

A genome-wide RNAi screen identifies factors required for distinct stages of *C. elegans* piRNA biogenesis

Wee-Siong Sho Goh,^{1,2} Jun Wen Eugene Seah,^{1,2} Emily J. Harrison,^{1,2} Caifu Chen,³ Christopher M. Hammell,¹ and Gregory J. Hannon^{1,2,4}

¹Watson School of Biological Sciences, ²Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA; ³Genetic Applications R&D, Life Technologies Corporation, Foster City, California 94404, USA

In animals, piRNAs and their associated Piwi proteins guard germ cell genomes against mobile genetic elements via an RNAi-like mechanism. In *Caenorhabditis elegans*, 21U-RNAs comprise the piRNA class, and these collaborate with 22G RNAs via unclear mechanisms to discriminate self from nonself and selectively and heritably silence the latter. Recent work indicates that 21U-RNAs are post-transcriptional processing products of individual transcription units that produce ~26-nucleotide capped precursors. However, nothing is known of how the expression of precursors is controlled or how primary transcripts give rise to mature small RNAs. We conducted a genome-wide RNAi screen to identify components of the 21U biogenesis machinery. Screening by direct, quantitative PCR (qPCR)-based measurements of mature 21U-RNA levels, we identified 22 genes important for 21U-RNA production, termed TOFUs (Twenty-One-u Fouled Ups). We also identified seven genes that normally repress 21U production. By measuring mature 21U-RNA and precursor levels for the seven strongest hits from the screen, we assigned factors to discrete stages of 21U-RNA production. Our work identifies for the first time factors separately required for the transcription of 21U precursors and the processing of these precursors into mature 21U-RNAs, thereby providing a resource for studying the biogenesis of this important small RNA class.

[*Keywords:* *C. elegans*; piRNA; genome-wide; RNAi screen; capped precursor; TOFU]

Supplemental material is available for this article.

Received November 27, 2013; revised version accepted February 20, 2014.

21U-RNAs emerged as a discrete class of *Caenorhabditis elegans* small RNAs (sRNAs) from data collected during early deep-sequencing studies (Ruby et al. 2006; Batista et al. 2008; Das et al. 2008). These RNAs are derived from clustered loci on chromosome IV and show a number of distinguishing features. They have a strong bias for U at their 5' position and a periodate-resistant 3' end, indicative of a modification of the 3' OH group. These features are shared with piRNAs characterized in other animals, and 21U-RNAs were indeed shown to partner with *C. elegans* Piwi family proteins, particularly PRG-1. Loss of 21U-RNA-PRG-1 complexes results in reduced fertility, just as loss of piRNA pathway genes results in sterility in many other animals [Batista et al. 2008; Das et al. 2008; Wang

and Reinke 2008). Thus, 21U-RNAs are considered to be analogous to piRNAs in other species.

In *Drosophila* and mammals, piRNAs have clearly established functions in the control of mobile genetic elements, and fly piRNAs and mammalian piRNAs that are expressed during embryonic germ cell development are enriched for transposon sequences (Siomi et al. 2011). An enigmatic class of mammalian piRNAs appears during meiosis, and although their partner proteins are essential for spermatogenesis, the targets of these piRNAs are unclear, as they stringently match only those loci from

⁴Corresponding author
E-mail hannon@cshl.edu

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.235622.113>.

© 2014 Goh et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genesdev.cshl.org/site/misc/terms.html>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

which they are derived. Similarly, the targets of worm 21U-RNAs were, until recently, unclear, as these ~15,000 distinct sequences lack close matches to annotated transposons or genic targets.

Recent studies have raised the possibility that the 21U-RNAs represent a complex sequence repertoire that permits PRG-1 to recognize foreign sequences that become integrated into the *C. elegans* genome (Bagijn et al. 2012). This targeting mechanism is independent of the PRG-1 “slicer” activity and tolerates up to four mismatches in the selection of sequences, which become subjected to transcriptional gene silencing. PRG-1 has been proposed to act only in the initiation phase of what becomes fixed as a heritable silencing event; once a sequence is recognized by a 21U-RNA as “non-self,” other pathways enforce its heritable repression (Ashe et al. 2012; Luteijn et al. 2012; Shirayama et al. 2012).

Although piRNA sequences themselves are not conserved among worm species, they all contain a strongly conserved proximal upstream motif, “CTGTTTCA.” This has been proposed to act as a regulatory element that controls autonomous transcription of 21U-RNA precursors (pre-21Us) that become individual 21U-RNAs. Forkhead family transcription factors were recently shown to bind 21U-RNA promoters (Cecere et al. 2012). In this regard, levels of mature 21U-RNAs were reduced upon loss of *unc-130*, a forkhead family protein, although a direct role for *unc-130* in promoting transcription of pre-21U-RNAs has yet to be demonstrated.

Recent work has suggested that the transcription of pre-21U RNAs begins 2 nucleotides (nt) upstream of the U that forms the mature 5' end to generate a capped transcript that terminates ~3 nt downstream from the mature 3' end (Gu et al. 2012). The mechanism by which pre-21Us are converted to mature piRNAs is unknown. Despite being a signature mark of piRNAs, the methyl mark on their 3' OH group is not required for piRNA maturation, as loss of the methylating enzyme HENN-1 barely affects 21U-RNA abundance (Montgomery et al. 2012). This leaves 5' decapping, removal of the 2-nt 5' overhang, and 3' end trimming as the three potential steps required for piRNA maturation.

To build a basis for understanding 21U-RNA biogenesis, we undertook a comprehensive RNAi screen, searching for genes that affect mature 21U-RNA levels. By directly quantifying piRNA levels upon silencing of individual genes in a genome-wide RNAi library, we identified 22 dsRNAs that resulted in reproducible 21U-RNA defects. Many of these dsRNAs mapped to novel and uncharacterized genes, which we refer to generally as Twenty-One-u Fouled Ups (TOFUs). Using a combination of both 5' monophosphate and 5' cap-dependent sRNA cloning, we were able to quantify differential effects on pre-21U levels upon the perturbation of individual TOFUs. This revealed that specific TOFUs were required for either the transcription and capping of pre-21U or the downstream processing of pre-21Us into mature 21U-RNAs. We were thus able to build the first genetic framework for piRNA biogenesis in *C. elegans*.

Results

A quantitative genome-wide RNAi screen for novel piRNA biogenesis factors

We used an RNAi-hypersensitive *rrf-3*-null strain to screen a collection of 19,680 dsRNAs (targeting ~81% of annotated *C. elegans* protein-coding genes) by quantifying the abundance of a mature 21U-RNA directly using quantitative RT-PCR (qRT-PCR) (Fig. 1A; Kamath et al. 2003). To control for effects on the integrity of the germ cells, which produce 21U-RNAs, we assayed in parallel the abundance of *cel-miR-35*, an embryo-enriched microRNA (miRNA) (Lau et al. 2001).

As an initial selection for further validation, we used a Z-score method with a Z-score cutoff of 2.5 to identify the dsRNAs that showed the strongest effects on 21U abundance (Zhang 2011). This yielded 209 candidate genes as a starting point for further validation (Fig. 1B). Individual, sequence-verified clonal bacterial strains drawn from the original library were used to retest effects on 21U levels. Results of biological triplicate assays narrowed our candidate list to 92 genes with reproducibly robust effects on 21U-RNA levels.

