

Maternally inherited siRNAs initiate piRNA cluster formation

Thesis by
Yicheng Luo

In Partial Fulfillment of the Requirements for
the Degree of
Doctor of Philosophy

The logo for the California Institute of Technology (Caltech), featuring the word "Caltech" in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
(Defended September 29, 2022)

© 2023

Yicheng Luo
ORCID: 0000-0003-3704-2389

All rights reserved

ACKNOWLEDGEMENTS

I feel so grateful in my research life since I have met so many talented and kind friends and advisors who are willing to help me and share their experiences. Looking back on my Ph.D. life, it is like a wonderful journey full of happiness, frustration, and appreciation.

First of all, I would like to thank my Ph.D. advisor Dr. Alexei Aravin. His patience, suggestions, and help supported me during my entire Ph.D. With his guidance, I have learned how to think about scientific questions and how to design experiments to answer these questions. He always encourages me to think independently, which is the most important lesson I have learned during my Ph.D. He is also a good friend with whom I can chat when I am frustrated. I also want to thank Dr. Katalin Fejes Tóth, who taught me how to visualize the data and always gives me very valuable ideas to make my projects better. I have also learned a lot from my lab mates and gotten many bits of help from them. It is a lot of fun to work with them. Thank all of you.

I also appreciate my committee, Dr. Mitchell Guttman, Dr. Viviana Gradinaru, and Dr. Shu-ou Shan for their valuable suggestions and help. Their suggestions helped me better understand my projects and think about science from the other side, as well as helped me to plan my career.

I want to express my appreciation to Dr. Jinbiao Ma, professor at Fudan University of China, who gave me the opportunity to do research in the US and shared his experiences with me. Without his help, I may not be here.

I also want to thank Dr. Yang Yu, professor at Chinese Academy of Sciences. He was my first mentor when I was in the Hannon Lab at Cold Spring Harbor Laboratory and taught me many experimental designs and techniques. I was so lucky that I had an experienced and talented postdoc who was willing to spend time teaching and guiding me on how to do research. More importantly, he always tried newly developed techniques, which gave me the chance to learn and know advanced techniques.

It was a wonderful time when I was a visiting student in the Hannon Lab. I have learned so much and gotten lots of help from all Hannon lab members. Especially, I want to thank Dr. Gregory Hannon who allowed me to do research in his lab. It was a valuable experience to see and learn how to be a great scientist. Even though I was just a visiting student, Dr. Gregory Hannon still showed his respect and was willing to help me to continue my study in the US. I deeply appreciate Dr. Gregory Hannon's help and recommendation. Without him, I could not have gone this far.

Good friends are necessary to make life successful and full of fun. I am so lucky that I have so many of them. I want to thank Dr. Haohong Li, Dr. Gang Cao, Dr. Junwei Shi, Dr. Bin Zhang, Dr. Yuan Hao, Dr. Minghui Wang, Dr. Qingyu Wu, Dr. Hongtao Qin, Dr. Zhiming Zhang, et al. They were so nice to me when I just started my journey in the US. They helped me to adapt to living in the US. They also kindly shared their experiences with me and helped me solve my problems. I also want to thank my friends during my Ph.D. I want to apologize for not listing all their names here. When I feel stressed, chatting with them can truly make me feel relaxed. They are so talented, and I have seen and learned so much from all of my friends.

Finally, I want to deeply thank my parents, Shuying Ma and Yunguang Luo. They gave me everything to support me since I was born. Their patience and love are always with me. As an overseas student, I feel so sorry I cannot be with them, especially when they need help. I feel helpless and sad watching them get old, and I wish I could spend more time being with them in the future. To my wife, Shiyu Gu, I have nothing else to say but thank you. Thank you for always being by my side for 16 years, thank you for your understanding of my busy work, thank you for sharing the happiness and staying together through all the darkness, and thank you for giving me a home. We have been together since we were high school students in 2006. Since then we have been separated in two cities and two countries for 10 years, and

the stories between us are just like loving movies. Thanks for trusting me and being by my side. My lovely daughter, Mia Anxuan Luo, thanks for letting us be your parents and bringing happiness to us. You really light up my life and motivate me. It is probably the best job to be your father. Thanks to my family, I can have a chance to be a boy, a man, and a father.

In the past three years, everyone has experienced an extraordinary life because of COVID-19. This is a tough period for everyone. I want to thank everyone who helped me get through this difficult time, as well as the people who fight the virus. I wish people could understand each other and work together to overcome all darkness in the future.

ABSTRACT

PIWI-interacting RNAs (piRNAs) guide repression of transposable elements in germlines of animals to protect genome integrity. In *Drosophila*, the majority of piRNAs are produced from heterochromatic genomic loci, called piRNA clusters, that act as repositories of information about genome invaders. piRNA generation by dual-strand clusters depends on the chromatin-bound Rhino-Deadlock-Cutoff (RDC) complex, a complex specifically enriched on a dual-strand cluster, which is deposited on clusters guided by piRNAs, forming a feed-forward loop in which piRNAs promote their own biogenesis. However, this rises a fundamental question about how piRNA clusters are formed initially before cognate piRNAs are present.

Here we report the spontaneous de novo formation of a Rhino-dependent piRNA cluster from repetitive transgenic sequences. We show that cluster formation occurs gradually over several generations and requires continuous trans-generational transmission of small RNAs from mothers to their progeny. Importantly, we discovered that maternally-supplied siRNAs are responsible for triggering de novo cluster activation in progeny. In contrast, the siRNA pathway is dispensable for piRNA cluster function and maintenance after its establishment. These results revealed an unexpected cross-talk between the siRNA and piRNA pathways

and suggested a mechanism for de novo formation of piRNA clusters triggered by production of siRNAs.

PUBLISHED CONTENT AND CONTRIBUTIONS

Luo, Y., He, P., Kanrar, N., Toth, K.F., Aravin, A.A. Maternally inherited siRNAs initiate piRNA cluster formation. bioRxiv 2022. doi: <https://doi.org/10.1101/2022.02.08.479612>

Y.L. participated in the conception of the project, performed experiments, prepared the data, and participated in the writing of the manuscript.

