Title

Maternally inherited siRNAs initiate piRNA cluster formation.

Journal

Molecular Cell, 83(21)

Authors

Luo, Yicheng
He, Peng
Kanrar, Nivedita
et al.

Publication Date

2023-11-02

DOI

10.1016/j.molcel.2023.09.033

Peer reviewed

HHS Public Access

Author manuscript

Mol Cell. Author manuscript; available in PMC 2024 November 02.

Published in final edited form as:

Mol Cell. 2023 November 02; 83(21): 3835–3851.e7. doi:10.1016/j.molcel.2023.09.033.

Maternally inherited siRNAs initiate piRNA cluster formation

Yicheng Luo[1], **Peng He**[1,2], **Nivedita Kanrar**[1], **Katalin Fejes Toth**[1], **Alexei A. Aravin**[1,&]

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA

[2]present address: European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Cambridge, UK

Abstract

PIWI-interacting RNAs (piRNAs) guide transposable element repression in animal germlines. In *Drosophila*, piRNAs are produced from heterochromatic loci, called piRNA clusters, which act as information-repositories about genome invaders. piRNA generation by dual-strand clusters depend on the chromatin-bound Rhino-Deadlock-Cutoff (RDC) complex, which is deposited on clusters guided by piRNAs, forming a positive feedback loop in which piRNAs promote their own biogenesis. However, how piRNA clusters are formed before cognate piRNAs are present, remains unknown. Here we report spontaneous *de novo* piRNA cluster formation from repetitive transgenic sequences. Cluster formation occurs over several generations and requires continuous trans-generational maternal transmission of small RNAs. We discovered that maternally supplied siRNAs trigger *de novo* cluster activation in progeny. In contrast, siRNAs are dispensable for cluster function after its establishment. These results revealed an unexpected interplay between the siRNA and piRNA pathways and suggest a mechanism for *de novo* piRNA cluster formation triggered by siRNAs.

eTOC Blurb

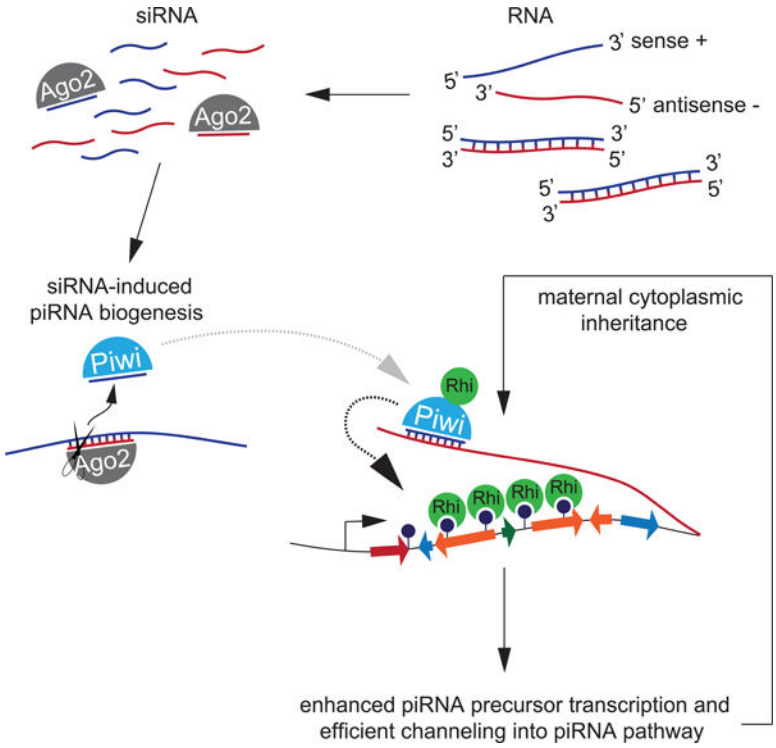
[&]To whom correspondence should be addressed: Alexei A. Aravin, aaa@caltech.edu.
The Lead Contact: Alexei A. Aravin, aaa@caltech.edu

AUTHOR CONTRIBUTIONS

Y.L., K.F.T. and A.A.A. designed experiments. Y.L. performed all experiments; N.K. helped to perform imaging and ChIP experiments; P.H. performed the computational analysis; Y.L. and P.H. analyzed the data; Y.L., K.F.T. and A.A.A. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.



Luo et al. elucidate the mechanism behind the genesis of piRNA clusters, highlighting the role of siRNAs originating from repetitive sequences as initiators in this process. This discovery reveals the unanticipated involvement of siRNAs and offers insights into the origin of piRNA clusters.

Introduction

Binary complexes of small non-coding RNAs and Argonaute (Ago) proteins play essential roles in regulating gene expression and suppressing foreign and selfish nucleic acids. Despite the common architecture of Ago-small RNA complexes, there are three distinct classes of small RNAs - siRNA, miRNA and piRNA - that differ in their biogenesis, functions. Both siRNA and piRNA were reported to suppress activity of endogenous and exogenous genetic elements in various animal species[1–7], however, the two pathways are believed to work independently of each other. siRNAs are processed from double-stranded or hairpin precursors by the Dicer nuclease and then loaded into their Ago protein partner[8–10]. siRNA/Ago complexes cleave complementary RNA targets in the cytoplasm[9,10]. In *Drosophila*, siRNAs associate exclusively with Ago2 and mutation of Ago2 abrogates siRNA-guided repression[11–15]. piRNA processing is independent of Dicer but involves multiple other proteins[13,16–20]. piRNAs are loaded into a distinct clade of Argonautes called Piwi proteins, which in flies consist of Piwi, Aub and Ago3[2,13,21].

siRNA precursors are recognized by their double-stranded nature. Similar to siRNAs, piRNAs are also processed from longer RNA precursors, however these transcripts lack distinct sequence and structure motifs. In *Drosophila*, the chromatin-bound Rhino-Deadlock-Cutoff (RDC) protein complex marks dual-strand piRNA clusters, genomic regions

that generate the majority of piRNAs in the germline. RDC complex is required for transcription of piRNA precursors by promoting initiation[16,18,22] and suppressing premature termination[17,19]. The process of RDC complex deposition on chromatin seems to be guided by piRNAs[18,23–25]. Several studies demonstrate the critical role of piRNA inheritance from the mother to the progeny in initiating piRNA production[26–28]. These findings suggest that piRNA biogenesis is governed by a trans-generational positive feedback loop in which piRNA biogenesis is promoted by RDC complex, which in turn is deposited on chromatin guided by piRNAs. This positive feedback loop explains how piRNA profiles are maintained through generations. However, in order to adapt to new transposon invasions, the pathway must be able to generate novel piRNAs.

Here we describe the *de novo* formation of a piRNA cluster over several generations. This process is accompanied by increasing piRNA levels and accumulation of the H3K9me3 mark and Rhi on cluster chromatin and requires continuous, maternal trans-generational cytoplasmic transmission of small RNAs. We found that cognate siRNAs trigger initial cluster activation, however, siRNA are dispensable after the cluster is established. Our results point to a tight cooperation between the siRNA and piRNA pathways in the fight against genome invaders and suggest that transposons are first detected by the siRNA pathway, which activates a robust piRNA response.

Results

Reporters inserted in the dual-strand 42AB cluster are repressed by piRNA, while reporters in the uni-strand 20A cluster disrupt cluster expression

To understand how new insertions into piRNA clusters are regulated, we integrated reporters into the two types of piRNA clusters – uni-strand and dual-strand clusters. We employed recombinase-mediated cassette exchange (RMCE) to integrate a reporter into specific genomic sites using a collection of *Minos*-mediated integration cassette (MiMIC) containing *D. melanogaster* stocks. The reporter encodes nuclear EGFP expressed under control of the ubiquitin (*ubi-p63E*) gene promoter, which drives expression in both somatic and germ cells (Fig. 1A). Using RMCE we inserted reporters into the major dual-strand cluster 42AB and the uni-strand cluster 20A (Fig. 1B). In the 20A cluster, the reporter was integrated 2.5 kb downstream of the cluster promoter and two reporter orientations were obtained. Reporters inserted in both clusters were expressed in somatic follicular cells of the fly ovary (Fig.1C) indicating that the transgenes are functional. Flies with reporters in the uni-strand 20A cluster also expressed GFP in the germline. Unlike 20A cluster transcripts, which localized to the nucleus, GFP mRNA was predominantly in the cytoplasm. In contrast, although both strands of the reporter sequence were transcribed, GFP protein was not expressed from the 42AB cluster insertion in the germline.

The exclusive repression of the 42AB reporter in the germline suggests that repression might occur in a piRNA-dependent fashion. To explore if reporter sequences generate piRNAs, we cloned small RNA libraries from ovaries of transgenic animals. Analysis of the libraries revealed that reporter-derived piRNAs were abundant in flies with 42AB cluster insertions but not in flies with reporters in cluster 20A and in a control non-cluster 66A6 region (Fig. 1D). The 42AB reporter-derived piRNAs have the expected bias (69.95%) for U (Uridine)

in the first position. Though piRNA were derived from both strands of the 42AB reporter sequence, 2.45-fold more piRNA are in antisense orientation relative to the GFP mRNA (Fig. 1E), indicating that they are not processed from reporter mRNA. In contrast, the few RNA reads derived from reporters inserted in 20A and the non-cluster region were predominantly in sense orientation and did not have a U bias, indicating that they likely represent mRNA degradation products.

To explore whether the repression of GFP reporter inserted into the 42AB cluster depends on piRNAs, we knock-down Zucchini (Zuc), a critical piRNA biogenesis factor [29,30], in the germline. Depletion of Zuc led to strong (26.4-fold) reduction in the level of piRNAs mapping to the 42AB reporter (Fig. 1E) and to its derepression (Fig.1F). GFP protein expression upon *Zuc* germline knockdown (GLKD) correlated with the appearance of sense reporter transcripts in the cytoplasm of nurse cells, while antisense RNA remained in the nucleus. Thus, insertion of a gene into the 42AB cluster leads to generation of abundant piRNAs that trigger its repression.

To understand why 20A reporters do not produce piRNAs, we analyzed expression of 20A cluster transcripts by RT-qPCR using primer sets designed to detect cluster transcripts upstream and downstream of the reporter insertion site. Surprisingly, we found that the abundance of 20A cluster transcripts was strongly (> 20-fold) reduced in flies with reporter insertions compared to both wild-type flies and the original MiMIC flies used for RMCE (Fig. 1G). In agreement with the decreased cluster transcript level, 20A piRNA levels dropped 84- and 321-fold in 20A[−] and 20A[+] flies, respectively, throughout the whole cluster as far as 38 kb from the insertion site (Fig. 1B, G). In contrast, the insertion in the 42AB cluster did not affect piRNA level from this cluster (data not shown). As the original MiMIC line contains a promoterless insertion in the same site as the reporter lines, this result suggests that cluster expression is disrupted by reporter transcription rather than insertion of a heterologous sequence *per se*. Thus, insertion of an actively transcribed gene in the 20A cluster close to its promoter disrupts cluster expression.

An unusual reporter behaves like a *bona fide* dual-strand piRNA cluster that is enriched in Rhi and depends on it to generate piRNA

Replacement of the MiMIC cassette with the reporter through recombinase-mediated cassette exchange (RMCE) leads to random orientation of the inserted sequence. Each replacement experiment generates several independent *Drosophila* lines. Unexpectedly, we found that one of the lines with insertion into the 20A cluster has distinct properties. This line, which we dubbed 20A-X, lost germline expression after several months of propagation (Fig. 2A). The somatic GFP expression, which we also confirmed by detecting abundant cytoplasmic GFP mRNA in follicular cells by in situ hybridization, argues against genetic damage of the reporter cassette. GFP RNA was also detected in germline nurse cells, however, transcripts from 20A-X localized exclusively to the nuclei (Fig. 2A). Thus, unlike other 20A insertion lines, 20A-X shows normal GFP expression in follicular cells and strong GFP repression in the germline.

Analysis of small RNA libraries cloned from ovaries of 20A-X flies revealed abundant piRNAs and siRNAs corresponding to the reporter sequence indicating that 20A-X is active

as a piRNA producing locus. In fact, 20A-X generates 20.3-fold more piRNAs than the 42AB reporter. Germline knockdown of the piRNA biogenesis factor *Zuc* led to loss of 20A-X piRNAs (Fig. 2B). *Zuc* GLKD also led to release of the germline reporter repression as demonstrated by GFP protein expression and detection of sense RNA in the cytoplasm (Fig. 2C) indicating that, similar to 42AB insertions, repression of 20A-X is piRNA-dependent.

Several studies revealed essential differences between uni-strand and dual-strand piRNA clusters[17,18,22,31]. Dual-strand clusters, such as 42AB, are active exclusively in the germline and their transcription, nuclear processing and export require the Rhino-Deadlock-Cutoff (RDC) complex, which is anchored to these regions by the H3K9me3 histone mark[18,19,26]. In contrast, uni-strand clusters, such as flamenco and 20A, do not depend on RDC complex and the H3K9me3 mark and can be active in the soma. To explore if 20A-X functions as a uni- or dual-strand piRNA cluster, we analyzed Rhino binding and H3K9me3 enrichment. ChIP-qPCR and ChIP-seq analyses revealed that, in contrast to the native 20A cluster and 20A[+] and 20A[−] reporters, 20A-X, as well as its flanking regions, are strongly enriched in both Rhi and the H3K9me3 mark (Fig. 2D, S1A), suggesting that it acts as a dual-strand piRNA cluster.

To test whether 20A-X activity depends on the RDC complex, we analyzed GFP repression and small RNA profile upon germline knockdown of *Rhi*. As expected, *Rhi* GLKD reduces the level of piRNAs generated from the 42AB reporter (Fig. 2E). *Rhi* GLKD also caused 4.2-fold reduction in 20A-X piRNA levels and released repression of GFP in the germline of 20A-X flies (Fig. 2E). In agreement with the piRNA cloning and GFP imaging results, FISH experiments detected a decrease in antisense GFP transcripts and an increase in sense GFP transcripts upon *Rhi* GLKD (Fig. S1B), suggesting that 20A-X transcription is Rhi-dependent.