Amongst these 92 candidates, we identified some false positives. The hexosamine pathway gene *gna-2* was initially identified as a strong biogenesis candidate (Fig. 1C; Johnston et al. 2006). However, strains with mutations in this gene were found to have no piRNA defect (Supplemental Fig. S1). Additionally, we scored as positives a number of miRNA biogenesis factors, including *dcr-1*, *pash-1*, and *alg-2* (Fig. 2A; Grishok et al. 2001; Denli et al. 2004). Prior studies showed that these proteins do not participate in piRNA biogenesis (Das et al. 2008). These analyses illustrated several scenarios with the potential to lead to false positive signals in our assay and allowed us to hone our hit-calling strategy. For example, one major class of false positives would be those affecting the levels of our control miRNA. Since our 21U-RNA signals were normalized to *cel-miR-35*, dsRNAs that either decrease or increase *cel-miR-35* levels would cause the corresponding genes to be called as repressors or biogenesis factors, respectively, for piRNAs.

Using a scatter plot of miRNA and 21U-RNA CT values, we noticed that the genes that we identified as false positives exhibited a characteristic pattern (Figs. 1C, 2A). True negatives (empty vector RNAi where no sequences were inserted between bidirectional T7 promoters in the RNAi construct or GFP dsRNA) (Figs. 1C, 2A, brown squares and green plus signs) are clustered along a diagonal, while true positives (*prg-1* dsRNA) (Figs. 1C, 2A, purple diamonds) mainly exhibited a vertical shift from the true negative cluster. While most of the 92 candidates (Figs. 1C, 2A, subset of x's) displayed the same vertical shift, the false positives identified by further validation exhibited a horizontal shift. Therefore, we used this horizontal shift pattern to denote and filter out likely false positives from among the remaining 92 candidates (see the Materials and Methods). While there are other criteria that we could have implemented, we found this strategy to successfully

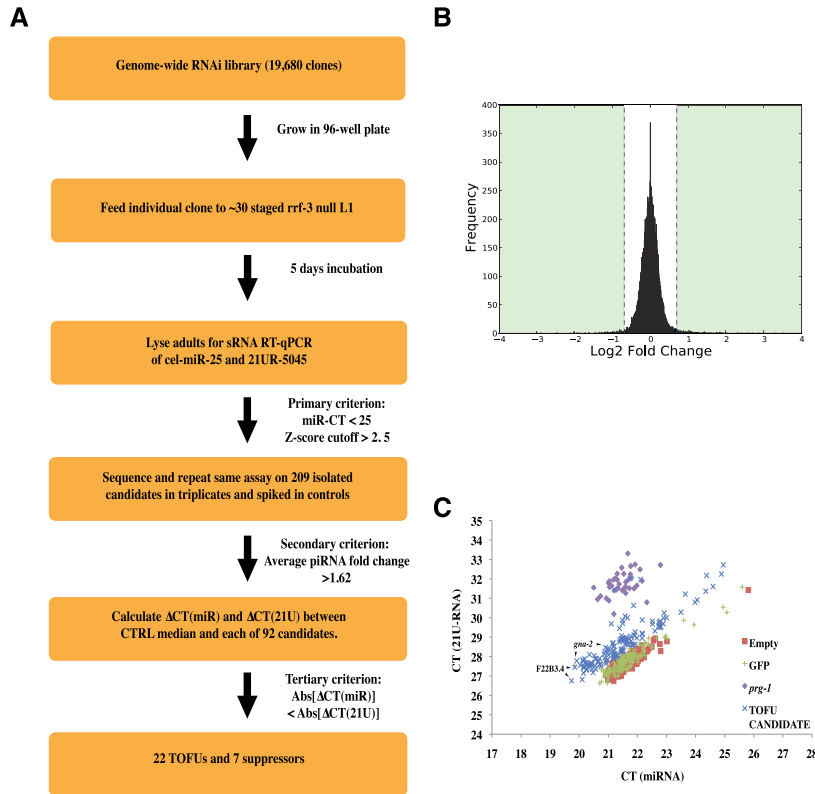


Figure 1. qRT-PCR-based genome-wide RNAi screen for TOFUs. (A) A work flow of the 21U-RNA biogenesis screen is presented. (B) Distribution of log₂ 21U-RNA fold change for all clones in the RNAi library. Highlighted regions depict those in which the 209 candidates passing the primary criterion of selection outlined in A fell. (C) Scatter plot of *21ur-5045* CT against *cel-miR-35* CT from the repeat RNAi assay denoted in A. Empty vector, GFP, and *prg-1* RNAi replicates are compared against RNAi of all TOFU candidates that passed the secondary filter. Two examples of false TOFU positives are indicated by arrowheads.

remove all obvious false positives, especially hexosamine pathway genes (*gna-2* and *F22B3.4*) and miRNA biogenesis factors. Using this approach, we arrived at a final list of 22 TOFUs and seven suppressors of piRNA accumulation (Tables 1, 2).

Suppressors of piRNA expression

Although the goal of our screen was to identify piRNA biogenesis factors, we also had the potential to identify genes that, when silenced, increased piRNA levels (Fig. 2A). Of particular interest in this category was the Argonaute protein CSR-1. *csr-1* RNAi caused an elevation in piRNA levels that was also seen in analyses of 2 *csr-1* mutants (Fig. 2B). *csr-1* loads 22-nt 22G endo-siRNAs that control holocentric chromosomal segregation (Claycomb et al. 2009; Gu et al. 2009). In support of a role for *csr-1* in regulating 21U levels, *ekl-1* and *ego-1*, which are members of the *csr-1* 22G pathway required for 22G RNA biogenesis, were also identified as genes that negatively regulate piRNA levels (Table 2).

Roles for *csr-1* and *ekl-1* in repressing piRNA levels were independently validated based on analysis of sRNA sequencing data sets (W Gu and C Mello, pers. comm.). *prg-1* is a known target of *csr-1*-bound 22G RNAs, and *prg-1* mRNA levels increased upon RNAi of *csr-1*, *ekl-1*, or *ego-1* (Fig. 2C). This is consistent with a scenario in which 22Gs and CSR-1 normally represses *prg-1* mRNA levels in wild-type conditions; in conditions of reduced CSR-1 or 22G-RNAs, *prg-1* repression is alleviated, resulting in an increased supply of PRG-1 proteins to load and

stabilize more 21U-RNAs. However, previous studies have proposed that CSR-1 either does not act to repress its targets or acts to identify “self” sequences and prevent their silencing by PRG-1-bound 21U-RNAs. (Claycomb et al. 2009; Conine et al. 2013; Seth et al. 2013; Wedeles et al. 2013). Thus, the simple scenarios proposed above may not hold true for all CSR-1 targets, and the relationship between the 22G pathway and piRNA levels might be more complex than thus far imagined.

Independent validation of TOFU-dependent piRNA defects

For all follow-up experiments, we focused on the seven TOFUs that had the most robust effects on 21U-RNA levels for validation and characterization. Given that we assayed only a single 21U-RNA in our initial screen, we first sought to test whether the effects of repressing TOFU genes extended to the entire 21U-RNA population. We therefore cloned 5′ monophosphorylated sRNA populations comprising 19- to 30-nt species and quantified the effects on 21U-RNAs in the various RNAi strains or null mutants as compared with an empty vector RNAi strain or N2 strain, respectively (Fig. 3). In all cases, we observed a global reduction in piRNAs measured by sRNA sequencing that was comparable with the reduction of the single 21U-RNA measured by qRT-PCR (Supplemental Fig. S2).