TABLE OF CONTENTS

Acknowledgements.....	iii
Abstract	vii
Published Content and Contributions.....	ix
Table of Contents.....	x
List of Illustrations and Tables.....	xi
Nomenclature.....	xii
Chapter I: Introduction	1
Chapter II: Maternally inherited siRNAs initiate piRNA cluster formation	
2.1 Reporters inserted in the dual-strand 42AB cluster are repressed by piRNA, while reporters in the uni-strand 20A cluster disrupt cluster expression.....	8
2.2 An unusual reporter behaves like a bona fide dual-strand piRNA cluster that is enriched in Rhi and depends on it to generate piRNA.....	14
2.3 The repetitive organization of the 20A-X locus correlates with its function as dual-strand piRNA cluster.....	18
2.4 piRNA-induced repression of the 20A-X reporter depends on maternal transmission of cognate piRNAs	27
2.5 piRNA cluster is established over several generations.....	31
2.6 Maternal siRNA triggers activation of piRNA cluster in the progeny	38
Chapter III: Discussion	
3.1 siRNAs can provide initial trigger to activate piRNA biogenesis	43
3.2 Genomic requirements for piRNA cluster function	47
Chapter IV: Materials and methods	50
Bibliography	64

LIST OF ILLUSTRATIONS AND TABLES

<i>Figures</i>	<i>Page</i>
2.1 Reporters integrated in uni- and dual-strand clusters have different expression and effects on cluster activity.....	12
2.2 An unusual reporter insertion in the 20A cluster is a dual-stand piRNA cluster that generates piRNAs in a Rhi-dependent manner.....	17
2.3 The 20A-X locus contains rearranged multi-copy reporter sequences.....	23
S2.3 Analysis of 20A-X locus organization.....	26
2.4 piRNA-induced repression of the 20A-X reporter depends on maternal transmission of cognate piRNAs.....	30
2.5 piRNA cluster is established over several generations.....	35
S2.5 Analysis of 20A locus over generations	37
2.6 Cytoplasmic inheritance of siRNAs activates piRNA biogenesis in the progeny.....	41
3.1 The model for siRNA-triggered activation of piRNA immunity.....	47
Table 1. <i>Drosophila melanogaster</i> stocks.....	63
Table 2. Primers and probes.....	64

NOMENCLATURE

ChIP. Chromatin immunoprecipitation

CRISPR. Clustered regularly interspaced short palindromic repeats

FISH. Fluorescence *in situ* hybridization

GFP. Green fluorescent protein

GLKD. Germline knock down

H3K9me3. Histone 3 lysine 9 trimethylation

HCR. Hybridization chain reaction

HP1. Heterochromatin protein 1

LINE. Long interspersed element

LTR. Long terminal repeat

miRNA. MicroRNA

MiMIC. *Minos*-mediated integration cassette

NLS. Nuclear localization signal

PCR. Polymerase chain reaction

piRNA. Piwi-interacting RNA

PolII. RNA polymerase II

RDC. Rhino-Deadlock-Cutoff

Rhi. Rhino

RMCE. Recombinase-mediated cassette exchange

RPM. Reads per million

RT-qPCR. Quantitative reverse transcription PCR

SDS. Sodium dodecyl sulfate

SINE. Short interspersed element

siRNA. Small interfering RNA

SSC. Saline-sodium citrate

TEs. Transposable elements

TREX. Transcription-export

Ubi. Ubiquitin

UTR. Untranslated region

Zuc. Zucchini

Chapter 1

INTRODUCTION

Transposable elements (TEs) were first discovered by Barbara McClintock in maize in 1940s (Mc, 1950), thereafter TEs have been extensively studied and many interesting phenomena have been found. TEs are ubiquitously and massively existing in most animal species and occupy a substantial portion of genome, for example, 60-70% of maize genome is comprised of LTR retrotransposons, a class of transposons; about 46% of human genome is from TEs (Lander et al., 2001; Warren et al., 2008). However, for decades after Barbara McClintock found TEs, most scientists dismissed TEs' importance and considered them as so called "junk" DNA. With the emergence of a number of evidences that prove the importance of TEs' biological functions, eventually, this thought has been changed. For instance, TEs can modify transcriptional network and regulate gene expression by dispersing a number of promoters and enhancers (Bejerano et al., 2006; Bourque et al., 2008; Chuong et al., 2013; Trizzino et al., 2017), expanding transcription factor binding sites (Bourque et al., 2008; Ito et al., 2017; Sun et al., 2018; Sundaram et al., 2014), and providing insulator sequences (Lunyak et al., 2007; Schmidt et al., 2012; Wang et al., 2015), and repressive elements (Lippman et al., 2004; Ninova et al., 2020b; Rebollo et al., 2011). More importantly,

TEs are considered to drive the evolution of the genome because their transpositions can potentially alter genome structure. (Bourque et al., 2008; Chen et al., 2022; Cordaux and Batzer, 2009; Feschotte, 2008; Khodosevich et al., 2002; Li et al., 2022; Mills et al., 2006; Warren et al., 2008). Although TEs are important and necessary in many aspects of biological functions, in fact, TEs originate from viruses and are able to self-copy and jump from one position to the other across the genome. Because of this, uncontrolled TEs are probably a disaster to their host. Derepression of TEs may disrupt and interfere with the transcription of the nearby genes (Daniel et al., 2015; Elbarbary et al., 2016); TEs encoded proteins such as the endonuclease can cause DNA breaks and genomic instability (Hedges and Deininger, 2007); accumulation of TEs' RNA transcripts and vast TEs derived genomic sequences can trigger an innate immune response leading to autoimmune diseases and sterile inflammation (Kassiotis and Stoye, 2016; Roulois et al., 2015). Based on current studies, more than 120 independent TEs are associated with human diseases (Chen et al., 2006; Hancks and Kazazian, 2016) such as cancers (Burns, 2017; Helman et al., 2014; Lee et al., 2012; Nguyen et al., 2018; Rodic et al., 2015; Scott et al., 2016; Shukla et al., 2013; Tang et al., 2017); moreover, the frequency of de novo germline transposition in humans is approximately one in 21 births for Alu (SINE) (Xing et al., 2009) and one in 95 births for L1 (LINE) (Ewing and Kazazian, 2010). These two are the main active TEs in human genome (Beck et al., 2010;

Brouha et al., 2003; Dewannieux et al., 2003; Ullu and Tschudi, 1984). In order to antagonize the harmfulness of uncontrolled TEs' transpositions, several small RNA based pathways have been adapted to compromising the active TEs.