At dual-strand piRNA clusters Maelstrom (Mael) was proposed to repress canonical, promoter-dependent transcription, allowing the RDC complex to initiate non-canonical transcription[32]. 20A-X showed GFP expression upon *Mael* GLKD, which suggests that Mael represses the ubiquitin promoter-driven canonical transcription of 20A-X (Fig. S1C). In contrast, GFP expression from the 42AB[−] reporter remained silenced in the *Mael* mutant (or GLKD) or the *Mael*/*Rhi* double mutant (double GLKD shows expression as a result of *Rhi* GLKD) (Fig. S1C), which implies that this reporter can be silenced independently of Mael[32]. Taken together, our results indicate that 20A-X acts as a genuine dual-strand piRNA cluster that generates piRNAs in Rhi- and Mael-dependent manner.

The repetitive organization of the 20A-X locus correlates with its function as dual-strand piRNA cluster

We employed several approaches to understand how 20A-X differs from 20A[+] and 20A[−] reporters. Genomic PCR of flanking regions suggested that 20A-X harbors the reporter cassette in the correct site in the 20A cluster (Fig. 3A). In fact, similar to other 20A reporters, expression of 20A cluster transcripts is decreased in 20A-X flies (Fig. S2A). To confirm the insertion site, we employed *in situ* hybridization on salivary gland polytene chromosomes using the Ubi-GFP reporter sequence as a probe. *In situ* hybridization revealed two signals: the expected signal in the 20A region of the X chromosome and additional

signal on chromosome 3L, which harbors the endogenous *ubi-p63E* gene (Fig. 3B). To further validate these findings, we performed whole-genome sequencing and searched for reads corresponding to junctions between the reporter and genomic sequences (Fig. S2B). We identified multiple reads corresponding to the two expected flanking regions in the 20A cluster, while no additional insertions were identified, corroborating results of the chromosome hybridization (Fig. S2C). Finally, we employed CRISPR/Cas9 to generate a deletion that removes sequences flanking the insertion site in the 20A-X line. We verified the deletion and concomitant loss of the reporter sequence from the genome by genomic qPCR and loss of GFP expression (Fig. 3C). Thus, the 20A-X line contains a reporter insertion in the single genomic site in the same position as other 20A lines.

We employed whole-genome DNA-seq and qPCR to analyze reporter copy number in the genome, which showed that while other 20A reporters harbor a single copy of the reporter sequence, 20A-X contains approximately 10 copies (Fig. 3D, E, S2D). In addition, both approaches revealed that, unlike the 20A[+] and 20[−] lines, the 20A-X insertion contains plasmid backbone sequence that is used for recombinase-mediated cassette exchange and is normally removed during this process (Fig. 1A, 3D, S2D). Furthermore, we also found multiple DNA-seq reads indicating three unexpected junctions (Fig. 3F). We confirmed all junctions by genomic PCR and Sanger sequencing. As in situ hybridization on polytene chromosomes, DNA-seq and CRISPR/Cas9 deletion all indicate a single insertion site in the genome, all reporter copies must reside in a single genomic site. Taken together, our results indicate that the 20A-X line contain multiple, rearranged copies of the reporter sequence in a single site. Together, analysis of 20A-X indicates that, unlike single-copy reporters in 20A, repetitive sequences inserted in the same site generate piRNAs that induce repression in the germline.

To explore if the ‘host’ 20A cluster is required for 20A-X to function as dual-strand piRNA cluster we employed CRISPR/Cas9 to delete the promoter of the 20A cluster (Fig. 3G). In wild-type flies, deletion of the putative promoter eliminated expression of long RNA (piRNA precursors) from the 20A cluster, indicating that deletion disrupts its function. However, deletion of the 20A cluster promoter in 20A-X flies did not release reporter silencing (Fig. 3G), indicating that piRNA-dependent repression of the 20A-X locus does not require activity of the ‘host’ 20A cluster.

20A-X is inserted in the same site as the 20A[+] and 20[−] reporters but differs in its repetitive nature as well as its structural rearrangements. Particularly, 20A-X harbors inversions that are expected to generate dsRNA upon their transcription and thus can give rise to siRNAs. We tested for the presence of such siRNAs by cloning and sequencing small RNA libraries from ovaries of flies that inherited the 20A-X locus paternally or upon GLKD of *Rhi*, both of which would eliminate piRNAs (Fig. 2E, 4C). We found 21–22 nt small RNAs mapping to the entire length of the 20A-X reporter (Fig. S2E). We also observed a prominent peak in the SV40 region, which due to recombination site 1 (Fig. 3F) presumably produces dsRNA. Thus, the inverted repeats in the 20A-X cluster lead to production of dsRNA that can be processed into siRNA. To explore if the presence of transcribed inverted repeats is sufficient to create a functional dual-strand piRNA cluster, we generated a dsGFP reporter, which consists of the same sequence fragments as the original reporter, but harbors

inverted GFP sequences that would form dsRNA upon transcription (Fig. 3H). We obtained flies with insertion of the dsGFP reporter into the same site as other 20A reporters as well as in a control (non-piRNA cluster) region of the genome (chr 3L: 7,575,013, region 66A6). Analysis of small RNAs in ovaries of flies carrying dsGFP constructs revealed the presence of abundant 21 nt siRNAs generated from the inverted GFP repeat, but no other portion of the construct (Fig. 3H). These 21 nt siRNAs are resistant to oxidation treatment (Fig. 3H, S2F, S2G). In contrast to 20A-X, the dsGFP insertions produce only miniscule amount of longer (24–30nt) RNA species and these RNAs lack a U-bias. There were no significant differences between small RNAs generated from the dsGFP reporter inserted into the 20A cluster in either orientation and in the non-cluster control genomic region. To confirm that the dsGFP reporter only gives rise to siRNAs and not piRNAs, we immunoprecipitated Piwi and sequenced associated small RNAs. The number of small RNA reads mapping to the dsGFP sequence was negligible (Fig. S2H). Overall, these results shows that inverted repeats trigger generation of siRNAs, but not piRNAs, indicating that the presence of inverted repeats might be necessary, but is not sufficient to make 20A-X an active dual-strand piRNA cluster and that the multi-copy nature and/or the extended lengths of the locus might play a critical role.

piRNA-induced repression of the 20A-X reporter depends on maternal transmission of cognate piRNAs

Previously we and others have shown that the activity of artificial dual-strand piRNA clusters in the progeny requires cytoplasmic inheritance of piRNAs from the mother[26,27,33], and proposed that all dual-strand clusters might depend on trans-generationally inherited piRNAs to maintain their activity[26,27]. Therefore, we explored expression of the 20A-X reporter in the progeny after paternal and maternal inheritance. Flies that inherited the 20A-X insertion from their mothers (maternal transmission) showed – similar to their mothers – GFP expression in follicular cells, but strong GFP repression in the germline with sense reporter RNA restricted to the nucleus. In contrast, females that inherited the 20A-X reporter from their fathers (paternal transmission) had robust GFP protein expression and cytoplasmic localization of GFP mRNA in the germline (Fig. 4A). We also observed GFP repression in the germline of males that inherited the 20A-X locus maternally, but not when they inherited it paternally. In agreement with FISH and immunofluorescence (IF) results, RT-qPCR showed ~5-fold increase of GFP RNA after paternal transmission (Fig. 4B). Thus, repression of the 20A-X reporter in both the male and female germline requires maternal inheritance of the reporter.

To understand if derepression of GFP after paternal transmission is caused by changes in piRNA expression, we cloned small RNA libraries from ovaries of progeny that inherited the locus maternally and paternally. Upon paternal transmission, piRNA level was 10.1-fold reduced (Fig. 4C). Thus, piRNA generation from the 20A-X locus requires its maternal inheritance and derepression of GFP upon paternal reporter transmission is explained by the dramatic decrease in reporter-targeting piRNA. Next, we determined enrichment of the H3K9me3 mark and Rhino protein on chromatin of the 20A-X reporter in the progeny upon paternal or maternal inheritance of the locus. Both Rhi and H3K9me3 were reduced on 20A-X chromatin after paternal transmission to levels comparable to those detected at

the control euchromatic region (Fig. 4D). Thus, loss of piRNA upon paternal transmission correlates with loss of H3K9me3 and Rhino from the 20A-X locus.

Progeny that inherits the 20A-X locus from their mothers receive two distinct contributions. First, the locus itself might have different chromatin imprints when inherited maternally or paternally. Second, mothers deposit piRNAs into the oocyte, while paternal progeny do not inherit piRNA from their fathers[26,27,33]. Therefore, it is important to discriminate if the parent-of-origin effects of 20A-X depend on inheritance of the genomic locus or cytoplasmic transmission of piRNAs to the next generation. To discriminate between these possibilities, we designed two different crosses: in both crosses the progeny inherited 20A-X paternally, however, in one cross the mothers also carried a copy of the 20A-X locus which, however, was not transmitted to the progeny (Fig. 4E). The presence of the 20A-X locus in mothers caused GFP repression and piRNA generation in the progeny even though the locus itself was not transmitted to the offspring. Indeed, progeny that inherited the 20A-X locus from their fathers but received cognate piRNAs from their mothers had similar level of 20A-X piRNAs as progeny that simply inherited the 20A-X locus maternally. It is worth noting that maternally inherited cognate piRNAs (in crosses where the mothers carried the 20A-X locus but the locus itself was not transmitted to the progeny) were not able to convert the 20A[+] and 20A[−] loci to piRNA-producing clusters, nor did they change the expression of GFP from these reporters in the adult progeny (data not shown). These results collectively indicate that the activity of 20A-X as a piRNA-generating locus requires both the extended, multi-copy nature of the locus and cytoplasmic inheritance of cognate piRNAs through the maternal germline.

piRNA cluster is established over several generations

The finding that maternal inheritance of piRNAs is required for the activity of the 20A-X locus in the progeny prompted us to re-examine the observation that initially, upon establishment of the 20A-X transgenic flies by RMCE, GFP was expressed in the germline but got repressed in later generations (Fig. 5A). First, we established that the age of flies did not influence GFP silencing, as repression was similar in young (5-days) and old (30-days) females of the 14 months old stock (Fig. 5A). Next, we analyzed small RNA profiles in ovaries of 20A-X flies 1, 11 and 21 months after establishment of the stock. piRNAs and siRNAs derived from 20A-X were already present at 1 months, but their abundance increased 4-fold and 3.7-fold, respectively, by 11 months (Fig. 5B). At 11 months 20A-X piRNAs also showed stronger sign of ping-pong processing as measured by complementary piRNA pairs that overlap by 10nt (Z-scores at 1 months and 11 months were 1.0 and 4.2, respectively) (Fig. 5C, S3A). No further increase in abundance of 20A-X piRNAs and ping-pong processing was observed when comparing 11 and 21 months-old stocks. Thus, transgene-derived piRNA abundance is increasing over multiple generations after transgenesis and this increase correlates with repression of GFP in the germline. The findings that 20A-X requires maternally supplied piRNAs in order to generate piRNAs suggests that 20A-X was not active as a piRNA cluster in the first generation after establishment of this stock.

The loss of 20A-X's ability to generate piRNAs upon paternal inheritance provides a unique opportunity to explore *de novo* establishment of a piRNA cluster. After paternal transmission the progeny (G0) generated very few piRNAs and siRNAs with levels similar to those of 20A[+] and 20A[−] flies (after normalization to transgene copy numbers) (Fig. S3B). We monitored whether 20A-X can recover its ability to generate piRNAs in future generations upon continuous maternal transmission (Fig. 5D). While no GFP repression occurred in the germline of G0, each subsequent generation showed decreased GFP expression, until complete repression was observed in G8 (Fig. 5D, S3C). Establishment of GFP repression over multiple generations also occurred in the male germline. Interestingly, ovaries of flies of intermediate generations (G2-G5) showed large variation in the extent of repression between individual egg chambers (Fig. 5D, E, S3C). For example, in G2 almost equal fractions of flies showed normal expression, complete silencing and a mixed phenotype (Fig. 5E).

We profiled small RNAs (piRNAs and siRNAs) in ovaries of G3 flies after separating them into three groups (active, mixed and silenced) based on GFP expression as well as from ovaries of G0 and G8 flies. Reporter piRNA and siRNA levels increased over generations. Importantly, the three groups of G3 ovaries with different levels of GFP repression had proportionally different levels of reporter piRNAs and siRNAs, indicating that repression correlates with small RNA abundance (Fig. 5F). Intriguingly, by analyzing the ratio of siRNAs and piRNAs, we observed that piRNA abundance increased more significantly over generations (Fig. 5F). This finding has led us to hypothesize that siRNAs may be produced first and serve as stimuli to initiate piRNA generation. To validate this hypothesis, we performed Degradome-seq, cloning and sequencing of long 5' monophosphorylated RNAs that form as a result of siRNA and piRNA-guided cleavage, as well as degradation of mRNAs by other nucleases. Analysis of Degradome-seq and small RNA libraries from G0 flies revealed the presence of a 5' cleavage product of GFP mRNA with 10 nt overlap to siRNAs. This indicates that siRNAs can induce cleavage of the reporter within the GFP sequence, which can lead to initiation of piRNA production (Fig. 5G).

piRNAs are generated by two mechanisms, each with unique roles: phased cleavage of precursors by Zucchini increases the diversity of piRNAs and the ping-pong amplification loop boosts piRNA abundance, as documented in previous studies [2,34–36]. We analyzed signatures of phasing (the distance from the 3′ end of each piRNA to the 5′ end of the next downstream piRNA) and ping-pong processing in G0, G3 (active, mixed and silenced), and G8 flies. We observed decrease in phased processing over generations: Z0 scores changed from 2.5 in G0 to 1.0 in G8). In contrast, ping-pong was not detectable in G0 flies and became prominent at G8: Z10 scores changed from 0.8 in G0 to 2.7 in G8 (Fig. S3D). Furthermore, we observed higher phasing and low ping-pong in the 'active' group of G3 flies, while the 'silenced' G3 group had lower phasing and higher ping-pong (as expected, the intermediate group had intermediate scores for both). Additional phasing analysis by measuring the distance from the 5′ ends of upstream piRNAs to the 5′ ends of downstream piRNAs also detected slightly increased phasing signal in G0 compared to G8 (Fig. S3E). These results suggest that initial cleavage by siRNAs triggers phased processing of target transcripts, while ping-pong processing becomes efficient later when piRNA levels are sufficiently high.