TOFUs could impact 21U-RNA levels in any number of ways. They could act directly in or regulate either pre-21U transcription or the processing of these precursors

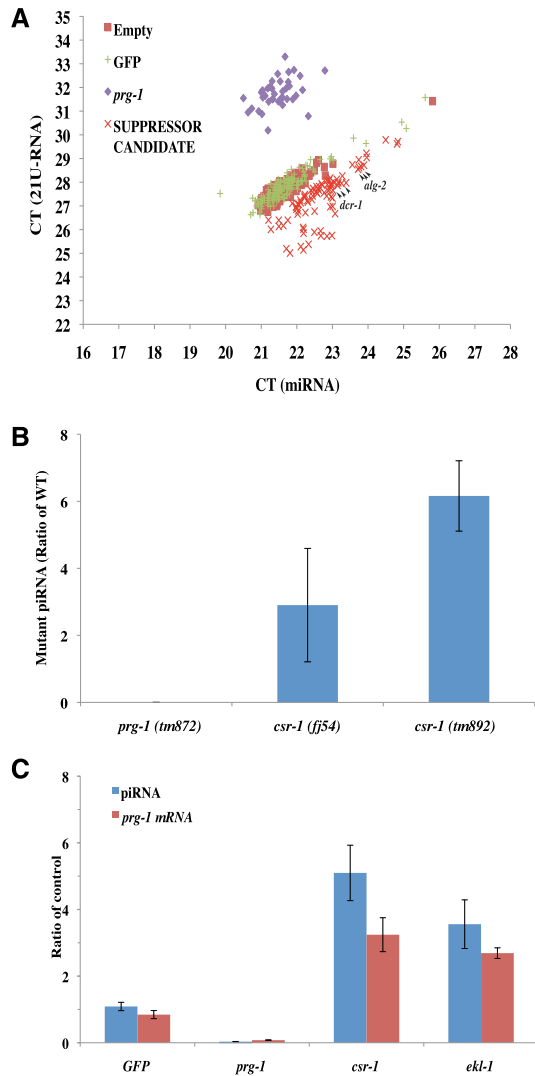


Figure 2. Suppressors of 21U-RNA expression. (A) A scatter plot is shown of *21ur-5045* CT against *cel-miR-35* CT from the repeat RNAi assay (denoted in Fig. 1A) for all suppressor candidates that passed the secondary filter. Two examples of false suppressor positives are indicated by arrowheads. (B) qRT-PCR of *21ur-5045* piRNA in two *csr-1* mutants were expressed as a ratio of N2 levels. Error bars represent standard deviation of at least three replicates. (C) qRT-PCR of *21ur-5045* piRNA (blue) and *prg-1* mRNA levels (red) in suppressor RNAi worms were expressed as a ratio of levels in empty vector RNAi worms. Error bars represent standard deviation of at least two replicates.

into mature species. They could also indirectly impact 21U-RNAs by affecting the expression levels of *prg-1*. To rule out this latter indirect effect, we measured *prg-1* mRNA levels in each TOFU RNAi strain or mutant and found these to be similar to those observed in wild-type or control animals (Fig. 4A).

The importance of PRG-1 is highlighted by the temperature-dependent brood size reduction observed in *prg-1* mutants. As piRNA mutants, *tofu-3* mutants were also expected to display this phenotype. While not sterile, both

tofu-3 mutants exhibited a brood size reduction to roughly half of that seen from wild-type worms at 25°C (Fig. 4B).

In accord with the ability of the 21U pathway to target foreign sequences that have integrated into the worm genome, transposable elements were previously shown to be derepressed in a *prg-1* mutant that lacks piRNAs (Bagijn et al. 2012). Therefore, derepression of these elements provides an alternative method for validating the piRNA defect in TOFU-null mutants. As expected, mutants for *tofu-1*, *tofu-2*, *tofu-3*, or *tofu-4* all exhibited derepression of the *line2h* transposon to a degree similar to that seen in the *prg-1* mutant (Fig. 4C). Certain protein-coding genes, especially *bath-45*, have also been observed to be derepressed in *prg-1* mutants. While the physiological relevance of *prg-1* silencing an endogenous gene is not obvious, we were surprised to find that besides the *tofu-1* mutant, all other assayed *tofu* mutants failed to exhibit *bath-45* derepression (Fig. 4D). This suggests that the mechanism of coding gene repression by *prg-1* may differ in some ways from that of transposon silencing, perhaps even being in some ways independent of the bulk of 21U-RNAs.

Analysis of pre-21U levels reveal specific roles for TOFUs in worm piRNA biogenesis

The recent identification of pre-21Us as *C. elegans* piRNA biogenesis intermediates has provided a biochemical framework by which 21U genomic loci are transcribed and processed into 21U-RNAs (Fig. 5). Using this framework, we envisioned two scenarios in which loss of an individual TOFU could result in a piRNA defect. If a TOFU is involved in the transcription of a pre-21U transcript or its subsequent 5' capping, loss of this TOFU would result in the reduction of pre-21Us, which in turn would lead to a deficit in mature 21U-RNAs. On the other hand, a TOFU could be required for the downstream processing of a pre-21U into a mature 21U-RNA. These processes include 5' decapping, 5' trimming, and 3' trimming. Therefore, loss of such a TOFU would result in the accumulation of pre-21Us because they would not flow into mature 21U-RNAs. We therefore sought to characterize how our validated TOFUs regulate piRNA biogenesis by quantifying changes in pre-21U species when TOFUs are perturbed.

Pre-21Us are distinctive in that they possess a 5' cap structure and begin 2 nt upstream of and end ~3 nt downstream from mature 21U-RNAs. Given these unique features, it is possible to quantify pre-21U levels by specifically cloning capped sRNAs (csRNA) using a CIP-TAP-based sRNA cloning protocol (Supplemental Fig. S3A; Gu et al. 2012). From our csRNA sequencing data, all reads >22 nt that mapped to 21U loci exhibited a 3U and one-purine bias and peaked in size at 26–27 nt, which are the expected pre-21U signatures (Fig. 5A; Supplemental Fig. S3B; Gu et al. 2012). We noticed that the csRNA sequencing protocol enriched for mature piRNAs and for miRNA sequences that lack the first nucleotide, which was unexpected based on our existing understanding of their structure. Otherwise, the csRNA libraries appeared as expected; mature piRNAs and miRNAs were virtually

Table 1. TOFU candidates

Sequence	Gene	Fold change	Domains/processes	Germline enrichment
C27H6.3	<i>tofu-1</i>	17	ATP binding	15.4
F20A1.9	<i>tofu-2</i>	3	SPRY of DDX1 and ATP binding	9.4
K02F2.4	<i>ulp-5/tofu-3</i>	5	Sumo-related peptidase	6.6
F57C9.7	<i>tofu-4</i>	3	—	3.8
W01A8.5	<i>tofu-5</i>	3	SANT	5.2
EEED8.1	<i>mel-47/tofu-6</i>	6	RRM and Tudor	22.9
ZK418.8	<i>tofu-7</i>	3	KH	7.5
F32H2.1	<i>gei-11</i>	3	SANT/RNA pol2/3 transcription	4.1
K06A5.7	<i>cdc-25.1</i>	3	RHOD/cell division cycle related	9.9
EGAP2.3	<i>pho-1</i>	2	Histidine acid phosphatase	—
C35E7.8	<i>C35E7.8</i>	2	ATP binding	7.4
Y23H5A.3	<i>Y23H5A.3</i>	2	RNA binding	7.4
F10D2.9	<i>fat-7</i>	2	Acyl-coA desaturase	—
ZK858.4	<i>mel-26</i>	2	Adaptor of E3 ubiquitin ligase	2.6
C06G3.10	<i>cogc-2</i>	2	Exocytosis	2.8
T20B12.2	<i>tbp-1</i>	2	TATA box/Transcription regulation	4.0
ZK809.4	<i>ent-1</i>	2	Nucleoside transporter	2.2
F57B10.12	<i>mei-2</i>	2	Microtubule disassembly	29.9
C32E8.8	<i>ptr-2</i>	2	SSD/cytokinesis	9.3
Y37D8A.16	<i>Y37D8A.16</i>	2	—	—
F57B10.3	<i>F57B10.3</i>	2	Phosphoglycerate mutase	—
D2092.3	<i>acr-11</i>	2	Nicotinic acetylcholine receptor	—

