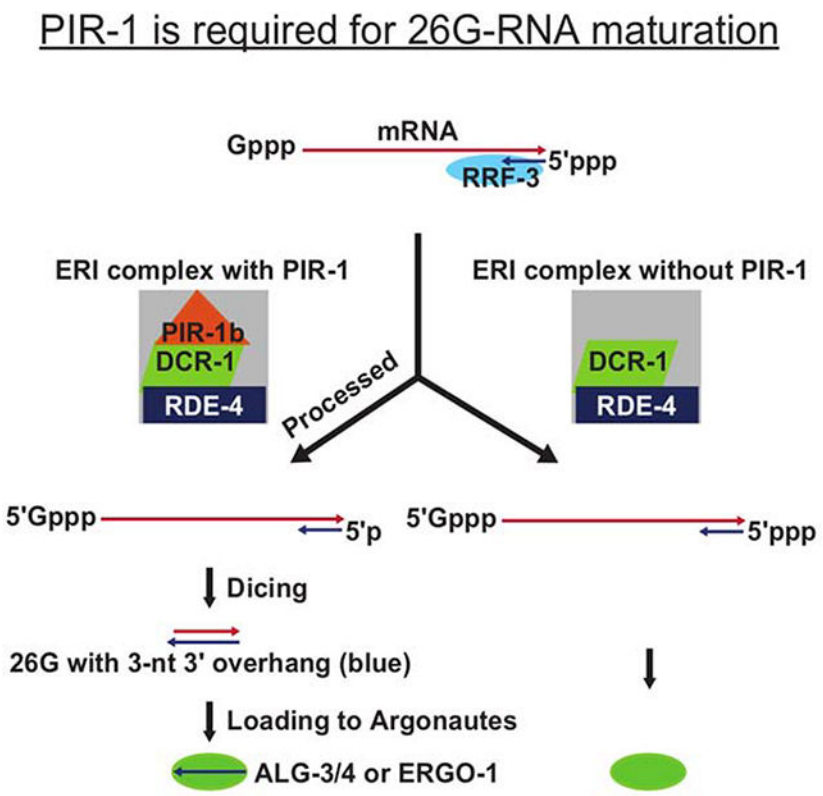
*These authors contributed equally.

AUTHOR CONTRIBUTIONS

Conceptualization, D.A.C., C.C.M., and W.G.; Methodology, D.A.C., C.C.M., and W.G.; Investigation, D.A.C., H.D., J.J.M., and W.G.; Writing—Original Draft, W.G., D.C., and C.C.M; Writing—Review & Editing, W.G., D.C., and C.C.M.; Funding Acquisition, J.R.Y., C.C.M., and W.G.; Resources—W.G.; Supervision, W.G. and C.C.M.

Declaration of Interests

The authors declare no competing interests.



eTOC Blurb:

Eukaryotic cells strictly regulate 5’-triphosphorylated RNAs by capping or other processes to promote cellular functions and prevent recognition by antiviral RNA sensors. Chaves *et al.* showed that PIR-1 modifies the triphosphate group of RNAs to promote the maturation of small RNAs which regulate thousands of genes during spermatogenesis and embryogenesis.

INTRODUCTION

Cells can modify and sense the phosphorylation status of RNA 5’ ends to regulate gene expression, control RNA stability, and mediate antiviral defense (Hornung et al., 2006; Kato et al., 2006; Shatkin, 1976). For example, the eukaryotic RNA polymerase, Pol II, recruits a capping enzyme that co-transcriptionally modifies the 5’ end of its RNA products. A key enzymatic modality in this capping enzyme is an RNA triphosphatase domain related to the cysteine phosphatase superfamily of protein and RNA phosphatases (Deshpande et al., 1999; Takagi et al., 1998; Yuan et al., 1998). After removing the γ phosphate from a nascent transcript, capping enzyme installs a guanine-nucleotide cap that masks the 5’ end from cellular nucleases and sensors that recognize RNAs made by viral polymerases (Shatkin, 1976). Cellular and viral homologs of the triphosphatase domain of capping enzymes include the PIR-1 (Phosphatase that Interacts with RNA and Ribonucleoprotein Particle 1) family of RNA polyphosphatases, which catalyze the removal of γ and β phosphates from triphosphorylated RNAs (ppp-RNA) *in vitro* (Deshpande et al., 1999; Takagi et al., 1998; Yuan et al., 1998). However, the cellular functions and targets of PIR-1 are largely unknown.

The *C. elegans* PIR-1 homolog was identified as a binding partner of the RNA interference (RNAi) factor, Dicer (Duchaine et al., 2006). RNAi plays important roles in regulating gene expression and viral immunity in diverse organisms (Baulcombe, 2004; Hannon, 2002; McCaffrey et al., 2002). Dicer encodes a multifunctional protein with double-stranded (ds)RNA-binding motifs, a DExH/D helicase motif, and a bidentate RNase III domain (Macrae et al., 2006); Dicer is known to bind and then process dsRNAs into short-interfering (si)RNAs or micro(mi)RNAs that guide Argonaute co-factors to mediate genetic silencing (Bernstein et al., 2001; Grishok et al., 2001). *In vitro* studies suggest that Dicer is not sensitive to the 5' phosphorylation status of its substrates (Welker et al., 2011), raising the question of how and why Dicer associates with PIR-1.

In *C. elegans*, Dicer (DCR-1) functions in several small RNA pathways, including the miRNA pathway and the endogenous (endo-) and exogenous (exo-) RNAi pathways triggered by dsRNAs (Duchaine et al., 2006; Grishok et al., 2001; Welker et al., 2010). Upon exposure to exogenous or viral dsRNAs, DCR-1 processes dsRNAs into short 23-nucleotide (nt) duplex siRNAs with monophosphorylated 5' ends and 3' 2-nt overhangs (Ashe et al., 2013; Coffman et al., 2017; Guo et al., 2013). These diced siRNAs are loaded onto the Argonaute RDE-1 (Tabara et al., 1999), which cannot silence targets alone (Gu et al., 2009; Steiner et al., 2009; Yigit et al., 2006). Instead, RDE-1 recruits cellular RNA-dependent RNA Polymerase (RdRP) to generate/amplify the silencing signal (Pak and Fire, 2007; Yigit et al., 2006). These RdRPs prefer to initiate transcription at C residues located 5' of a purine on template RNAs, and thus produce 22-nt products that contain a 5'-triphosphorylated (ppp-)G residue, so called 22G-RNAs (Claycomb et al., 2009; Gu et al., 2012; Pak and Fire, 2007). 22G-RNAs are then loaded onto worm-specific Argonautes (WAGOs) (Gu et al., 2009). Unlike other Argonautes, which usually bind monophosphorylated (p-)RNA guides, WAGOs directly accommodate the ppp-RNA guides synthesized by RdRP (Gu et al., 2009).

In addition to its key role in the exo-RNAi pathway, DCR-1 also functions in endogenous small RNA pathways (Duchaine et al., 2006; Fire et al., 1998; Grishok et al., 2001; Ruby et al., 2006; Welker et al., 2010). Several genes that function in non-essential endo-RNAi pathways were identified as mutants with enhanced exo-RNAi (ERI mutants), perhaps because these endo-RNAi pathways compete for downstream components that are limiting for robust exo-RNAi (Duchaine et al., 2006; Fischer et al., 2011; Kennedy et al., 2004; Simmer et al., 2002). The ERI genes and their associated factors define two major endo-RNAi pathways both employing 26-nt antisense RNAs usually starting with G, 26G-RNAs. The ERGO-1 Argonaute and two redundant Argonautes, ALG-3 and ALG-4 (ALG-3/4), engage 26G-RNAs during embryogenesis and spermatogenesis respectively, and like RDE-1 can trigger the biogenesis of the RdRP-mediated WAGO-dependent 22G-RNAs (Conine et al., 2010; Gent et al., 2010; Han et al., 2009; Vasale et al., 2010; Zhang et al., 2011). Interestingly, 26G-RNAs are also RdRP products themselves, but unlike 22G-RNAs, 26G-RNAs are monophosphorylated and are processed by Dicer in the context of the Dicer/ERI protein complex (Duchaine et al., 2006; Thivierge et al., 2011). Precisely how 26G-RNAs are processed by the Dicer/ERI complex and how 26G-RNAs acquire their 5' mono-phosphorylated state are largely unknown.

The CSR-1 Argonaute, which has been proposed to promote or modulate rather than silence germline gene expression (Kirino and Mourelatos, 2007; Langmead et al., 2009; Lee et al., 2012; Lee et al., 2010; Seth et al., 2013), engages 22G-RNAs targeting the majority of germline-expressed mRNAs, including many spermatogenesis mRNAs regulated by the ALG-3/4 26G-RNA pathway (Conine et al., 2010; Conine et al., 2013; Han et al., 2009). However, the vast majority of mRNAs targeted by CSR-1 22G-RNAs are expressed outside of spermatogenesis (Claycomb et al., 2009), and 26G-RNAs targeting these non-spermatogenesis CSR-1 targets have not been identified. Thus, it is not known whether 26G-RNAs and Dicer regulate non-spermatogenesis CSR-1 22G-RNAs.

Here we show that *C. elegans* PIR-1 is an RNA polyphosphatase required for germline development and for endogenous small RNA pathways. *pir-1* mutants exhibit a strong depletion of 22G-RNA species that depend on ALG-3/4 for their amplification, but also exhibit a striking more than 2-fold reduction in nearly all CSR-1 22G-RNAs. Recombinant PIR-1, like its vertebrate and viral homologs, removes γ and β phosphates from ppp-RNAs. Catalytically dead PIR-1 binds ppp-RNAs but not p-RNAs *in vitro*. Null and catalytically dead *pir-1* mutants exhibit dramatically delayed larval development and male and hermaphrodite infertility. PIR-1 copurifies with the DCR-1/ERI complex and PIR-1 activity is essential to make ERI-dependent 26G-RNAs that engage ALG-3/4. Our analyses suggest a model whereby 26G-RNAs are made in a unique phased manner by successive rounds mRNA processing by the Dicer/ERI complex, and that PIR-1 promotes this mechanism by removing a diphosphate group from the 5'-end of 26G-RNA precursors, likely to facilitate loading into Argonautes. Our findings implicate PIR-1 as a regulator of endogenous Argonaute pathways that process their small-RNA cofactors from RdRP products.

RESULTS

C. elegans PIR-1 is an RNA polyphosphatase

Previous studies have shown that vertebrate and viral homologs of PIR-1 have polyphosphatase activity that depends on a conserved cysteine in the catalytic motif HCX$_5$RXG (Figure 1A; (Deshpande et al., 1999; Takagi et al., 1998; Yuan et al., 1998). To characterize the enzymatic activity of *C. elegans* PIR-1, we purified recombinant wild-type (WT) PIR-1 protein, as well as recombinant mutant PIR-1(C150S), in which the catalytic cysteine is replaced with serine (Figure S1A). We then incubated ppp-RNAs with these proteins, and assessed the 5' phosphorylation status of reaction products using Terminator® exonuclease, which degrades 5' p-RNA but not diphosphorylated (pp-) or ppp-RNAs. WT PIR-1, but not PIR-1(C150S), efficiently converted ppp-RNAs into substrates that were degraded by Terminator exonuclease, including substrates ranging in size from 26 to 110 nts (Figures 1B and S1B). WT PIR-1 also dephosphorylated ppp-RNAs duplexed with RNA or DNA (Figures 1C and S1C). Notably, PIR-1(C150S), but not WT PIR-1 protein, remained bound to ppp-RNA but not p-RNA substrates in electrophoretic mobility shift assays (Figures 1D and S1D); this shift was caused by PIR-1(C150S) rather than by contamination, as confirmed by western blot analyses (Figure 1E). This mobility shift was detected using 50 mM Tris-Cl buffer (pH 8.0) but not when native protein gel buffer containing 25 mM Tris and 192 mM Glycine (pH 8.3) was used (Figure S1E), indicating that PIR-1(C150S) binds

ppp-RNA non-covalently. Thus, like its vertebrate and viral homologs PIR-1 is an RNA polyphosphatase that converts ppp-RNA to p-RNA. Whereas WT PIR-1 rapidly releases p-RNA products, the catalytically dead PIR-1(C150S) selectively binds and remains bound to ppp-RNA substrates.

PIR-1 associates with the ERI complex

C. elegans PIR-1 was previously identified as a DCR-1-interacting protein (Duchaine et al., 2006). To characterize PIR-1 complexes, we performed PIR-1 immunoprecipitation (IP) and analyzed the immunoprecipitates using Multidimensional Protein Identification Technology (MudPIT; (Wolters et al., 2001). To facilitate the identification of proteins that specifically interact with PIR-1, we rescued a *pir-1*(*tm3198*) null mutant with a *pir-1::gfp* transgene and labeled the *pir-1::gfp* worms with light nitrogen (^{14}N). In parallel, we labeled control WT worms with heavy nitrogen (^{15}N). We then mixed ^{14}N-labeled *pir-1::gfp* worms with an equal number of ^{15}N-labeled control worms, prepared worm lysates, and immunoprecipitated PIR-1::GFP using anti-GFP antibodies. We analyzed the GFP immunoprecipitates by MudPIT and identified candidate PIR-1 interactors as proteins with a minimum of 10 spectral counts for ^{14}N-labeled peptides and no spectral counts for ^{15}N-labeled peptides. These studies revealed that PIR-1 interacts with the core proteins of the ERI complex (Tables 1 and S1; (Kennedy et al., 2004; Pavelec et al., 2009; Simmer et al., 2002; Thivierge et al., 2011; Timmons, 2004). Similar results were obtained using a *pir-1::3×flag*-rescued strain with FLAG IP (Table S1).