Finally, we determined enrichment of the H3K9me3 mark and Rhino protein on chromatin of the 20A-X reporter in ovaries of G0 and G8 progeny. While Rhi and H3K9me3 were lost after paternal transmission in G0, both Rhi and H3K9me3 were enriched on 20A-X chromatin in G8 (Fig. 5H). Previous studies have shown that Piwi can genetically interact with Rhi and promote Rhi binding to piRNA clusters [23]. Therefore, we tested the role of Piwi during 20A-X piRNA cluster formation. We found that GLKD of *Piwi* leads to partial derepression of GFP expression from 20A-X (Fig. S3F), implying that silencing of the 20A-X GFP reporter is, at least in part, mediated by Piwi. Overall, our results indicate that 20A-X gradually establishes its ability to generate piRNAs over eight generations if maternal 20A-X piRNAs are transmitted to the progeny in each generation.

Maternal siRNA triggers activation of piRNA cluster in the progeny

To test the role of siRNAs in piRNA cluster activation, we crossed males carrying the 20A-X locus with heterozygous females carrying dsGFP constructs harboring simple inverted repeats that generate siRNAs (Fig. 6A). As seen before, paternal transmission of 20A-X led to release of GFP repression in the germline of the progeny (Fig. 6B). However, GFP remained repressed in progeny that carried maternally inherited dsGFP constructs. Remarkably, a similar level of GFP repression was also observed in sibling progeny that did not inherit the dsGFP construct from their mothers, indicating that the presence of cognate siRNAs in the mothers was sufficient to activate repression in progeny (Fig. 6B).

To further explore the role of siRNAs, we abrogated the siRNA pathway either in the mothers or in the progeny using Ago2 mutation (Fig. 6C). In flies, Ago2 is required for the stability and function of siRNAs and its mutation completely disrupts the siRNA pathway [15]. GFP repression was strongly disrupted and piRNAs mapping to the reporter decreased in the progeny of Ago2-deficient mothers, confirming that piRNA production and reporter silencing depends on trans-generational cytoplasmic transmission of siRNA (Fig. 6D). In contrast, Ago2-deficient progeny that inherited siRNA from their heterozygous mothers show strong GFP repression. We note that in the absence of maternal Ago2, the presence of zygotic dsGFP led to slightly reduced GFP expression (Fig 6D), implying that that zygotic siRNAs are capable of triggering piRNA biogenesis and repression albeit less efficiently than maternal siRNAs. Taken together, these results indicate that initiation of 20A-X repression in the progeny requires trans-generational inheritance of cytoplasmic siRNAs, while the siRNA pathway is dispensable for maintenance of the repression.

To further analyze the effect of siRNA on activation of 20A-X we cloned and sequenced small RNAs. As expected, progeny that inherited the dsGFP construct had high level of siRNAs targeting GFP (>100-fold increase compared to progeny with only paternal 20A-X) (Fig. 6E). These abundant siRNAs were restricted to the GFP sequence, which forms inverted repeats in the dsGFP construct. Interestingly, sibling progeny that inherited the balancer chromosome instead of dsGFP also showed moderate (3~7-fold) increase in GFP siRNA level compared to flies with paternal 20A-X only. Remarkably, both the progeny that inherited the dsGFP construct and their siblings that inherited the balancer chromosome had elevated levels of piRNA mapping to the 20A-X reporter (Fig. 6E). piRNAs mapping to the GFP sequence were 15–26-fold more abundant in progeny that inherited the dsGFP

constructs and 5–12-fold more in progeny with the balancer chromosome when compared to flies that only had the paternal 20A-X. However, even more remarkably, both progenies that inherited dsGFP and those with the balancer chromosome had similar, 3 to 8-fold elevated levels of piRNA produced from regions of 20A-X that are not targeted by GFP siRNAs. This means that maternally contributed GFP siRNAs that target a portion of the 20A-X locus were sufficient to induce piRNA generation from the entire 20A-X locus in the progeny. We noted the presence of some piRNAs in G0 flies that mapped to the SV40 part of the 20A-X reporter in sense orientation and were unaffected by *Rhi* GLKD (Fig. 2E, 5F), suggesting that they arise from the GFP-SV40 mRNA. It is unlikely that these piRNAs are involved in triggering 20A-X piRNA production as repression and piRNA production only occurred in progeny that inherited double-stranded GFP (dsGFP) (Fig. S4A). Finally, using ChIP-qPCR we found that cytoplasmic inheritance of GFP siRNAs from the mother was sufficient to trigger accumulation of H3K9me3 and Rhi on chromatin of paternally inherited 20A-X (Fig. 6F). These results suggest that siRNAs can provide the initial trigger that converts the 20A-X locus into a dual-strand piRNA cluster.

Maternal siRNA triggers activation of another, unrelated piRNA cluster

The complexity of the 20A-X structure and its genomic location may make it unique. To understand whether siRNA-induced cluster initiation is a general phenomenon and not a peculiarity of the 20A-X locus, we took advantage of the only other *de novo* piRNA cluster described, T1/BX2[27]. Strains T1 and BX2 contain seven tandem copies of the *P-lacZ-White (pLacW)* transgene inserted in the same genomic region (Fig. S5A), yet T1 produces abundant piRNAs, while BX2 does not[27]. However, maternal piRNAs derived from the T1 locus can convert the inactive BX2 region into a piRNA cluster, which after this initial conversion is maintained through multiple generations. We first determined whether paternal transmission of the T1 piRNA cluster leads to loss of piRNA production (Fig. 6G). Consistent with our observations for the 20A-X locus, paternal transmission of T1 led to loss of T1 piRNAs and siRNAs effectively inactivating its function as a piRNA cluster (Fig. 6H) and allowing us to use it to study *de novo* cluster establishment. Next, we crossed male T1 flies with the ‘inactivated’ T1 piRNA cluster to females expressing siRNAs against the *white* gene in the germline (using a construct that expresses *white* double-stranded RNA under control of the germline MT-GAL4 driver) and analyzed piRNA production in the progeny (Fig. 6G). The presence of *white* siRNA in the mothers led to more than 5-fold increase in T1-mapping piRNA levels in the progeny compared to flies with only the paternal T1 (Fig. 6H). Remarkably, this effect was independent of the inheritance of the construct expressing the *white* double-stranded RNA in the progeny, i.e., the presence of maternal siRNAs was sufficient to trigger piRNA biogenesis in the progeny. Furthermore, we observed similar increase in piRNA level when females expressing the *white* siRNAs in the germline were crossed to BX2 males, even though BX2 was never active as a piRNA cluster (Fig. 6H). These results indicate that induction of piRNA production by maternally inherited siRNAs is a general phenomenon that can occur in different genomic and sequence contexts, supporting our conclusions that siRNAs provide the initial trigger for conversion of homologous regions into dual-strand piRNA clusters.

Discussion

siRNAs can provide initial trigger to activate piRNA biogenesis

As a system that protects the genome against selfish genetic elements, the piRNA pathway has to be able to adapt to new invader elements. Previous studies have revealed mechanisms to store information about genome invaders in piRNA clusters and to maintain piRNA biogenesis through a positive feedback loop that involves trans-generational cytoplasmic transmission of piRNAs[26,27]. However, the question of how the pathway adapts to new TEs and starts creating piRNAs against novel threats remained unresolved. One possibility for adaptation is the integration of new transposons into pre-existing piRNA clusters, which would lead to generation of novel piRNAs (Fig 7), a process that has been modeled experimentally[26,37] and observed naturally[38,39]. Other studies suggest that entire new piRNA-generating regions can arise in evolution, providing another mechanism for acquiring immunity against novel elements[40]. piRNA cluster regions are extremely labile in *Drosophila* evolution, suggesting frequent acquisition and loss of piRNA clusters[40]. The spontaneous formation of novel piRNA clusters from transgenic sequences has been observed[27].

The finding that siRNAs can activate piRNA biogenesis provides an explanation for how immunity to new transposons can be established through initial detection of new element by the siRNA pathway, which then triggers a stable piRNA response. In contrast to the piRNA pathway that relies on genetic and epigenetic memory to recognize its targets, the siRNA pathway uses a simple rule for self/non-self-discrimination. Transposons often generate both sense and antisense transcripts that form dsRNAs, which are recognized by Dicer and processed into siRNAs, providing a simple yet efficient mechanism to discriminate mobile genetic elements from host genes. Indeed, the siRNA pathway has well-established functions in recognizing and suppressing both endogenous (transposons) and exogenous (viruses) invader genetic elements in all eukaryotic lineages. As both pathways target foreign genetic elements[41,42], siRNAs provide an ideal signal to activate piRNA biogenesis against novel invaders.

Our results indicate that siRNAs are important to jump-start piRNA cluster activity but are dispensable later (Fig. 6). Consistent with siRNAs being dispensable for maintenance of piRNA biogenesis, a previous study showed that siRNAs are not required for the activity of an artificial piRNA cluster[27]. Our results and the previous findings suggest a two-step model of cluster activation (Fig. 7): during the first step siRNAs activate piRNA generation from cognate genomic regions, while during the second step continuous generation and maternal inheritance of piRNAs reinforces piRNA biogenesis making siRNAs dispensable.

The precise molecular mechanism by which siRNAs trigger piRNA biogenesis remains to be understood. The cleavage of complementary transcripts by siRNAs (Fig. 5G) creates aberrant RNAs with 5’-monophosohorylated ends, which are good substrates for the cytoplasmic piRNA processing machinery suggesting that siRNA-induced cleavage of complementary transcripts generates substrates for piRNA processing. Cytoplasmic piRNA processing, in turn, generates piRNAs that are loaded into the three piwi proteins, including the nuclear Piwi protein[43–45]. As the nuclear Piwi/piRNA complex guides establishment of

the H3K9me3 mark and Rhi deposition[23], our model provides a plausible mechanism for how cytoplasmic siRNAs are capable of inducing chromatin changes that are associated with piRNA cluster activation (Fig. 7).

Genomic requirements for piRNA cluster function

Not every region that generates self-targeting siRNAs turns into a piRNA cluster, indicating that siRNAs might be necessary but not sufficient to convert a region into a piRNA cluster. The extended length and repetitive nature seems to be important for *de novo* establishment of clusters. The extended repetitive organization might be required for maintenance of the RDC chromatin compartment. piRNA clusters are enriched in the heterochromatic H3K9me3 mark and the RDC complex that binds this mark[18,19,22,26,46]. Like HP1, Rhi is capable of self-interactions through its chromo and chromo-shadow domains and these interactions are required for formation of RDC compartments and the function of piRNA clusters[26,46]. We postulate that the combination of the extended length and the presence of tandem repeats might be necessary to form a stable RDC chromatin compartment. Establishment of such a region that is capable of maintaining RDC-rich heterochromatin might be the first step in developing piRNA immunity to a new element (Fig. 7).

LIMITATIONS OF THE STUDY

To understand *de novo* formation of piRNA clusters we used artificial (transgenic) clusters and propose that the mechanisms by which natural clusters form are similar. We had to limit our studies to artificial clusters because it is impossible to eliminate pre-existing piRNAs that map to natural clusters. As we only know of two examples of artificial clusters, it is difficult to understand the exact structural requirements for piRNA clusters. The two artificial clusters used in this study are large and repetitive loci that were generated by spontaneous sequence re-arrangements rather than deliberate design. Hopefully, further development of transgenic technologies will enable the design large transgenes to comprehensively dissect the structural requirements for the formation of new piRNA loci.

STAR Methods

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by Alexei Aravin (aravin@caltech.edu).

Materials availability—*D. Melanogaster* strains and plasmids generated by this study (key resources table) are available on request from the lead contact.

Data and code availability

- Libraries generated from this study are deposited in GEO under accession codes GSE193091. Pol ll ChIP-seq data analyzed in this study were from GSE43829[47] and GSE97719[16]. The original images have been deposited at Mendeley: DOI: 10.17632/d65h9pttby.1.

- The scripts and code are available on DOI:10.5281/zenodo.8351702.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

***Drosophila* stocks—**All fly lines used in this study are listed in key resources table. Flies were raised at 25°C.

Transgenic flies—To make the Ubi-GFP-NLS-SV40 reporter, Ubiquitin promoter, GFP-NLS and SV40 were PCR amplified and PCR products were assembled into the EcoR1 and BamH1 digested pBS-KS-attB1–2 vector by Gibson Assembly. The recombinant vector was integrated into three genomic sites chrX: 21522657 dm6 (20A, BDSC #50496), chr2R: 6338399 dm6 (42AB, BDSC #43121) and chr3L: 7575013 dm6 (66A6 control, BDSC #38579). To make the Ubi-GFP(sense)-NLS-GFP(antisense)-SV40 fly, antisense GFP was amplified by PCR and digested with BglII and EagI, then ligated into the BglII and EagI double-digested Ubi-senseGFP vector. Recombinant vectors were integrated into genomic site chrX: 21522657 dm6 (20A, BDSC #50496) and chr3L: 7575013 dm6 (66A6 control, BDSC #38579). All constructs were injected by Bestgene.

20A-X and 20A promoter deletion gRNAs were designed using CRISPR Optimal Target Finder and synthesized by IDT. Oligos were cloned into the pCFD5 vector by Gibson Assembly as described [48]. The DNA oligos sequences are shown below:

20A-X deletion gRNA sequence

Forward:
GCGGCCCGGGTTCGATTCCCGGCCGATGCATTGAAGCTCCCACGAAGTTAGTTTTAGAGCTAGAAATAGCAAG

Reverse:
ATTTTAACTTGCTATTTCTAGCTCTAAAACTAGTTGACGAGTGTCCGCTTTGCACCAGCCGGGAATCGAACCC

20A promoter gRNA sequence

Forward:

GCGGCCCGGGTTCGATTCCCGGCCGATGCAACTACGTTACTAAGCATTTGGTTTTAGAGCTAGAAATAGCAAG

Reverse:
ATTTTAACTTGCTATTTCTAGCTCTAAAACGATGTCCAAACTTGCAATTTTGCACCAGCCGGGAATCGAACCC

All transgenic constructs were inserted into the attP40 landing site at 25C6 (y1w67c23; P[CaryP]attP40) on the 2nd chromosome and attP2 landing site at 68A4 (y1w67c23; P[CaryP]attP2) on the 3rd chromosome, unless specifically mentioned. To obtain 20A

promoter deletion flies, flies carrying gRNAs were crossed with Nos-Cas9 flies (CAS-0001, NIG-FLY). Individual progenies were screened to verify the promoter deletion by genomic PCR followed by sanger sequencing. Transgenic flies used in this study are listed in key resources table.

