



Published in final edited form as:

Science. 2018 February 02; 359(6375): 587–592. doi:10.1126/science.aao2840.

The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes

Donglei Zhang^{1,4,†}, Shikui Tu^{2,5,†}, Michael Stubna¹, Wei-Sheng Wu³, Wei-Che Huang³, Zhiping Weng², and Heng-Chi Lee^{1,*}

¹Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, IL 60637 USA

²Program in Bioinformatics and Integrative Biology, University of Mass. Medical School, Worcester, MA 01605 USA

³Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan

⁴Department of Biochemistry and Molecular Biology, School of Basic Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

⁵Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

Abstract

piRNAs silence transposons to safeguard genome integrity in animals. However, the functions of the many piRNAs that do not map to transposons remain unknown. Here we showed that piRNA targeting in *C. elegans* can tolerate a few mismatches but prefer perfect pairing at the seed region. The broad targeting capacity of piRNAs underlies the germline silencing of transgenes in *C. elegans*. Transgenes engineered to avoid piRNA recognition are stably expressed. Interestingly, many endogenous germline-expressed genes also contain predicted piRNA targeting sites, and periodic An/Tn clusters (PATCs) are an intrinsic signal that provides resistance to piRNA silencing. Together, our study revealed the piRNA targeting rules and highlights a unique strategy that *C. elegans* uses to distinguish endogenous from foreign nucleic acids.

PIWI and its associated piRNAs function as a guardian of animal genomes through transposon silencing in various animals (1–5). However, many animals produce piRNAs that do not match transposon sequences. For example, the vast majority of the 15,000 piRNAs encoded by the *C. elegans* genome do not exhibit extensive complementarity to transposons (3, 4, 6). In mice, tens of thousands of distinct piRNAs produced at the pachytene stage

*Correspondence to: hengchilee@uchicago.edu (H.-C. L.).

†These authors contributed equally to this work

Supplementary Materials:

Materials and Methods

Figs. S1 to S9

Tables S1 to S4

Full-length sequences of modified GFP, mCherry and Cas9

Program package with algorithms of piRNA target prediction

References (40–45)

during spermatogenesis do not map to transposons (7). These observations suggest additional targets and functions of piRNAs.

Identification of piRNA targets and the piRNA targeting rules has proven to be rather difficult. Crosslinking immunoprecipitation (CLIP) analyses of PIWIs suggest that they associate with diverse mRNAs (8–10). However, because diverse piRNAs engage with many mRNAs, it is difficult to infer the target of a given piRNA from these CLIP analyses. Therefore, additional approaches are required to identify piRNA sites *in vivo*. In some cases, targets of piRNAs can be inferred if the mRNA target is cleaved by PIWI (11, 12). However, these cleaved mRNAs likely present only a fraction of piRNA targets *in vivo*, since the slicer activity of PIWI is dispensable for silencing in some animals, including *C. elegans* (13–16). As only few piRNA targets other than transposons have been identified, the piRNA targeting rules remain undefined and both sequence-specific and sequence-nonspecific functions of PIWI/piRNA complex have been proposed (8–10, 12, 17, 18).

To gain insight into the piRNA targeting mechanism, we identified the targets of a single piRNA and examine how the piRNA recognizes its targets. In *C. elegans*, piRNA targeting leads to the recruitment of RNA dependent RNA polymerases (RdRPs) that produce secondary small RNAs named 22G-RNAs (3, 13, 15) (fig. S1A). These 22G-RNAs are loaded onto worm-specific Argonautes (WAGOs) to induce gene silencing (19–21). Because these 22G-RNAs are produced around the targeting site, the 22G-RNAs can serve as a “signature” for piRNA targeting sites *in vivo* (13, 15). Therefore, we identified the targets of a piRNA by examining the 22G-RNA species gained in animals expressing a synthetic piRNA or the 22G-RNA species lost in animals carrying a deletion of a specific piRNA (fig. S1B). We obtained animals expressing a synthetic piRNA or losing an endogenous piRNA through a CRISPR Cas9-based genome editing strategy that modified the locus of an endogenous piRNA (fig. S1B). Small RNA sequencing confirmed the expression or loss of specific piRNAs in these animals (fig. S1C-F) and was used to identify changes in 22G-RNA levels. Together, we identified six RNA targets in the animals producing the synthetic piRNAs and eleven RNA targets in animals lacking the endogenous piRNAs (Fig. 1A–C and table S1). We noticed that a region of the piRNAs, from the 2nd to 7th nucleotide, pairs well to the identified targeting sites (Fig. 1D). This implies a critical role for the pairing of this region in piRNA targeting, which we define as the piRNA seed. The piRNA seed is reminiscent of the miRNA seed, which is essential for miRNA target recognition (22). In addition, we observed significant pairing outside of the piRNA seed region (Fig. 1D and fig. S1G). These observations suggest that base pairing outside of the seed region also contributes to piRNA target recognition, but a few mismatches can be tolerated. Furthermore, we noticed that GU wobble pairs are over-represented, relative to other non-Watson-Crick pairs, in these targeting events (table S1). Finally, the first nucleotide does not appear to contribute to piRNA targeting (Fig. 1D).

In light of these findings, we developed a piRNA reporter assay to gain further insights into the piRNA targeting rules. In this assay, we examined whether synthetic *gfp*-targeting piRNAs with various mismatches to the GFP sequence can trigger the silencing of an expressed GFP^{dpiRNA}::CDK-1 transgene (dpiRNA stands for depletion of piRNA targeting sites, in which the GFP sequence has been re-coded to avoid silencing by endogenous

piRNAs) (Fig. 2A left panel, fig. S2A, and see below). As we noticed that the synthetic piRNAs can be produced from animals carrying extrachromosomal arrays with synthetic piRNA loci, we chose this method to systemically produce various *gfp*-targeting piRNAs (fig. S2A). We observed that GFP^{dpiRNA}::CDK-1 was silenced in the animals expressing synthetic piRNAs that are perfectly complementary to GFP mRNAs or contain two mismatches outside of the piRNA seed region (Fig. 2A–C, fig. S2B). On the contrary, we failed to observe the silencing of GFP^{dpiRNA}::CDK-1 when one or two mismatches were located at the piRNA seed region (Fig. 2, B and C, and fig. S2B). In addition, our reporter assay suggests that piRNAs tolerate up to three non-seed mismatches but not RNA bulges (fig. S2C–E). We also observed that one GU wobble pair is tolerated in the seed region, and GU pairs are moderately more tolerated than mismatches in the non-seed region (fig. S2C and S2D). Finally, we obtained consistent results in our reporter analyses using gene-edited worms expressing synthetic *gfp*-targeting piRNAs from an endogenous piRNA locus (fig. S2F). Overall, our reporter assay revealed a similar but more stringent piRNA-targeting logic than that from our analyses of synthetic piRNA targets. Together, our analyses suggest that piRNA targeting in *C. elegans* prefers near-perfect pairing at the piRNA seed region. In addition, supplementary pairing outside of the seed region also contributes to piRNA targeting, but few mismatches are tolerated.