Despite the common architecture of Argonautes-small RNA complexes, there are three distinct classes of small RNAs-siRNA, miRNA and piRNA-that differ in their biogenesis, functions as well as the specific members of the Ago family they partner with. However, all small RNAs share a similar mechanism which guide the Argonautes (Ago) to complementary RNA targets upon which Agos either cleave the target or recruit additional factors to repress them by other mechanisms. RNA interference was first discovered by Andrew Fire and Craig C. Mello in 1998 (Fire et al., 1998), henceforth, RNAi has been studied and used extensively in biological research and disease therapeutic treatments. RNAi is a biological technique where double stranded RNA induces gene silencing by targeting complementary RNA and leading to RNA degradation. Importantly, later on scientists found that in many species cells can produce endogenous siRNA from genomic loci (Czech et al., 2008; Ghildiyal et al., 2008; Hamilton and Baulcombe, 1999; Kawamura et al., 2008; Tam et al., 2008; Tang et al., 2003; Vagin et al., 2006). Similar to RNAi processing, endogenous siRNAs are transcribed from genomic loci where bearing double-stranded RNA or hairpin precursors, then cleaved by the Dicer nuclease and loaded into their Ago protein partner (Bernstein et al., 2001; Matranga et

al., 2005; Rand et al., 2005). siRNA/Ago complexes cleave complementary RNA targets in the cytoplasm leading to their degradation (Matranga et al., 2005; Rand et al., 2005).

Both siRNA and piRNA were reported to suppress activity of endogenous (transposable elements and other selfish genes) and exogenous (viruses) elements in various animal species (Aravin et al., 2006; Brennecke et al., 2007; Gammon and Mello, 2015; Gitlin et al., 2002; Lindbo et al., 1993; Vance and Vaucheret, 2001; Voinnet et al., 1999), however, the two pathways are believed to work independently of each other. Despite the complete abrogation of the siRNA pathway in Ago2 deficient flies, these flies are viable and fertile and show only mild activation of a few transposons in somatic tissues (Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Pelisson et al., 2007). In contrast, flies deficient for piRNA pathway components demonstrate strong activation of multiple TE families in the germline associated with DNA damage and complete sterility (Brennecke et al., 2007; Sienski et al., 2015; Vagin et al., 2006; Yu et al., 2015b).

In *Drosophila*, siRNAs associate exclusively with Ago2 and if Ago2 is mutated will abrogate siRNA-guided repression (Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura et al., 2004; Vagin et al., 2006). The biogenesis and function of piRNAs, however, is much more complex than that of siRNAs. piRNA processing is independent of Dicer but involves multiple other proteins (Andersen et al., 2017; Chen et al., 2016;

ElMaghraby et al., 2019; Mohn et al., 2014; Vagin et al., 2006; Zhang et al., 2014). piRNAs are expressed in the germline and closely associated somatic cells of the ovary and testis and are loaded into a distinct clade of Argonautes called Piwi proteins, which in flies consist of Piwi, Aub, and Ago3 (Brennecke et al., 2007; Saito et al., 2006; Vagin et al., 2006). Similarly to siRNAs, cytoplasmic Piwi/piRNA complexes cleave complementary RNA targets, however, in addition to target degradation, this process also amplifies piRNAs through the so-called ping-pong cycle (Aravin et al., 2008; Brennecke et al., 2007; Gunawardane et al., 2007), thereby connecting their function to their own biogenesis. Furthermore, both flies and mice harbor a nuclear Piwi/piRNA complex capable of transcriptional repression through recognition of nascent transcripts followed by recruitment of a chromatin modifying machinery (Aravin et al., 2008; Batki et al., 2019; Fabry et al., 2019; Iwasaki et al., 2021; Le Thomas et al., 2013; Ninova et al., 2020a; Sienski et al., 2015; Sienski et al., 2012; Yu et al., 2015b; Zhao et al., 2019).

siRNA precursors are recognized by their double-stranded nature. Similarly to siRNAs, piRNAs are also processed from longer RNA precursors that come from centromere and telomere regions called piRNA cluster (Brennecke et al., 2007), however these transcripts are single-stranded and lack distinct sequence and structure motifs, raising the question of how they are recognized and channeled into the processing machinery. In *Drosophila*, the

chromatin-bound Rhino-Deadlock-Cutoff (RDC) protein complex marks dual-strand piRNA clusters (Klattenhoff et al., 2009), genomic regions that generate the majority of piRNAs in the germline. RDC complex is required for transcription of piRNA precursors by promoting initiation (Andersen et al., 2017; Klattenhoff et al., 2009; Mohn et al., 2014) and suppressing splicing (Zhang et al., 2014) and premature termination (Chen et al., 2016; Zhang et al., 2014). RDC complex also promotes loading of the RNA-binding TREX (Hur et al., 2016; Zhang et al., 2018) and Nxf3-Nxt1 (ElMaghraby et al., 2019; Kneuss et al., 2019) RNA export complexes on pre-piRNAs. Loading of Nxf3-Nxt1 is proposed to channel precursors to the cytoplasmic processing machinery (ElMaghraby et al., 2019; Kneuss et al., 2019). Thus, RDC complex binding to chromatin of piRNA clusters might be both necessary and sufficient to sustain piRNA biogenesis.

Though the process of RDC complex deposition on chromatin is not completely understood, it seems to be guided by piRNAs, as the nuclear Piwi/piRNA complex directs deposition of the H3K9me_{2/3} mark, which in turn provides an anchor for RDC complex binding (Akkouche et al., 2017; Mohn et al., 2014; Sienski et al., 2015; Yu et al., 2015b). Several studies demonstrate the critical role of cytoplasmic piRNA inheritance from the mother to the progeny in initiating piRNA production (Casier et al., 2019; de Vanssay et al., 2012; Le Thomas et al., 2014b). Together these findings suggest that piRNA biogenesis is

governed by a trans-generational feed-forward loop in which piRNA biogenesis is promoted by RDC complex, which in turn is deposited on chromatin guided by cytoplasmically inherited piRNAs. This feed-forward 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. How novel piRNA clusters arise is poorly understood.