Using western blot analyses, we confirmed that DCR-1, DRH-3, RRF-3, ERI-1b, and RDE-8 interact with PIR-1 at all developmental stages (Figures 2A–2B and S2A). PIR-1 did not co-IP with ERI-1a (Figure 2A), an isoform of ERI-1 known to processes the 3′ end of 5.8S rRNA (Gabel and Ruvkun, 2008). Several Argonaute-dependent small RNA pathway factors that are not part of the ERI complex, including the 3′-to-5′ exonuclease MUT-7, the RdRPs RRF-1 and EGO-1, and the Argonautes CSR-1 and WAGOs, were not detected in PIR-1 immunoprecipitates (Figures 2A–2B and S2A, Tables 1 and S1). These western blot studies identified two PIR-1 isoforms, PIR-1a and PIR-1b, which differ in size by approximately 2 to 4 kDa on denaturing polyacrylamide gels (Figures 2 and S2). The molecular basis for this difference remains to be identified. Both isoforms were detected at all larval and adult stages. However, only PIR-1b was detected in embryos, where it associated with several components of the ERI complex (Figure 2B). The association of PIR-1b with the ERI complex was confirmed by gel filtration chromatography in which PIR-1b associated with a >440-kDa complex that included DCR-1, DRH-3, ERI-1b, and RDE-8 (Figure 2C; (Thivierge et al., 2011). Genetic analyses revealed that the interaction between PIR-1 and the ERI complex depends on DCR-1 and DRH-3 but not on ERI-1 or RDE-4 (Figures S2B–S2D). As expected, reciprocal immunoprecipitation of DRH-3 or DCR-1 pulled down PIR-1 (Figure S2E; (Duchaine et al., 2006; Gu et al., 2009). Interestingly, PIR-1a::GFP expression, as detected by GFP IP, appeared to depend on *drh-3(+)* activity, while PIR-1b::GFP expression required *dcr-1(+)* activity (Figures S2B–S2C).

PIR-1 is an essential protein broadly localized to nucleus and cytoplasm

A previous study identified PIR-1 as a Dicer interactor and described a mutation, *pir-1(tm1496),* which causes a fully penetrant larval lethal phenotype at the early L4 stage (Duchaine et al., 2006). The *tm1496* deletion also removes the promoter and part of the neighboring essential gene *sec-5* (Figure 3A), perhaps contributing to the early L4 arrest phenotype. To further explore the function of PIR-1, we generated a second deletion allele (*tm3198*) and a catalytic C150S mutant allele (*wg1000*; Figure 3A). These new alleles caused identical fully penetrant phenotypes. Homozygotes matured more slowly than wild-type animals, arresting at late larval and adult stages and were invariably sterile (Figure 3 and see below). The *tm3198* allele deletes the first intron and most of the second exon of *pir-1* (Figure 3A), which is expected to shift the *pir-1* open reading frame and causes premature translation termination. Moreover, the lethal phenotypes associated with *tm3198* were fully rescued by a single-copy *pir-1::gfp* fusion gene driven by the *pir-1* promoter and 3' UTR, indicating that the *tm3198* phenotypes result from loss of *pir-1*(+) activity.

Analysis of PIR-1::GFP revealed nuclear and cytoplasmic staining in most germline and somatic cells (Figures 3B and S3A–S3B). In the germlines of L4-stage hermaphrodite worms (i.e., during spermatogenesis), PIR-1::GFP was uniformly present in germ cells from the proliferative mitotic zone to the meiotic mid-pachytene region (Figure 3B). PIR-1::GFP fluorescence was reduced in germ cells transitioning through diplotene and meiosis I and II (i.e., through the bend in the ovotestis), and then increased again just before cells begin spermatogenesis. In adult hermaphrodites (i.e., during oogenesis), we detected PIR-1::GFP in the distal germline and through the bend in the ovotestis, but we did not detect PIR-1::GFP signal in maturing oocytes nor in the embryonic germline. PIR-1::GFP fluorescence was detected in most somatic nuclei throughout development, exhibiting the highest level in the large polyploid nuclei of intestinal cells (Figure S3A).

The majority of *tm3198* homozygotes (64%) arrested as sterile adults (Figure S3C), frequently with a protruding vulva and occasionally ruptured at the vulva (Figures 3C and S3D). Approximately 21% of worms made deformed oocytes, but none made progeny (Figure S3D). Approximately one quarter of *tm3198* animals arrested as viable L4-like larvae that survived for nearly a normal life span with apparently normal motility. Close examination of the germlines of these L4-like arrested larvae revealed features typical of normal L4 germline including a mitotic zone, a transition zone, an extended zone of meiotic nuclei undergoing pachytene, and a spermatogenic zone including spermatocytes and spermatids (Figure 3D). A PGL-1::RFP reporter was localized in a WT pattern throughout the distal germline but not in the proximal spermatogenic region of these arrested L4-like worms, suggesting that they transitioned properly to spermatogenic gene expression (Figure 3E). We noticed that many dividing spermatocytes in *pir-1* germlines exhibited abnormal meiotic figures, indicative of DNA-bridging (Figure 3F). Similar defects were previously described for mutants in ERI components. For example, loss-of-function mutations in *rrf-3*, *eri-1*, and *eri-3*, and the helicase-domain mutant *dcr-1(mg375)* have all been reported to cause similar DNA-bridging phenotypes when grown at 25°C (Figure 3F). These ERI pathway mutants all make defective spermatids (Conine et al., 2010; Han et al., 2009; Simmer et al., 2002). To summarize, *pir-1* mutants exhibit a spectrum of defects at larval and

adult stages similar to, and in some respects, such as the larval arrest and oogenesis defects, more severe than other Dicer-ERI complex co-factors.

PIR-1 is not required for miRNA or piRNA biogenesis

We next explored how *pir-1* mutations affect endogenous small RNA levels. To obtain large numbers of *pir-1* homozygotes, we used a strategy to select against heterozygotes in which *pir-1* is covered by the inversion balancer *mnC1.* Three redundant glutamate-gated chloride channels (AVR-14, AVR-15, and GLC-1) render *C. elegans* sensitive to the nematicidal drug ivermectin (Dent et al., 2000). We crossed *pir-1* into an *avr-14(ad1302); avr-15(ad1051); glc-1(pk54)* triple mutant (*avr3x*) background, and balanced *pir-1* with an *mnC1* balancer that also carries a rescuing *avr-15(+)* transgene (Table S2). In the presence of ivermectin, the *pir-1/mnC1* heterozygotes (expressing AVR-15) arrest as L1 larvae, but *pir-1* homozygotes (not expressing AVR-15) grow to late larval stages and adulthood. We grew synchronized populations of *pir-1* homozygous or control (*avr3x* or N2) worms to extract RNA and generate small RNA libraries for high throughput sequencing (see Experimental Procedures). We noted that *pir-1* mutants grew more slowly, both in size and developmental landmarks (e.g., adult cuticle and vulval differentiation), so we prepared samples from *pir-1* mutants grown for 3 days or for 7 days to attain parity in developmental stage with WT populations. To obtain a snapshot of all the different classes of Argonaute-associated small RNAs, we pretreated the small RNA samples with Tobacco Acid Pyrophosphatase (TAP) or with purified recombinant PIR-1 protein, both of which convert ppp-RNAs to p-RNAs. This approach allowed us to simultaneously recover p-RNAs (26G-RNAs, miRNAs, and piRNA/21U-RNAs) and ppp-RNAs (22G-RNAs).

Analysis of the small RNA sequencing data revealed that miRNA and piRNA species were largely unaffected in *pir-1* mutants (Figure 4A). Comparing small RNAs from temporally matched *pir-1* and control populations (i.e., *pir-1* and *avr3x* on ivermectin for 3 days), we found that *pir-1(tm3198)* expressed more miRNAs but fewer piRNAs and 22G-RNAs (when normalized to total genome mapping reads, including all authentic small RNA species). These findings are likely caused by the developmental delay of *pir-1(tm3198)* worms which causes a relatively smaller germline-to-soma ratio in the 3-day old *pir-1* worms, artificially enriching miRNAs since miRNAs are abundant in the soma, whereas 21U-RNAs and most 22G-RNAs are expressed in the germline (Figure 4A). Consistent with this idea, in 7-day *pir-1(tm3198)* animals, which appear developmentally similar to 3 day WT or *avr3x* worms, piRNAs and miRNAs were increased to a similar level, and in proportion to the corresponding decrease in 22G-RNA levels (Figure 4A). Moreover, developmentally matched (7-day) *pir-1* and control (3-day) *avr3x* worms expressed similar levels of DCR-1 and PRG-1 proteins, factors required for generating miRNAs and binding piRNAs respectively. These observations suggest that the biogenesis of miRNAs and 21Us is temporally delayed in *pir-1* mutants but is not likely to be directly regulated by PIR-1(+) activity (Figure 4B). Indeed, when we normalized our small RNA data to piRNA levels, we observed similar levels of miRNAs in control worms and developmentally matched *pir-1* mutants, but 22G-RNA levels were significantly lower in *pir-1* worms (Figure 4C). Thus *pir-1* mutants do not exhibit defects in miRNA levels, consistent with our finding that the

seam cell numbers (16 on each side of the worm) and adult alae differentiation (hallmarks of miRNA function) are not perturbed in *pir-1* mutants.

PIR-1 is required for ERI pathway 26G- and 22G-RNAs

Among the most dramatically affected small RNA species in *pir-1* mutants were 26G-RNAs that depend on the ALG-3/4 Argonautes. These 26G-RNAs are templated from the mRNAs of 1683 target genes, including many genes that play critical roles during spermatogenesis (Table S3;(Conine et al., 2010). We found that 26G-RNAs were approximately 10-fold less abundant in *pir-1* mutants than in WT populations (normalized to 21U-RNA levels; one-tailed t-test, $P < 0.005$; Figures 4D and S4A). Moreover, 22G-RNAs that are amplified downstream of ALG-3/4 targeting (Conine et al., 2010) were also significantly lower in *pir-1* mutants (~2.4-fold; one-tailed t-test, $P < 0.0029$; Figures 4D and S4B). ALG-3/4-independent WAGO 22G-RNAs were not significantly downregulated in *pir-1* mutants (one-tailed t-test, $P < 0.109$, or two-tailed $P < 0.218$, Figures 4D and S4C; (Conine et al., 2010; Han et al., 2009; Pavelec et al., 2009).

All CSR-1-bound 22G-RNAs are reduced in *pir-1* mutants

The CSR-1 Argonaute engages 22G-RNAs targeting thousands of germline mRNAs. Roughly 11% (~423) of CSR-1 target genes are also targeted by the ALG-3/4-dependent ERI pathway (Table S3;((Conine et al., 2010). However, most CSR-1 target genes have no known upstream Argonautes. We found that compared to WT worms, *pir-1* mutants make significantly (~3-fold) fewer 22G-RNAs for both categories of CSR-1 target genes (one-tailed t-test, $P < 0.034$ and 0.033, respectively Figures 4D and S4D). Both classes of CSR-1 22G-RNAs exhibited similar ratios of reads in the mutant to total reads in the mutant and WT (mutant/[mutant+WT]), with the same medians and similar variances (Figure 4E). Moreover, both classes of CSR-1 22G-RNAs were significantly and dramatically reduced in the *pir-1(C150S)* catalytic mutant ($P<0.01$ or lower; Figure S4E). We also observed a mild (~20%, same as the level in the *tm3198*) but statistically significant reduction of WAGO 22G-RNAs in the *pir-1(C150S)* mutant. Together, these results suggest that the catalytic activity of PIR-1 is required for the biogenesis of CSR-1 small RNAs.

26G-RNAs are generated in a phased manner

26G-RNAs are unique among *C. elegans* small RNA species in that their biogenesis depends on both RdRP and Dicer. However, why they are longer than typical Dicer products and how their 5’ ends become mono- instead of triphosphorylated as is typical of other *C. elegans* RdRP products remains mysterious. To investigate the role of PIR-1 in 26G-RNA biogenesis we used bioinformatics to analyze the distribution patterns of small RNAs associated with 26G-RNA target sites. To do this we compiled a metagene analysis of all available ALG-3/4 and ERGO-1 26G-RNA target sequences centered on the 26G-RNA and including about 40-nt upstream and downstream sequences. We then analyzed small RNAs mapping to this interval including both antisense small RNAs (RdRP-derived) and sense small RNAs (from mRNA cleavage). The frequency of each small RNA species was plotted according to its 5’ nt position, and color-coded according to its length (Figures 5 and S5A). The position of the C-residue of the mRNA corresponding to the 5’ G of the antisense 26G-RNA was defined as −1. As expected, for both ALG-3/4 and ERGO-1 26G-RNA pathways, the most abundant

antisense species were 26G-RNAs located at the −1 position (Figures 5 and S5A). Consistent with a previous study (Blumenfeld and Jose, 2016), this analysis revealed additional phased 26G-RNA peaks located at approximately 23-nt intervals upstream and downstream of −1 (Figures 5A and S5A). Mirroring the central and phased 26G-RNAs, we observed an identical distribution pattern of mRNA (i.e., sense-stranded) fragments that likely correspond to Dicer products (Figures 5 and S5A). For the ALG-3/4 pathway, most of these mRNA fragments were 22 nucleotides long (sense 22mer-RNA) with their 5' ends at −23 and their 3' ends at −2, just upstream of the −1 C residue (Figure 5A and S5A). For the ERGO-1 pathway the most abundant sense-stranded small RNAs were 19 nts long, with 5' ends at −23 and 3' ends at −5 (Figure 5B). These findings suggest that associated nucleolytic activities removes the −1 C residue (and a few additional nucleotides for ERGO-1 templates) after it templates 26G-RNA initiation (see Discussion). For both the ALG-3/4 and ERGO-1 pathways, the sense RNA 5' ends align 3 nucleotides downstream of the 26G-RNA 3' ends. Taken together, these findings suggest that the sense RNAs positioned at −23 in the metagene analysis represent a signature of Dicer processing on duplex 26G-RNA precursors (see Discussion).