METHOD DETAILS

RNA HCR-FISH—The HCR-FISH RNA protocol was adapted from a previous protocol[49,50]. Fly ovaries were dissected in cold PBS and subsequently fixed in 300 μl of fixation solution containing 4% paraformaldehyde and 0.15% Triton X-100 in PBS at room temperature. After fixation, the samples underwent three washes with PBX solution (PBS with 0.3% Triton X-100) for a duration of 5 min each at room temperature. The samples were then dehydrated in 500 μl of 70% ethanol and were allowed to permeabilize overnight at 4°C while placed on a nutator. Following permeabilization, the samples were rehydrated in 500 μl of wash buffer (consisting of 2 × SSC, 10% [v/v] formamide) for 5 min at room temperature. Subsequently, they were pre-hybridized in 500 μl of hybridization buffer (containing 50% formamide, 5x SSC, 9 mM citric acid pH 6.0, 0.1% Tween 20, 50 μg/ml heparin, 1x Denhardt's solution, 10% dextran sulfate) for a duration of 30 min at 37°C. After pre-hybridization, the samples were exposed to the hybridization solution that included 2 pmol of each probe and incubated for a period of 12–16 hours at 37°C. The samples underwent a series of washes: four times with 500 μl of probe wash buffer (comprising 50% formamide, 5x SSC, 9 mM citric acid pH 6.0, 0.1% Tween 20, 50 μg/ml heparin) for 15 min each at 37°C, three washes with 5 x SSCT (5x SSC with 0.1% Tween 20) for 5 min each at room temperature. Samples were then incubated in 500 μl of amplification buffer (5x SSC, 0.1% Tween 20, 10% dextran sulfate) for 30 min at room temperature. Hairpin H1 and hairpin H2 were each prepared separately by incubating them at 95°C for 90 seconds and then allowing them to cool to room temperature in the dark for 30 min. Samples were incubated with the hairpin solution for a duration of 12–16 hours in the dark at room temperature. After hairpin incubation, samples were washed with 500 μl of 5x SSCT at room temperature in the following sequence: 2 times for 5 min each, 2 times for 30 min each, and 1 time 5 min.The probes were designed and synthesized by Molecular Technologies and Alexa594 was used for probe detection. Images were acquired using the ZEISS LSM880 and data was processed using Zen software. Scale bar is 20μm and 2μm for egg chamber and single nurse cell nuclei, respectively. Images were taken using the same settings for all constructs.

ChIP-seq and ChIP-qPCR—All ChIP experiments were performed based on previous description[17,51]. Approximately 100 dissected fly ovaries were fixed with 1.8% formaldehyde in PBS for 10 min at room temperature. Subsequently, glycine (25 mM final concentration) was used to quench the fixation for 5 min at room temperature. The samples were then dounced and sonicated in lysis buffer (20 mM Tris-Cl (pH 7.4), 150 mM NaCl, 0.35% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, EDTA-free protease inhibitor, 10 mM NaF, and 0.2 mM Na_3VO_4). Sonication was conducted using a Bioruptor (Diagenode) at high power for 30 cycles (30 seconds on, 30 seconds off). Subsequently, two volumes of lysis dilution buffer (20 mM Tris-Cl, pH 7.4, 150 mM NaCl, EDTA-free protease inhibitor, 10 mM NaF, and 0.2 mM Na_3VO_4) were added. The samples were then centrifuged, with

inputs saved, and the remaining supernatants were incubated with the appropriate antibody at 4°C overnight. After this incubation, pre-cleared beads (pre-treated with 10 mg/mL BSA and 10 μg salmon sperm DNA) were added to the supernatants and further incubated at 4°C for 3 hours. Subsequently, the samples were washed five times with LiCl buffer (10 mM Tris-Cl, pH 7.5, 500 mM LiCl, 1% NP-40, 1% sodium deoxycholate). The purified DNA was eluted with 100 μg of proteinase K in proteinase K buffer (200 mM Tris-Cl, pH 7.4, 25 mM EDTA, 300 mM NaCl, 2% SDS) at 55°C for 3 hours, followed by an additional incubation at 65°C overnight. Finally, the eluted DNA was purified using phenol-chloroform extraction followed by ethanol precipitation. SYBR Green qPCR was performed by using MyTaq HS Mix (BioLine). CT values were calculated from technical duplicates. All ChIP-qPCR were normalized to respective inputs and to control region rp49. ChIP-qPCR were performed on a Mastercycler®ep Realplex PCR thermal cycler machine (Eppendorf), All qPCR primers are listed in key resources table. ChIP-seq libraries were generated using the NEBNext ChIP-Seq Library Prep Master Mix Set. All libraries were sequenced on the Illumina HiSeq 2500 platform (SE 100 bp reads).

Small RNA-seq—Total RNA was isolated from dissected ovaries using TRIzol (ThermoFisher #15596018). 4μg total RNA was loaded onto a 15% polyacrylamide gel and small RNA between 19 and 29 nt in length was excised and isolated, treated with 200 mM Sodium periodate at 25°C for 30min for the oxidized samples shown in Fig. 3H, S2F and S2G. Size selected small RNA was ethanol-precipitated and small RNA library constructed using the NEBNext small RNA library preparation set (#E7330S). Libraries were sequenced on the Illumina HiSeq 2500 platform (SE 50-bp reads).

RT-qPCR—Around 20 ovaries were dissected and homogenized in 1mL TRIzol (ThermoFisher #15596018) and total RNA was extracted following the manufacturer's recommendation. DNase I treatment and reverse transcription was performed from 1μg total RNA starting material, using DNase I and SuperScript III (Invitrogen) following the manufacturer's recommendation. qPCR was performed by using MyTaq HS Mix (BioLine) contain SYBR Green on a Mastercycler®ep Realplex PCR thermal cycler machine (Eppendorf). CT values were calculated from technical duplicates. All qPCR data were normalized to the *rp49* mRNA expression. All qPCR primers are listed in key resources table.

DNA FISH—Polytene chromosomes DNA FISH was performed as previously described[52,53] with the following modifications. Salivary glands were dissected from 3rd instar larvae and fixed in fixation buffer (3.7% Formaldehyde, 1% Triton X-100 in PBS, pH 7.5) for 5 min, then transferred into solution (3.7% Formaldehyde, 50% acetic acid) for 2 min on the cover slip. Cover slip was put on poly-L-lysine coated microscope slide and chromosomes were spread and quashed by gently moving the cover slip back and forth followed by pressure applied to the cover slip by thumb. Slides were flash frozen in liquid nitrogen to remove cover slip and submerged in PBS for 10 min followed by three 5 min washes in 2x SSC. Samples were dehydrated by 5 min incubations twice in 70% ethanol and twice in 96% ethanol, followed by air-drying slides. Slides were incubated in 2x SSC for 45 min at 70°C and dehydrated again as described above. To denature the DNA,

slides were incubated in 100 mM NaOH for 10 min, washed three times with 2x SSC and dehydrate as described above. Slides were incubated in hybridization buffer (2X SSC, 10% dextran sulfate, 50% formamide, 0.8 mg/mL salmon sperm DNA) for 5 min at 80°C and snap cooled on ice. DNA FISH probes were prepared following the manufacturer's recommendations (ThermoFisher # F32947 and F32949) using the BAC construct (BACPAC Resources #CH322–184J4) as probe template for 20A (Alexa 594) and the original reporter vector (non-RMCE) as probe template for the GFP reporter (Alexa 488). Probes pre-warmed to 37°C were loaded on the slides, covered with cover slip, sealed with rubber cement and incubated in a dark and humid chamber at 37°C overnight. Slides were washed in 2x SSC three times at 42°C and once at RT, 5 min each time followed by DAPI staining for 10 min and two washes in PBS. Slides were mounted with mounting medium (Vector Labs #H-1000). Images were acquired using the ZEISS LSM880.

Bioinformatic analysis—ChIP-seq processing and mapping. Trimmomatic (version 0.33)[54] and cutadapt (version 1.15)[55] were used to trim off adaptors and filter out those shorter than 50 nt after trimming. The first 50 nt from each read were mapped to dm3 genome and vector sequence respectively, using Bowtie[56] (version 1.0.1, parameters: -v 2 -k 1 -m 1 -t –best -y --strata). After mitochondria reads were removed, aligned reads were then used to generate piled-up RPM signals and enrichment profiles by our customized scripts and deepTools[57]. Regions blacklisted by ENCODE[58] were excluded from enrichment analysis. Read counts over equal-sized bins were calculated using deepTools2 and BEDOPS[59], and figures were made using MATLAB.

To map fusion read, the first 20 nt and full length of the reads were mapped to vector sequences using the aforementioned bowtie settings. Reads where the first 20nt mapped to vector sequences but the full length did not were selected. The last 20 nt of such reads was mapped to the reference genome with the same settings. Mappable reads among these were considered fusion reads between the vector and genome which were used to identify insertion location.

For small RNA-seq analysis, Trimmomatic and cutadapt were used to trim off adaptors and filter out reads shorter than 20 nt after trimming. We then extracted reads of have specific lengths were extracted: 21–22 nt (siRNA), 23–29 nt (piRNA) and 21–30 nt (small RNA). The selected reads were mapped to the dm6 genome using Bowtie (parameters: -v 0 -a -m 1 -t --best --strata). After mitochondrial reads were removal, deepTools2 and BEDOPS were used to calculate read counts over equal-sized bins. Ping-pong signature was inferred using a published method[60]. Reporter coverage was calculated based on 10 nt bin size, and figures were made using MATLAB.

QUANTIFICATION AND STATISTICAL ANALYSIS

Phasing analysis was performed as described previously[36] with the following modifications. The selected 23–29bp reads were mapped to the vector sequence. Considering only pairs of nonoverlapping ($d \geq 0$) piRNA reads on the same strand, the distribution of head-to-tail distances ($d_{i,j}$) were calculated. We focused on the distance distributions between 0bp and 50bp and normalized by linearly scaling (P) or standardization (Z).

$$d_{i,j} = R_{j,1} - R_{i,2}$$

$$C_d = \sum_{i,j} [d_{i,j} = d]$$

$$P(d = X) = \frac{C_X}{\sum_{X=0}^{50} C_X}$$

$$Z(d = X) = \frac{C_X - \overline{C_{X|0<X\leq 20}}}{\frac{\sqrt{\sum_{0<X\leq 20}(C_X - \overline{C_{X|0<X\leq 20}})^2}}{20}}$$

Where $R_{i,1}$ and $R_{i,2}$ are the left and right coordinates of a mapped read and C_d is the frequency of head-to-tail distances.

Then, considering only the pairs of piRNA reads that are not sharing the 5’ end, within 200bp range ($0 < D \leq 200$), and on the same strand, the distribution of head-to-head distances (Di,j) were calculated.

$$D_{i,j|+} = R_{j,1} - R_{i,1}$$

$$D_{i,j|-} = R_{j,2} - R_{i,2}$$

$$C_{d|+} = \sum_{i,j} [D_{i,j|+} = D]$$

$$C_{d|-} = \sum_{i,j} [D_{i,j|-} = D]$$

Where $R_{i,1}$ and $R_{i,2}$ are the left and right coordinates of of a mapped read, C_d is the frequency of head-to-head distances.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank members of the Aravin and Fejes Toth labs for discussion and comments. We thank Julius Brennecke for providing the Rhino antibody. We thank Igor Antoshechkin (Caltech) for help with sequencing. This work was supported by grants from the National Institutes of Health (R01 GM097363 to AA and R01 GM110217 to KFT) and by the HHMI Faculty Scholar Award to AAA.

REFERENCES

1. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, et al. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature 442, 203–207. 10.1038/nature04916. [PubMed: 16751777]
2. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, and Hannon GJ (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128, 1089–1103. 10.1016/j.cell.2007.01.043. [PubMed: 17346786]
3. Gitlin L, Karelsky S, and Andino R (2002). Short interfering RNA confers intracellular antiviral immunity in human cells. Nature 418, 430–434. 10.1038/nature00873. [PubMed: 12087357]
4. Vance V, and Vaucheret H (2001). RNA silencing in plants--defense and counterdefense. Science 292, 2277–2280. 10.1126/science.1061334. [PubMed: 11423650]
5. Lindbo JA, Silva-Rosales L, Proebsting WM, and Dougherty WG (1993). Induction of a Highly Specific Antiviral State in Transgenic Plants: Implications for Regulation of Gene Expression and Virus Resistance. Plant Cell 5, 1749–1759. 10.1105/tpc.5.12.1749. [PubMed: 12271055]
6. Voinnet O, Pinto YM, and Baulcombe DC (1999). Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. Proc Natl Acad Sci U S A 96, 14147–14152. 10.1073/pnas.96.24.14147. [PubMed: 10570213]
7. Gammon DB, and Mello CC (2015). RNA interference-mediated antiviral defense in insects. Curr Opin Insect Sci 8, 111–120. 10.1016/j.cois.2015.01.006. [PubMed: 26034705]
8. 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. 10.1038/35053110. [PubMed: 11201747]
9. Rand TA, Petersen S, Du F, and Wang X (2005). Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. Cell 123, 621–629. 10.1016/j.cell.2005.10.020. [PubMed: 16271385]
10. Matranga C, Tomari Y, Shin C, Bartel DP, and Zamore PD (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell 123, 607–620. 10.1016/j.cell.2005.08.044. [PubMed: 16271386]
11. Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler EL, Zapp ML, Weng Z, and Zamore PD (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science 320, 1077–1081. 10.1126/science.1157396. [PubMed: 18403677]
12. Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, and Siomi H (2008). Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature 453, 793–797. 10.1038/nature06938. [PubMed: 18463636]
13. Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, and Zamore PD (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science 313, 320–324. 10.1126/science.1129333. [PubMed: 16809489]
14. Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, et al. (2008). An endogenous small interfering RNA pathway in Drosophila. Nature 453, 798–802. 10.1038/nature07007. [PubMed: 18463631]
15. Okamura K, Ishizuka A, Siomi H, and Siomi MC (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. Genes Dev 18, 1655–1666. 10.1101/gad.1210204. [PubMed: 15231716]
16. Andersen PR, Tirian L, Vunjak M, and Brennecke J (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. Nature 549, 54–59. 10.1038/nature23482. [PubMed: 28847004]