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.

*Chapter 2***MATERNALLY INHERITED SIRNAS INITIATE PIIRNA CLUSTER FORMATION****2.1 Reporters inserted in the dual-strand 42AB cluster are repressed by piRNAs, 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. 2.1A). Using RMCE we inserted reporters into the major dual-strand cluster 42AB and the uni-strand cluster 20A (Fig. 2.1B). 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.2.1C), as well as in other somatic tissues, indicating that the transgenes are functional and that integration into the repeat-rich cluster environments is compatible with reporter expression in somatic cells. Flies with reporters in the uni-strand 20A cluster,

inserted in either orientation, 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 fluorescent in situ hybridization revealed that both strands of the 42AB reporter sequence were transcribed, GFP protein was not expressed from the 42AB cluster insertion in the germline. Similar to native 42AB cluster transcripts, RNA transcribed from both strands of the 42AB reporter was concentrated in the nucleus.

The exclusive repression of the 42AB reporter in the germline, where the 42AB cluster is active, suggest 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. 2.1D). The 42AB reporter-derived piRNAs have the expected bias (69.95%) for U 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. 2.1E), 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 (Ipsaro et al., 2012; Nishimasu et al., 2012), in the germline. Depletion of Zuc led to strong (26.4-fold) reduction in the level of piRNAs mapping to the 42AB reporter (Fig. 2.1E) and to its derepression (Fig.2.1F). GFP protein expression upon Zuc 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. 2.1G). 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. 2.1B, 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.

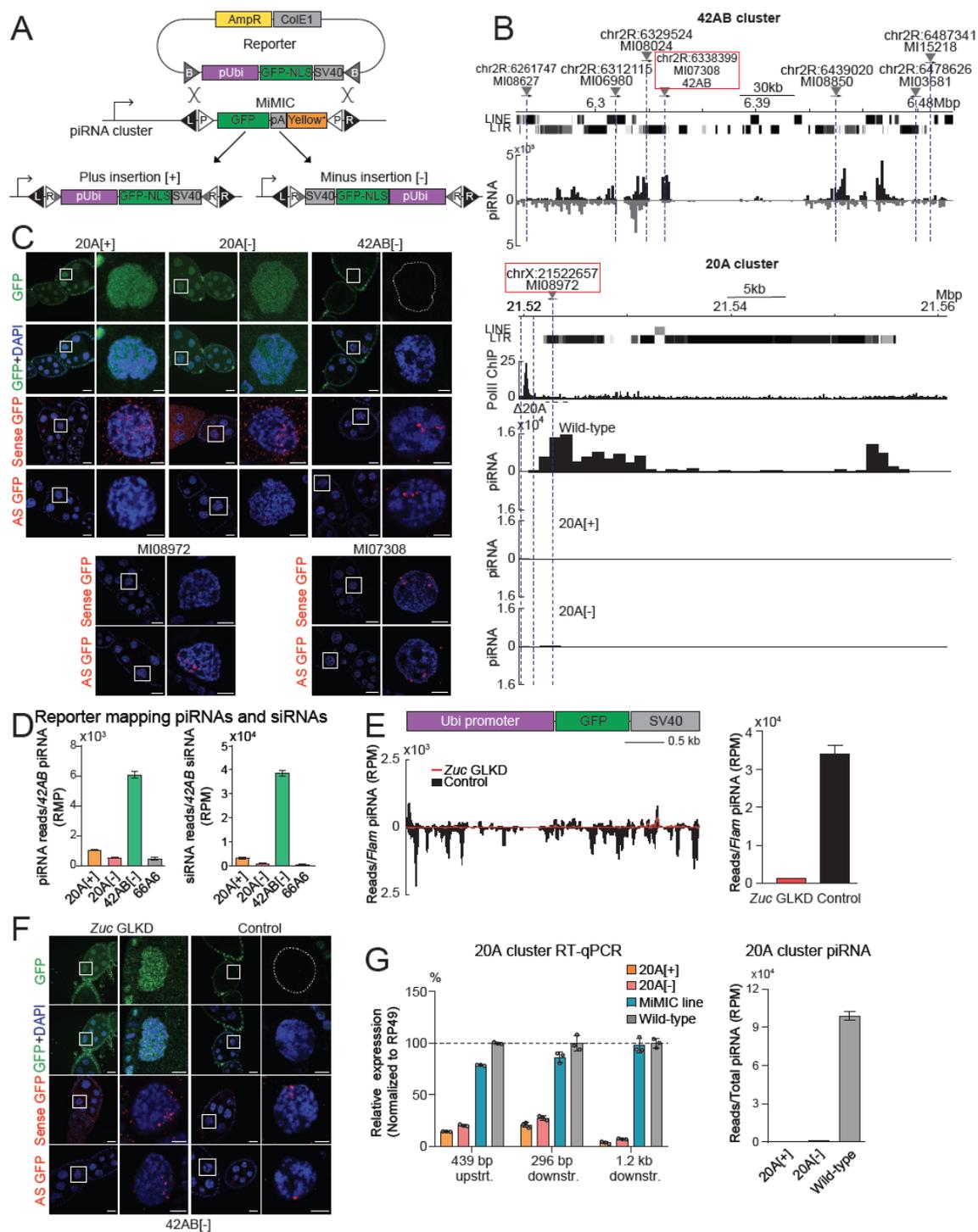


Figure 2.1 Reporters integrated in uni- and dual-strand clusters have different expression and effects on cluster activity.

(A) Scheme of reporter integration into piRNA clusters using recombinase-mediated cassette exchange (RMCE) to replace Minos-mediated integration cassettes (MiMIC). MiMICs contain GFP sequence but lack a promoter. In contrast, the reporter encodes nuclear EGFP expressed under control of the Ubiquitin gene promoter. RMCE can result in integration of the reporter in either orientation, which can be discriminated by genomic PCR.