List of 22 TOFU candidates. Fold change denotes *21ur-5045* fold reduction relative to *cel-miR-35* as measured by qRT-PCR and then normalized to empty vector RNAi strain fold change. The value given is the average of individual clones, as described in Figure 1A. Domains and processes represent a summary of predictions and descriptions from Phyre2, NCBI Blastp, and WormBase (Kelley and Sternberg 2009). Germline enrichment was derived from a previous report that compared N2 (wild-type) versus a *glp-4 (bn2)* germlineless mutant, where gene expression was measured using reverse-transcribed poly(A)⁺ RNA hybridized onto DNA microarrays (Reinke et al. 2000).

nonexistent, and capped pri-miRNA 5' ends were enriched. sRNA libraries cloned from the same total RNA with methods that require a 5' monophosphate enriched for mature piRNAs and miRNAs with correct lengths as expected.

Comparisons of csRNA data revealed that loss of individual TOFUs affected pre-21U levels differently. The patterns of differential pre-21U levels were consistent regardless of our method of normalization, be it to total mapped reads or to pri-miRNA 5' ends that were enriched in csRNA sequences (Fig. 5B; Supplemental Fig. S3C). Upon silencing of *tofu-1* or *tofu-2*, we observed a twofold to threefold increase in pre-21Us. This accumulation was further enhanced in the *tofu-1 (tm6424)*- and *tofu-2 (tm6145)*-null mutants. Therefore, TOFU-1 and TOFU-2 appeared to play roles in the downstream processing of pre-21Us into mature 21U-RNAs. On the other

hand, perturbations of *tofu-3*, *tofu-4*, or *tofu-5* caused severe reductions in pre-21U levels in both RNAi strains and null mutants. This strongly implicated these TOFUs in pre-21U synthesis in either transcription or 5' capping.

Discussion

Since 21U-RNAs were first identified in *C. elegans*, this enigmatic class of sRNAs has been implicated in maintaining fertility, surveying the genome for foreign elements, and initiating the multigenerational transcriptional silencing of these elements. Despite our expanding knowledge of the function of 21U-RNAs, we are only just beginning to uncover the mechanisms by which these piRNAs are made in nematodes. By identifying a new class of ~26-nt capped transcriptional units as the intermediate between the 21U genomic locus and the mature 21U-RNA,

Table 2. piRNA suppressors

Sequence	Gene	Fold change	Domains/processes	Germline enrichment
C54D1.5	<i>lam-2</i>	8	Laminin	—
W03F8.5	<i>lam-1</i>	8	Laminin	—
K08C7.3	<i>epi-1</i>	7	Laminin	—
F20D12.1	<i>csr-1</i>	6	PAZ and PIWI	3.7
F26A3.3	<i>ego-1</i>	3	RNA binding/RNAi	4.4
F22D6.6	<i>ekl-1</i>	3	Tudor	2.3
ZK520.4	<i>cul-2</i>	2	Ubiquitin protein ligase binding	—

List of eight suppressors of 21U-RNA expression. Same format as Table 1, except that fold change denotes *21ur-5045* fold increase relative to *cel-miR-35* as measured by qRT-PCR and then normalized to empty vector RNAi strain fold change.

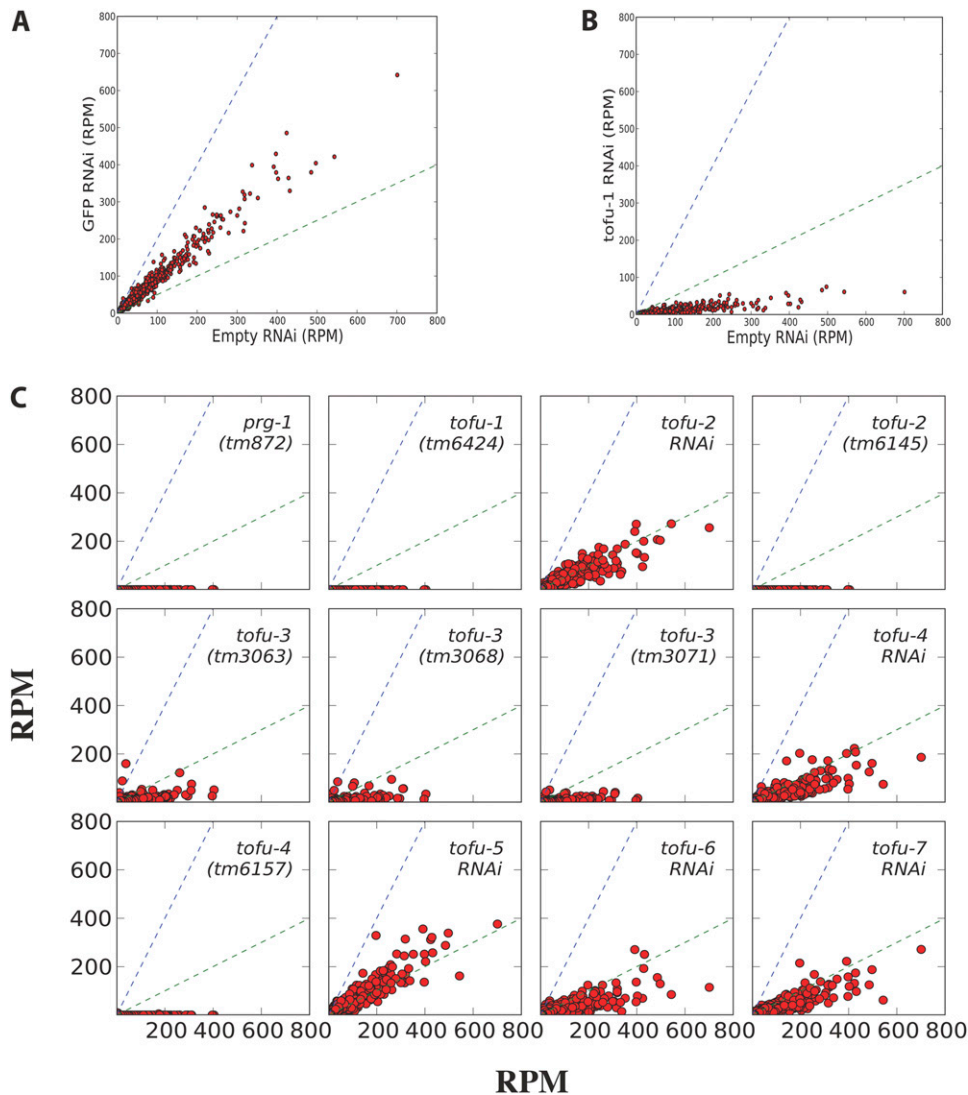


Figure 3. 5' monophosphate-dependent sRNA quantification of 21U-RNA defect upon TOFU loss. (A,B) Scatter plots of piRNA levels from 5' monophosphate-dependent sRNA cloning data of VT1012 worms are shown. Each dot represents a unique piRNA. Sample libraries—GFP RNAi (A) and *tofu-1* RNAi (B)—are both plotted on the Y-axis against empty vector RNAi control on the X-axis. All dots above the blue dashed line and below the green dashed line represent piRNAs that are more than twofold increased or decreased, respectively, in the sample library as compared with empty vector RNAi library. (C) Scatter plots of all piRNAs from all 5' monophosphate-dependent sRNA cloning data are shown for the indicated RNAi strains and null mutants. The format is the same as in A and B. The control for RNAi worms is the empty vector RNAi of VT1012 worms, while the control for mutant worms is N2. RPM denotes piRNA read counts per million mapped unstructured read counts.