17. Chen YA, Stuwe E, Luo Y, Ninova M, Le Thomas A, Rozhavskaya E, Li S, Vempati S, Laver JD, Patel DJ, et al. (2016). Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. Mol Cell 63, 97–109. 10.1016/j.molcel.2016.05.010. [PubMed: 27292797]

18. Mohn F, Sienski G, Handler D, and Brennecke J (2014). The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. Cell 157, 1364–1379. 10.1016/j.cell.2014.04.031. [PubMed: 24906153]

19. Zhang Z, Wang J, Schultz N, Zhang F, Parhad SS, Tu S, Vreven T, Zamore PD, Weng Z, and Theurkauf WE (2014). The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. Cell 157, 1353–1363. 10.1016/j.cell.2014.04.030. [PubMed: 24906152]

20. ElMaghraby MF, Andersen PR, Puhringer F, Hohmann U, Meixner K, Lendl T, Tirian L, and Brennecke J (2019). A Heterochromatin-Specific RNA Export Pathway Facilitates piRNA Production. Cell 178, 964–979 e920. 10.1016/j.cell.2019.07.007. [PubMed: 31398345]

21. Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, and Siomi MC (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. Genes Dev 20, 2214–2222. 10.1101/gad.1454806. [PubMed: 16882972]

22. Klattenhoff C, Xi H, Li C, Lee S, Xu J, Khurana JS, Zhang F, Schultz N, Koppetsch BS, Nowosielska A, et al. (2009). The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell 138, 1137–1149. 10.1016/j.cell.2009.07.014. [PubMed: 19732946]

23. Akkouche A, Mugat B, Barckmann B, Varela-Chavez C, Li B, Raffel R, Pelisson A, and Chambeyron S (2017). Piwi Is Required during Drosophila Embryogenesis to License Dual-Strand piRNA Clusters for Transposon Repression in Adult Ovaries. Mol Cell 66, 411–419 e414. 10.1016/j.molcel.2017.03.017. [PubMed: 28457744]

24. Yu Y, Gu J, Jin Y, Luo Y, Preall JB, Ma J, Czech B, and Hannon GJ (2015). Panoramix enforces piRNA-dependent cotranscriptional silencing. Science 350, 339–342. 10.1126/science.aab0700. [PubMed: 26472911]

25. Sienski G, Batki J, Senti KA, Donertas D, Tirian L, Meixner K, and Brennecke J (2015). Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. Genes Dev 29, 2258–2271. 10.1101/gad.271908.115. [PubMed: 26494711]

26. Le Thomas A, Stuwe E, Li S, Du J, Marinov G, Rozhkov N, Chen YC, Luo Y, Sachidanandam R, Toth KF, et al. (2014). Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. Genes Dev 28, 1667–1680. 10.1101/gad.245514.114. [PubMed: 25085419]

27. de Vanssay A, Bouge AL, Boivin A, Hermant C, Teysset L, Delmarre V, Antoniewski C, and Ronsseray S (2012). Paramutation in Drosophila linked to emergence of a piRNA-producing locus. Nature 490, 112–115. 10.1038/nature11416. [PubMed: 22922650]

28. Casier K, Delmarre V, Gueguen N, Hermant C, Viode E, Vaury C, Ronsseray S, Brasset E, Teysset L, and Boivin A (2019). Environmentally-induced epigenetic conversion of a piRNA cluster. Elife 8. 10.7554/eLife.39842.

29. Ipsaro JJ, Haase AD, Knott SR, Joshua-Tor L, and Hannon GJ (2012). The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. Nature 491, 279–283. 10.1038/nature11502. [PubMed: 23064227]

30. Nishimasu H, Ishizu H, Saito K, Fukuhara S, Kamatani MK, Bonnefond L, Matsumoto N, Nishizawa T, Nakanaga K, Aoki J, et al. (2012). Structure and function of Zucchini endoribonuclease in piRNA biogenesis. Nature 491, 284–287. 10.1038/nature11509. [PubMed: 23064230]

31. Goriaux C, Desset S, Renaud Y, Vaury C, and Brasset E (2014). Transcriptional properties and splicing of the flamenco piRNA cluster. EMBO Rep 15, 411–418. 10.1002/embr.201337898. [PubMed: 24562610]

32. Chang TH, Mattei E, Gainetdinov I, Colpan C, Weng Z, and Zamore PD (2019). Maelstrom Represses Canonical Polymerase II Transcription within Bi-directional piRNA Clusters in Drosophila melanogaster. Mol Cell 73, 291–303 e296. 10.1016/j.molcel.2018.10.038. [PubMed: 30527661]

33. Hermant C, Boivin A, Teysset L, Delmarre V, Asif-Laidin A, van den Beek M, Antoniewski C, and Ronsseray S (2015). Paramutation in Drosophila Requires Both Nuclear and Cytoplasmic Actors of the piRNA Pathway and Induces Cis-spreading of piRNA Production. Genetics 201, 1381–1396. 10.1534/genetics.115.180307. [PubMed: 26482790]
34. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, and Siomi MC (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science 315, 1587–1590. 10.1126/science.1140494. [PubMed: 17322028]
35. Mohn F, Handler D, and Brennecke J (2015). Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science 348, 812–817. 10.1126/science.aaa1039. [PubMed: 25977553]
36. Han BW, Wang W, Li C, Weng Z, and Zamore PD (2015). Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science 348, 817–821. 10.1126/science.aaa1264. [PubMed: 25977554]
37. Muerdter F, Olovnikov I, Molaro A, Rozhkov NV, Czech B, Gordon A, Hannon GJ, and Aravin AA (2012). Production of artificial piRNAs in flies and mice. RNA 18, 42–52. 10.1261/rna.029769.111. [PubMed: 22096018]
38. Zhang S, Pointer B, and Kelleher ES (2020). Rapid evolution of piRNA-mediated silencing of an invading transposable element was driven by abundant de novo mutations. Genome Res 30, 566–575. 10.1101/gr.251546.119. [PubMed: 32238416]
39. Khurana JS, Wang J, Xu J, Koppetsch BS, Thomson TC, Nowosielska A, Li C, Zamore PD, Weng Z, and Theurkauf WE (2011). Adaptation to P element transposon invasion in Drosophila melanogaster. Cell 147, 1551–1563. 10.1016/j.cell.2011.11.042. [PubMed: 22196730]
40. Gebert D, Neubert LK, Lloyd C, Gui J, Lehmann R, and Teixeira FK (2021). Large Drosophila germline piRNA clusters are evolutionarily labile and dispensable for transposon regulation. Mol Cell 81, 3965–3978 e3965. 10.1016/j.molcel.2021.07.011. [PubMed: 34352205]
41. Mondal M, Mansfield K, and Flynt A (2018). siRNAs and piRNAs collaborate for transposon control in the two-spotted spider mite. RNA 24, 899–907. 10.1261/rna.065839.118. [PubMed: 29678924]
42. Mondal M, Brown JK, and Flynt A (2020). Exploiting somatic piRNAs in Bemisia tabaci enables novel gene silencing through RNA feeding. Life Sci Alliance 3. 10.26508/lsa.202000731.
43. Rogers AK, Situ K, Perkins EM, and Toth KF (2017). Zucchini-dependent piRNA processing is triggered by recruitment to the cytoplasmic processing machinery. Genes Dev 31, 1858–1869. 10.1101/gad.303214.117. [PubMed: 29021243]
44. Pandey RR, Homolka D, Chen KM, Sachidanandam R, Fauvarque MO, and Pillai RS (2017). Recruitment of Armitage and Yb to a transcript triggers its phased processing into primary piRNAs in Drosophila ovaries. PLoS Genet 13, e1006956. 10.1371/journal.pgen.1006956. [PubMed: 28827804]
45. Huang X, Fejes Toth K, and Aravin AA (2017). piRNA Biogenesis in Drosophila melanogaster. Trends Genet 33, 882–894. 10.1016/j.tig.2017.09.002. [PubMed: 28964526]
46. Yu B, Cassani M, Wang M, Liu M, Ma J, Li G, Zhang Z, and Huang Y (2015). Structural insights into Rhino-mediated germline piRNA cluster formation. Cell Res 25, 525–528. 10.1038/cr.2015.10. [PubMed: 25613572]
47. Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, and Toth KF (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev 27, 390–399. 10.1101/gad.209841.112. [PubMed: 23392610]
48. Port F, and Bullock SL (2016). Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. Nat Methods 13, 852–854. 10.1038/nmeth.3972. [PubMed: 27595403]
49. Choi HMT, Schwarzkopf M, Fornace ME, Acharya A, Artavanis G, Stegmaier J, Cunha A, and Pierce NA (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development 145. 10.1242/dev.165753.
50. Luo Y, Fefelova E, Ninova M, Chen YA, and Aravin AA (2020). Repression of interrupted and intact rDNA by the SUMO pathway in Drosophila melanogaster. Elife 9. 10.7554/eLife.52416.

51. Hur JK, Luo Y, Moon S, Ninova M, Marinov GK, Chung YD, and Aravin AA (2016). Splicing-independent loading of TREX on nascent RNA is required for efficient expression of dual-strand piRNA clusters in Drosophila. Genes Dev 30, 840–855. 10.1101/gad.276030.115. [PubMed: 27036967]
52. Lavrov S, Dejardin J, and Cavalli G (2004). Combined immunostaining and FISH analysis of polytene chromosomes. Methods Mol Biol 247, 289–303. 10.1385/1-59259-665-7:289. [PubMed: 14707354]
53. Cai W, Jin Y, Girton J, Johansen J, and Johansen KM (2010). Preparation of Drosophila polytene chromosome squashes for antibody labeling. J Vis Exp 10.3791/1748.
54. Bolger AM, Lohse M, and Usadel B (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. 10.1093/bioinformatics/btu170. [PubMed: 24695404]
55. Martin M (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads 2011 17, 3. 10.14806/ej.17.1.200.
56. Langmead B, Trapnell C, Pop M, and Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10, R25. 10.1186/gb-2009-10-3-r25. [PubMed: 19261174]
57. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, and Manke T (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 44, W160–165. 10.1093/nar/gkw257. [PubMed: 27079975]
58. Amemiya HM, Kundaje A, and Boyle AP (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci Rep 9, 9354. 10.1038/s41598-019-45839-z. [PubMed: 31249361]
59. Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J, Thomas S, et al. (2012). BEDOPS: high-performance genomic feature operations. Bioinformatics 28, 1919–1920. 10.1093/bioinformatics/bts277. [PubMed: 22576172]
60. Antoniewski C (2014). Computing siRNA and piRNA overlap signatures. Methods Mol Biol 1173, 135–146. 10.1007/978-1-4939-0931-5_12. [PubMed: 24920366]

Highlights

- A transcribed repetitive transgene is spontaneously converted into a piRNA cluster
- Establishment of piRNA cluster requires inheritance of cognate small RNA from mothers
- siRNAs initiate the formation of piRNA cluster, but are dispensable after its establishment

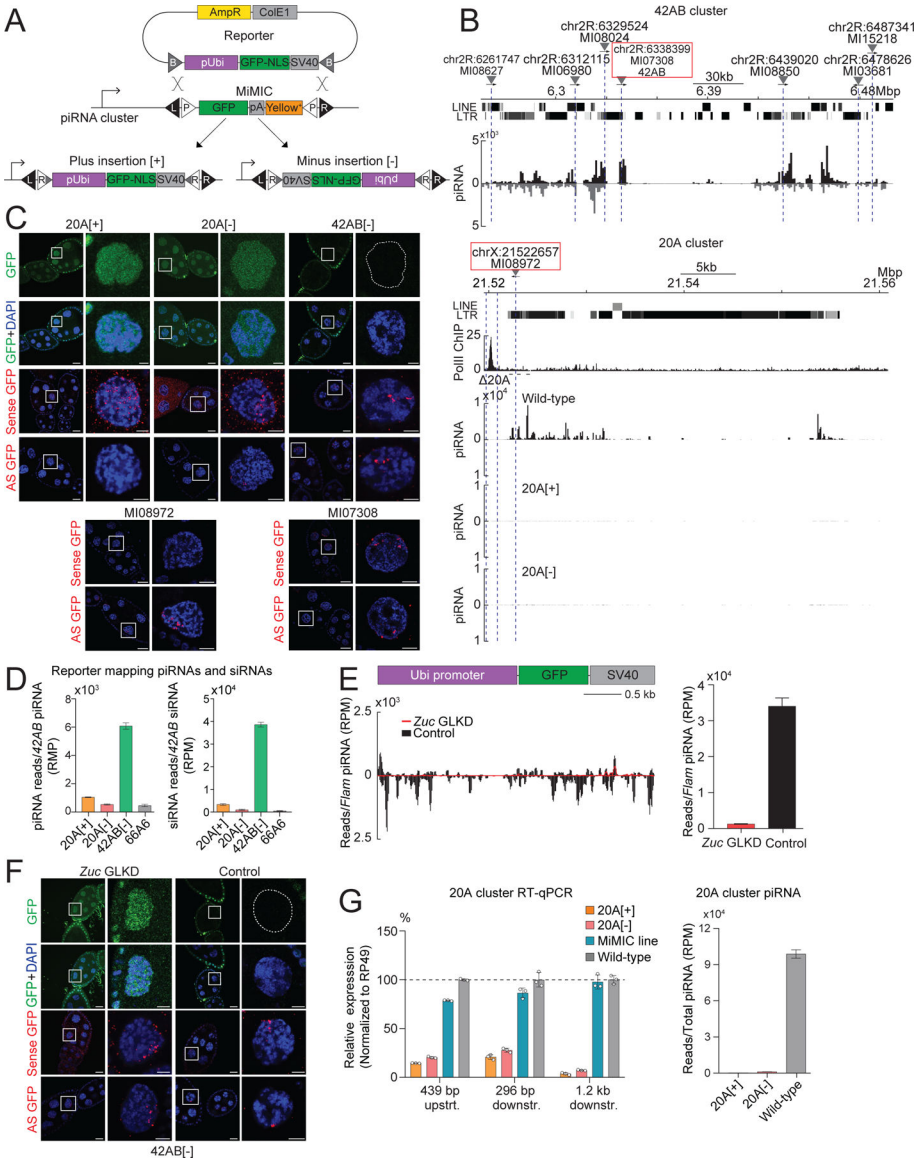


Figure 1. Reporters integrated in uni- and dual-strand clusters have different expression and effects on cluster activity.