The above analysis suggests that template mRNAs are processed stepwise by RdRP and Dicer, with RdRP initiating at a C residue and then re-initiating recursively at the first available C residue after each Dicer cleavage event. We further tested this idea by simulating 26G-RNA biogenesis on a computer-generated transcriptome containing random RNA sequences and 26G-RNA densities similar to those in our data sets. Simulated 26G-RNAs were generated on targets by initiating at a randomly selected C residue and then recursively at the first C residue at least 23-nt upstream of the initial template C, propagating the 26G-RNA synthesis toward the 5' end of target mRNAs in a unidirectional manner. Strikingly, this simulation produced exactly the same metagene pattern observed in our experimental data including the overall shape, the symmetric phased distribution, the loss of phasing at distances greater than 40 nts, and other minor details (Figure S5B–S5D). More details and findings of this simulation analysis are provided in Figure S6A–B.

pir-1 mutants exhibit defects in 26G-RNA maturation

A metagene analysis based solely on the *pir-1* mutant data proved impractical, as the depletion of 26G-RNAs made the signal very weak (Figure S6C). Instead, we mapped sense and antisense RNAs cloned from the *pir-1* mutants to the metagene intervals defined by the wild-type data. Both 26G-RNAs and complementary sense-stranded 22mer RNAs (positioned at −23 in the metagene) were dramatically reduced in *pir-1* mutants (Figure 6A). Interestingly, 26G-RNAs were disproportionately reduced compared to the sense-stranded 22mer RNAs (Figure 6B). The ratio of 26G-RNAs to −23 22mer-RNAs was 31:1 for WT animals, 3:1 for the *pir-1* null mutant and 11:1 for the *pir-1(C150S).* Phased 26G-RNAs were not detected. These finding suggest that PIR-1 activity is required both for the processivity of the Dicer ERI complex and for the maturation of antisense 26G-RNAs.

Since PIR-1 is an RNA phosphatase, one possible explanation for the above finding is that dephosphorylation of the 26G-RNA precursors promote maturation. If so, we reasoned that unprocessed ppp-26G-RNAs should increase relative to p-26G-RNAs in *pir-1* mutants, and

that this difference might be enhanced by binding and protection of the unprocessed ppp-RNA by catalytically dead PIR-1(C150S), as predicted by our *in vitro* studies above. To explore this possibility, we generated small RNA sequencing libraries using a ligation-dependent method that requires a 5' monophosphate for efficient cloning (Gu et al., 2011; Li et al., 2019). For each mutant and WT sample, we prepared libraries with or without pretreating the RNA with recombinant PIR-1. As expected, most (94%) 26G-RNAs in WT worms were recovered without PIR-1 digestion when normalized to those with recombinant PIR-1 treatment, suggesting that these 26G-RNAs bear 5' monophosphate. In contrast, we found that ~40% of 26G-RNAs present in the *pir-1* mutants were resistant to ligation-dependent cloning unless treated with recombinant PIR-1, suggesting that they contain a 5' triphosphate group (Figures 6C–6D). Taken together these findings suggest that PIR-1 dephosphorylates 26G-RNA precursors and is required for efficient 26G-RNA maturation by the Dicer ERI complex.

DISCUSSION

Eukaryotic cells can sense and modify structural features of RNAs to regulate their stability and functions, and to distinguish self- from viral-RNAs. For example, the Dicer protein binds dsRNAs and processes them into duplexed siRNAs and miRNAs that engage Argonaute proteins to mediate sequence-specific viral immunity and mRNA regulation. Conversely, the human RIG-I protein, which contains a Dicer-related helicase domain, detects duplex ppp-RNAs produced by viral RdRPs and then initiates a non-sequence specific cascade of secondary signals that promote viral immunity (Hornung et al., 2006; Kato et al., 2006). Here we have shown that the Dicer-interacting protein PIR-1, like its human and insect virus homologs, removes the β and γ phosphates from ppp-RNAs *in vitro*, generating 5' p-RNAs. *In vivo,* PIR-1 is required for fertility and for the accumulation of 26G-RNAs antisense to hundreds of spermatogenesis mRNAs.

26G-RNAs are an enigmatic species of Dicer product best understood for their role in spermatogenesis where along with their AGO-related Argonaute co-factors, ALG-3/4, they promote spermatogenesis-specific gene regulation and epigenetic inheritance (Conine et al., 2010; Conine et al., 2013; Han et al., 2009). During embryonic development 26G-RNAs engage the Argonaute ERGO-1 to regulate a group of repetitive RNAs of unknown functions (Gent et al., 2010; Vasale et al., 2010). Mutations that inactivate the ERGO-1 pathway cause enhanced RNAi (ERI phenotypes; (Kennedy et al., 2004; Simmer et al., 2002). While our genetic studies only revealed a role for PIR-1 in the larval-stage ALG-3/4 26G-RNA pathway, it is likely, as previously shown for other RNAi components including Dicer and RDE-1 (Parrish and Fire, 2001; Tabara et al., 1999; Tabara et al., 2002), that the embryonic functions of PIR-1, including its possible function in the ERI pathway are rescued in embryos of heterozygous mothers by maternally provided PIR-1(+) activity.

A model for 26G-RNA biogenesis

Our findings are consistent with the idea originally proposed by Blumenthal and Jose (2016) that 26G-RNAs are produced in a phased manner along mRNAs through successive cycles of antisense transcription by RRF-3 and cleavage by Dicer. Our findings allow us to add

details to this model for phased biogenesis of 26G-RNAs, and to propose where PIR-1 functions in this process (Figure 6E). After transcription by RRF3, at least 23 nucleotides are removed from the 3′ end of the template RNA. Then, before Dicer cleaves, a 3′-to-5′ exonuclease, possibly ERI-1b, digests the template RNA, removing the transcription start site C residue (–1 in the model; Figure 6E) to generate a dsRNA with a 1-nt recessed 3′ end. Binding of the recessed 3′ end and engagement of the Dicer helicase domain positions Dicer to process the duplex into a 22mer-RNA passenger strand (by cleavage at −23 relative to the initiator C residue) and a 26G-RNA strand with a 3-nt 3′ overhang (Figure 6E; see also Welker et al., 2011). The cycle is repeated when RRF-3 re-initiates transcription at the C residue closest to the processed 3′ end of the template. Our analysis failed to uncover evidence for processive dicing of longer RdRP-derived dsRNA substrates. We cannot, however, rule out the possibility that longer substrates are generated and diced processively at much lower frequencies.

PIR-1 could remove the diphosphate from ppp-26G-RNA before or after dicing (Figure 6E). Indeed, *in vitro* studies suggest that Dicer is not sensitive to the phosphorylation status of the substrate 5’ end (Welker et al., 2011; Zhang et al., 2002). Moreover, the levels of sense-stranded 22-mer RNA fragments (presumptive Dicer products) and ppp-26G-RNAs were increased in *pir-1* mutants compared to WT worms. Thus, dicing still occurs in *pir-1* mutants but maturation into p-26G-RNAs appears to be reduced. Perhaps diphosphate removal is required for efficient transfer of diced 26G-RNA products to the Argonautes ALG-3/4, whose homologs prefer mono-phosphorylated guide RNAs.

PIR-1 exhibits ppp-RNA-specific binding activity

In vitro studies on PIR-1 revealed a surprising activity associated with the presumptive catalytically dead C150S lesion. This mutation behaved like a strong loss of function allele, causing small-RNA and developmental defects identical to those caused by a *pir-1* null mutation. However, we found that PIR-1(C150S) nevertheless bound specifically to ppp-RNAs in our gel-shift assays. Structural studies on members of the cysteine phosphatase superfamily to which PIR-1 belongs have shown that during catalysis the cysteine motif generates a covalent cysteinyl-S-phosphate intermediate that is later hydrolyzed in a two-step reaction (Sankhala et al., 2014; Takagi et al., 1998). The substitution of serine for cysteine in PIR-1 C150S replaces the reactive sulfhydryl group of cysteine with a hydroxyl group, preventing formation of the covalent linkage. The finding that this catalytically dead protein retains its ppp-RNA-specific binding activity suggests that substrate recognition is separable from catalysis in PIR-1. Thus it is possible that PIR-1 utilizes its affinity for ppp-RNAs to recognize RRF-3 products and to help recruit Dicer and other ERI complex co-factors to the nascent duplex.

It is interesting to note that a baculovirus-encoded PIR-1 homolog, PTP, functions as a virulence factor that promotes a fascinating behavioral change in infected host caterpillars (Katsuma et al., 2012). Ingested virus spreads to the brain, and the infection eventually causes the caterpillar to migrate to upper foliage, where the dying animal ‘liquifies’—a process thought to maximize dispersal of the virus. Interestingly, *ptp* null mutants were partially defective in brain infectivity and behavioral modification, but PTP C119S mutants

supported both activities, suggesting that PTP provides a purely structural capacity to promote virulence, e.g., through its interaction with viral capsid protein (Katsuma et al., 2012). However, if PTP C119S selectively binds ppp-RNA—similar to PIR-1 C150S—then it remains possible PTP C119S interacts with and promotes viral packaging of cellular or viral ppp-RNAs that function as small-RNA cues that alter host behavior. This possibility is particularly intriguing as a growing number of reports have described the modulation of neural and behavioral activity by small RNAs originating in other tissues (Bharadwaj and Hall, 2017; Cai et al., 2018; Hou et al., 2019; Posner et al., 2019).

PIR-1 is required for robust levels of CSR-1 22G-RNAs

We were surprised to find that *pir-1* mutants exhibit significantly reduced levels of all CSR-1 22G-RNAs. The biogenesis of 22G-RNAs does not require Dicer. Instead 22G-RNAs appear to be produced directly by the RdRP EGO-1 and are then loaded, without further processing, as ppp-RNAs onto their downstream Argonaute co-factors. It is therefore intriguing that CSR-1 22G-RNAs but not WAGO 22G-RNA levels were depleted in *pir-1* mutants. The upstream events in the WAGO 22G-RNA pathway differ from events involved in the CSR-1 pathway. For example, WAGO 22G-RNA biogenesis is initiated by RDE-1 guided by an siRNA processed by Dicer or by Piwi Argonaute (PRG-1) guided by a piRNA. When RDE-1 and PRG-1 bind target mRNAs, they recruit cellular RdRPs that synthesizing WAGO 22G-RNAs (Ashe et al., 2012; Bagijn et al., 2012; Grentzinger et al., 2012; Lee et al., 2012; Pak and Fire, 2007; Shen et al., 2018; Shirayama et al., 2012; Yigit et al., 2006; Zhang et al., 2018). Whether an upstream Argonaute functions in the CSR-1 pathway is unknown. Although 26G-RNAs have not been detected for most CSR-1 targets, perhaps they are short lived, developmentally restricted, for example to larvae, or are simply very low abundance, and have been missed. Further investigation will be required to understand this connection between PIR-1 and CSR-1.

Limitations of Study

Here we have shown that PIR-1 is an RNA phosphatase and that null and catalytic alleles exhibit delayed development, larval arrest, and reduced levels of small RNAs in two Argonaute pathways, the ALG-3/4 pathway and the CSR-1 pathway. A striking feature of the PIR-1 mutant phenotype is the dramatically slowed development of homozygous larvae. Those *pir-1* homozygotes that do reach adulthood take nearly twice as long as WT animals to do so, while their developmentally retarded siblings behave like otherwise active and healthy larvae for the course of an approximately WT lifespan of 16 to 18 days. The developmental delay of *pir-1* limits our ability to compare its patterns of RNA and protein expression to that of WT animals. We must either use chronologically matched or developmentally matched populations, and some developmental stages, such as embryos are entirely absent due to infertility. Previous studies have shown that the loss of ALG-3/4 and CSR-1 pathways cause defects in male and female fertility respectively. However, mutants that perturb these Argonaute pathways do not exhibit delayed development (Claycomb et al., 2009; Conine et al., 2010; Gent et al., 2009; Han et al., 2009; Vasale et al., 2010). Moreover, the *pir-1* sterile phenotype is more severe than that of *alg-3/4* or *csr-1* mutants. For example, *alg-3/4* and *csr-1* males are partially fertile at the permissive temperature of 20°C (Conine et al., 2013), but *pir-1* males are unconditionally sterile. *csr-1* hermaphrodites make some

embryos, but most *pir-1* adult hermaphrodites do not even make oocytes. Unlike the ALG-3/4 and CSR-1 Argonautes, PIR-1 is expressed throughout the worm, and thus PIR-1 may have functions outside the germline that are required for both developmental progression and fertility. Thus we do not know why *pir-1* mutants exhibit arrested development and this defect is unlikely to stem from its role in regulating germline Argonautes. Perhaps the presence of cellular RdRPs in *C. elegans* makes RNA phosphatase activity essential in order to ensure that accumulating ppp-RNA products do not compromise RNA homeostasis or activate heretofore unknown innate immunity mechanisms. Conceivably, the absence of PIR-1 activity could trigger a diapause that is normally triggered only when an excessive cytoplasmic accumulation of viral ppp-RNAs overwhelms the capacity of PIR-1 and Dicer mediated immunity. A diapause in response to ppp-RNA might allow animals to postpone reproduction until after the viral infection is cleared. Understanding the essential role(s) of PIR-1 in development and its possible function in anti-viral immunity will require further investigation beyond the scope of this paper.