It has been known for decades that transgenes carrying various foreign sequences, such as green fluorescent protein (GFP) or mCherry, are frequently silenced in the germline of *C. elegans* (23, 24). A previous study has shown that PIWI protein PRG-1 is required for the silencing of the transgene GFP::CDK-1 (19). If piRNAs recognize GFP sequences, then removal of piRNA targeting sites from the GFP sequences should allow transgene GFP::CDK-1 expression in the germline. To predict more piRNA targeting sites on transgenes, we employed the relaxed piRNA targeting criteria similar to those derived from our analyses of synthetic piRNA targets (see supplementary materials for algorithms of target prediction). These criteria predicted 17 piRNA targeting sites on GFP mRNA (Fig. 3A and table S2). We introduced silent mutations in the GFP sequences such that we no longer identified piRNA targeting sites, yielding the recoded-GFP^{dpiRNA} sequences. Remarkably, while the GFP::CDK-1 transgene is always silenced in the germline of wild type animals, we observed strong GFP expression from all five independent transgenic strains we obtained with recoded GFP^{dpiRNA}::CDK-1 inserted at the same locus (Fig. 3B).

To test if our approach can be generally applied to other transgenes to avoid gene silencing, we chose to modify the mCherry-tagged C-terminal region of Anillin (mCherry::ANI-1⁶⁸¹⁻¹¹⁵⁹), another transgene that is always silenced in the germline (25). We predicted 10 piRNA targeting sites in mCherry mRNA and introduced silent mutations to disrupt predicted piRNA targeting sites (Fig. 3C, table S2). Whereas the original mCherry::ANI-1⁶⁸¹⁻¹¹⁵⁹ transgene was silenced in all six transgenic lines, the modified mCherry^{dpiRNA}::ANI-1⁶⁸¹⁻¹¹⁵⁹ was robustly expressed at the cleavage furrow of the one-cell embryo in all six transgenic lines we obtained (Fig. 3D).

As a final test, we applied this approach to modify Cas9 sequences. Transgenic *C. elegans* strains stably expressing Cas9 have not been successfully obtained (26). Again, we introduced silent mutations to remove all predicted piRNA targeting sites (Fig. 3E and table

S2) and obtained transgenic animals carrying the original or the modified Cas9 transgene. To test if Cas9 is stably expressed, we injected the transgenic animals with an *unc-22* sgRNA-expressing plasmid and a *rol-6(su1006)* plasmid as a dominant transformation marker (Fig. 3F). The animals carrying *unc-22* mutations exhibit a visible twitcher phenotype and can be easily identified (27). Importantly, out of thirty F1 transformed progeny (roller), nine animals produced F2 twitcher progeny from animals carrying the modified Cas9 transgene, while no F2 twitcher progeny were observed from animals carrying the original Cas9 transgene. DNA sequencing of these F2 twitcher animals confirmed that they carry various *unc-22*-edited alleles (Fig. 3F). These observations functionally demonstrate that the modified Cas9 transgene is stably expressed and thus can create edited alleles. Taken together, these experiments verify that our predictions of piRNA targeting sites encompass the critical sites that trigger gene silencing.

We next wondered whether endogenous germline genes have evolved to avoid piRNA recognition. Previous studies have shown that most *C. elegans* germline transcripts are targeted by either WAGO Argonaute-associated 22G-RNA, which correlates with silencing of the transcript, or CSR-1 Argonaute-associated 22G-RNA, which correlates with expression of the transcript (28, 29). Surprisingly, using stringent piRNA targeting criteria corresponding to the ones derived from our reporter analyses, we predicted that around half of germline-expressed genes (CSR-1 targets), as well as germline-silenced genes (WAGO-1 targets), contain at least one piRNA targeting site (Fig. 4A and fig. S3A), which is sufficient for silencing at least in our *gfp* reporter assay. In addition, the density of piRNA targeting sites in germline-expressed genes is only slightly less than that of somatic-specific genes and control sequences (fig. S3B). While the germline-silenced genes contain more predicted piRNA sites than germline-expressed genes, such differences alone cannot explain why only one set of genes is silenced (fig. S3B). Taken together, these predictions implied that some endogenously expressed germline genes are resistant to piRNA silencing. To test this hypothesis, we obtained animals that produce synthetic piRNAs that are perfectly complementary to several of these genes, including *pie-1*, *nop-1*, *cdk-1* and *oma-1*. We engineered these piRNAs using the same locus (21U-5499) we used for producing *gfp*-targeting piRNAs and these synthetic piRNAs were expressed at similar levels (Fig. 4B). Remarkably, in the animals expressing synthetic piRNAs targeting endogenous genes, we did not observe a reduction of mRNA levels or the phenotypes associated with silencing of these genes (Fig. 4B and fig. S4, A–D). In addition, no phenotype associated with silencing was observed in animals expressing any of six additional synthetic piRNAs that target various regions of the *pie-1* or *nop-1* genes (fig. S4A and S4B). This is in stark contrast to the animals expressing seven distinct *gfp* or mCherry-targeting piRNAs, which all trigger potent silencing of GFP^{dpRNA}::CDK-1 or mCherry^{dpRNA} ANI-1, respectively (Fig 4B, fig S4E and S4F). Together, our results suggested that at least some endogenous germline genes exhibit resistance to piRNA-mediated gene silencing in *C. elegans*.