(A) Scheme of reporter integration into piRNA clusters using recombinase-mediated cassette exchange (RMCE) to replace Minos-mediated integration cassettes (MiMIC).

(B) Profiles of the dual-strand cluster 42AB (top) and uni-strand cluster 20A (bottom). Shown are profiles of uniquely-mapping piRNAs, positions of the putative promoter (Pol II ChIP peak) of the 20A cluster and positions of the reporters. Position of primers used in Fig. 1G are indicated by solid lines.

(C) Expression of reporters integrated in 20A and 42AB clusters. Expression of GFP protein and sense and antisense RNA in flies with reporter insertions in the 20A and 42AB clusters. The 42AB reporter generates transcripts from both strands and both sense and antisense RNA are localized in the nucleus. Scale bar is 20μm and 2μm for egg chamber and single nurse cell nuclei, respectively.

(D) piRNAs and siRNA are generated from the reporter in the 42AB cluster, but not the reporter inserted into 20A or the control locus. Error bars indicate standard deviation of two biological replicates.

(E) Knockdown of *Zuc* eliminates 42AB reporter piRNA. Error bars indicate standard deviation of two biological replicates.

(F) Derepression of 42AB reporter expression upon *Zuc* GLKD. Shown are GFP protein expression and FISH signal for both strands of the reporter.

(G) Ovarian expression of the 20A cluster is suppressed upon reporter integration. (Left) 20A cluster transcripts were measured by RT-qPCR. Error bars indicate the standard deviation of three biological replicates. (Right) Expression of 20A cluster piRNA. Error bars indicate standard deviation of two biological replicates.

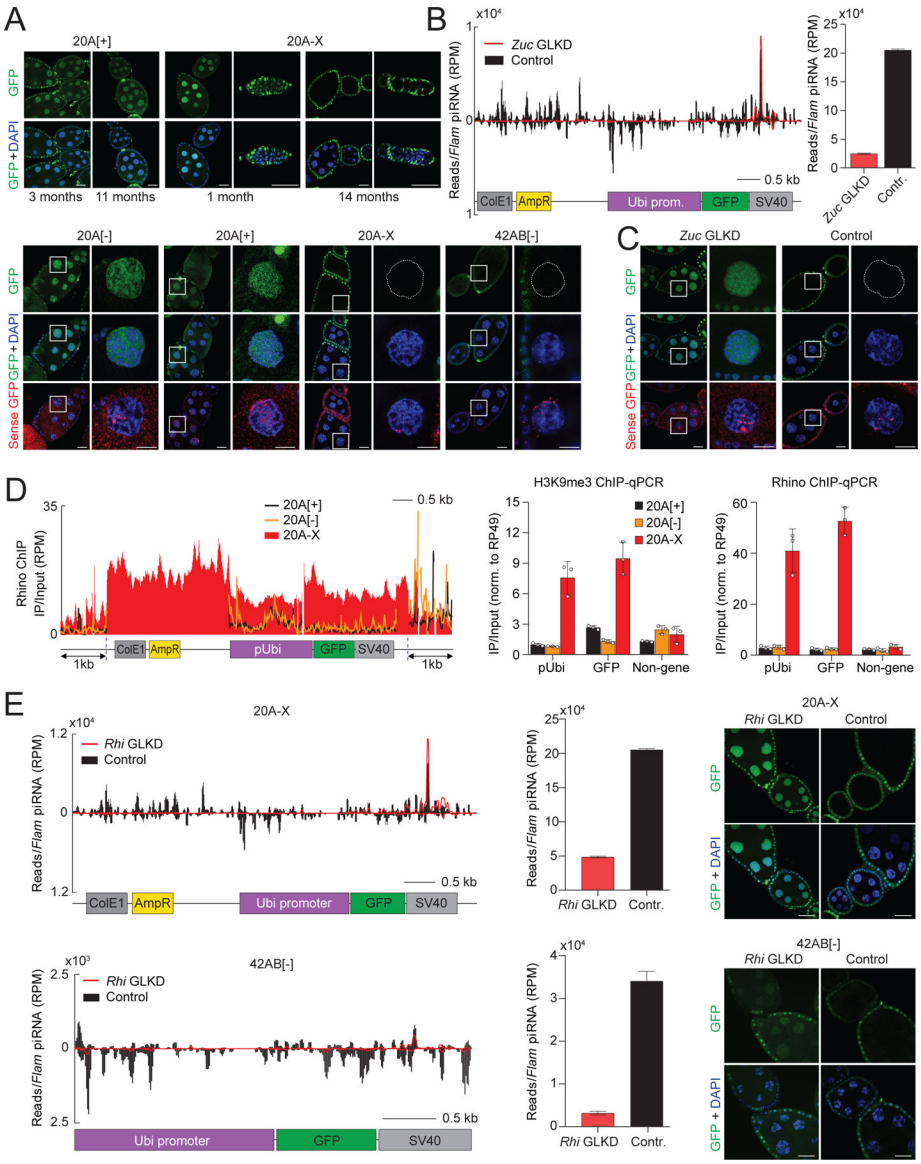


Figure 2. An unusual reporter insertion in the 20A cluster is a dual-stand piRNA cluster that generates piRNAs in a Rhi-dependent manner

(A) The 20A-X reporter is silenced in the germline after several months of maintaining this line. Top: GFP protein in ovaries of 20A[+] and 20A-X flies at different times after the establishment of lines. Bottom: GFP protein and sense RNA expression in flies 14 months (20A reporters) and 1 month (42AB[−]) after the establishment of lines.

(B) 20A-X generates piRNAs in a Zuc-dependent manner. Error bars indicate standard deviation of two biological replicates.

(C) *Zuc* knockdown releases 20A-X repression in the germline. Shown are expression of GFP protein and sense RNA in control (*white* GLKD) and upon *Zuc* GLKD driven by nos-GAL4.

(D) 20A-X, but not other 20A reporters are enriched in Rhino and the H3K9me3 mark. Rhino ChIP-seq profiles (mean of two biological replicates) on the reporters and 1

kb flanking regions (left). Rhino and H3K9me3 enrichment were determined by ChIP-qPCR (right). Error bars indicate the standard deviation of three biological replicates.

(E) *Rhino* knockdown reduces 20A-X piRNA and releases its silencing in the germline. piRNA profiles in control flies (*white* GLKD) and upon *Rhi* GLKD driven by nos-GAL4 (left). Error bars indicate the standard deviation of two biological replicates. GFP protein expression upon *Rhi* GLKD (right).

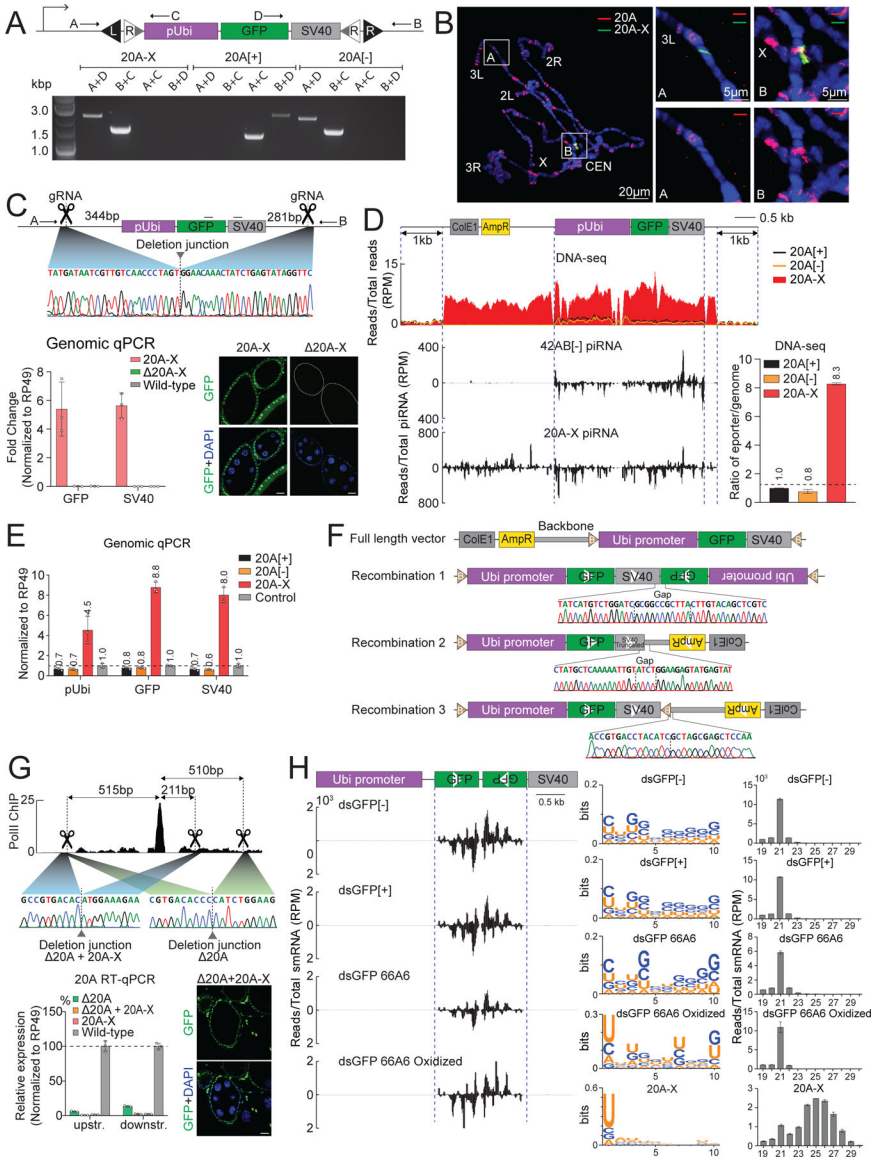


Figure 3. The 20A-X locus contains rearranged, multi-copy reporter sequences

(A) Determining reporter orientation by genomic PCR. 20A-X has the same minus-strand orientation as the 20A[−] reporter.

(B) 20A-X is located in a single site in the 20A region. DNA FISH on polytene chromosomes using probes against the reporter (green, Cy488) and the 20A cluster (red, Cy594). Probes against the reporter detected two locations: one co-localizes with the 20A cluster signal, the other signal is localized on chromosome 3L where the native ubiquitin gene resides.

(C) Verification of the 20A-X insertion site by CRISPR deletion. Detection of reporter sequences using qPCR (left) shows absence of reporter in flies with the deletion. No GFP expression is detected in flies with the deletion (right).

(D) 20A-X includes plasmid backbone sequence. (Top) Profiles of whole-genome DNA-seq reads. Only 20A-X flies harbor plasmid backbone sequence. (Bottom) piRNA profiles over the reporter sequence. Profiles represent a mean of two biological replicates.

(E) 20A-X contains multiple copies of reporter sequence. Different portions of the reporter sequence were measured by genomic qPCR. Error bars indicate standard deviation of three biological replicates.

(F) Reporter sequence rearrangements in 20A-X. Three abnormal sequence junctions were detected in the 20A-X sequence by DNA-seq, Sanger sequencing PCR-amplified genomic DNA.

(G) Deletion of the 20A cluster promoter does not affect activity of 20A-X. The putative promoter of the 20A cluster was deleted using CRISPR/Cas9 (Δ20A) in wild-type and 20A-X flies. RT-qPCR shows a dramatic reduction of cluster transcripts in flies with the deletion. Error bars indicate standard deviation of three biological replicates.

(H) Inverted repeat reporters generate abundant endo-siRNA, but not piRNA. Scheme of the inverted repeat dsGFP reporter. Shown are small RNA profiles (19–30 nt) along the reporter sequence in ovaries of flies with reporter integration into the 20A cluster and the control region (66A6). Shown on the right are size distributions and nucleotide composition of reporter mapping small RNAs.