Gu et al. (2012) and Cecere et al. (2012) provided an outline for piRNA biogenesis in *C. elegans*. Furthermore, the latter group linked the “CTGTTTCA” sequence found in the 21U-RNA upstream region to forkhead proteins that could recognize promoters for individual pre-21Us. However, neither group has tested the role of forkhead proteins in pre-21U synthesis.

In *Drosophila*, combined genetic and biochemical analyses have provided an extensive list of factors, which mediate the processing of piRNA intermediates into mature piRNAs (Guzzardo et al. 2013). To provide a similar substrate for understanding 21U biogenesis, we undertook a quantitative, genome-wide, reverse genetic screen for

factors that play roles in 21U-RNA production. The screen, which used the direct quantification of a 21U-RNA as its readout, yielded a list of 22 TOFUs for which corresponding dsRNAs could reproducibly induce a piRNA deficit. We did not identify forkhead family proteins as positives in our screen, but this was not surprising given their potential for redundancy.

Previously, two large-scale screens were carried out in our laboratory for components of the *Drosophila* piRNA pathway (Czech et al. 2013; Muerdter et al. 2013). Orthologs of TOFU genes identified in *C. elegans* were not found in lists of piRNA pathway components from either the *Drosophila* germ lineage or its somatic follicle

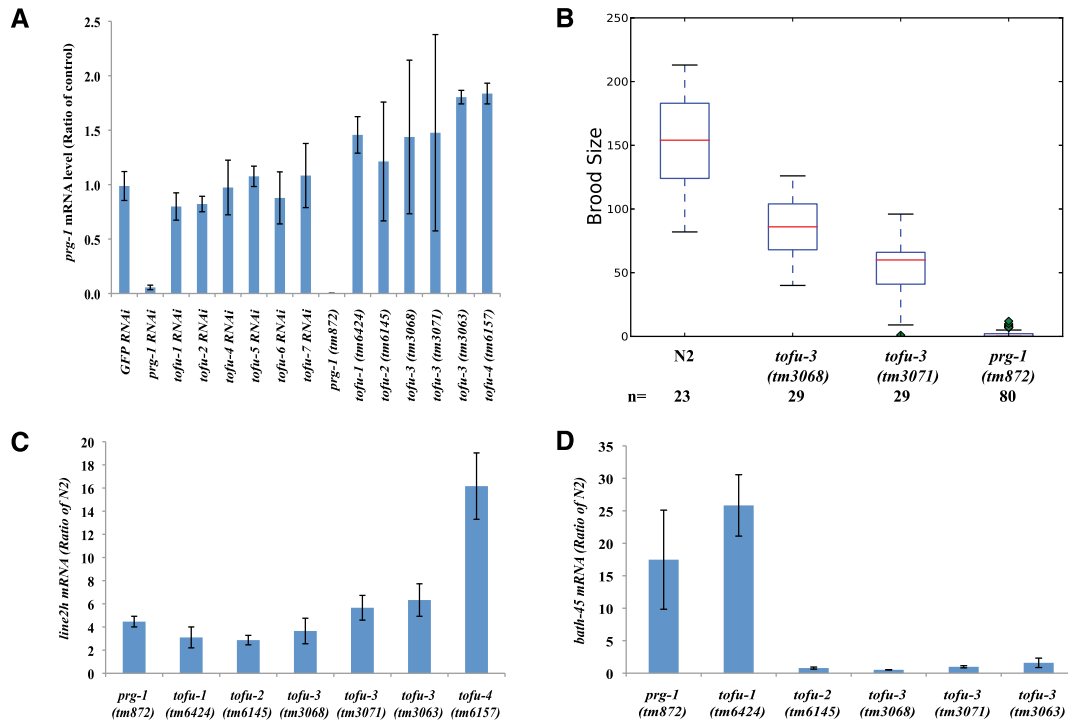


Figure 4. Independent validations of TOFU-dependent piRNA defects. (A) qRT-PCR of *prg-1* mRNA normalized to respective control levels are shown. Error bars represent standard deviation of at least two replicates. The control for RNAi worms is the empty vector RNAi of VT1012 worms, while the control for mutant worms is N2. (B) Brood sizes of individual unmated hermaphrodites at the nonpermissive temperature of 25°C are plotted as box and whisker plots. Green diamonds represent outliers. (C) qRT-PCR of *line2h* transposon levels in *prg-1* (tm872), *tofu-1* (tm6424), *tofu-2* (tm6145), *tofu-3* (tm3068, tm3071, and tm3063), and *tofu-4* (tm6157) relative to *act-3* mRNA and normalized to N2 levels are displayed. Error bars represent standard deviation of at least two replicates. (D) qRT-PCR of *bath-45* mRNA levels in *prg-1* (tm872), *tofu-1* (tm6424), *tofu-2* (tm6145), and *tofu-3* (tm3068, tm3071, and tm3063) relative to *act-3* mRNA and normalized to N2 levels are displayed. Error bars represent standard deviation of at least two replicates.

cells. Based on this observation, we believe that the production of piRNAs in nematodes proceeds via a biochemical mechanism distinct from that in *Drosophila melanogaster*. This conclusion is supported by differences in the genome organization of piRNAs in worms as compared with many other animals that have been studied.

Our highest-confidence TOFUs are mostly genes that have not previously been studied. We short-listed the seven highest-confidence TOFUs for further validation and found that none of them regulate piRNA levels through impacts on *prg-1* mRNA, as would be expected of core biogenesis factors. In order to dissect the role of TOFUs in piRNA biogenesis, we sought to quantify any changes in pre-21Us upon the loss of individual TOFUs, which might in turn lead to a 21U-RNA deficit. Using csRNA cloning, we detected an accumulation of pre-21Us despite a deficit of mature 21U-RNAs in *tofu-1*- or *tofu-2*-deficient worms. To our knowledge, this is the first time that any gene required for the processing steps that occur between pre-21Us and mature 21U-RNAs has been identified. However, whether *tofu-1* and *tofu-2* function in 5' decapping, removal of the 2-nt 5' overhang, or 3' trimming remains to be determined (Fig. 5C, middle panel). It is noteworthy that TOFU-2, with its SPRY- and ATP-binding

domains, resembles the DEAD-box RNA helicase DDX1, suggesting that TOFU-2 could participate in an RNA processing reaction (D'Cruz et al. 2012).