Previous studies have proposed that CSR-1 Argonaute-associated 22G-RNAs may form an epigenetic memory of “self” to promote gene expression in the germline (30, 31). We therefore examined whether the *nop-1*-targeting piRNA can trigger *nop-1* silencing in *csr-1* mutants. In either *csr-1* F2 one-cell embryos with or without the treatment of *csr-1* RNAi, we did not observe that the synthetic piRNA conferred the phenotype associated with *nop-1*

silencing (fig S5A). To further test if the piRNA resistance of endogenous genes is mediated by epigenetic signals, we used Cas9-based gene editing strategy to delete the *nop-1* gene and its untranslated regions (UTRs) from the genome, which would remove all chromatin-based signals as well as the DNA/RNA templates that are required to produce CSR-1 22G-RNAs targeting *nop-1* (fig S5B). We then re-inserted the *nop-1* gene back to its original locus or to the locus where our silenced transgenes are inserted. Interestingly, the re-inserted *nop-1* gene remained resistant to silencing by endogenous or synthetic *nop-1*-targeting piRNA (fig. S6C–D). While these data represented negative results, our analyses provided no evidence for epigenetic mechanisms in licensing germline gene expression.

We therefore investigated if intrinsic sequences of germline genes provide resistance to piRNA silencing. One such candidate is 10-base periodic An/Tn clusters (PATCs), an intrinsic DNA sequence element found in the introns or promoters of some germline genes in *C. elegans* (32). A recent study has reported that PATCs can promote the expression of transgenes inserted at heterochromatin in *C. elegans*, but whether PATCs can provide resistance to piRNA silencing has not been explored (33). Interestingly, we found that PATCs are enriched in germline genes, and particularly enriched in the germline-expressed genes (Fig. 4C, fig. S6A and S6B). To examine the global effect of PATCs on piRNA silencing, we compared the local 22G-RNA distribution at predicted piRNA sites in germline-specific transcripts with high or low PATC density. Remarkably, we observed that local 22G-RNAs accumulated around the targets only for genes with low, but not high, PATC density (Fig. 4D and fig. S6C and S6D). These observations imply that PATCs negatively affect the ability of the piRNA pathway to induce and/or maintain 22G-RNAs at the piRNA targeting sites. Furthermore, we observed that germline-specific genes of higher PATC density contain more predicted piRNA targeting sites than genes with lower PATC density (Fig 4E and fig. S7A and S7B), implying that PATCs allow the expression of germline genes despite harboring piRNA targeting sites. Finally, if PATCs can provide resistance to piRNA silencing, insertion of PATC introns to a silencing-prone transgene should license its expression. Indeed, replacing the mCherry introns of the mCherry::ANI-1 transgene with PATC-containing introns from the *smu-1* gene led to the stable expression of the transgene (Fig. 4F). Small RNA sequencing further shows that dramatically fewer mCherry antisense 22G-RNAs are produced in the worms carrying mCherry^{PATC}::ANI-1⁶⁸¹⁻¹¹⁵⁹ than in those carrying the original mCherry::ANI-1⁶⁸¹⁻¹¹⁵⁹ (Fig. 4G). Together, our findings suggest that PATCs act as a licensing signal that provides resistance to piRNA silencing.

Overall, our study revealed the piRNA targeting logic in *C. elegans*. In addition, our research suggested that diverse piRNAs can recognize and silence various foreign nucleic acids due to their broad targeting capacity. Since several different modes of miRNA targeting have been described in animals and plants (22, 34–38), additional modes of piRNA-targeting are likely to exist as well. Nonetheless, our study demonstrated that piRNA-mediated gene silencing underlies the transgene silencing phenomenon in the germline of *C. elegans* and provided a simple solution to achieve transgene expression by avoiding piRNA recognition.

Our study showed that many endogenous genes also contain piRNA targeting sites but exhibit resistance to piRNA silencing. Our analyses suggested PATCs to be a licensing

signal protecting endogenous genes from piRNA silencing. How PATCs counter against piRNA silencing remains unknown. A recent study showed that PATCs are enriched in germline genes within repressive chromatin domain, suggesting that PATCs may prevent piRNAs from establishing heterochromatin at their target (33). Interestingly, our data suggest that PATCs function not simply by promoting euchromatin formation, but also by inhibiting the production of 22G-RNA at piRNA targeting sites (Fig 4D and 4G). If so, it will suggest that the formation of heterochromatin may feedback to promote the production of 22G-RNAs. Such a relationship between chromatin and small RNA production is reminiscent of RNA induced transcriptional silencing described in *S. pombe*, in which small RNA-guided heterochromatin recruits RdRPs to produce more small RNAs (39). In addition, as our data suggested that some endogenous genes, such as *nop-1*, *cdk-1* or *oma-1*, exhibit resistance to piRNA silencing despite low PATC density (Fig. 4B) (33), other mechanisms may exist to license self genes for expression. Taken together, our studies revealed a strategy by which *C. elegans* defends its genome against foreign nucleic acids, whereby diverse piRNAs silence foreign genes that are not licensed for expression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Craig Mello for his guidance of H.-C. L. during his postdoctoral training and for resources to initiate the project; Dr. Jonathan Staley, Dr. Edwin Ferguson, Dr. Alex Ruthenburg, and Jordan Brown for critical comments on the manuscripts; Elaine Xiao for assistance on designing and cloning transgenes; Dr. Michael Glotzer and Katrina Longhini for sharing reagents and unpublished results; members of the Lee lab and Staley lab for helpful discussions. The deep sequencing data described in this manuscript are available at Sequence Read Archive (SRA) of NCBI (accession number SRP108932). This work is supported by Ministry of Science of Technology of Taiwan grants (MOST-105-2918-I-006-002, MOST-105-2221-E-006-203-MY2, MOST-106-2628-E-006—006-MY2) to W.-S. W., a NIH P01 grant (HD078253) to Z.W., and a NIH R00 grant (GM108866) to H.-C. L.