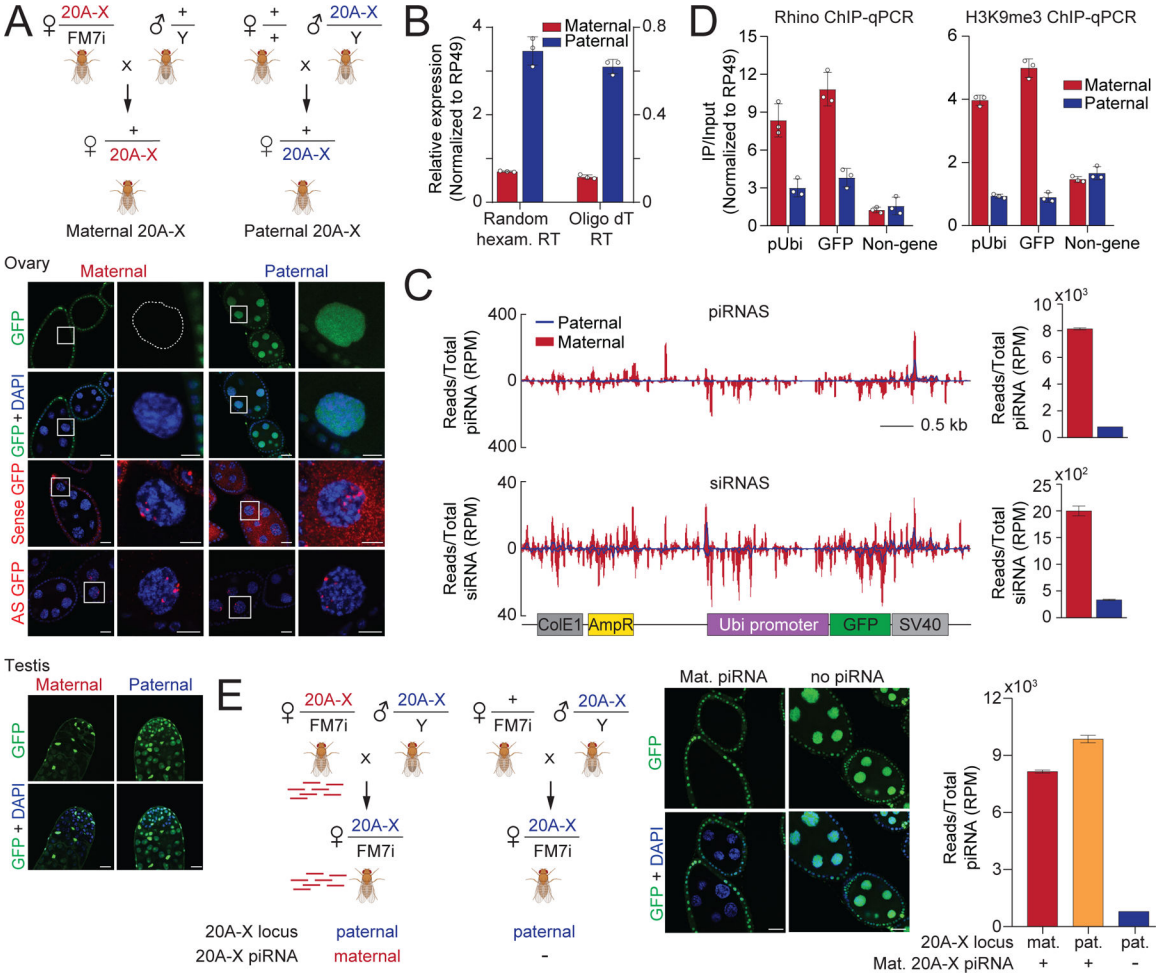


Figure 4. piRNA-induced repression of the 20A-X reporter depends on maternal transmission of cognate piRNAs.

(A) 20A-X repression is released after paternal transmission. Scheme of crosses to test effects of maternal and paternal transmission of 20A-X. Shown are GFP and sense and antisense RNA expression in ovaries and testes of the progeny.

(B) GFP RNA expression increases after paternal transmission of 20A-X. RT-qPCR of GFP RNA in progenies from the two crosses in (A). Error bars indicate the standard deviation of three biological replicates.

(C) 20A-X piRNA level drops after paternal inheritance. Shown are piRNA and siRNA profiles along the reporter sequence in progeny from the two crosses shown in (A). Bargraphs on the right show number of piRNA and siRNA reads mapping to the reporter. Error bars indicate standard deviation of two biological replicates.

(D) Rhino and H3K9me3 are lost on 20A-X chromatin after paternal transmission. The levels of Rhino and H3K9me3 on chromatin were measured by ChIP-qPCR. Error bars indicate the standard deviation of three biological replicates.

(E) Cytoplasmic piRNA inheritance is sufficient for repression of paternally transmitted 20A-X. Scheme of crosses to test the effect of maternal piRNA inheritance on repression of paternally transmitted 20A-X. Error bars indicate standard deviation of two biological replicates.

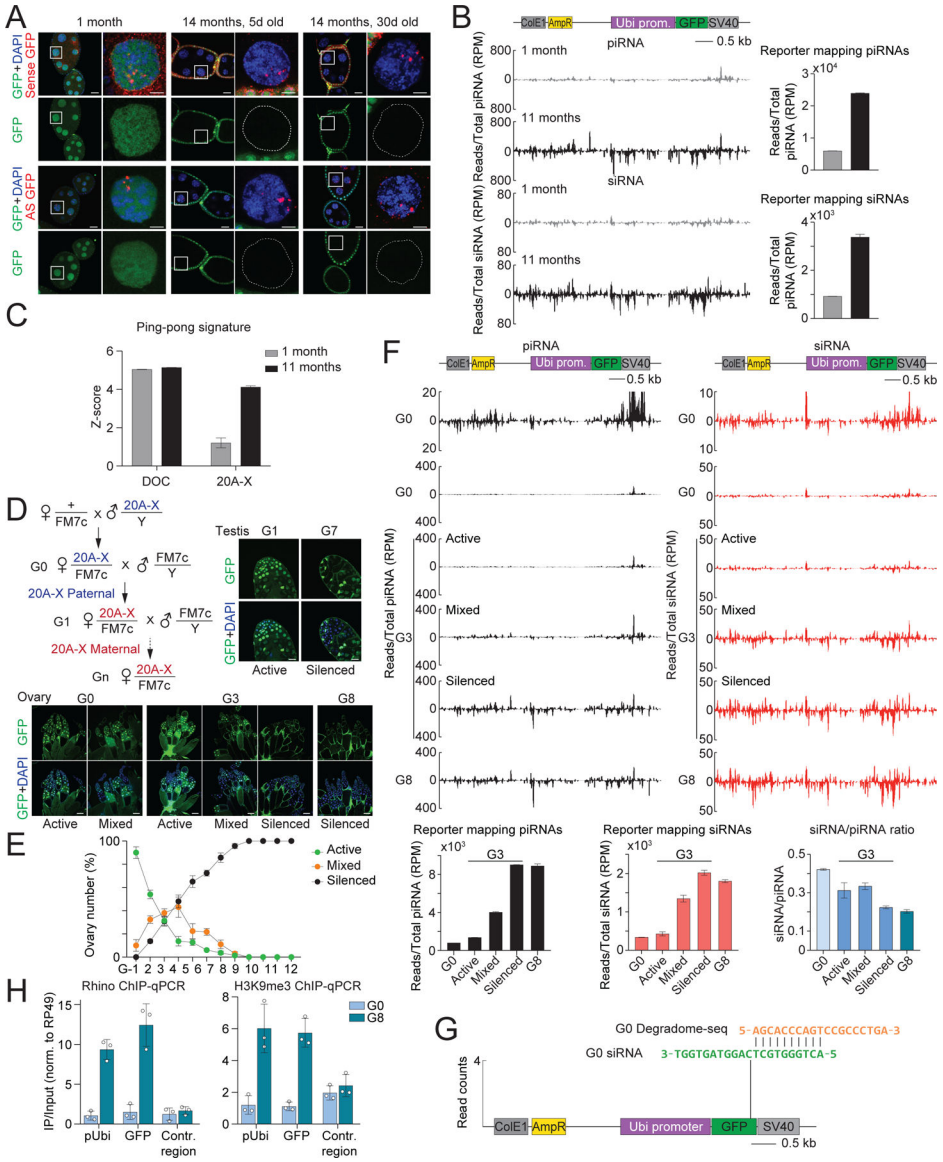


Figure 5. piRNA cluster is established over several generations.

(A) Establishment of 20A-X repression over several generations. Shown are GFP protein and RNA expression in 20A-X flies 1 and 14 months after establishment of the line.

(B) 20A-X piRNA level increases over several generations. Shown are ovarian reporter-mapping piRNA and siRNA profiles 1 and 11 months after establishing the 20A-X line. Error bars indicate standard deviation of two biological replicates.

(C) Ping-pong signature of 20A-X-mapping piRNAs increases over several generations. Shown are Z-scores indicating ping-pong signature of 20A-X piRNAs. Z-scores of DOC transposon piRNA are shown for comparison. Error bars indicate standard deviation of two biological replicates.

(D) Recovery of 20A-X repression after paternal transmission. Scheme of crosses to monitor 20A-X after its paternal transmission (left). After paternal transmission in the first cross (G0), 20A-X is inherited maternally in each subsequent generation (G1-G8).

(E) Accumulation of 20A-X repression over several generations. In each fly germline GFP expression was determined and assigned one of three values: ‘silenced’ indicates complete lack of expression, ‘active’ indicates expression in the majority of germline nuclei, while ‘mixed’ indicates variable expression between individual egg chambers. Plotted is the fraction of ovaries with corresponding expression pattern in each generation. Error bars indicate the standard deviation of three biological replicates.

(F) Accumulation of 20A-X piRNAs over several generations. Shown are profiles of 20A-X piRNAs and siRNAs in different generations. In G3, ovaries were separated in three groups according to GFP expression as described in (E). Error bars indicate standard deviation of two biological replicates.

(G) siRNAs recognize and cleave reporter transcripts in the G0 generation. Degradome-seq and siRNA data were analyzed to search for cleavage sites. One example of a 10 nt overlap between siRNAs and a cleavage product was detected.

(H) Accumulation of Rhino and H3K9me3 on 20A-X chromatin. Rhino and H3K9me3 levels on chromatin were measured by ChIP-qPCR in ovaries of G0 and G8 generation. Error bars indicate the standard deviation of three biological replicates.

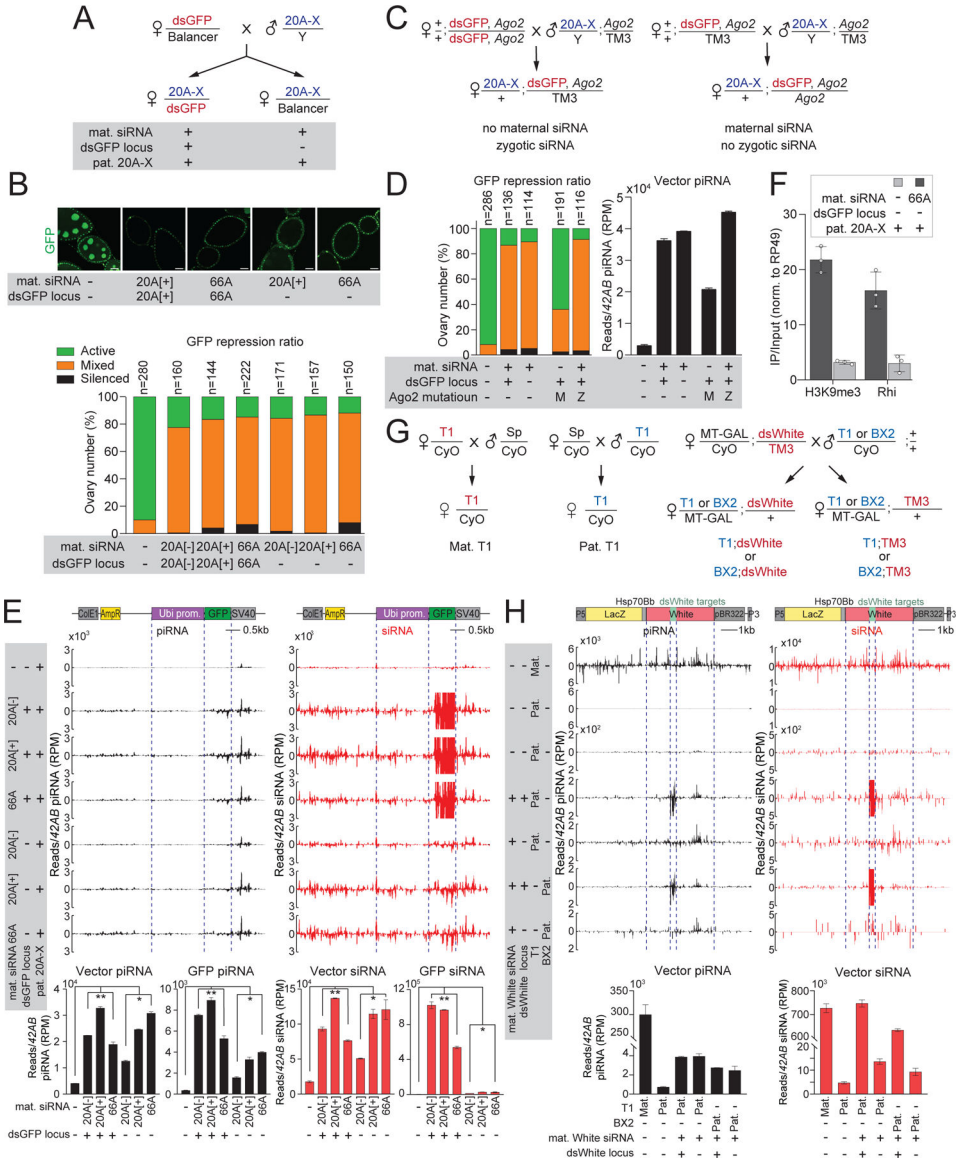


Figure 6. Cytoplasmic inheritance of siRNAs activates piRNA biogenesis in the progeny

(A) Crossing scheme to test the role of siRNAs in triggering 20A-X repression. Progeny that inherited the dsGFP locus and those that did not were compared.

(B) Cytoplasmic inheritance of siRNAs triggers 20A-X repression in the progeny. GFP expression in progenies of each genotype was assessed by fluorescent microscopy (top) and assigned one of the three values as described in Fig. 5E (bottom). N indicates the number of ovaries analyzed for each genotype.

(C) Crossing scheme to test the role of Ago2 in 20A-X repression. Crosses are similar to crosses shown in (A) except of the presence *Ago2* mutation either in mothers or the progeny.

(D) Triggering of 20A-X repression by trans-generational siRNAs depends on a functional siRNA pathway in the mothers, but not in the progeny. Analysis of GFP expression in progenies of crosses shown in (C).

(E) Cytoplasmic inheritance of siRNAs activates piRNA production in the progeny. Shown are profiles of piRNAs and siRNAs in progenies of crosses shown in (A). Error bars indicate standard deviation of two biological replicates. Statistical significance is estimated by two-tailed Student's t-test; *p<0.05, **p<0.01.

(F) Cytoplasmic siRNA inheritance is required for accumulation of H3K9me3 and for Rhino recruitment. Rhino and H3K9me3 levels were measured by ChIP-qPCR in progenies of crosses shown in (A). Error bars indicate the standard deviation of three biological replicates.

(G) Crossing scheme to test the role of siRNAs in triggering T1/BX2 piRNA cluster formation.

(H) Inheritance of *White* siRNAs triggers T1/BX2 piRNA production in the progeny. Shown are profiles of piRNAs and siRNAs in progenies from crosses shown in (G). The green boxes indicate the dsWhite target region. Error bars indicate standard deviation of two biological replicates.