We found TOFU-3–5 to be required for the expression of pre-21Us. TOFU-3/ULP-5 is homologous to the mammalian SUMO-specific protease SENP7. This protease is known to modulate heterochromatin protein 1 retention at pericentric heterochromatin (Maison et al. 2012). TOFU-5 contains SANT-like domains that can bind to histone tails and participate in chromatin remodeling (Boyer et al. 2004). It is thus possible that TOFU-3 and TOFU-5 might regulate transcription of pre-21Us through modification of chromatin near pre-21U transcription start sites. Although TOFU-4 has no known domains, given the strong similarity in the 21U and pre-21U phenotype with TOFU-3 and TOFU-5, we imagine that TOFU-4 may also have a role in regulating pre-21U transcription or capping (Fig. 5C, right panel). In the *prg-1* mutant, pre-21Us were modestly reduced, roughly by approximately threefold as compared with levels seen in wild-type animals. This highlights an unexpected coupling of steps within the pathway, the nature of which remains to be determined.

TOFU-6 and TOFU-7 affect mature piRNAs without any substantial effect on precursor levels. This implies that they act downstream from primary biogenesis and

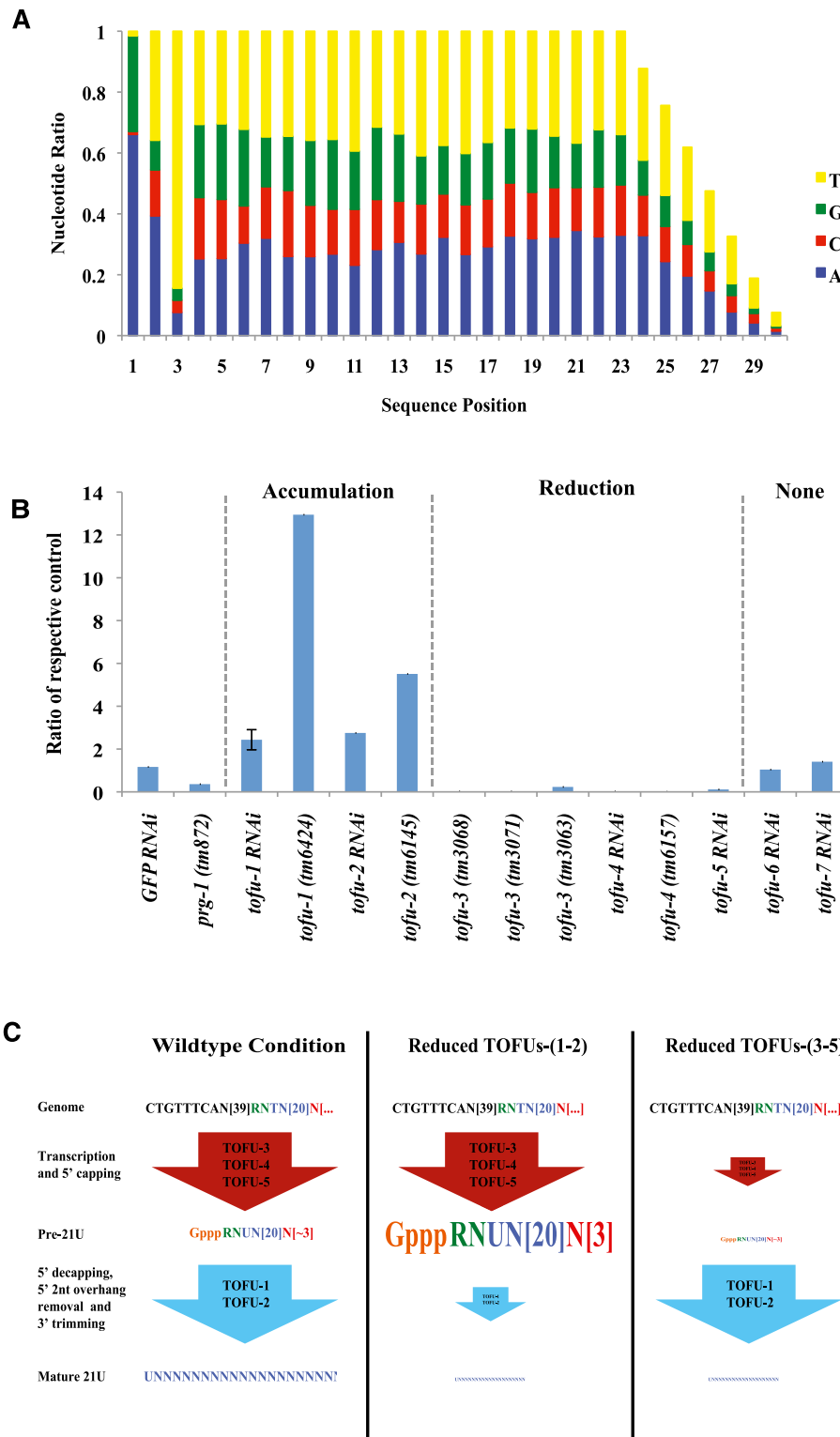


Figure 5. Characterizing the requirement of TOFUs in piRNA biogenesis through 5' cap-dependent sRNA cloning. (A) Nucleotide distribution at each position of >22-nt capped sequences that overlap with 21U-RNA loci in the N2 csRNA library are shown. Nucleotide ratio was calculated as the read count of each nucleotide at each position as a fraction of the total read counts of all nucleotides at position 1. (B) Pre-21U reads relative to total mapped unstructured reads are plotted normalized to respective control levels. The control for RNAi worms is the empty vector RNAi of VT1012 worms, while the control for mutant worms is N2. Dotted lines separate groups of TOFUs based on their effects on pre-21U levels—accumulation, reduction, and none—as indicated. Error bars are indicated only for *tofu-1* RNAi and represent a standard deviation of two replicates. (C, left) Schematic of piRNA biogenesis model is presented. (Right) The same model is also depicted under conditions of perturbed TOFU-1–2 (middle) or perturbed TOFU-3–5. (A,C,G,T,U) DNA or RNA nucleotides; (N) any DNA or RNA nucleotide; (R) any purine nucleotide; (p) phosphate group.

upstream of mature, stable PRG-1 complexes. The RRM domain of TOFU-6 and the KH domain of TOFU-7 suggest that they may function as RNA-binding proteins. Perhaps these factors act during loading of 21U-RNAs into PRG-1, which is likely important for their stability, based on PRG-1 loss-of-function phenotypes. Alternatively, they could function upstream of piRISC loading by helping to ensure proper localization of processed 21U-RNAs.

Through a genome-wide RNAi screen, we established the first genetic framework for *C. elegans* 21U-RNA biogenesis by identifying 22 TOFU genes. For seven of the strongest, integration of measurements of levels of mature sRNAs and their precursors has provided insights into the step of 21U biogenesis at which each protein acts. These efforts provide an important step toward a detailed biochemical understanding of *C. elegans* piRNA biogenesis and afford the opportunity to compare its mechanism with our emerging understanding of piRNA production in other animals.

Materials and methods

Worm strains

The Bristol N2 strain was used as the standard reference wild-type strain. Other strains/alleles used in this study were WM161 *prg-1(tm872)* I, TM6424 *tofu-1(tm6424)* V, TM6145 *tofu-2(tm6145)* V, TM3068 *tofu-3/ulp-5(tm3068)* I, TM3071 *tofu-3/ulp-5(tm3071)* I, TM3063 *tofu-3/ulp-5(tm3063)* I, TM6157 *tofu-4(tm6157)* I, VC632 *gna-2(ok867)* I/hT2[*bli-4(e937)* *let-1(q782)* *qls48*] (I;III), VC729 *gna-2(gk308)* I/hT2[*bli-4(e937)* *let-1(q782)* *qls48*] (I;III), WM182 *csr-1(tm892)* IV/nT1[*unc-1(n754)* *let-1*] (IV;V), ZT3 *csr-1(fj54)* IV/nT1[*qls51*] (IV;V), and VT1012 *rrf-3(pk1426)* II; *mals105* V. All RNAi experiments were carried out in VT1012. Only non-*Unc csr-1* homozygous null WM182 worms and non-GFP *csr-1* homozygous null ZT3 worms were picked and lysed for *21ur-5045* quantification. All strains were cultured essentially as described (Brenner 1974).