REFERENCES AND NOTES

1. Brennecke J, et al. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell*. 2007; 128:1089–1103. [PubMed: 17346786]
2. Saito K, et al. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. *Genes & development*. 2006; 20:2214–22. [PubMed: 16882972]
3. Batista PJ, et al. PRG-1 and 21U-RNAs Interact to Form the piRNA Complex Required for Fertility in *C. elegans*. *Molecular Cell*. 2008; 31:67–78. [PubMed: 18571452]
4. Das PP, et al. Piwi and piRNAs Act Upstream of an Endogenous siRNA Pathway to Suppress Tc3 Transposon Mobility in the *Caenorhabditis elegans* Germline. *Molecular Cell*. 2008; 31:79–90. [PubMed: 18571451]
5. Carmell MA, et al. MIWI2 Is Essential for Spermatogenesis and Repression of Transposons in the Mouse Male Germline. *Developmental Cell*. 2007; 12:503–514. [PubMed: 17395546]
6. Ruby JG, et al. Large-Scale Sequencing Reveals 21U-RNAs and Additional MicroRNAs and Endogenous siRNAs in *C. elegans*. *Cell*. 2006; 127:1193–1207. [PubMed: 17174894]
7. Aravin AA, Hannon GJ, Brennecke J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science (New York, NY)*. 2007; 318:761–4.
8. Vourekas A, et al. Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. *Nature Structural & Molecular Biology*. 2012; 19:773–781.

9. Gou LT, et al. Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Research*. 2014; 24:680–700. [PubMed: 24787618]
10. Toombs JA, et al. *Xenopus* Piwi proteins interact with a broad proportion of the oocyte transcriptome. *RNA (New York, NY)*. 2017; 23:504–520.
11. Reuter M, et al. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature*. 2011; 480:264–267. [PubMed: 22121019]
12. Goh WSS, et al. piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. *Genes & Development*. 2015; 29:1032–1044. [PubMed: 25995188]
13. Bagijn MP, et al. Function, targets, and evolution of *Caenorhabditis elegans* piRNAs. *Science (New York, NY)*. 2012; 337:574–8.
14. Darricarrère N, Liu N, Watanabe T, Lin H. Function of Piwi, a nuclear Piwi/Argonaute protein, is independent of its slicer activity. *Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110:1297–302. [PubMed: 23297219]
15. Lee HC, et al. *C. elegans* piRNAs mediate the genome-wide surveillance of germline transcripts. *Cell*. 2012; 150:78–87. [PubMed: 22738724]
16. Saito K, et al. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes & development*. 2010; 24:2493–8. [PubMed: 20966047]
17. Rajasethupathy P, et al. A Role for Neuronal piRNAs in the Epigenetic Control of Memory-Related Synaptic Plasticity. *Cell*. 2012; 149:693–707. [PubMed: 22541438]
18. Vourekas A, Alexiou P, Vrettos N, Maragkakis M, Mourelatos Z. Sequence-dependent but not sequence-specific piRNA adhesion traps mRNAs to the germ plasm. *Nature*. 2016; 531:390–394. [PubMed: 26950602]
19. Shirayama M, et al. PiRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. *Cell*. 2012; 150:65–77. [PubMed: 22738726]
20. Buckley BA, et al. A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. *Nature*. 2012; 489:447–51. [PubMed: 22810588]
21. Ashe A, et al. PiRNAs can trigger a multigenerational epigenetic memory in the germline of *C. elegans*. *Cell*. 2012; 150:88–99. [PubMed: 22738725]
22. Bartel DP. MicroRNAs: Target Recognition and Regulatory Functions. *Cell*. 2009; 136:215–233. [PubMed: 19167326]
23. Kelly WG, Fire A. Chromatin silencing and the maintenance of a functional germline in *Caenorhabditis elegans*. *Development*. 1998; 125:2451–2456. [PubMed: 9609828]
24. Merritt C, Gallo CM, Rasoloson D, Seydoux G. Transgenic solutions for the germline. *WormBook*. 2010:1–21.
25. Tse YC, Piekny A, Glotzer M. Anillin promotes astral microtubule-directed cortical myosin polarization. *Molecular biology of the cell*. 2011; 22:3165–75. [PubMed: 21737681]
26. Waaijers S, et al. CRISPR/Cas9-Targeted Mutagenesis in *Caenorhabditis elegans*. *Genetics*. 2013; 195
27. Kim H, et al. A co-CRISPR strategy for efficient genome editing in *Caenorhabditis elegans*. *Genetics*. 2014; 197:1069–80. [PubMed: 24879462]
28. Gu W, et al. Distinct Argonaute-Mediated 22G-RNA Pathways Direct Genome Surveillance in the *C. elegans* Germline. *Molecular Cell*. 2009; 36:231–244. [PubMed: 19800275]
29. Claycomb JM, et al. The Argonaute CSR-1 and Its 22G-RNA Cofactors Are Required for Holocentric Chromosome Segregation. *Cell*. 2009; 139:123–134. [PubMed: 19804758]
30. Seth M, et al. The *C. elegans* CSR-1 argonaute pathway counteracts epigenetic silencing to promote germline gene expression. *Developmental cell*. 2013; 27:656–63. [PubMed: 24360782]
31. Wedeles C, Wu M, Claycomb J. Protection of germline gene expression by the *C. elegans* argonaute CSR-1. *Developmental Cell*. 2013; 27:664–671. [PubMed: 24360783]
32. Fire A, Alcazar R, Tan F. Unusual DNA Structures Associated With Germline Genetic Activity in *Caenorhabditis elegans*. *Genetics*. 2006; 173
33. Frøkjær-Jensen C, et al. An Abundant Class of Non-coding DNA Can Prevent Stochastic Gene Silencing in the *C. elegans* Germline. *Cell*. 2016; 166:343–57. [PubMed: 27374334]