STAR*METHODS

RESOURCE AVAILABILITY

Lead Contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Weifeng Gu (weifeng.gu@ucr.edu).

Materials Availability—Worm strains generated in this study (Table S2) are available from the Caenorhabditis Genetics Center (CGC) and can also be requested from the lead contact. The E. coli strains for PIR-1 expression can be requested from the lead contact or obtained from Addgene.

Data and Code Availability—High-throughput sequencing data is available from the GEO DataSets under the series number GSE150690. Original data for figures in this paper is available at Mendeley data http://dx.doi.org/10.17632/g3z3k2fppy.1.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

C. elegans Strains and Genetics—The *C. elegans* Bristol N2 strain and its derivatives used in this study were cultured essentially as described (Brenner, 1974). NGM plates containing 10–25 μg/L ivermectin were used to select for *pir-1* homozygous worms. Worms strains used in this study are listed in Table S2.

METHOD DETAILS

Cloning, Expression, and Purification of Recombinant PIR-1—Wild-type (WT) or mutant PIR-1 cDNA sequences lacking the first ATG was inserted between the *Nde*I site and *Bam*HI sites of pET-28a (Novagen) in fusion with the 6× Histidine tag N-terminally. The resulting plasmid was transformed into BL21 (DE3) RIL *E. coli* cells, which were grown in 1 liter of LB medium at 37°C to an OD_{600} of 0.4, and induced for 4 hr with 1 mM IPTG at room temperature. Cells were pelleted at 5,000× *g* for 10 min at 4°C and lysed by sonication in 25 ml of lysis/binding buffer (50 mM Tris-HCl pH 7.5, 700 mM NaCl, 5 mM β-mercaptoethanol, 5% glycerol, 15 mM imidazole, 0.01% NP-40). S100 fractions were

prepared by ultracentrifugation at 100,000 × *g* at 4°C for 1 hr. In a 15-ml conical tube, 2 ml of HisPur beads (Thermo Scientific) were washed 3 times with the lysis/binding buffer and centrifuged at 3,000 × *g* between washes. The beads were mixed with the S100 supernatant, transferred to a 50-ml conical tube for rotation at 4°C for 1 hr. Beads were transferred to an empty Poly-Prep chromatography column (Bio-Rad) and washed at 4°C with at least 200 bead volumes of the lysis/binding buffer. Elution was performed at 4°C with 500 μl of imidazole buffer per fraction (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 5 mM β-mercaptoethanol, 5% glycerol, 400 mM imidazole, 0.01% NP-40). Peak fractions were analyzed by 10% SDS-PAGE followed by Coomassie Blue staining. Proteins were dialyzed using 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 50% glycerol, 0.01% Triton X-100.

PIR-1 Activity Assays—To examine the dephosphorylation activity of recombinant PIR-1, a 26-nt long ppp-RNA1 (ppp-GGAUCCUUGAAAUGGAACAUCUGAAU) and a 103-nt long ppp-RNA2 (GUUGUAGUGUUCCCGCUCCAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU) were transcribed *in vitro* with T7 RNA polymerase followed by gel-purification using 15% PAGE/6M urea (two bands were co-purified for both RNA1 and RNA2). In Figure1B (RNA1) and S1B (RNA2), 1 μM of ppp-RNA was co-digested with ~0.25 μM of recombinant WT or mutant PIR-1 and 0.25 U of Terminator exonuclease (Epicentre) in 10 μl 1X PIR-1 reaction buffer containing 50 mM Tris-HCl (pH 8.0) and 0.1 M NaCl, 2 mM DTT, and 2 mM $MgCl_2$ at 30°C for 1 hr. The reaction was stopped by adding formamide gel loading buffer II (Ambion), and run on a 15% PAGE/6 M urea with 0.5X TBE buffer. The RNA was visualized with UV light after staining with SYBR Gold (Thermo Fisher Scientific).

The above *in vitro* transcription predominantly generates byproduct RNAs of much bigger size likely due to template switching when T7 RNA polymerase runs off a template. This prompted us to generate a precursor RNA ppp-GUCAUUCAGAUGUUCCAUUUCAAGGAGGGUCGGCAUGGCAUCUCCACCUCCUCGCGGUCCGACCUGGGCUACUUCGGUAGGCUAAGGGAGAAG, which contains a Hepatitis delta virus (HDV) ribozyme (underlined) to self-cleave the precursor, generating ppp-RNA3 co-transcriptionally (ppp-GUCAUUCAGAUGUUCCAUUUCAAGGA; (Schurer et al., 2002). In Figure1C, ppp-RNA3 alone, ppp-dsRNA generated using ppp-RNA3 annealed with an RNA oligo 5'OH-UUGAAAUGGAACAUCUGAAUGAC (the oligo is smaller than ppp-RNA3 and thus can be separated from ppp-RNA3 in gel purification) and ppp-RNA/DNA hybrid generated using ppp-RNA3 annealed with a DNA oligo 5'OH-TTGAAATGGAAC ATCTGAATGAC in 1X PIR-1 reaction buffer (the annealing rate is close to 100% as shown in Figure S1C), digested with recombinant PIR-1 using the above reaction condition, and gel-purified to obtain processed ppp-RNA3. Then these processed RNAs were subjected to digestion with 0.05 U of Terminator in a 10 μl PIR-1 reaction buffer at 30°C for 30 minutes, resolved on a 15% PAGE/6M urea, and visualized using SYBR Gold staining.

In the binding assay, recombinant PIR-1 (no Terminator) was incubated with ppp-RNA1 (Figure 1D and 1E) or double stranded nucleic acids including ppp-RNA3/RNA oligo or ppp-RNA3/DNA oligo (Figure S1D and S1E) using 1X PIR-1 reaction buffer at 20°C for 40 minutes. The reaction was resolved using a 10% native PAGE gel containing 50 mM Tris-HCl (pH 8.0; Figure S1D) or 25 mM Tris and 192 mM Glycine (pH 8.3; Figure S1E) at room temperature and visualized using SYBR Gold staining.

Worm Transgenics—GFP- and FLAG-tagged *pir-1* strains were generated by microparticle bombardment (Praitis et al., 2001) or Mos1-mediated single-copy insertion (Frokjaer-Jensen et al., 2008). Transformants were identified using an *unc-119* transformation rescue strategy. Integrated lines were crossed with *pir-1(tm3198)/mnC1** males, and hermaphrodite progeny not carrying the mnC1 balancer chromosome (which was marked with a nuclear P*sur-5*::GFP transgene) were allowed to self-fertilize. Several F2 were allowed to produce offspring and then analyzed by PCR to worms positive for the transgene but homozygous for the *pir-1* deletion allele.

Gene Editing by CRISPR—The *pir-1(C150S)* mutant was generated by injecting a pre-assembled Cas9 ribonucleoprotein complex and oligos to template homology directed repair, essentially as described (Paix et al., 2015), but injections included *rol-6 (su1006)* as an injection marker. Genome editing events were identified among the F1 rollers.

Immunoprecipitation and Western Blot Analyses—Immunoprecipitation and western blots were performed as described previously (Gu et al., 2009).

MudPIT Analysis

1. Identification of PIR-1 Binding Partners (Table 1 and Table S1): Proteins were precipitated in 23% trichloroacetic acid final concentration and rinsed with cold acetone. Air-dried pellets were dissolved in 60 ul of 8 M urea 100 mM Tris pH 8.5 and reduced with TCEP (5 mM final concentration) for 20 min, followed by alkylation by iodoacetamide (10 mM final concentration) for 15 min. Sample is diluted to 2 M urea 100 mM Tris pH 8.5. Calcium chloride to 1 mM final concentration and 1 ug of trypsin (Promega, product V5111) is added. The sample is placed in a 37 °C shaker for 18 hrs. Formic acid to 5% is added and the sample is centrifuged at 16000 rcf for 15 min. The supernatant is transferred to a new tube for loading onto the MudPit column.

2. Mass Spectrometry analysis by MudPIT (Table 1 and Table S1 Bottom; (Wolters et al., 2001)): Digested protein was pressure loaded onto an in-house biphasic microcapillary column (250 μm id/360 μm od capillary of 30 cm length) packed with a strong cation exchanger (SCX Luna, Phenomenex, Torrance, CA, USA) and RP resin (Aqua C18, Phenomenex, Ventura, CA, USA). Subsequently, an analytical microcapillary column packed with RP resin was attached to the biphasic column in line with an Eksignet HPLC on an LTQ Orbitrap (Thermo Fisher Scientific, San Jose, CA, USA). Samples were analyzed using a 3 step separation with the first step corresponding to a desalting and transfer step. The second step used a 30% pulse of 500 mM ammonium acetate for 5 min followed by a 155 min gradient to 79% ACN. The third step had a 100 % 500 mM ammonium acetate

pulse. Peptides eluted from the microcapillary column (100 μm id/360 μm od capillary of 12 cm length) were electrosprayed directly into the mass spectrometer with the application of distal 2.5 kV spray voltage at an inlet capillary temperature of 200° C. From one full-scan of mass spectrum (400–1800 *m/z*), 5 most intense ions were sequentially isolated and fragmented by CID with 35% normalized collision energy repeating continuously through each step of the multidimensional separation. The *m/z* ratios selected for MS/MS were dynamically excluded for 120 s.

3. Mass Spectrometry (Table S1 Top): Digested protein was pressure loaded onto an in-house biphasic microcapillary column (250 μm id/360 μm od capillary of 30 cm length) packed with a strong cation exchanger (SCX Luna, Phenomenex, Torrance, CA, USA) and RP resin (Aqua C18, Phenomenex, Ventura, CA, USA). Subsequently, an analytical microcapillary column packed with RP resin was attached to the biphasic column in line with an Agilent 1100 quaternery pump coupled to an LTQ (Thermo Fisher Scientific, San Jose, CA, USA). Samples were analyzed using a 7 step separation with the first step corresponding to a desalting and transfer step. Steps 2–7, used 500 mM ammonium acetate pulses of 20%,.30%, 40%, 70%, 100% and 100%, respectively. Peptides eluted from the microcapillary column (100 μm id/360 μm od capillary of 12 cm length) were electrosprayed directly into the mass spectrometer with the application of distal 2.5 kV spray voltage at an inlet capillary temperature of 200° C. From one full-scan of mass spectrum (300–2000 *m/z*), 8 most intense ions were sequentially isolated and fragmented by CID with 35% normalized collision energy repeating continuously through each step of the multidimensional separation. The *m/z* ratios selected for MS/MS were dynamically excluded for 60 s.

4. Mass Spectrometry Data Processing: All mass spectra were converted to ms2 files by RawConverter (Version 1.1.0.23; (He et al., 2015), with monoisotopic peak selection for Orbitrap files. All files were searched against the protein database from www.wormbase.org (WS266), with common contaminants and reversed decoy sequences added (Peng et al., 2003). The searches were done with ProLuCID, version 1.4 (Xu et al., 2015).

For Orbitrap data shown in Table 1 and Table S1-Bottom, the search parameters were; unlimited missed cleavages, precursor ion tolerance 0.3 Da, fragment mass tolerance 600 pm, no variable modification, half or fully tryptic, and fixed cysteine modification of 57.02146. A “heavy” search with the same parameters with 100% ^{15}N incorporation was done also. The LTQ data in Table S1-Top, was searched with; unlimited missed cleavages, precursor ion tolerance 50 ppm, fragment mass tolerance 600 pm, no variable modification, half or fully tryptic, and fixed cysteine modification of 57.02146.

Search results were filtered with DTASelect version v2.1.12 (Tabb et al., 2002). For Orbitrap data- -p 2 -y 1 --trypstat --pfp .01 --extra --pI -DM 5 --DB --dm -in -t 1 --brief –quiet. For LTQ data- -p 2 -y 1 --trypstat --pfp .01 --extra --pI --DB --dm -in -t 1 --brief –quiet.

Preparation of Tissues for Microscopy—To visualize live animals, washed worms were mounted on slides with a 2% agarose pad with M9 buffer containing 0.4% levamisole to paralyze the animals.