(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. 20A cluster piRNA levels are dramatically reduced in ovaries of 20A[+] and 20A[-] flies. Number of piRNA reads mapped to the 20A cluster is normalized to total piRNA read count.

(C) Expression of reporters integrated in 20A and 42AB clusters. Expression of GFP protein and sense and antisense RNA (by RNA FISH) in ovaries of flies with reporter insertions in the 20A and 42AB clusters. In all reporter lines GFP protein is expressed in somatic follicular cells, however, GFP is not expressed in germline (nurse) cells in the 42AB reporter line. Sense GFP RNA is produced in both 20A[+] and 20A[-] flies and it is predominantly localized in the cytoplasm of nurse cells. This cytoplasmic localization is different from the nuclear localization of exclusively antisense RNA produced by the promoterless MiMIC sequence integrated in the same site (MI08972). The 42AB reporter generates transcripts from both strands and both sense and antisense RNA are localized in the nucleus, similar to transcripts produced by the promoterless MiMIC sequence integrated in the same site (MI07308). 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 (non-cluster) locus. Number of the small RNA reads mapping to the reporter were normalized to reads mapping to the 42AB cluster. Error bars indicate standard deviation of two biological replicates.

(E) Knockdown of Zuc eliminates 42AB reporter piRNA. Shown are piRNA profiles along the reporter in ovaries of control flies (white GLKD, black) and upon GLKD of Zuc GLKD driven by nos-Gal4 (red). Bar graph shows number of piRNA reads mapped to the reporter normalized to flamenco-derived piRNAs, which are not affected by Zuc knockdown. 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. Note appearance of sense GFP transcripts in the cytoplasm upon Zuc GLKD. Scale bar is 20 μ m and 2 μ m for egg chambers and single nurse cell nuclei, respectively.

(G) Ovarian expression of the 20A cluster is suppressed upon reporter integration. (Left) 20A cluster transcripts were measured by RT-qPCR using primers 439 bp upstream as well as 296 bp and 1217 bp downstream of the reporter insertion site. Cluster transcripts were normalized to rp49 mRNA. Error bars indicate the standard deviation of three biological replicates. (Right) Expression of 20A cluster piRNA. piRNA reads uniquely-mapping to 20A cluster were normalized to total piRNA read count. Error bars indicate standard deviation of two biological replicates.

2.2 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 that we maintained and in which we determined cassette orientation. Unexpectedly, we found that one of the lines with insertion into the 20A cluster has distinct properties. Unlike other lines that maintained reporter expression in both the soma and the germline, this line, which we dubbed 20A-X, lost germline expression after several months of propagation (Fig. 2.2A). 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, unlike 20A[+] and 20A[-] reporters, transcripts from 20A-X localized exclusively to the nuclei (Fig. 2.2A). Thus, unlike other 20A insertion lines and similar to insertions in 42AB, 20A-X shows normal GFP expression in follicular cells and strong GFP repression in the germline.

To test involvement of piRNAs in repression of the 20A-X reporter in the germline, we cloned and analyzed small RNA libraries from ovaries of 20A-X flies. This analysis 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. 2.2B). 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. 2.2C) 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 (Chen et al., 2016; Goriaux et al., 2014; Mohn et al., 2014). 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 (Le Thomas et al., 2014b; Mohn et al., 2014; Zhang et al., 2014). 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 is strongly enriched in both Rhi and the H3K9me3 mark (Fig. 2.2D), suggesting that it acts as a dual-strand piRNA cluster.

To test whether 20A-X activity depends on 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. 2.2E). Rhi GLKD also caused 4.2-fold reduction in 20A-X piRNA levels and released repression of GFP in germline of 20A-X flies (Fig. 2.2E). Taken together, our results indicate that 20A-X acts as a genuine dual-strand piRNA cluster that generates piRNAs in Rhi-dependent manner.