34. Shin C, et al. Expanding the microRNA targeting code: functional sites with centered pairing. *Molecular cell*. 2010; 38:789–802. [PubMed: 20620952]
35. Ha I, Wightman B, Ruvkun G. A bulged lin-4/lin-14 RNA duplex is sufficient for *Caenorhabditis elegans* lin-14 temporal gradient formation. *Genes & development*. 1996; 10:3041–50. [PubMed: 8957004]
36. Lal A, et al. miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3' UTR microRNA recognition elements. *Molecular cell*. 2009; 35:610–25. [PubMed: 19748357]
37. Allen E, Xie Z, Gustafson AM, Carrington JC. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell*. 2005; 121:207–21. [PubMed: 15851028]
38. Jones-Rhoades MW, Bartel DP. Computational Identification of Plant MicroRNAs and Their Targets, Including a Stress-Induced miRNA. *Molecular Cell*. 2004; 14:787–799. [PubMed: 15200956]
39. Noma K, et al. RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nature Genetics*. 2004; 36:1174–1180. [PubMed: 15475954]
40. Dickinson DJ, Ward JD, Reiner DJ, Goldstein B. Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. *Nature Methods*. 2013; 10:1028–1034. [PubMed: 23995389]
41. Zhang Y, Werling U, Edelmann W. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Research*. 2012; 40:e55–e55. [PubMed: 22241772]
42. Zhang D, Glotzer M. The RhoGAP activity of CYK-4/MgcRacGAP functions non-canonically by promoting RhoA activation during cytokinesis. *eLife*. 2015; 4doi: 10.7554/eLife.08898
43. Frøkjær-Jensen C, Davis MW, Ailion M, Jorgensen EM. Improved Mos1-mediated transgenesis in *C. elegans*. *Nature Methods*. 2012; 9:117–118. [PubMed: 22290181]
44. Ortiz MA, Noble D, Sorokin EP, Kimble J. A New Dataset of Spermatogenic vs. Oogenic Transcriptomes in the Nematode *Caenorhabditis elegans*. *G3: Genes, Genomes, Genetics*. 2014; 4:1765–1772. [PubMed: 25060624]
45. Chen C, et al. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Research*. 2005; 33:e179–e179. [PubMed: 16314309]

One Sentence Summary

We revealed how piRNAs recognize their RNA targets and showed that piRNAs defend the *C. elegans* genome by silencing sequences that are not marked as self.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

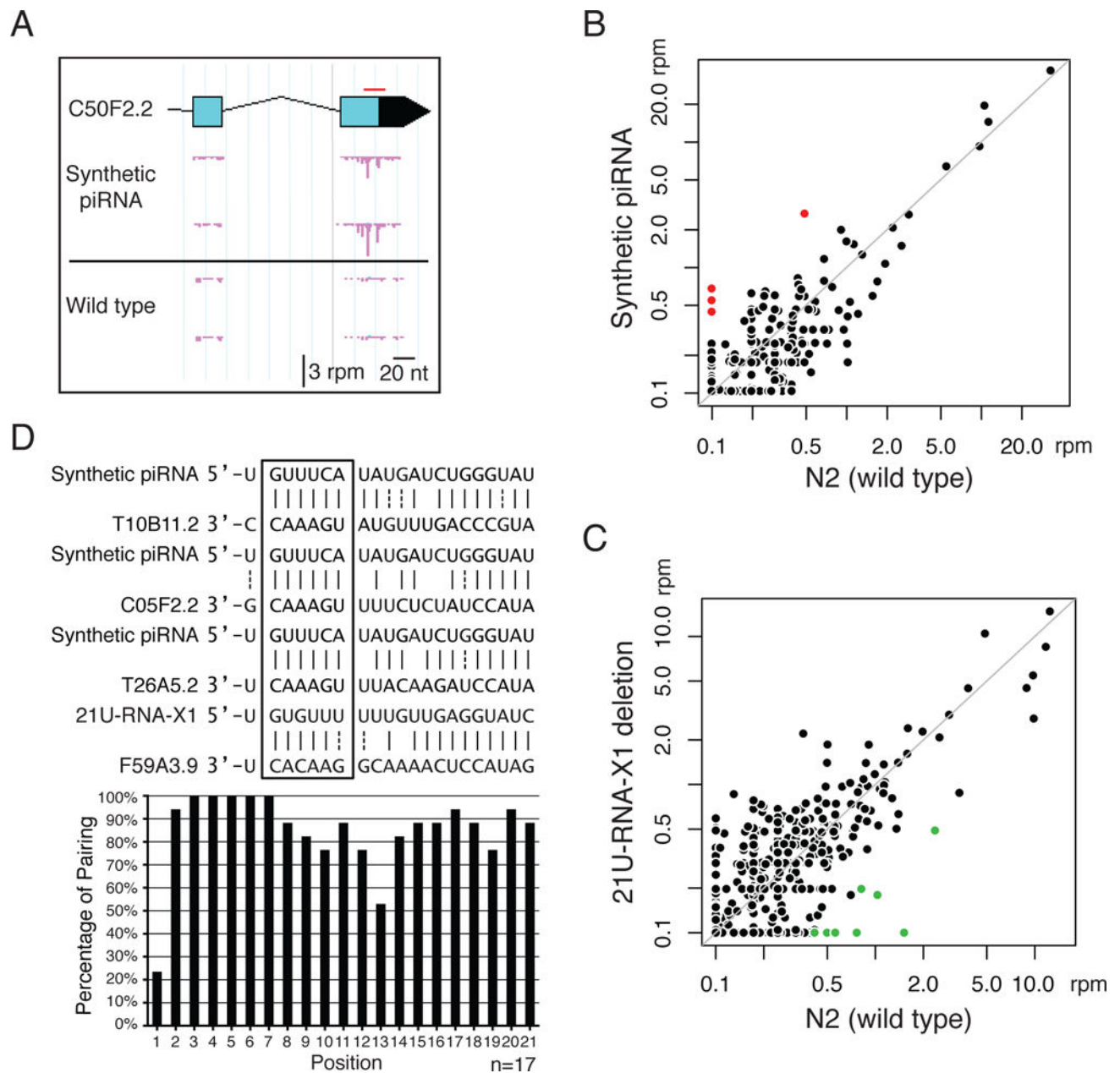


Fig. 1. 22G-RNA loci as a proxy to identify the targets of specific piRNAs

(A) An example of 22G-RNA distributions at one of the RNA targets of the synthetic piRNA (GFP-targeting piRNA#1) in the indicated strains with biological replicates. Each pink bar indicates the 1st nucleotide position and abundance of 22G-RNAs. The red bar marks the position targeted by the synthetic piRNA. rpm: reads per million.