Tissues were prepared for DAPI staining and immunofluorescence were carried out as described in Claycomb et al. (2009). For gonad dissection 40 to 50 L4 to young adult worms were picked from plates and washed extensively with 1× Egg Buffer (25 mM HEPES-NaOH, pH 7.4, 118 mM NaCl, 2 mM EDTA, 0.5 mM EGTA, 0.1% Tween-20) to eliminate bacteria. The buffer was replaced with Egg Buffer containing 0.4 mM levamisole (15–30 µl) and transferred onto an 18×18-mm coverslip. Animals were cut with the tip of a fine hypodermic needle at either the head below the pharynx or at the tail to release the gonads (and intestines). An equal volume of fixative solution (3.7% formaldehyde in 1× Egg Buffer without Tween-20) was added and pipetted up and down to further extrude and dissociate germline tissue from the rest of the animals. Fixation was allowed to occur for 5 min at room temperature. All but about 10 µl of solution were removed from the coverslip. The coverslip was picked up by touching the drop at the center of a positively charged slide (VWR VistaVision HistoBond), with one corner of the coverslip slightly protruding from the edge of the slide. To promote adherence of the tissues to the slide, excess fixative was removed from the edge of the coverslip using torn strips of absorbent filter paper. The tissue was freeze-cracked by placing the slide on a pre-cooled aluminum block on dry-ice for at least 10 min and quickly flicking the protruding corner of the coverslip, removing it from the slide. The slide was immediately dipped in cold (−20°C) methanol in a Coplin jar for 1 min, and then transferred to 1× PBS buffer (10 mM phosphate, pH 7.4, 137 mM NaCl, 2.7 mM KCl) containing 0.1% Tween-20 (PBST) at room temperature.

For DAPI staining only, slides were washed in PBST for 10 min, PBST containing 0.5 µg/ml DAPI for 10 min, and PBST for 30 min, at room temperature. Excess buffer was removed from the slides without letting the sample dry completely, and slides were mounted with 10 µl of Vectashield mounting medium placed at the center of a 22×22-mm coverslip. Excess medium was removed by inverting and pressing the mounted slide on a paper towel, and the edges were sealed with transparent nail polish.

For immunofluorescence staining of PIR-1::GFP, slides were immersed in (−20°C) methanol for 1 min, then washed three times in PBST, 10 min each. Slides were blocked by adding 100 µl 0.5% BSA in PBST onto the worms, covering with a square Parafilm coverslip, and incubating in a humid chamber at room temperature for at least 30 min. The slides were dipped in PBST to remove the Parafilm. Mouse monoclonal anti-GFP antibody (Wako) was diluted 1:100 in blocking solution and 100 µl was placed on the sample and covered with a Parafilm coverslip, and slides were incubated in a humid chamber for 2 hr at room temperature or overnight at 4°C. After three 10-min washes in PBST at room temperature, slides were incubated for 2 hr at room temperature with a 1:500 dilution of FITC-conjugated donkey anti-mouse (Jackson), and then washed, stained DAPI, and mounted as described above.

Images of live or fixed samples were acquired with a Zeiss Axioplan 2 microscope using Zeiss AxioVision software.

Small RNA extraction, cloning, and sequencing—RNA was extracted from worms or from Argonaute immunoprecipitates with TRI Reagent (MRC, Inc.), according to the manufacturer’s protocol. Small RNA libraries were prepared essentially as described (Gu et

al., 2011; Li et al., 2019). Briefly, ~1 μg of total RNA was used for cloning small RNAs either via the conventional ligation-based method or the one-pot cloning method; Tobacco Acid Pyrophosphatase (Epicentre, discontinued) or recombinant PIR-1 was used to dephosphorylate ppp-RNAs for cloning ppp-RNAs when needed while no such treatment was required for cloning p-RNAs. Libraries were sequenced using Illumina NextSeq, HiSeq 4000, and Genome Analyzer II at UMass Medical School and UC Riverside.

Quantification and Statistical analysis—High-throughput sequencing reads were processed and mapped to *C. elegans* genome and annotations (WormBase release WS215) using Bowtie 0.12.7 (Langmead et al., 2009) and further analyzed using custom PERL scripts, which have been deposited in GitHub at https://github.com/guweifengucr/WGlab_small_RNA_analysis for free access (Dai and Gu, 2020; Gu et al., 2009; Li et al., 2019). The Generic Genome Browser was used to visualize the alignments (Stein et al., 2002). All the statistics in Figure 4C–D, Figure 6 A–D, and Figure S4E were obtained using unpaired student's t-test provided in LibreOffice 6.1 Calc based on two replicas of each sample; we used F-test from LibreOffice 6.1 Calc to determine if the data follows student's t-test with equal variance or unequal variance. The bar height (Figure 4C–D, Figure 6 A–D, and Figure S4E) and the line (Y-axis, Figure S4A–D) represents the mean; in all the above figures, one standard error of mean (SEM) was indicated; 'n' represents the gene number in the indicated group for counting the total reads. All the above information as well as the normalization standard was provided in the figure legends.

The metagene analysis obtained the distribution of small RNAs (represented by their 5' nt positions) of various sizes flanking 26G-RNAs (also represented by their 5' nt positions, i.e., the −1 template C's). It first obtained the small RNA profile flanking each 26G-RNA and then the accumulative profile using all 26G-RNAs, as shown in Figure 5A–B and Figure S5A. The custom PERL scripts for these analyses were deposited in GitHub, as shown above.

A simulation algorithm was developed to verify the metagene results using the simulated small RNA data and parameters obtained from the experimental data. In each round of simulation, one thousand of RNA molecules, each with a 1000-nt random sequence containing 21% C (frequency in template mRNAs), are generated; 20 C's are randomly selected as −1 C's and used to generate initial 26G-RNAs; if the −24 (23 nts upstream of −1) is C, a phased 26G-RNA is generated; otherwise, the next available upstream template C is selected; the next round of phased 26G-RNAs starting at −47 regions are generated using the same rule and so on (Figure S5B). The selected C's for both the initial 26G-RNAs and phased 26G-RNAs in the ALG-3/4 pathway are limited to the 5' and 3' 10% of mRNAs since most ALG-3/4-bound 26G-RNAs are located there (Conine et al., 2010). To achieve the best result, each C could fail to generate a 26G-RNA at 30–40% rate, and if a failure occurs, next upstream C can serve as a template nt also with a 30–40% failure rate. This failure rate, the only parameter not obtained from the experimental data, simply mimics RNA degradation or other competing processes, and was empirically determined based on the best fitting results. To minimize variations, the average results of 100 rounds of simulations were obtained, as shown in Figure S5C&D. The control utilizes the same

parameters and algorithm but only allows for generating initial 26G-RNAs (no recursive mode).

To examine if phased 26G-RNAs may be caused by a higher frequency of nt C in the −24 template region, the frequencies of nt C's flanking any specified C's (designated as −1) on template mRNAs were examined (Figure S6A–B). This is basically a metagene motif analysis, since each genomic C locus has a weight 1 instead of the read numbers in the metagene analysis (Figure 5 and S5). We also included a motif analysis only using template C's which generate 26G-RNAs in the experimental data (Figure S6A–B). If a 26G-RNA locus is selected, it could represent a phased 26G-RNA, meaning there is an initial 26G-RNA locus at the 23 position and therefore the 23 position enriches C nts. By contrast, the −24 position won't enrich C nts since for any given −1 C the biogenesis of phased 26G-RNAs at −24 just follows the genomic C frequency, i.e., 21%. As expected, this motif analysis exhibits a sub-peak at the 23 position and no obvious sub-peak at −24 (Figure S6A–B).

Table S3. *C. elegans* genes targeted by Argonautes. Related to Figure 4.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by NIH grants GM124349 (W.G.), P41 GM103533 for (J.R.Y.), and GM058800 (C.C.M.). C.C.M. is a Howard Hughes Medical Institute Investigator.

REFERENCES

Ashe A, Belicard T, Le Pen J, Sarkies P, Frezal L, Lehrbach NJ, Felix MA, and Miska EA (2013). A deletion polymorphism in the Caenorhabditis elegans RIG-I homolog disables viral RNA dicing and antiviral immunity. Elife 2, e00994. [PubMed: 24137537]

Ashe A, Sapetschnig A, Weick EM, Mitchell J, Bagijn MP, Cording AC, Doebley AL, Goldstein LD, Lehrbach NJ, Le Pen J, et al. (2012). piRNAs can trigger a multigenerational epigenetic memory in the germline of C. elegans. Cell 150, 88–99. [PubMed: 22738725]

Bagijn MP, Goldstein LD, Sapetschnig A, Weick EM, Bouasker S, Lehrbach NJ, Simard MJ, and Miska EA (2012). Function, Targets, and Evolution of Caenorhabditis elegans piRNAs. Science 337, 574–578. [PubMed: 22700655]

Baulcombe D (2004). RNA silencing in plants. Nature 431, 356–363. [PubMed: 15372043]

Bernstein E, Caudy AA, Hammond SM, and Hannon GJ (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409, 363–366. [PubMed: 11201747]

Bharadwaj PS, and Hall SE (2017). Endogenous RNAi Pathways Are Required in Neurons for Dauer Formation in Caenorhabditis elegans. Genetics 205, 1503–1516. [PubMed: 28122825]

Blumenfeld AL, and Jose AM (2016). Reproducible features of small RNAs in C. elegans reveal NU RNAs and provide insights into 22G RNAs and 26G RNAs. RNA 22, 184–192. [PubMed: 26647462]

Brenner S (1974). The genetics of Caenorhabditis elegans. Genetics 77, 71–94. [PubMed: 4366476]

Cai Q, Qiao L, Wang M, He B, Lin FM, Palmquist J, Huang SD, and Jin H (2018). Plants send small RNAs in extracellular vesicles to fungal pathogen to silence virulence genes. Science 360, 1126–1129. [PubMed: 29773668]

Claycomb JM, Batista PJ, Pang KM, Gu W, Vasale JJ, van Wolfswinkel JC, Chaves DA, Shirayama M, Mitani S, Ketting RF, et al. (2009). The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. Cell 139, 123–134. [PubMed: 19804758]

Coffman SR, Lu J, Guo X, Zhong J, Jiang H, Broitman-Maduro G, Li WX, Lu R, Maduro M, and Ding SW (2017). Caenorhabditis elegans RIG-I Homolog Mediates Antiviral RNA Interference Downstream of Dicer-Dependent Biogenesis of Viral Small Interfering RNAs. mBio 8, e00264–17. [PubMed: 28325765]

Conine CC, Batista PJ, Gu W, Claycomb JM, Chaves DA, Shirayama M, and Mello CC (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. [PubMed: 20133686]

Conine CC, Moresco JJ, Gu W, Shirayama M, Conte D, Yates JR, and Mello CC (2013). Argonautes promote male fertility and provide a paternal memory of germline gene expression in C. elegans. Cell 155, 1532–1544. [PubMed: 24360276]

Dai H, and Gu W (2020). Strategies and Best Practice in Cloning Small RNAs. Gene Technology 9, 151. [PubMed: 32953938]

Dent JA, Smith MM, Vassilatis DK, and Avery L (2000). The genetics of ivermectin resistance in Caenorhabditis elegans. Proc Natl Acad Sci U S A 97, 2674–2679. [PubMed: 10716995]

Deshpande T, Takagi T, Hao L, Buratowski S, and Charbonneau H (1999). Human PIR1 of the protein-tyrosine phosphatase superfamily has RNA 5'-triphosphatase and diphosphatase activities. J Biol Chem 274, 16590–16594. [PubMed: 10347225]

Duchaine TF, Wohlschlegel JA, Kennedy S, Bei Y, Conte D, Pang K, Brownell DR, Harding S, Mitani S, Ruvkun G, et al. (2006). Functional proteomics reveals the biochemical niche of C. elegans DCR-1 in multiple small-RNA-mediated pathways. Cell 124, 343–354. [PubMed: 16439208]

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, and Mello CC (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature 391, 806–811. [PubMed: 9486653]

Fischer SE, Montgomery TA, Zhang C, Fahlgren N, Breen PC, Hwang A, Sullivan CM, Carrington JC, and Ruvkun G (2011). The ERI-6/7 helicase acts at the first stage of an siRNA amplification pathway that targets recent gene duplications. PLoS Genet 7, e1002369. [PubMed: 22102828]

Frokjaer-Jensen C, Davis MW, Hopkins CE, Newman BJ, Thummel JM, Olesen SP, Grunnet M, and Jorgensen EM (2008). Single-copy insertion of transgenes in Caenorhabditis elegans. Nat Genet 40, 1375–1383. [PubMed: 18953339]

Gabel HW, and Ruvkun G (2008). The exonuclease ERI-1 has a conserved dual role in 5.8S rRNA processing and RNAi. Nat Struct Mol Biol 15, 531–533. [PubMed: 18438419]

Gent JI, Lamm AT, Pavelec DM, Maniar JM, Parameswaran P, Tao L, Kennedy S, and Fire AZ (2010). Distinct phases of siRNA synthesis in an endogenous RNAi pathway in C. elegans soma. Mol Cell 37, 679–689. [PubMed: 20116306]

Gent JI, Schvarzstein M, Villeneuve AM, Gu SG, Jantsch V, Fire AZ, and Baudrimont A (2009). A Caenorhabditis elegans RNA-directed RNA polymerase in sperm development and endogenous RNA interference. Genetics 183, 1297–1314. [PubMed: 19805814]

Grentzinger T, Armenise C, Brun C, Mugat B, Serrano V, Pelisson A, and Chambeyron S (2012). piRNA-mediated transgenerational inheritance of an acquired trait. Genome research 22, 1877–1888. [PubMed: 22555593]

Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, and Mello CC (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. [PubMed: 11461699]

Gu W, Claycomb JM, Batista PJ, Mello CC, and Conte D (2011). Cloning Argonaute-associated small RNAs from Caenorhabditis elegans. Methods in Molecular Biology 725, 251–280. [PubMed: 21528459]

Gu W, Lee H-C, Chaves D, Youngman EM, Pazour GJ, Conte D, and Mello CC (2012). CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as C. elegans piRNA precursors. Cell 151, 1488–1500. [PubMed: 23260138]