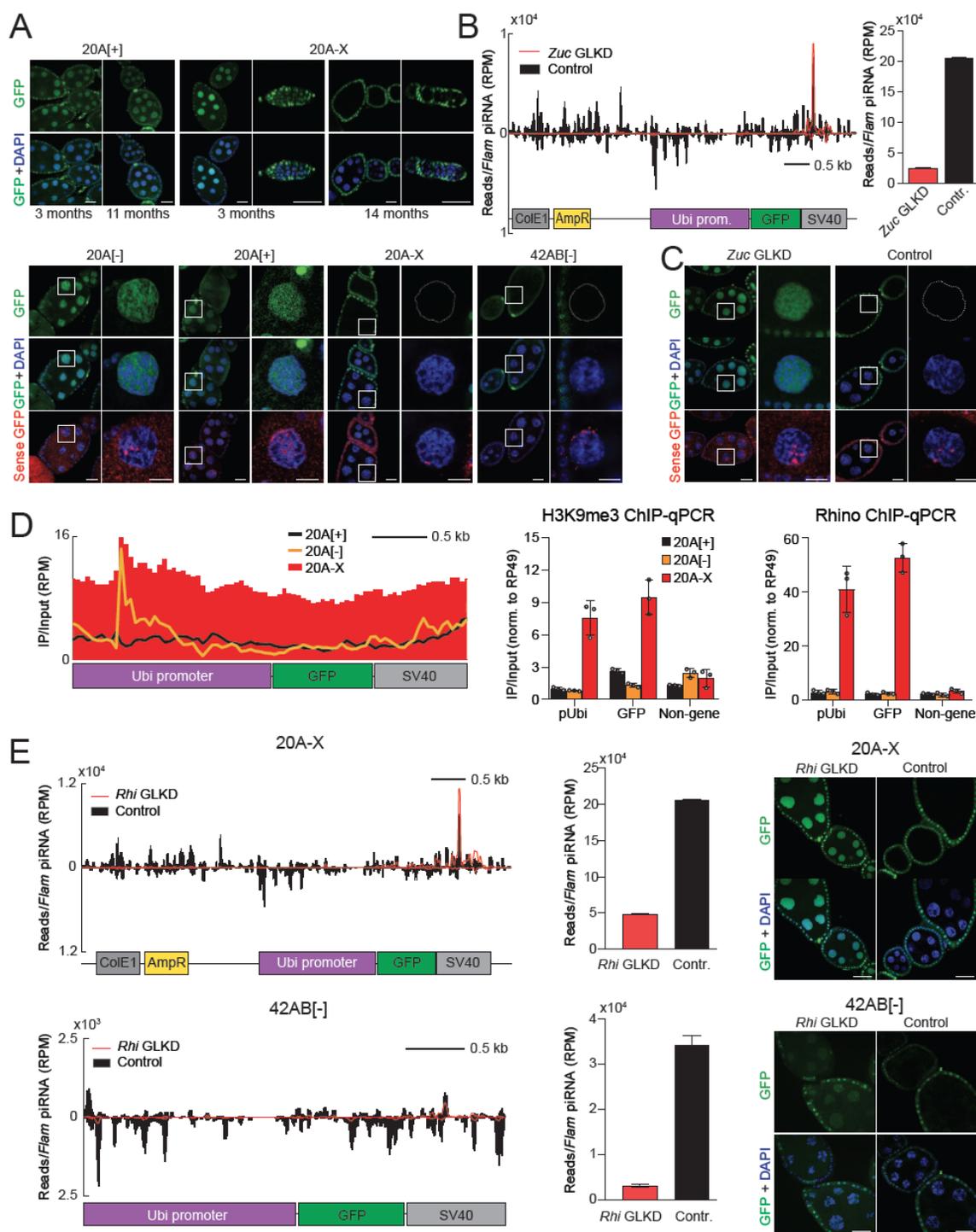


Figure 2.2 An unusual reporter insertion in the 20A cluster is a dual-stand piRNA cluster that generates piRNAs in a Rhi-dependent manner.

- (A) The 20A-X reporter is silenced in the germline after several months of maintaining this line. Top: GFP protein and sense RNA expression in ovaries of 20A[+] and 20A-X flies at different times after establishment of lines by RMCE. Bottom: GFP protein and sense RNA expression in ovaries of 20A[+], 20A[-], 20A-X and 42AB[-] flies 14 months (20A[+], 20A[-] and 20A-X) and 1 month (42AB[-]) xxx after establishment of lines. Scale bar is 20 μ m and 2 μ m for egg chambers and single nurse cell nuclei, respectively.
- (B) 20A-X generates piRNAs in a Zuc-dependent manner. Shown are profiles of piRNAs mapping to the 20A-X reporter in control (white GLKD) and upon Zuc GLKD driven by nos-GAL4. 20A-X piRNAs were normalized to piRNA reads mapping to the flam cluster. 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. Note that although GFP protein is not expressed in control flies, sense RNA is present in the nucleus. Zuc GLKD causes accumulation of RNA in the cytoplasm and leads to GFP protein expression. Scale bar is 20 μ m and 2 μ m for egg chambers and single nurse cell nuclei, respectively.
- (D) 20A-X, but not other 20A reporters are enriched in Rhino and the H3K9me3 mark. Left: Rhino ChIP-seq profiles (mean of two biological replicates) on 20A-X, 20A[+] (black) and 20A[-] (yellow) reporters. Right: Rhino and H3K9me3 enrichment on 20A reporter (Ubi and GFP region) as well as control locus (chr 2L: 968088- 968187, dm6) were determined by ChIP-qPCR. Only the 20A-X reporter is enriched in Rhi and H3K9me3. ChIP signals were normalized to the rp49 gene. Error bars indicate the standard deviation of three biological replicates.
- (E) Rhino knockdown reduces 20A-X piRNA and releases its silencing in the germline. Left: Shown are piRNA profiles over 20A-X and 42AB reporters in ovaries of control flies (white GLKD) and upon Rhi GLKD driven by nos-GAL4. Reporter piRNAs were normalized to piRNA reads mapping to the flam cluster. Error bars indicate standard deviation of two biological replicates. Right: GFP protein expression in 20A-X and 42AB ovaries upon Rhi GLKD. Scale bar is 20 μ m.

2.3 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. 2.3A). In fact, similar to other 20A reporters, expression of 20A cluster transcripts is decreased in 20A-X flies (Fig. S2.3A). 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. 2.3B). The absence of signals in other sites indicates that 20A-X line does not harbor additional transgene insertions at other genomic regions. To further validate these findings, we performed whole-genome sequencing and searched for reads corresponding to junctions between the reporter and genomic sequences. 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. S2.3B, 2.3B). 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. 2.3C). 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 of reporter (Fig. 2.3D, 2.3E, S2.3D). In addition, both approaches revealed that, unlike 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. 2.1A, 2.3E, S2.3D). Furthermore, we also found multiple DNA-seq reads indicating three unexpected junctions: (1) between the SV40 3' UTR and the GFP sequence, (2) between the SV40 3' UTR and the plasmid backbone and (3) between the attB site and the plasmid backbone (Fig. 2.3F). 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 are located 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 within the 20A cluster. 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. 2.3G). 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. 2.3G), 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 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. 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. 2.3H). 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). 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. 2.3H). In contrast to 20A-X, dsGFP insertions produce only miniscule amount of larger 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. 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 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.