(B) A scatter plot showing the abundance of 22G-RNAs around each potential targeting site (100nt window centered with each target site) of the synthetic piRNA (GFP-targeting #1) in the control strain and in the strain expressing the synthetic piRNA. Note that the potential targeting sites are sites of RNA transcripts that pair to the specific piRNA with six or fewer mismatches. Marked in red are sites at which 22G-RNA levels increased over 4 fold in the strain expressing the synthetic piRNA relative to the control strain.

(C) A scatter plot showing the abundance of 22G-RNAs at each potential targeting site of 21U-RNA-X1 in the N2 (wild-type) strain and in the strain containing a deletion of the 21U-RNA-X1 coding loci. Marked in green are sites at which 22G-RNA levels decreased over 4 fold in the strain loss of 21U-RNA-X1 relative to the N2 wild type.

(D) The pairing between piRNAs and identified targets. Examples of pairings between the piRNAs and their targets (Top). A bar graph showing the percentage of base pairing at each position within the piRNAs with all 17 identified targets (Bottom). GU wobble pairing is considered as paired here to highlight the near-perfect pairing at the seed region when GU pair is allowed.

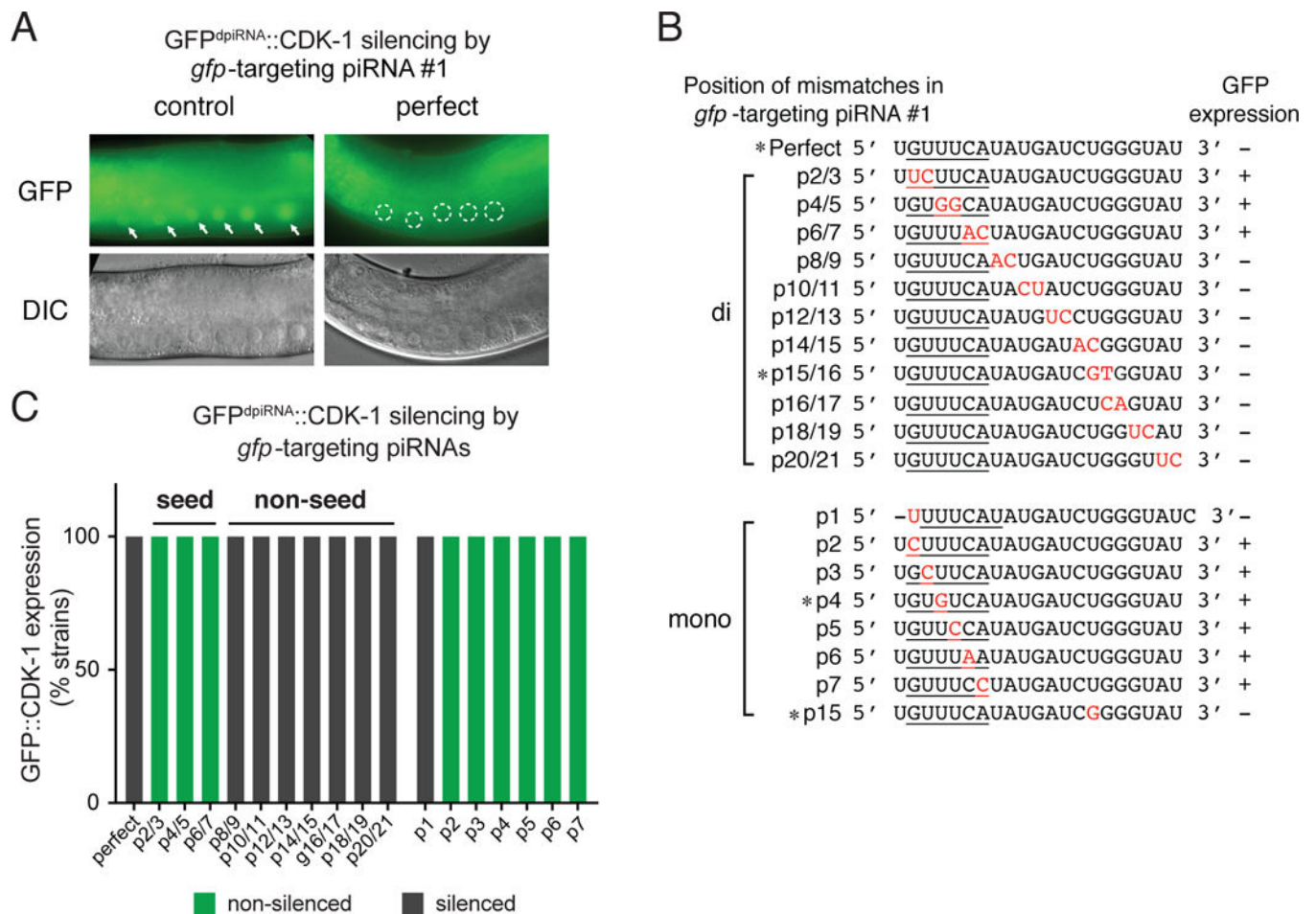


Fig. 2. A piRNA reporter assay to investigate the piRNA targeting rules

(A) Fluorescence micrographs showing the expression of transgene GFP^{dpiRNA::}CDK-1 in worms carrying an extrachromosomal array to express the *gfp*-targeting piRNA with perfect pairing or in the control strain that does not express the synthetic piRNA. Arrows mark the germline nuclei with expressed transgene. Circles mark the germline nuclei with silenced transgene. Note that the unmarked green fluorescent signals are auto-fluorescent signals generated from worm intestinal granules.

(B) The sequences of the *gfp*-targeting piRNAs, the positions of the mismatches (red), and their effects on the expression of GFP^{dpiRNA::}CDK-1. *: *gfp*-targeting piRNAs produced by gene-edited animals modified at an endogenous piRNA locus (21U-5499).

(C) Percentage of transgenic animals that exhibit the silencing of GFP^{dpiRNA::}CDK-1 in animals expressing specific *gfp*-targeting piRNAs. At least 8 independent strains carrying extrachromosomal arrays (roller) are examined for each piRNA.