Gu W, Shirayama M, Conte D, Vasale J, Batista PJ, Claycomb JM, Moresco JJ, Youngman EM, Keys J, Stoltz MJ, et al. (2009). Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. Mol Cell 36, 231–244. [PubMed: 19800275]

Guo X, Zhang R, Wang J, Ding SW, and Lu R (2013). Homologous RIG-I-like helicase proteins direct RNAi-mediated antiviral immunity in C. elegans by distinct mechanisms. Proc Natl Acad Sci U S A 110, 16085–16090. [PubMed: 24043766]

Han T, Manoharan AP, Harkins TT, Bouffard P, Fitzpatrick C, Chu DS, Thierry-Mieg D, Thierry-Mieg J, and Kim JK (2009). 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in Caenorhabditis elegans. Proc Natl Acad Sci U S A 106, 18674–18679. [PubMed: 19846761]

Hannon GJ (2002). RNA interference. Nature 418, 244–251. [PubMed: 12110901]

He L, Diedrich J, Chu YY, and Yates JR 3rd (2015). Extracting Accurate Precursor Information for Tandem Mass Spectra by RawConverter. Anal Chem 87, 11361–11367. [PubMed: 26499134]

Hornung V, Ellegast J, Kim S, Brzozka K, Jung A, Kato H, Poeck H, Akira S, Conzelmann KK, Schlee M, et al. (2006). 5'-Triphosphate RNA is the ligand for RIG-I. Science 314, 994–997. [PubMed: 17038590]

Hou Y, Zhai Y, Feng L, Karimi HZ, Rutter BD, Zeng L, Choi DS, Zhang B, Gu W, Chen X, et al. (2019). A Phytophthora Effector Suppresses Trans-Kingdom RNAi to Promote Disease Susceptibility. Cell Host Microbe 25, 153–165 e155. [PubMed: 30595554]

Kato H, Takeuchi O, Sato S, Yoneyama M, Yamamoto M, Matsui K, Uematsu S, Jung A, Kawai T, Ishii KJ, et al. (2006). Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. Nature 441, 101–105. [PubMed: 16625202]

Katsuma S, Koyano Y, Kang W, Kokusho R, Kamita SG, and Shimada T (2012). The baculovirus uses a captured host phosphatase to induce enhanced locomotory activity in host caterpillars. PLoS Pathog 8, e1002644. [PubMed: 22496662]

Kennedy S, Wang D, and Ruvkun G (2004). A conserved siRNA-degrading RNase negatively regulates RNA interference in C. elegans. Nature 427, 645–649. [PubMed: 14961122]

Kirino Y, and Mourelatos Z (2007). Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. Nature Structural & Molecular Biology 14, 347–348.

Langmead B, Trapnell C, Pop M, and Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10, R25. [PubMed: 19261174]

Lee HC, Gu W, Shirayama M, Youngman E, Conte D, and Mello CC (2012). C. elegans piRNAs mediate the genome-wide surveillance of germline transcripts. Cell 150, 78–87. [PubMed: 22738724]

Lee HC, Li L, Gu W, Xue Z, Crosthwaite SK, Pertsemlidis A, Lewis ZA, Freitag M, Selker EU, Mello CC, et al. (2010). Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. Mol Cell 38, 803–814. [PubMed: 20417140]

Li L, Dai H, Nguyen AP, and Gu W (2019). A convenient strategy to clone modified/unmodified small RNA and mRNA for high throughput sequencing. RNA 26, 218–227. [PubMed: 31754076]

Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, and Doudna JA (2006). Structural basis for double-stranded RNA processing by Dicer. Science 311, 195–198. [PubMed: 16410517]

McCaffrey AP, Meuse L, Pham TT, Conklin DS, Hannon GJ, and Kay MA (2002). RNA interference in adult mice. Nature 418, 38–39. [PubMed: 12097900]

Paix A, Folkmann A, Rasoloson D, and Seydoux G (2015). High Efficiency, Homology-Directed Genome Editing in Caenorhabditis elegans Using CRISPR-Cas9 Ribonucleoprotein Complexes. Genetics 201, 47–54. [PubMed: 26187122]

Pak J, and Fire A (2007). Distinct populations of primary and secondary effectors during RNAi in C. elegans. Science 315, 241–244. [PubMed: 17124291]

Parrish S, and Fire A (2001). Distinct roles for RDE-1 and RDE-4 during RNA interference in Caenorhabditis elegans. RNA 7, 1397–1402. [PubMed: 11680844]

Pavelec DM, Lachowiec J, Duchaine TF, Smith HE, and Kennedy S (2009). Requirement for the ERI/DICER complex in endogenous RNA interference and sperm development in Caenorhabditis elegans. Genetics 183, 1283–1295. [PubMed: 19797044]

Peng J, Elias JE, Thoreen CC, Licklider LJ, and Gygi SP (2003). Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J Proteome Res 2, 43–50. [PubMed: 12643542]

Posner R, Toker IA, Antonova O, Star E, Anava S, Azmon E, Hendricks M, Bracha S, Gingold H, and Rechavi O (2019). Neuronal Small RNAs Control Behavior Transgenerationally. Cell 177, 1814–1826.e15. [PubMed: 31178120]

Praitis V, Casey E, Collar D, and Austin J (2001). Creation of low-copy integrated transgenic lines in Caenorhabditis elegans. Genetics 157, 1217–1226. [PubMed: 11238406]

Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, and Bartel DP (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell 127, 1193–1207. [PubMed: 17174894]

Sankhala RS, Lokareddy RK, and Cingolani G (2014). Structure of human PIR1, an atypical dual-specificity phosphatase. Biochemistry 53, 862–871. [PubMed: 24447265]

Schurer H, Lang K, Schuster J, and Morl M (2002). A universal method to produce in vitro transcripts with homogeneous 3' ends. Nucleic Acids Res 30, e56. [PubMed: 12060694]

Seth M, Shirayama M, Gu W, Ishidate T, Conte D, and Mello CC (2013). The C. elegans CSR-1 argonaute pathway counteracts epigenetic silencing to promote germline gene expression. Dev Cell 27, 656–663. [PubMed: 24360782]

Shatkin AJ (1976). Capping of eucaryotic mRNAs. Cell 9, 645–653. [PubMed: 1017010]

Shen EZ, Chen H, Ozturk AR, Tu S, Shirayama M, Tang W, Ding YH, Dai SY, Weng Z, and Mello CC (2018). Identification of piRNA Binding Sites Reveals the Argonaute Regulatory Landscape of the C. elegans Germline. Cell 172, 937–951.e18. [PubMed: 29456082]

Shirayama M, Seth M, Lee HC, Gu W, Ishidate T, Conte D, and Mello CC (2012). piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline. Cell 150, 65–77. [PubMed: 22738726]

Simmer F, Tijsterman M, Parrish S, Koushika SP, Nonet ML, Fire A, Ahringer J, and Plasterk RH (2002). Loss of the putative RNA-directed RNA polymerase RRF-3 makes C. elegans hypersensitive to RNAi. Curr Biol 12, 1317–1319. [PubMed: 12176360]

Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, et al. (2002). The generic genome browser: a building block for a model organism system database. Genome Research 12, 1599–1610. [PubMed: 12368253]

Steiner FA, Okihara KL, Hoogstrate SW, Sijen T, and Ketting RF (2009). RDE-1 slicer activity is required only for passenger-strand cleavage during RNAi in Caenorhabditis elegans. Nat Struct Mol Biol 16, 207–211. [PubMed: 19151723]

Tabara H, Sarkissian M, Kelly WG, Fleenor J, Grishok A, Timmons L, Fire A, and Mello CC (1999). The rde-1 gene, RNA interference, and transposon silencing in C. elegans. Cell 99, 123–132. [PubMed: 10535731]

Tabara H, Yigit E, Siomi H, and 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. [PubMed: 12110183]

Tabb DL, McDonald WH, and Yates JR 3rd (2002). DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J Proteome Res 1, 21–26. [PubMed: 12643522]

Takagi T, Taylor GS, Kusakabe T, Charbonneau H, and Buratowski S (1998). A protein tyrosine phosphatase-like protein from baculovirus has RNA 5'-triphosphatase and diphosphatase activities. Proc Natl Acad Sci U S A 95, 9808–9812. [PubMed: 9707557]

Thivierge C, Makil N, Flamand M, Vasale JJ, Mello CC, Wohlschlegel J, Conte D Jr., and Duchaine TF (2011). Tudor domain ERI-5 tethers an RNA-dependent RNA polymerase to DCR-1 to potentiate endo-RNAi. Nat Struct Mol Biol 19, 90–97. [PubMed: 22179787]

Timmons L (2004). Endogenous inhibitors of RNA interference in Caenorhabditis elegans. Bioessays 26, 715–718. [PubMed: 15221853]

Vasale JJ, Gu W, Thivierge C, Batista PJ, Claycomb JM, Youngman EM, Duchaine TF, Mello CC, and Conte D (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. [PubMed: 20133583]

Welker NC, Maity TS, Ye X, Aruscavage PJ, Krauchuk AA, Liu Q, and Bass BL (2011). Dicer's helicase domain discriminates dsRNA termini to promote an altered reaction mode. Mol Cell 41, 589–599. [PubMed: 21362554]

Welker NC, Pavelec DM, Nix DA, Duchaine TF, Kennedy S, and Bass BL (2010). Dicer's helicase domain is required for accumulation of some, but not all, C. elegans endogenous siRNAs. RNA 16, 893–903. [PubMed: 20354150]

Wolters DA, Washburn MP, and Yates JR 3rd (2001). An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem 73, 5683–5690. [PubMed: 11774908]

Xu T, Park SK, Venable JD, Wohlschlegel JA, Diedrich JK, Cociorva D, Lu B, Liao L, Hewel J, Han X, et al. (2015). ProLuCID: An improved SEQUEST-like algorithm with enhanced sensitivity and specificity. J Proteomics 129, 16–24. [PubMed: 26171723]

Yigit E, Batista PJ, Bei Y, Pang KM, Chen CC, Tolia NH, Joshua-Tor L, Mitani S, Simard MJ, and Mello CC (2006). Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. Cell 127, 747–757. [PubMed: 17110334]

Yuan Y, Li DM, and Sun H (1998). PIR1, a novel phosphatase that exhibits high affinity to RNA ribonucleoprotein complexes. J Biol Chem 273, 20347–20353. [PubMed: 9685386]

Zhang C, Montgomery TA, Gabel HW, Fischer SE, Phillips CM, Fahlgren N, Sullivan CM, Carrington JC, and Ruvkun G (2011). mut-16 and other mutator class genes modulate 22G and 26G siRNA pathways in Caenorhabditis elegans. Proc Natl Acad Sci U S A 108, 1201–1208. [PubMed: 21245313]

Zhang D, Tu S, Stubna M, Wu WS, Huang WC, Weng Z, and Lee HC (2018). The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. Science 359, 587–592. [PubMed: 29420292]

Zhang H, Kolb FA, Brondani V, Billy E, and Filipowicz W (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. EMBO J 21, 5875–5885. [PubMed: 12411505]

Highlights:

RNA phosphatase, RNAi, germline gene regulation, regulation of triphosphorylated RNA.

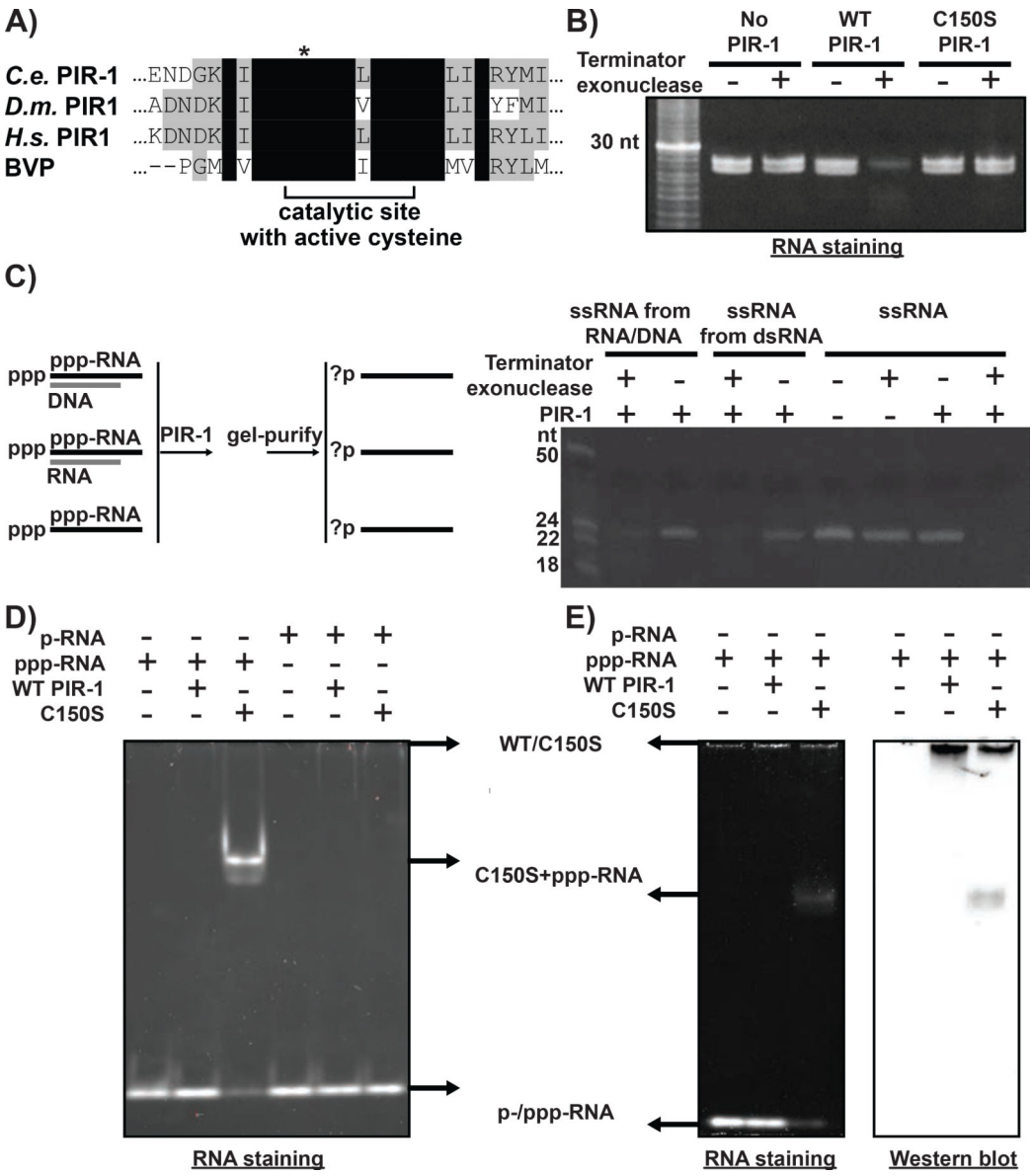


Figure 1. PIR-1 is an RNA polyphosphatase.