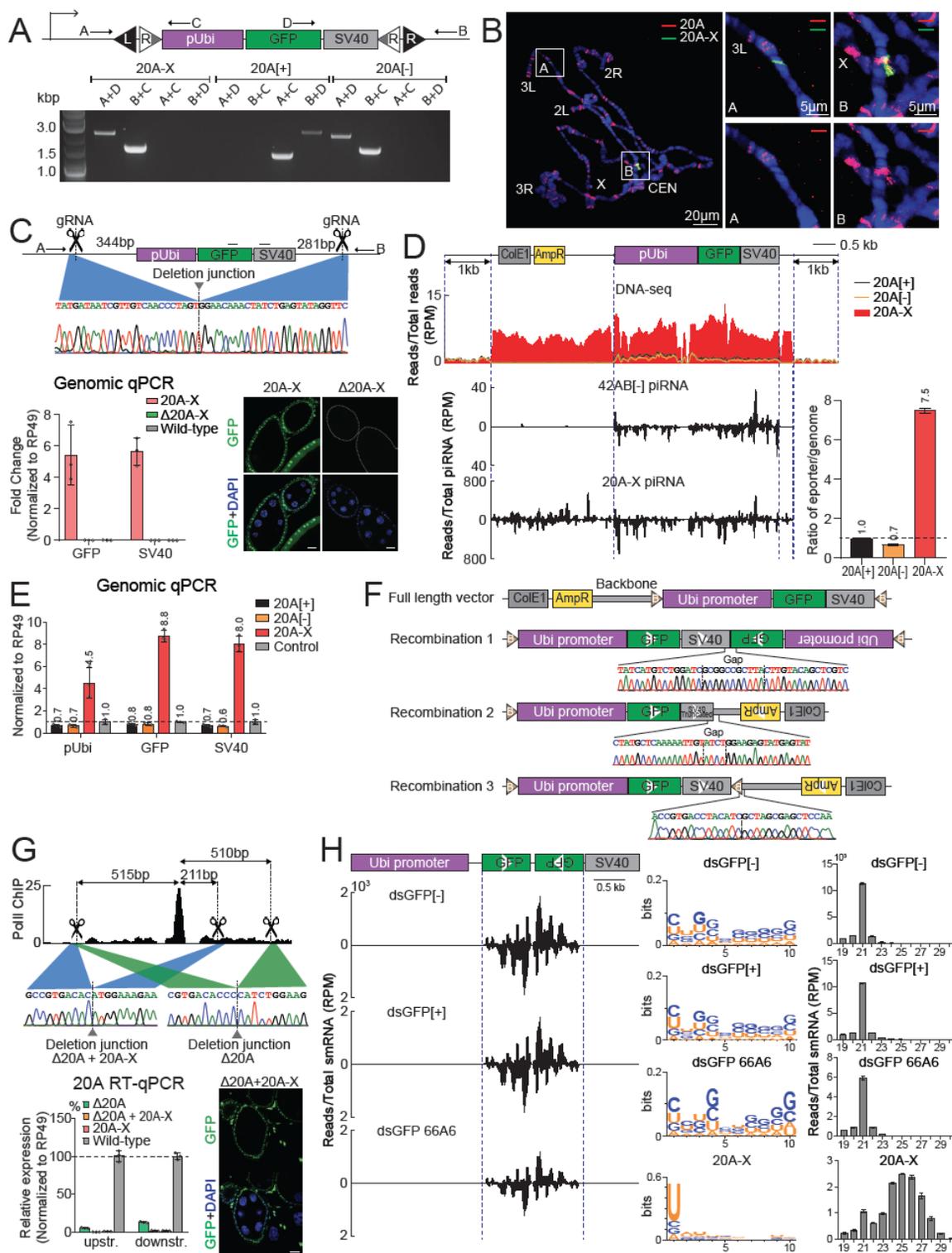


Figure 2.3 The 20A-X locus contains rearranged, multi-copy reporter sequences.

(A) Determining reporter orientation by genomic PCR. Positions of the primers are shown. 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 salivary gland polytene chromosomes was performed using probes against the Ubi-GFP reporter (green, Cy488) and the 20A cluster (red, Cy594). Probes against the reporter detected two locations: one co-localizes with the 20A cluster signal on the X chromosome, the other signal is localized on chromosome 3L where the native ubiquitin gene (Ubi-p63E) resides (chr3L:3899259-3903184, dm6).

(C) Verification of the 20A-X insertion site by CRISPR deletion. The region of the 20A cluster that includes the reporter insertion site was deleted in the 20A-X line using CRISPR/Cas9. Shown are positions of guide RNAs and primers used to verify the deletion using genomic PCR and Sanger sequencing (top). Detection of reporter sequences using qPCR of genomic DNA (left) shows absence of reporter sequences in flies with the deletion. No GFP expression is detected in flies with the deletion (right). Scale bar is 20 μ m.

(D) 20A-X includes plasmid backbone sequence (Top) Profiles of whole-genome DNA-seq reads over the sequence of the plasmid used for RMCE and the flanking 1 kB genomic sequences in flies with different 20A reporters. 20A-X, but not 20A[+] and 20A[-] flies harbor plasmid backbone sequence, which is normally not integrated during RMCE. Bar graph shows the ratio of reads derived from the reporter to reads from flanking genome sequences in different reporter lines. (Bottom) piRNA profiles over the reporter sequence in 20A-X and 42AB reporter flies. Note that 20A-X generates more piRNA even after normalization to sequence copy number. Profiles represent a mean of two biological replicates and were normalized to total number of piRNA reads in each library.

(E) 20A-X contains multiple copies of reporter sequence. Different portions of the reporter sequence were measured by genomic qPCR in flies with 20A reporters as well as a control line in which the reporter was integrated into non-cluster region (chr 3L: 7,575,013, dm6). qPCR values were normalized to the rp49 gene region. Fold-differences compared to homozygous flies with the control reporter are indicated above the bars. 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 (determined by a prominent Pol II ChIP-seq peak) was deleted using CRISPR/Cas9 in wild-type and 20A-X flies. RT-qPCR shows dramatic reduction of 20A cluster transcripts in flies with the deletion, indicating that it disrupts the cluster promoter. RT-qPCR primers were positioned 2030 and 3769 bp downstream of the putative promoter (upstream and downstream of the reporter insertion, respectively). Error bars indicate standard deviation of three biological replicates. GFP remained repressed in the germline of 20A-X flies upon promoter deletion, indicating that piRNA-mediated silencing remains unaffected. Scale bar is 20 μ m.

(H) Inverted repeat reporters generate abundant endo-siRNA, but not piRNA. Scheme of the inverted repeat dsGFP reporter. The dsGFP reporter harbors the same sequence fragments as other reporters, however, the GFP sequence forms an inverted repeat that generates hairpin dsRNA after transcription from the ubiquitin promoter. Shown are small RNA profiles (19-30 nt) along the reporter sequence in ovaries of flies with reporter integration into the 20A cluster in both orientations as well as integration into the control non-cluster region (66A6, chr 3L: 7575013, dm6). Shown on the right are size distributions of reporter mapping small RNAs and nucleotide composition of 23-29 nt (piRNA size range) RNAs. Size profile and nucleotide bias of 20A-X small RNAs are shown for comparison. dsGFP reporters generate abundant endo-siRNAs exclusively from the inverted repeat sequence. The miniscule amount of 23-29nt small RNA generated from dsGFP reporters does not show a 1U-bias expected from genuine piRNAs.