Worm lysis and reverse transcription for qPCR of *cel-mir-35* and *21ur-5045*

Approximately 20 gravid adults were collected and resuspended in 25 μ L of lysis buffer containing 50 mM KCl, 10 mM Tris (pH 8), 10 mM Tris (pH 8), 2.5 mM MgCl₂, 0.45% NP-40, 0.45% Tween-20, 0.01% gelatin, and 0.06 mg/mL New England Biolabs proteinase K. Worms were lysed for 1 h at 60°C and for 20 min at 94°C. Worm lysates were diluted 10 \times in distilled water before being used for reverse transcription. Custom TaqMan sRNA assays from Applied Biosystems were designed for the reverse transcription and qPCR of the embryonic-specific endogenous control miRNA (*cel-mir-35*) and piRNA (*21ur-5045*), which were used for all sRNA qRT-PCRs in this study.

qPCR screen for TOFUs and suppressors

We used the *C. elegans* RNAi library from Source BioScience, which contains HT115 bacteria that express dsRNAs homologous to *C. elegans* genes (Kamath et al. 2003). dsRNAs were expressed from gene sequences inserted between bidirectional IPTG-inducible T7 promoters. Using an adapted 96-well liquid-based RNAi screening protocol, ~30 staged VT1012 L1s were fed individual clones from the RNAi library (Timmons and Fire 1998; Lehner

et al. 2006). After 5 d, adult worms were lysed, and qRT-PCR (40 cycles) of *21ur-5045* and *cel-mir-35* was performed on the lysate. For each group of up to 96 clones, the piRNA fold change of each RNAi clone was calculated relative to the median fold change of its group (Supplemental Table S1). All RNAi clones that exhibited a poor miR-CT (>25) were considered to cause either lethality or germline defects as RNAi phenotypes and were excluded from further analyses. A Z-score cutoff of 2.5 was used as a primary criterion to identify the initial group of 209 candidates, with the mean and standard deviation calculated based on the piRNA fold changes of all dsRNAs that were analyzed. Counterparts of the candidates from an ORF-based RNAi library and other genes of interest were then spiked into this group of candidates (Rual et al. 2004). Frozen glycerol stocks of the candidates from the RNAi libraries were streaked onto ampicillin plates. Three single clones were picked and sequenced to confirm the identity of each candidate. The clones were grown in shakers overnight, then sent for rolling circle amplification and sequencing. These three single clones were also used for validation assays. An empty vector RNAi control as well as RNAi against GFP and *prg-1* were also included controls at this point. The same qRT-PCR assay was then repeated with the piRNA fold change calculated relative to an empty vector RNAi construct (Supplemental Table S2). On the basis of reproducibility, as a secondary criterion, only candidates whose average fold change exceeded 1.62 were retained, leaving 92 candidates. In order to remove potential false positives from within this list of 92 candidates, miR-CT and 21U-CT values from the last qRT-PCR assay were plotted as a scatter plot (Figs. 1C, 2A). By comparing the scatter pattern of established false positives with that of empty vector RNAi control, GFP, and *prg-1*, a tertiary criterion was established to filter out potential false positives. First, the median miR-CT and 21U-CT of the combined group of empty vector control and GFP were determined. For each candidate, the absolute shift in its miR-CT and 21U-CT relative to the control/GFP median was calculated. Any candidate with even one replicate that exhibited a horizontal shift (greater absolute miRNA-CT shift than its absolute 21U-CT shift) was denoted as a false positive and filtered out, leaving 22 TOFUs and seven suppressors.

RNA preparation

Unless otherwise stated, all mRNA qRT-PCRs, sRNA cloning, and csRNA cloning and some sRNA qRT-PCRs were carried out on total RNA prepared as described here. Staged N2 or N2-based mutant L1s were grown on NGM plates for 3 d at 20°C. Staged VT1012 L1s were grown in S Basal buffer (with 100 μ g/mL ampicillin, 4 mM IPTG) for 4 d at 20°C. The resultant staged 12- to 36-h adults were harvested and washed with M9 buffer until the supernatant was clear. Worms were homogenized in Trizol reagent and subjected to three cycles of freezing in liquid nitrogen and thawing at room temperature before RNA extraction by phase separation and isopropanol precipitation.

mRNA reverse transcription and qPCR

For mRNA and transposon RT-qPCR, 1–2 μ g of total RNA was DNase-treated with Promega RQ1 and then reverse-transcribed using random hexamers and Superscript III reverse transcriptase. The resultant cDNAs were quantified using specific qPCR primer pairs and SYBR Green PCR master mix.

qPCR primers

The endogenous gene control *act-3* cDNA was amplified using 5'-TGCTGATCGTATGCAGAAAG-3' and 5'-ATCTTGATCTT

CATGGTTGATGG-3', and *prg-1* cDNA was amplified using 5'-GTGGGCACGTGGCGTGAAGG-3' and 5'-ACCAAGACGACGAGCCTCCTCAATG-3'. *line2h* and *bath-45* cDNAs were amplified using forward and reverse primers as listed (Bagijn et al. 2012).

sRNA cloning and csRNA cloning

Unless otherwise stated, 20 µg of total RNA from 12- to 36-h adults was used for cloning. sRNA libraries were cloned as described (Malone et al. 2012). csRNA cloning was modified from the sRNA cloning protocol by including CIP and TAP steps before 5' ligation of sequencing adaptors. Libraries were sequenced in-house using the Illumina HiSeq or Genome Analyzer II SE36 platform. Deep-sequencing data can be found at Gene Expression Omnibus using accession number GSE53970.

Bioinformatics analysis

FASTQ reads had their 3' adaptor sequences clipped off before mapping to the *ce6* genome using Bowtie, allowing no mismatches and suppressing reads that mapped >10 times. Mapped reads were annotated with the following priority: piRNA loci > miRNA loci > genic regions > repeat regions. All other bioinformatics analyses were carried out using custom home-built Python scripts. piRNA coordinates and sequences are as listed (Batista et al. 2008). Pri-miRNA 5' ends are as listed (Gu et al. 2012). Pre-21Us were denoted as >22-nt sequences that began exactly 2 nt upstream of mature piRNA coordinates. Unless otherwise stated, piRNA and pre-21U read counts are always calculated as raw piRNA reads per million nonstructural mapped reads.

Brood size analysis

Mixed staged worms were grown overnight in 25°C. On the following day ($T = -1$ d), single hermaphrodite L2s or L3s were picked onto individual NGM plates and allowed to grow for 48 h at 25°C. On $T = 1, 2, 3$ d, P0 adults were transferred to new plates at 25°C, leaving the F1s behind. On $T = 4$ d, P0 adults were removed. All F1 offspring that remained on the plates were counted 2 d after removal of P0 adults. The individual brood size of each P0 adult is the total sum of F1s for all four plates. All P0 adults that died before $T = 4$ d were discounted from the brood size analysis.