(A) Alignment of PIR-1 orthologs from *C. elegans* (*C.e.*), *Drosophila* (*D.m.*), and human (*H.s.*), asterisk indicates the catalytic cysteine.

(B) Terminator exonuclease assays on ppp-RNA substrates with and without pretreatment by WT or C150S recombinant PIR-1.

(C) Terminator exonuclease assays on ppp-RNA duplexed with DNA and RNA (schematic of pretreatments, left).

(D) Gel-shift assays on single-stranded ppp-RNA and p-RNA substrates using recombinant WT and C150S PIR-1, visualized by 15% native PAGE and SYBR Gold staining.

(E) Gel shift assays followed by SYBR Gold staining (left) and western blot (for detection of His-tagged WT and C150S PIR-1, right) on RNA substrates (as indicated).

See also Figure S1.

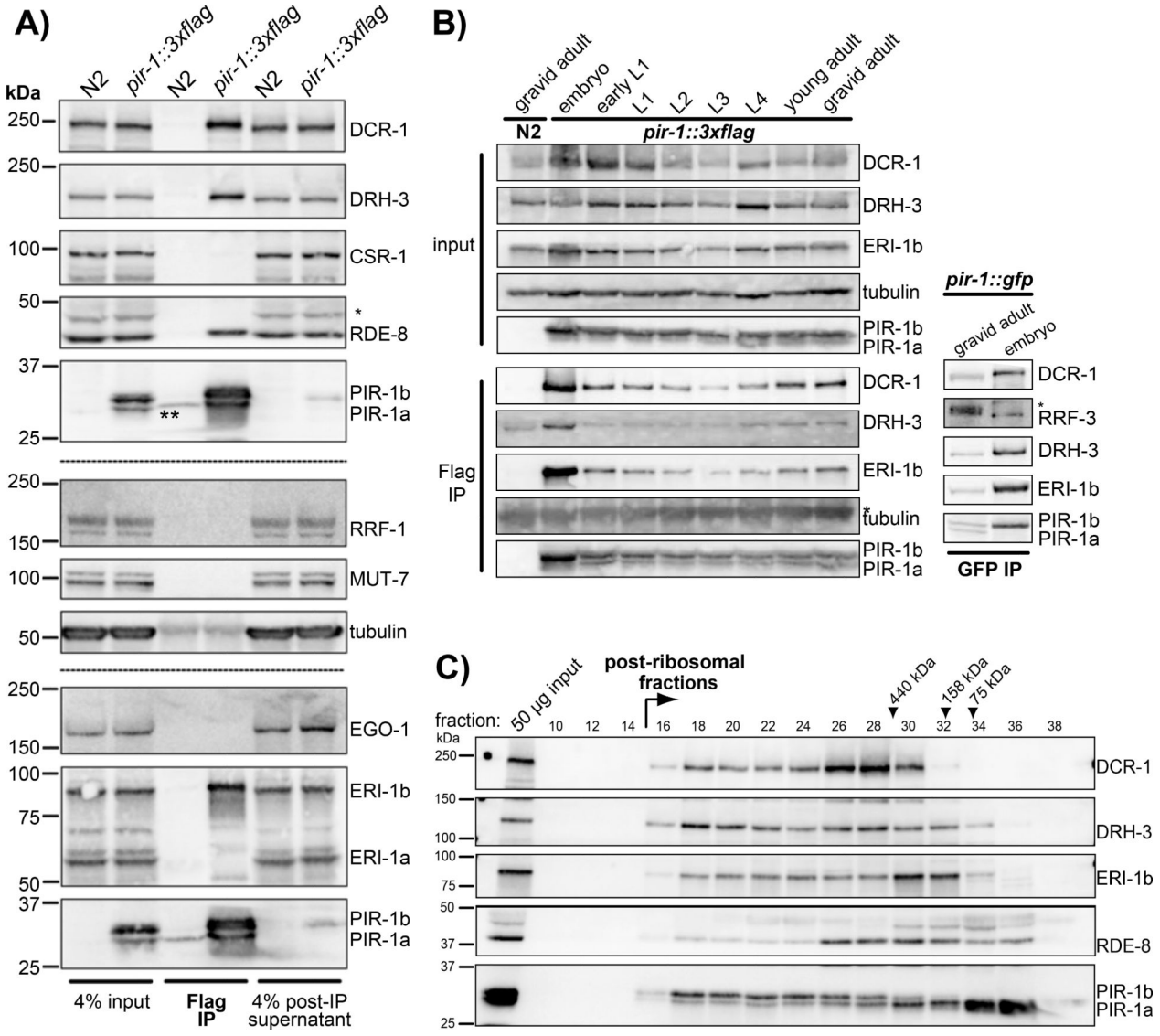


Figure 2. PIR-1 interacts with the ERI complex.

(A) Western blot analyses on WT (N2) and *pir-1::3xflag*-rescued young adults showing proteins present in input (lysate), FLAG IP, and post-IP supernatant. The * and ** indicate unspecified bands.

(B) Western blot analyses of PIR-1-associated proteins across developmental stages in N2 and PIR-1::3xFLAG (left) and in PIR-1::GFP lysates (as indicated). The * indicates a background signal that co-migrates with tubulin from binding of the secondary antibody to the heavy chain of the anti-Flag antibody.

(C) Gel filtration analysis of *pir-1::3xflag* lysates followed by western blot analyses (as indicated). Arrowheads indicate molecular weights of size standards.

See also Figure S2, Table 1 and Table S1.

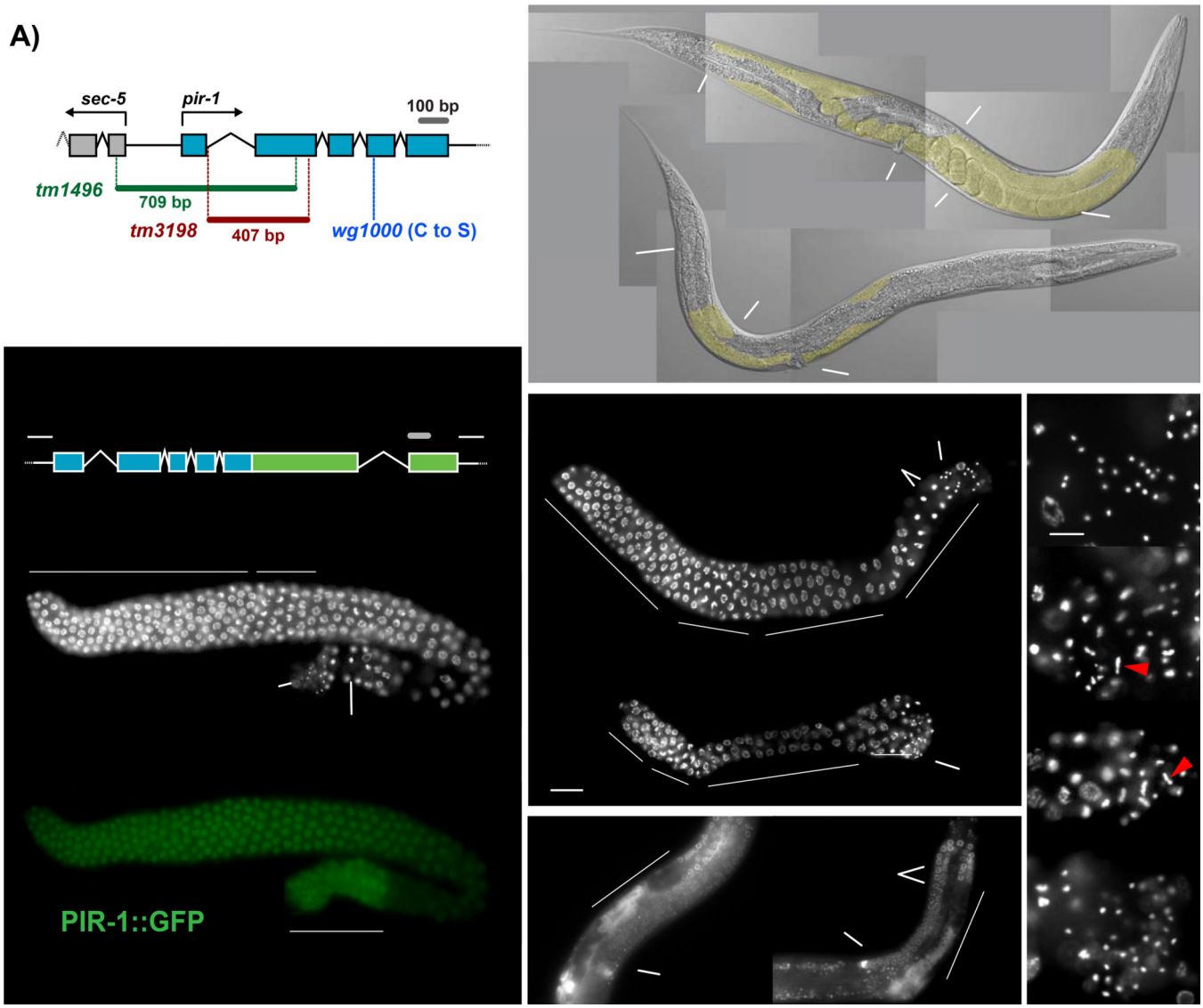


Figure 3. PIR-1 is essential for somatic and germline development.

(A) Schematic of the *pir-1* locus indicating genetic lesions used.

(B) Schematic of a rescuing transgene with GFP exons indicated in green (top panel) and fluorescence micrographs of an L4 germline stained with DAPI and with anti-GFP (bottom panels).

(C) DIC images of a *pir-1(+/tm3198)* heterozygote and an arrested *tm3198* homozygote cultured at 20°C for 96 hours with germlines indicated by yellow highlighting (partly concealed by intestine).

(D–F) Fluorescence micrographs of WT (N2) and mutant germlines visualized by DAPI in (D–F) and by PGL-1::RFP fluorescence in (E). Distal germline oriented to left. ‘m’, mitotic zone; ‘tz’, transition zone; ‘p’, pachehytene; and ‘sp’, spermatids indicated in (D). Abnormal chromosome bridging is indicated with red arrows in (F).

See also Figure S3.