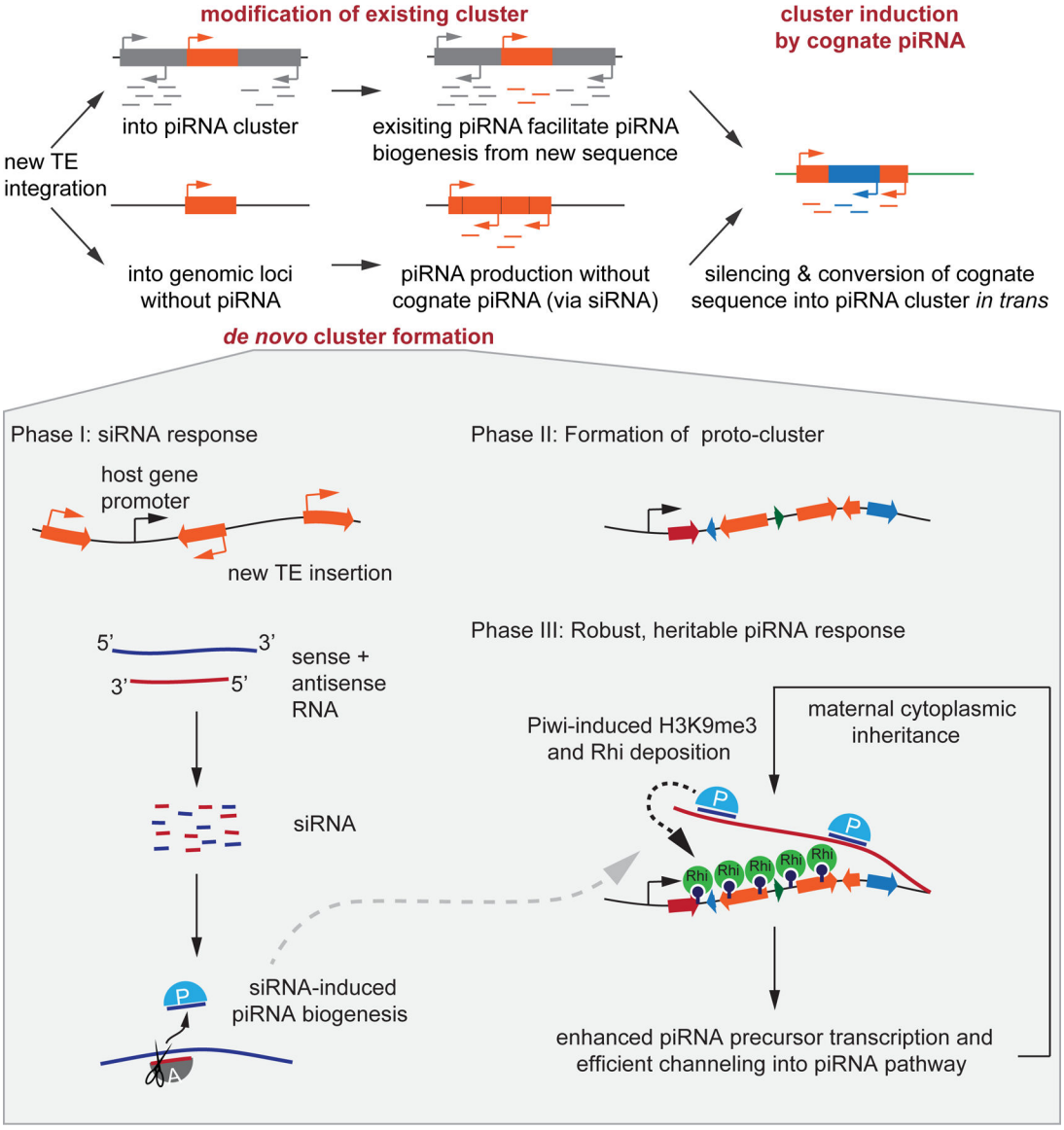


Figure 7. A model for siRNA-triggered activation of piRNA immunity

Top: piRNA production to a new TE invader can be induced by different mechanisms. TEs can integrate into existing clusters. Alternatively, new clusters can arise from TE insertions *de novo* without pre-existing piRNAs. Bottom: Proposed model for siRNA-induced *de novo* piRNA cluster formation. Newly inserted transposable elements generate double-stranded RNA leading to siRNA production. Ago2-bound siRNAs can then recognize and cleave targets, generating the 5' end of RNAs that are processed into piRNAs (left). Transposition events can lead to genomic rearrangements and formation of proto-clusters. Piwi-piRNA complexes recognize cognate sequences in the proto-cluster, leading to the recruitment and deposition of Rhino and conversion into a *bona fide* cluster that can sustain piRNA production in an siRNA-independent manner (right).

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal to Histone H3 (tri methyl K9)	Abcam	Cat# ab8898; RRID:AB_306848
Rabbit polyclonal to Rhino	Mohn et al.[18]	N/A
Bacterial and virus strains		
Biological samples		
Chemicals, peptides, and recombinant proteins		
TRIzol™ Reagent	Thermo Fisher Scientific	Cat# 15596018
16% Formaldehyde (w/v), Methanol-free	Thermo Fisher Scientific	Cat# 28908
Poly-D-Lysine	Thermo Fisher Scientific	Cat# A3890401
Formamide	Thermo Fisher Scientific	Cat# 17899
Dextran sulfate sodium salt	MilliporeSigma	Cat# D8906-5G
Denhardt's Solution	Thermo Fisher Scientific	Cat# 750018
Heparin sodium salt	MilliporeSigma	Cat# H3149-10KU
Triton X-100	MilliporeSigma	Cat# X100-100ML
TWEEN 20	MilliporeSigma	Cat# P1379-100ML
NP-40 Surfact-Amps Detergent Solution	Thermo Fisher Scientific	Cat# 85124
Sodium deoxycholate	MilliporeSigma	Cat# D6750-10G
Sodium dodecyl sulfate	MilliporeSigma	Cat# L3771-100G
Sodium fluoride	MilliporeSigma	Cat# 201154-5G
Sodium orthovanadate	MilliporeSigma	Cat# S6508-10G
Salmon Sperm DNA Solution	Thermo Fisher Scientific	Cat# 15632011
EDTA-free Protease Inhibitor Cocktail	MilliporeSigma	Cat# 11873580001
Antifade Mounting Medium	VECTASHIELD	Cat# H-1000-10
Critical commercial assays		

REAGENT or RESOURCE	SOURCE	IDENTIFIER
FISH Tag DNA Green Kit, with Alexa Fluor 488 dye	Thermo Fisher Scientific	Cat# F32947
FISH Tag DNA Red Kit, with Alexa Fluor 594 dye	Thermo Fisher Scientific	Cat# F32949
SuperScript III Reverse Transcriptase	Thermo Fisher Scientific	Cat# 18080093
NEBNext Small RNA Library Prep Set for Illumina	New England Biolabs	Cat# E7330S
NEBNext ChIP-seq Sample Prep Master Mix Set	New England Biolabs	Cat# E6240
Deposited data		
Original images for Figures	This paper	Mendeley Data: DOI: 10.17632/d65h9pttby.1
Raw and analyzed sequence data	This paper	GSE193091
Scripts and code	This paper	DOI:10.5281/zenodo.8351702
Experimental models: Cell lines		
Experimental models: Organisms/strains		
D. melanogaster integration strain: pUbi > eGFP-NLS 20A[−]	This paper	N/A
D. melanogaster integration strain: pUbi > eGFP-NLS 20A[+]	This paper	N/A
D. melanogaster integration strain: pUbi > eGFP-NLS 20A-X	This paper	N/A
D. melanogaster integration strain: pUbi > eGFP-NLS 42AB[−]	This paper	N/A
D. melanogaster integration strain: pUbi > eGFP-NLS 66A6	This paper	N/A
D. melanogaster integration strain: pUbi > eGFP(sense)-NLS-GFP(antisense) 20A[−]	This paper	N/A
D. melanogaster integration strain: pUbi > eGFP(sense)-NLS-GFP(antisense) 20A[+]	This paper	N/A
D. melanogaster integration strain: pUbi > eGFP(sense)-NLS-GFP(antisense) 66A6	This paper	N/A
D. melanogaster: y[1] w[*]; Mi{y[+mDint2]=MIC}MI07308	Bloomington Drosophila Stock Center	BDSC #43121
D. melanogaster: y[1] w[*] Mi{y[+mDint2]=MIC}MI08972	Bloomington Drosophila Stock Center	BDSC #50496
D. melanogaster: y[1] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.HMS00017}attP2	Bloomington Drosophila Stock Center	BDSC #33623
D. melanogaster: P{y[+t7.7] v[+t1.8]=TRiP.GL00041}attP2	Bloomington Drosophila Stock Center	BDSC #35171
D. melanogaster: y[1] sc[*] v[1] sev[21]; P{y[+t7.7] v[+t1.8]=TRiP.HMS00606}attP2	Bloomington Drosophila Stock Center	BDSC #33724
D. melanogaster: w; pW22>zuc_sh1[attp40]/CyO; Sb/TM3,Ser	Vienna Drosophila Resource Center	#313693
D. melanogaster: w; ; pGLKD>mael_sh1[attP2]/TM3,Sb;	Vienna Drosophila Resource Center	#313154

REAGENT or RESOURCE	SOURCE	IDENTIFIER
D. melanogaster: UASt > double strand White	Vienna Drosophila Resource Center	#30033
D. melanogaster: 20A-X deletion gRNA	This paper	N/A
D. melanogaster: Δ20A-X	This paper	N/A
D. melanogaster: 20A promoter gRNA	This paper	N/A
D. melanogaster: Δ20A promoter	This paper	N/A
D. melanogaster: Δ20A promoter 20A-X	This paper	N/A
D. melanogaster: attP40{nos-Cas9}	NIG-FLY	CAS-0001
D. melanogaster: w[*]; P{w[+mC]=matalpha4-GAL-VP16}V37	Bloomington Drosophila Stock Center	BDSC #7063
D. melanogaster: w[1118]; P{w[+mC]=GAL4::VP16-nanos.UTR}CG6325[MVD1]	Bloomington Drosophila Stock Center	BDSC #4937
D. melanogaster: w[*]; AGO2[414]	Kyoto Stock Center	#109027
D. melanogaster: Rhi[2]	Klattenhoff et al.[22]	N/A
D. melanogaster: Rhi[KG]	Klattenhoff et al.[22]	N/A
D. melanogaster: w[*]; mael[r20]/TM3, Sb[1]	Bloomington Drosophila Stock Center	BDSC #8516
D. melanogaster: Mael[391]	Gift of Hannon lab	N/A
D. melanogaster: T1/CyO	de Vanssay et al.[27]	N/A
D. melanogaster: BX2/CyO	de Vanssay et al.[27]	N/A
Oligonucleotides		
20A-X upstream qPCR F GCCAGTAGTCGTCTCTCATTTATGC	IDT	N/A
20A-X upstream qPCR R GCTGAAGCACTTGATTGCCAAC	IDT	N/A
20A-X downstream qPCR F GCTTCCCATAAACCTCCCATGTG	IDT	N/A
20A-X downstream qPCR R TCGTGGGAGCTTCAAGAGTATTGG	IDT	N/A
20A qPCR F GCCTACGCAGAGGCCTAAGT	IDT	N/A
20A qPCR R CAGATGTGGTCCAGTTGTGC	IDT	N/A
20A-X genomic A ATGAGTTCAATTCGCTACTGCGAG	IDT	N/A
20A-X genomic B ACTTCAACAGGAGCATACCGCTAC	IDT	N/A
20A-X genomic C GTGCTTTCCCCGTGTGTGG	IDT	N/A
20A-X genomic D CCGACAACCACTACCTGAGC	IDT	N/A
20A-X deletion A GCCAGTAGTCGTCTCTCATTTATGC (Same as 20A-X upstream qPCR F)	IDT	N/A
20A-X deletion B GACCGTTCCAGATTCGCTGC	IDT	N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pUbi qPCR F TGCATTTCAAGGTCTTTGTTCGG	IDT	N/A
pUbi qPCR R GCGAAAATCAACACGCAAGTTTT	IDT	N/A
GFP qPCR F TACAACAGCCACAAGGTCTATATCA	IDT	N/A
GFP qPCR R GGTGTTCTGCTGGTAGTGGTC	IDT	N/A
SV40 qPCR F TGGTGGAATGCCTTTAATGAGGA	IDT	N/A
SV40 qPCR R CCTTGGGGTCTTCTACCTTTCTC	IDT	N/A
ColE qPCR F AACTATCGTCTTGAGTCCAACCC	IDT	N/A
ColE qPCR R GTAGTTAGGCCACCACTTCAAGA	IDT	N/A
AmpR qPCR F ACGATCAAGGCGAGTTACATGAT	IDT	N/A
AmpR qPCR R TACGGATGGCATGACAGTAAGAG	IDT	N/A
Non-gene qPCR F CCCCATTTCCAGACGAGTCC	IDT	N/A
Non-gene qPCR R TGACGGCAATAAGGATGCGA	IDT	N/A
White qPCR 1 F CACAATATGGACATCTTTGGGGC	IDT	N/A
White qPCR 1 R CTGCGAATAGAAACTCACCGTTC	IDT	N/A
White qPCR 2 F GAACGGTGAGTTTCTATTCGCAG	IDT	N/A
White qPCR 2 R GATCGAAAGGCAAGGGCATTC	IDT	N/A
rp49 qPCR F CCGCTTCAAGGGACAGTATCTG	IDT	N/A
rp49 qPCR R ATCTCGCCGCAGTAAACGC	IDT	N/A
GFP sense probes	Molecular Technologies	3066/A160
GFP antisense probes	Molecular Technologies	4017/D577
GFP sense probes (co-FISH)	Molecular Technologies	3743/D811
GFP antisense probes (co-FISH)	Molecular Technologies	3743/D813
Recombinant DNA		
BAC construct: 20A	BACPAC Resources	#CH322-184J4
pUbi-GFP-NLS-SV40	This paper	N/A
Software and algorithms		

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ZEN lite	Zeiss	https://www.zeiss.com/microscopy/en/products/software/zeisszen-lite.html
MATLAB	MathWorks	https://www.mathworks.com/products/matlab.html
ImageJ	National Institutes of Health	https://imagej.net/downloads
Other		