Acknowledgments

We are grateful to Sabrina Boettcher for handling the logistics component of the screen. We thank Assaf Gordon, Dario Bressan, and Emily Lee for their respective help with bioinformatics, robotics, and sequencing. We acknowledge Weifeng Gu and Craig Mello for advice on the csRNA cloning protocol. We also thank Jon Preall, Felix Muerdter, Paloma Guzzardo, Ben Czech, Weifeng Gu, and Craig Mello for sharing data prior to publication. We thank all past and present members of the Hannon laboratory for helpful discussions. We acknowledge the National Bioresource Project of Japan as well as the *Caenorhabditis* Genetics Center, which is funded by the National Institutes of Health (NIH) Office of Research Infrastructure Programs (P40 OD010440), for providing worm strains. W.-S.S.G. is a Delbrück fellow of the Watson School of Biological Sciences and is supported by a National Science Scholarship (PhD) from the Agency for Science, Technology, and Research, Singapore. This work was supported by NIH grant 5R01 GM062534 to G.J.H. G.J.H. is a Howard Hughes Medical Institute Investigator.

References

- Ashe A, Sapetschnig A, Weick E-M, Mitchell J, Bagijn MP, Cording AC, Doebley A-L, 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.
- Bagijn MP, Goldstein LD, Sapetschnig A, Weick E-M, Bouasker S, Lehrbach NJ, Simard MJ, Miska EA. 2012. Function, targets, and evolution of *Caenorhabditis elegans* piRNAs. *Science* **337**: 574–578.
- Batista PJ, Ruby JG, Claycomb JM, Chiang R, Fahlgren N, Kasschau KD, Chaves DA, Gu W, Vasale JJ, Duan S, et al. 2008. PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol Cell* **31**: 67–78.
- Boyer LA, Latek RR, Peterson CL. 2004. The SANT domain: a unique histone-tail-binding module? *Nat Rev Mol Cell Biol* **5**: 158–163.
- Brenner S. 1974. The genetics of *Caenorhabditis elegans*. *Genetics* **77**: 71–94.
- Cecere G, Zheng GXY, Mansisidor AR, Klymko KE, Grishok A. 2012. Promoters recognized by forkhead proteins exist for individual 21U-RNAs. *Mol Cell* **47**: 734–745.
- 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.
- Conine CC, Moresco JJ, Gu W, Shirayama M, Conte D Jr, Yates JR 3rd, Mello CC. 2013. Argonautes promote male fertility and provide a paternal memory of germline gene expression in *C. elegans*. *Cell* **155**: 1532–1544.
- Czech B, Preall JB, McGinn J, Hannon GJ. 2013. A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway. *Mol Cell* **50**: 749–761.
- D'Cruz AA, Babon JJ, Norton RS, Nicola NA, Nicholson SE. 2012. Structure and function of the SPRY/B30.2 domain proteins involved in innate immunity. *Protein Sci* **22**: 1–10.
- Das PP, Bagijn MP, Goldstein LD, Woolford JR, Lehrbach NJ, Sapetschnig A, Buhecha HR, Gilchrist MJ, Howe KL, Stark R, et al. 2008. Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. *Mol Cell* **31**: 79–90.
- Denli AM, Tops BBJ, Plasterk RHA, Ketting RF, Hannon GJ. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**: 231–235.
- Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, 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.
- Gu W, Shirayama M, Conte D Jr, 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.
- Gu W, Lee H.-C, Chaves D, Youngman EM, Pazour GJ, Conte D Jr, 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.
- Guzzardo PM, Muerdter F, Hannon GJ. 2013. The piRNA pathway in flies: highlights and future directions. *Curr Opin Genet Dev* **23**: 44–52.
- Johnston WL, Krizus A, Dennis JW. 2006. The eggshell is required for meiotic fidelity, polar-body extrusion and polarization of the *C. elegans* embryo. *BMC Biol* **4**: 35.

- Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, et al. 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**: 231–237.
- Kelley LA, Sternberg MJE. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**: 363–371.
- Lau NC, Lim LP, Weinstein EG, Bartel DP. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**: 858–862.
- Lehner B, Tischler J, Fraser AG. 2006. RNAi screens in *Caenorhabditis elegans* in a 96-well liquid format and their application to the systematic identification of genetic interactions. *Nat Protoc* **1**: 1617–1620.
- Luteijn MJ, van Bergeijk P, Kaaij LJT, Almeida MV, Roovers EF, Berezikov E, Ketting RF. 2012. Extremely stable Piwi-induced gene silencing in *Caenorhabditis elegans*. *EMBO J* **31**: 3422–3430.
- Maison C, Romeo K, Bailly D, Dubarry M, Quivy J-P, Almouzni G. 2012. The SUMO protease SENP7 is a critical component to ensure HP1 enrichment at pericentric heterochromatin. *Nat Struct Mol Biol* **19**: 458–460.
- Malone C, Brennecke J, Czech B, Aravin A, Hannon GJ. 2012. Preparation of small RNA libraries for high-throughput sequencing. *Cold Spring Harb Protoc* **2012**: 1067–1077.
- Montgomery TA, Rim Y-S, Zhang C, Dowen RH, Phillips CM, Fischer SEJ, Ruvkun G. 2012. PIWI associated siRNAs and piRNAs specifically require the *Caenorhabditis elegans* HEN1 ortholog henn-1. *PLoS Genet* **8**: e1002616.
- Muerdter F, Guzzardo PM, Gillis J, Luo Y, Yu Y, Chen C, Fekete R, Hannon GJ. 2013. A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in *Drosophila*. *Mol Cell* **50**: 736–748.
- Reinke V, Smith HE, Nance J, Wang J, Van Doren C, Begley R, Jones SJ, Davis EB, Scherer S, Ward S, et al. 2000. A global profile of germline gene expression in *C. elegans*. *Mol Cell* **6**: 605–616.
- Rual J-F, Ceron J, Koreth J, Hao T, Nicot A-S, Hirozane-Kishikawa T, Vandenhaute J, Orkin SH, Hill DE, van den Heuvel S, et al. 2004. Toward improving *Caenorhabditis elegans* phenome mapping with an ORFeome-based RNAi library. *Genome Res* **14**: 2162–2168.
- Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP. 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* **127**: 1193–1207.
- Seth M, Shirayama M, Gu W, Ishidate T, Conte D Jr., Mello CC. 2013. The *C. elegans* CSR-1 argonaute pathway counteracts epigenetic silencing to promote germline gene expression. *Dev Cell* **27**: 656–663.
- Shirayama M, Seth M, Lee H-C, Gu W, Ishidate T, Conte D, Mello CC. 2012. piRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. *Cell* **150**: 65–77.
- Siomi MC, Sato K, Pezic D, Aravin AA. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* **12**: 246–258.
- Timmons L, Fire A. 1998. Specific interference by ingested dsRNA. *Nature* **395**: 854.
- Wang G, Reinke V. 2008. A *C. elegans* Piwi, PRG-1, regulates 21U-RNAs during spermatogenesis. *Curr Biol* **18**: 861–867.
- Wedeles CJ, Wu MZ, Claycomb JM. 2013. Protection of germline gene expression by the *C. elegans* Argonaute CSR-1. *Dev Cell* **27**: 664–671.
- Zhang XD. 2011. Illustration of SSMD, z score, SSMD*, z* score, and t statistic for hit selection in RNAi high-throughput screens. *J Biomol Screen* **16**: 775–785.