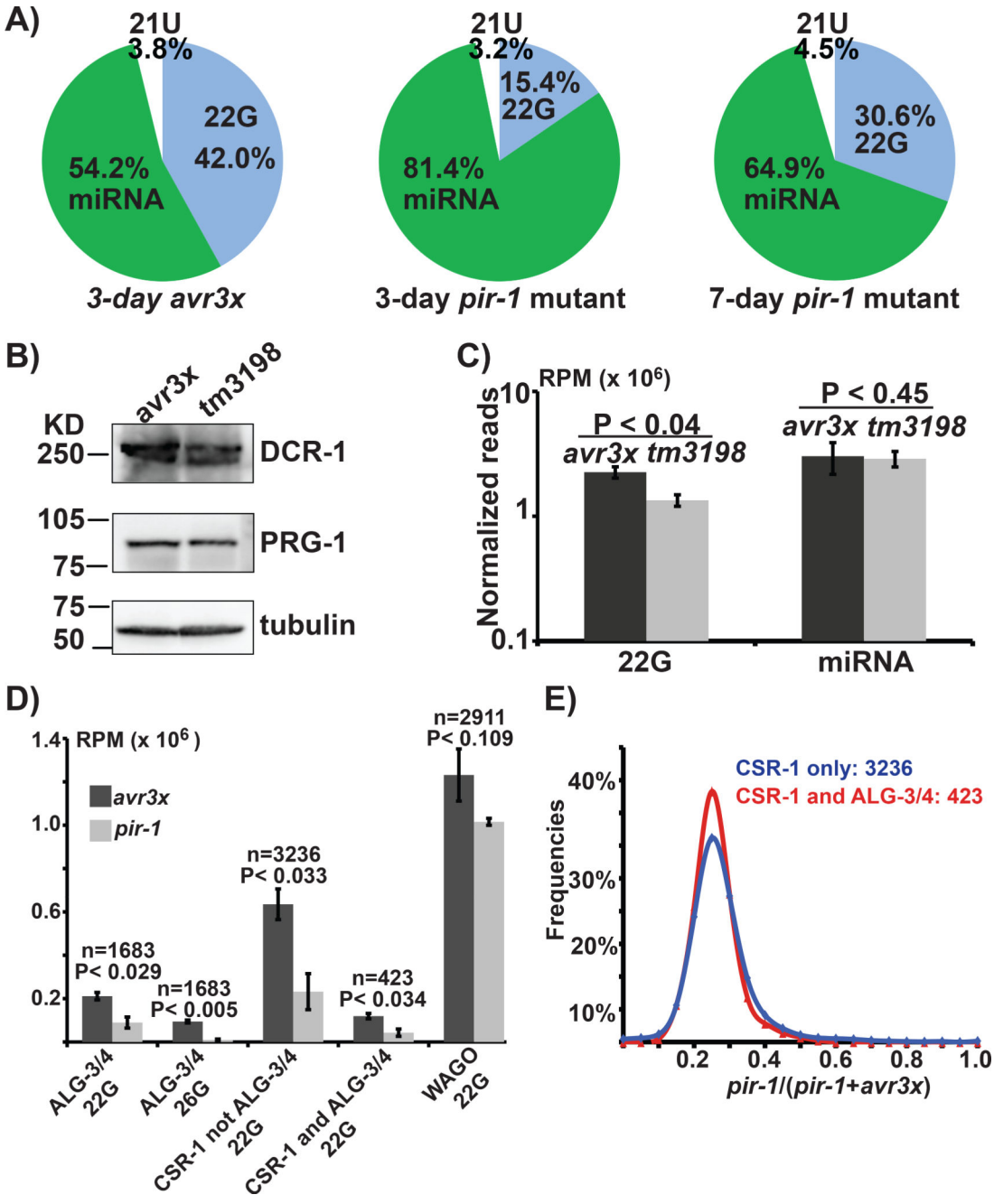


Figure 4. PIR-1 is required for the biogenesis of 26G-RNAs and non-WAGO-bound 22G-RNAs.

(A) Venn diagrams showing relative abundance of small RNA species in *pir-1*(*tm3198*) mutants and control *avr3x* animals. For each strain, small RNA composition was calculated as the average of two replicas.

(B) Western blot analyses of DCR-1 and PRG-1 in control avr3x and *pir-1* mutants, normalized to tubulin.

(C-D) Bar graphs comparing abundance of small RNA species in reads per million (RPM) in arrested *pir-1* mutants (7 days old) and L4 stage *avr3x* animals (as indicated). The error bar represents one standard error. P values were calculated for two replicas using unpaired

student's t-test (one-tailed for 22G-RNAs and 26G-RNAs, and two-tailed for miRNA). In (D), 'n' indicates the number of target genes in each category.

(E) Histogram showing ratios of 22G-RNAs (*pir-1/pir-1* + control avr3x) (x-axis) calculated for each individual gene in the two CSR-1 target categories and binned into 20 intervals plotted against frequency for each ratio (y-axis).

See also Figure S4 and Table S3.

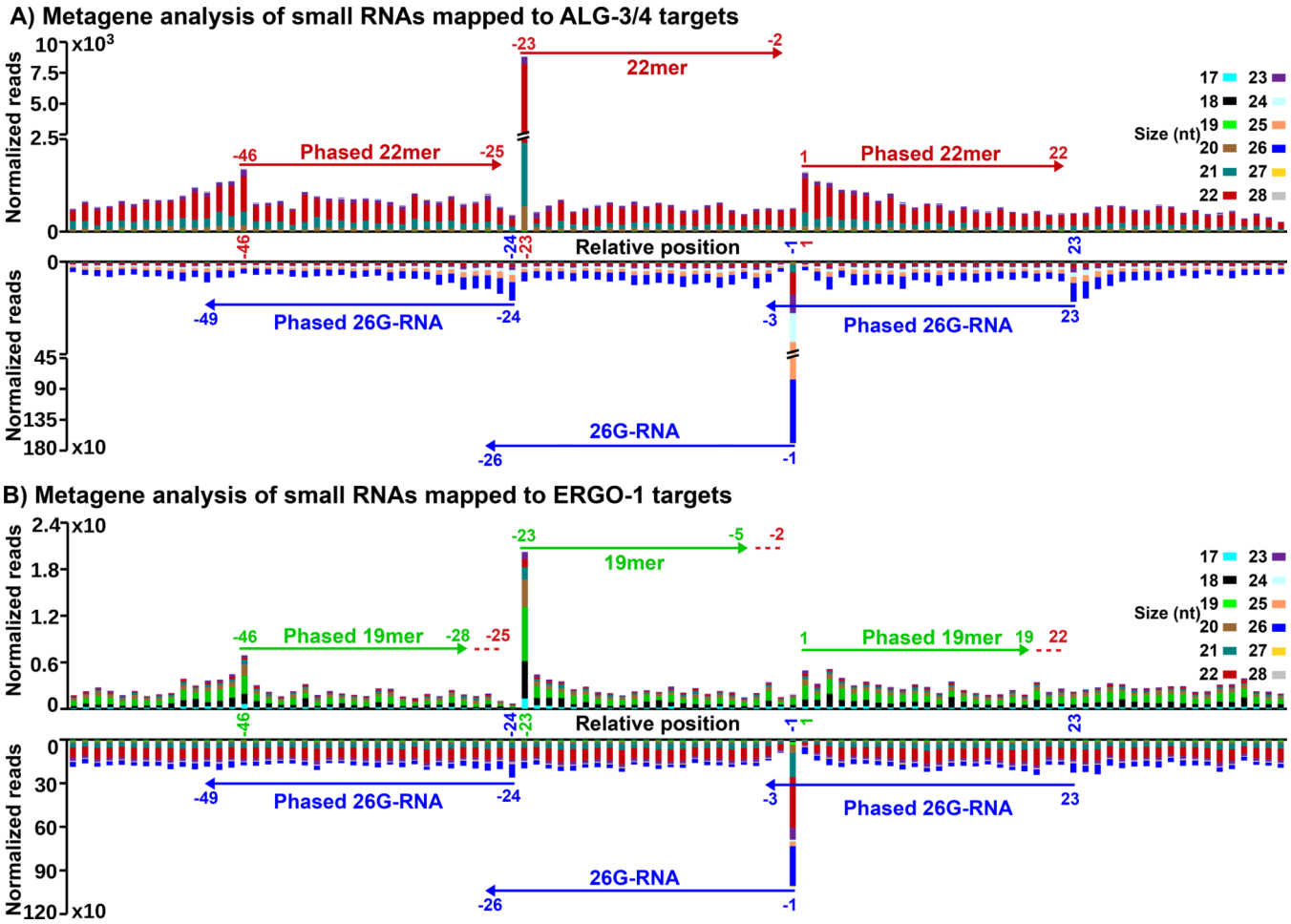


Figure 5. Metagene analysis of 26G-RNA loci.

(A and B) Bar graphs plotting small RNA levels across mRNA intervals that template 26G-RNAs in the ALG-3/4 (A) and ERGO-1 pathways (B). Frequencies of mRNA-derived species (upper) and RdRP-derived species (lower) are plotted according to the position of their 5' nt. Length is color-coded. Coordinates are defined relative to the C-nucleotide (–1) used to template 26G production. RNA was prepared from (*fog-2*) male-enriched populations (A) or WT embryos (B).

See also Figure S5.

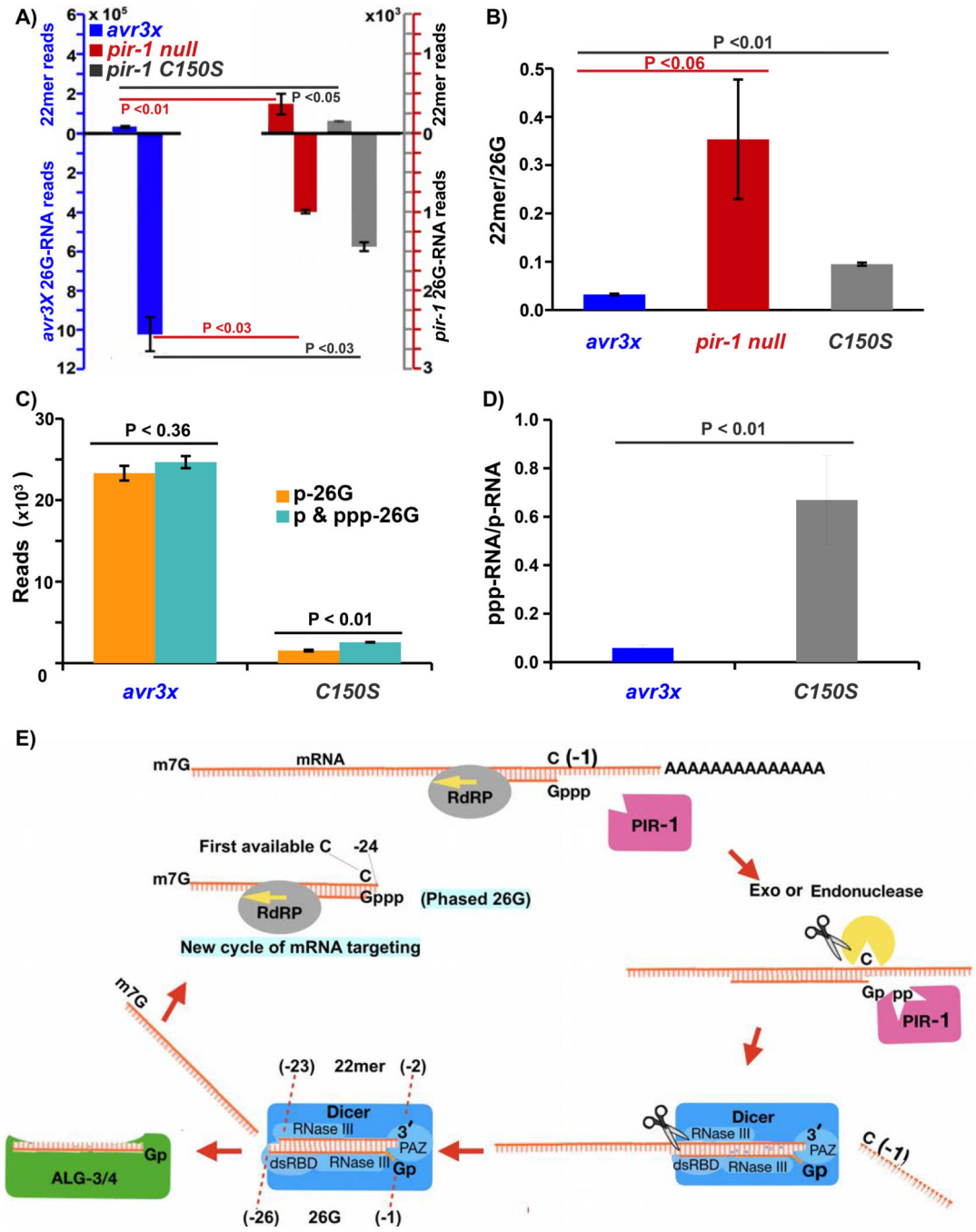


Figure 6. *pir-1* mutants are defective in 26G-RNA maturation.

(A) Bar graph comparing the levels of antisense 26G-RNAs (located at −1 in the metagene space) and sense-stranded 22mers (located at −23) cloned from of WT (avr3x), *pir-1* null, and *pir-1(C150S)* respectively. Small RNAs were cloned using TAP or recombinant PIR-1 pretreatment to prevent cloning bias against ppp-RNA species (see Experimental Procedures).

(B) Bar graph showing the ratio of 22mer to 26G-RNA in each strain using the data in panel (A).

(C) Bar graph comparing the levels of p- and ppp-26G-RNA cloned from WT (avr3x) or *pir-1(C150S)* worms. Small RNAs were directly ligated to clone p-26G-RNAs (yellow), or they pretreated with recombinant PIR-1 to remove gamma and beta phosphates before ligation to clone p- and ppp-26G-RNAs (cyan). Reads were normalized to total 21U-RNAs.
(D) Bar graph showing the ratio of ppp-26G-RNA to p-26G-RNA in WT and *pir-1(C150S)* data from (C).
(E) Model of 26G-RNA biogenesis.
P values were obtained using an unpaired Student's t-test (one-tailed for A, B, and D; two-tailed for C) based on two replicas of each sample; error bars represent one standard error.
See also Figure S6.

Table 1.

PIR-1 Interactors identified in PIR-1::GFP IP Using Young Adult Worms

Protein	Amino Acid Number	Spectral Counts[a]	Protein Coverage
PIR-1	233	100	47.2%
DCR-1	1910	115	22.3%
RRF-3	1765	63	21.8%
DRH-3	1119	52	26.2%
ERI-3 (W09B6.3a)[b]	578	24	16.1%
RDE-4	385	16	19.0%
ERI-1 (T07A9.5b)[b]	582	12	13.1%
ERI-5 (Y38F2AR.1a)[b]	531	11	11.9%

[a] The number of tandem mass spectra matching peptides derived from each protein.

[b] Only the isoform with the most counts is shown.