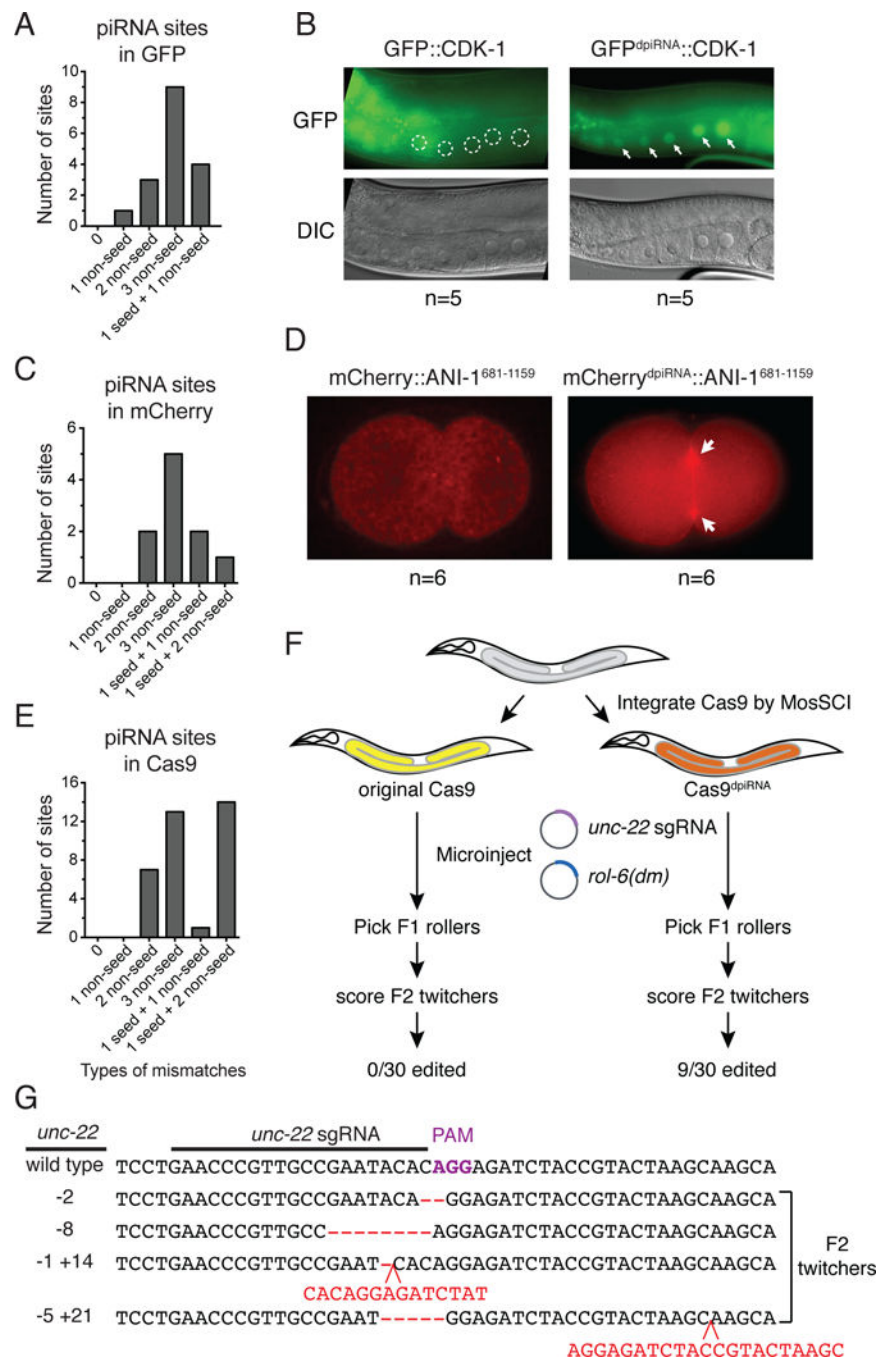


Fig. 3. Silencing-prone transgenes can be expressed in the germline by avoiding piRNA targeting

(A) Predicted piRNA sites in GFP mRNA sequence. The numbers of piRNA sites that contain different types of mismatches are shown. The relaxed criteria are used to predict piRNA sites on transgenes: all GU wobble pairing is allowed (considered as paired), and up to 3 non-GU mismatches are allowed when sites have perfect seed pairing, or up to 1 non-GU mismatches are allowed when sites have 1 non-GU mismatches in the seed region. The mismatch at the first nucleotide of a piRNA is not counted/considered.

- (B) The expression of original GFP::CDK-1 that contains the predicted piRNA targeting sites, or the modified GFP^{dpiRNA}::CDK-1 where all predicted piRNA sites have been removed by introducing silent mutations (right). Arrows mark the germline nuclei with expressed GFP^{dpiRNA}::CDK-1. Circles mark the germline nuclei with silenced GFP::CDK-1.
- (C) Predicted piRNA sites in mCherry mRNA sequence (left).
- (D) The expression of original mCherry::ANI-1⁶⁸¹⁻¹¹⁵⁹ that contains the predicted piRNA targeting sites, or the modified mCherry^{dpiRNA}::ANI-1⁶⁸¹⁻¹¹⁵⁹ where the predicted piRNA sites have been removed by introducing silent mutations (right). Arrows mark the expression of mCherry^{dpiRNA}::ANI-1⁶⁸¹⁻¹¹⁵⁹ at cleavage furrows of the one-cell embryo.
- (E) Predicted piRNA sites in Cas9 mRNA sequence.
- (F) A schematic showing the procedure followed to examine if genome editing occurs in transgenic animals that carry the original or modified Cas9 transgenes. Plasmids containing *unc-22* sgRNA and *rol-6(su1006)* dominant transformation marker plasmid are co-injected into transgenic animals that have been carrying the Cas9 transgene for over 5 generations. F1 transformed roller animals are picked and their F2 progeny are scored for *unc-22* gene editing through twitcher phenotype.
- (G) Sequences of various *unc-22* edited alleles obtained in the animals carrying the modified Cas9 transgene injected with plasmid encoding *unc-22* sgRNA. Indels are highlighted in red.

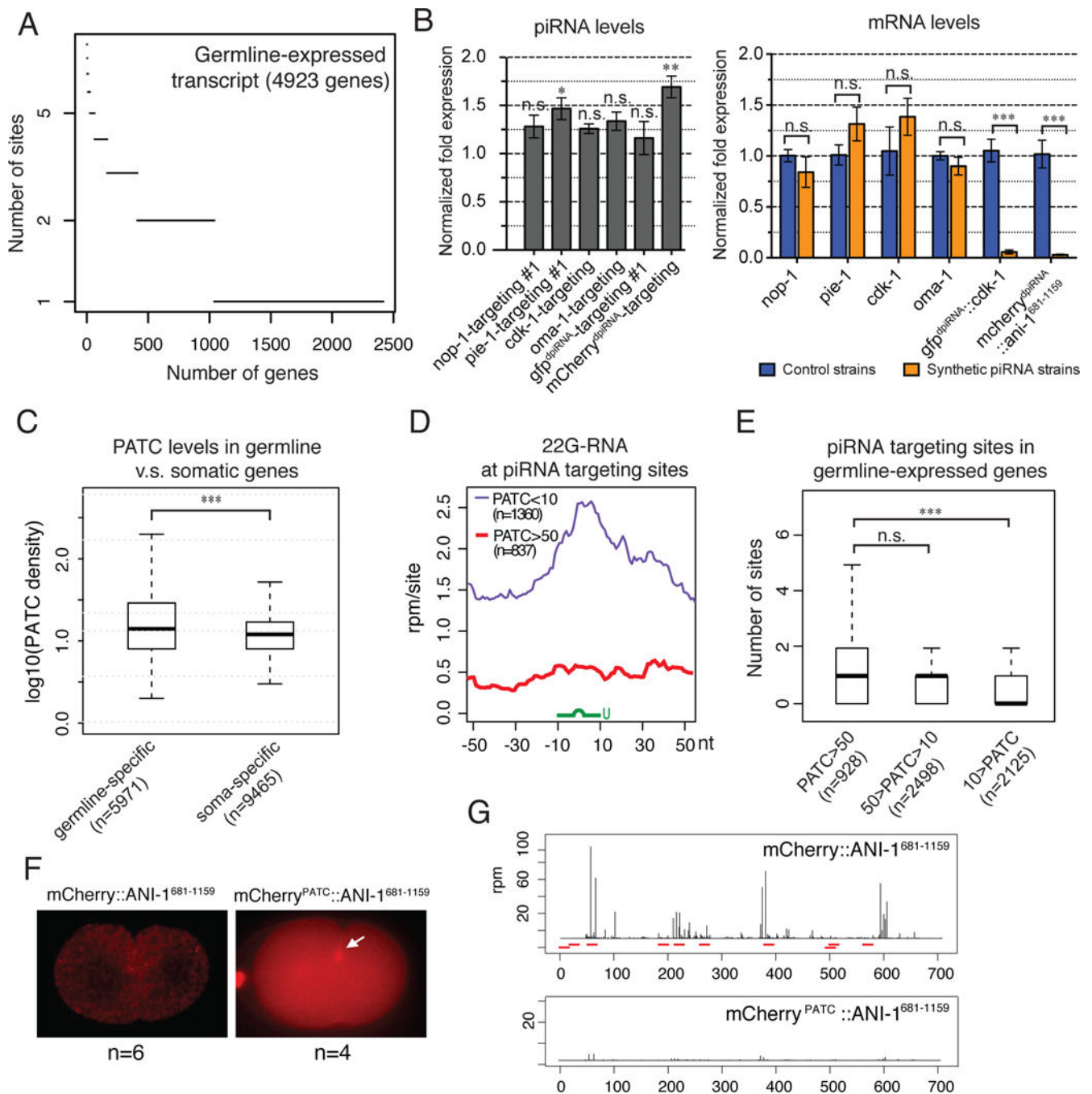


Fig. 4. Germline-expressed genes exhibit resistance to piRNA silencing through their intrinsic signals, such as PATCs

(A) Numbers of predicted piRNA sites on germline-expressed RNA transcripts. To predict more confident targeting sites, the stringent piRNA targeting criteria are used here, where up to one GU wobble pair was allowed in the seed region, and overall only up to two mismatches plus an additional GU mismatch were allowed. In addition, the mismatch at the first nucleotide of the piRNA is not counted/considered. The RNA targets of CSR-1 Argonaute (CSR-1 targets) are used to define the germline-expressed genes.

(B) qRT-PCR measurements of the abundance of the synthetic piRNAs in comparison to the level of endogenous 21U-5499 (value=1) in the control strain (left) and the expression levels of corresponding mRNA targets in the indicated strains (right). Note that *nop-1*, *cdk-1*, and *oma-1* targeting piRNAs were produced by gene-edited animals, whereas *pie-1*-targeting piRNAs were produced by animals carrying extrachromosomal arrays. Error bars represent standard error of the mean from biological duplicated samples. The statistics for synthetic piRNA expression were calculated by comparing the levels of specific piRNAs and 21U-5499. n.s.: not significant, *: p -value<0.05, **: p -value <0.01, ***: p -value <0.001, t -test.

(C) A box and whisker blot showing the density of PATC in the germline-specific and somatic specific genes. ***: p -value <0.001, t -test.

(D) The density of 22G-RNAs within a 100 nt window around predicted piRNA target sites of germline-specific transcripts with high PATC density (PATC>50) or low PATC density (PATC<10). The plots are centered at sequence targeted by piRNAs (green). The stringent piRNA targeting criteria were used here to predict piRNA target sites. n = number of predicted piRNA sites.

(E) The box-and-whisker plots showing the number of predicted piRNA targeted sites on germline-expressed genes that contain the indicated range of PATC density. The stringent piRNA targeting criteria were used here to predict piRNA target sites. n.s.: not significant, ***: p -value <0.001, t -test.

(F) Fluorescence micrographs showing the expression of the original mCherry::ANI-1⁶⁸¹⁻¹¹⁵⁹ harboring synthetic introns (no PATC) and mCherry^{PATC}::ANI-1⁶⁸¹⁻¹¹⁵⁹ harboring PATC-containing introns.

(G) 22G-RNA distribution at mCherry coding sequence of the indicated transgenes. Each bar indicates the 1st nucleotide position and abundance of 22G-RNAs. The red bars mark the location of piRNA targeting sites predicted by using the relaxed piRNA targeting criteria.