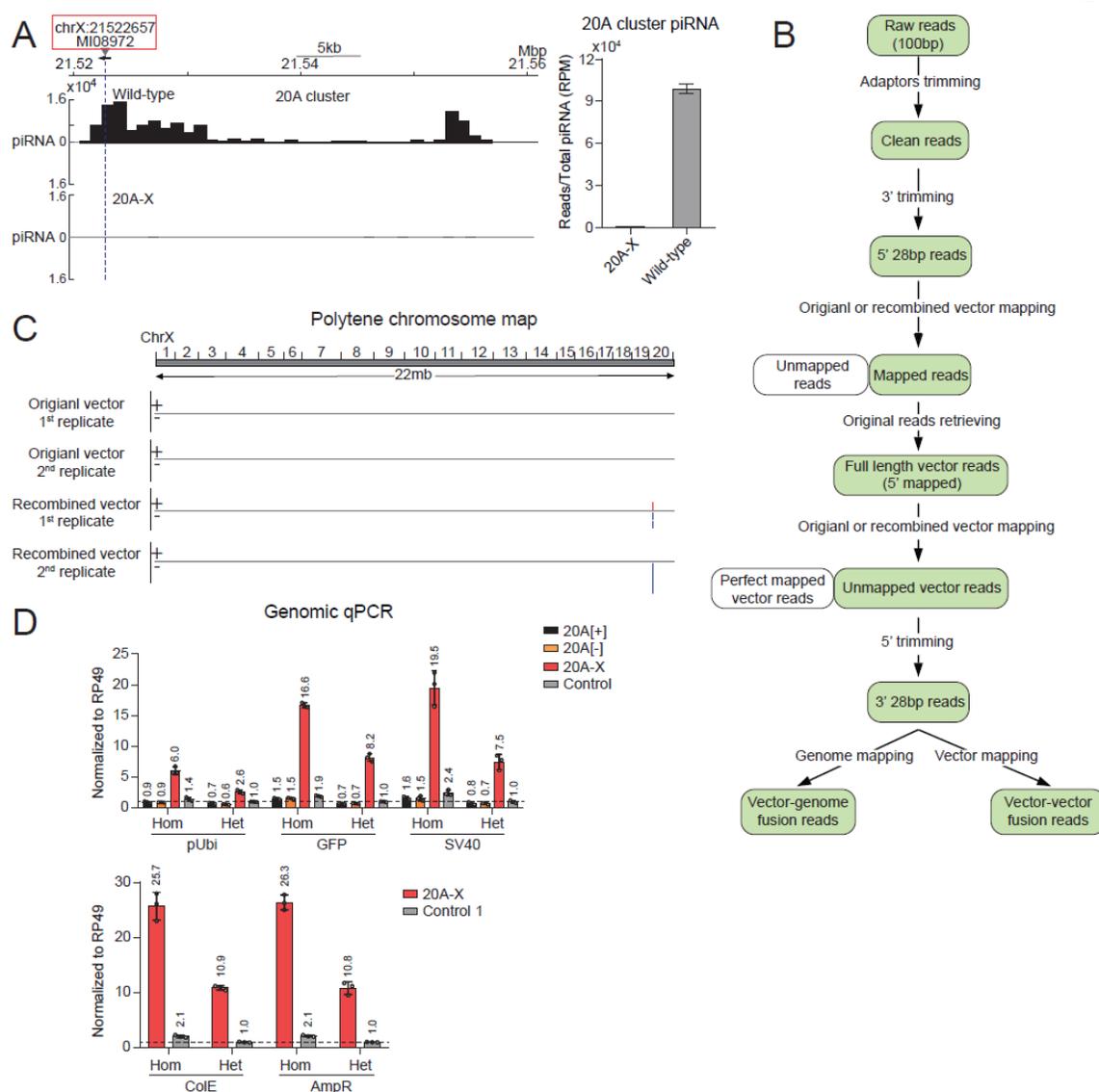


Figure S2.3 Analysis of 20A-X locus organization.

(A) (Left) Profiles of uniquely-mapping piRNAs of 20A cluster piRNA showed reduced in ovaries of 20A-X flies. Number of piRNA reads mapped to the 20A cluster is normalized to total piRNA read count. (Right) Total piRNA reads uniquely-mapping to 20A cluster were normalized to total piRNA read count. Error bars indicate standard deviation of two biological replicates.

(B) Pipeline to analyze genomic DNA-seq data to find reporter fusion reads.

(C) Sequencing of the 20A-X genome revealed a single site of reporter sequence integration in the expected position on the X chromosome. Original vector is the vector prior to recombinase-mediated

cassette exchange (including the vector backbone); recombined vector is after recombinase-mediated cassette exchange (without the vector backbone). Analysis was done in two replicates.

(D) Top: Different portions of the reporter sequence were measured by genomic qPCR in flies with homozygous (without X chromosome balancer) and heterozygous (with X chromosome balancer) 20A reporters as well as control reporter integrated into a non-cluster region (chr3L:7575013, dm6). Bottom: control1 reporter contains a single copy of ColE and AmpR (chr3L:11070538, dm6). The copy number is about two fold more in 20A reporter flies that are homozygous compared to heterozygous flies. qPCR values were normalized to rp49 gene region. Fold-differences compared to heterozygous control reporter are indicated above the bars. Error bars indicate standard deviation of three biological replicates.

2.4 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 (de Vanssay et al., 2012; Hermant et al., 2015; Le Thomas et al., 2014b), and proposed that all dual-strand clusters might depend on trans-generationally inherited piRNAs to maintain their activity (de Vanssay et al., 2012; Le Thomas et al., 2014b). 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. 2.4A). 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 IF results, RT-qPCR showed ~5-fold increase of GFP RNA after paternal transmission (Fig. 2.4B). Thus, repression of the 20A-X reporter in both in 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. 2.4C). 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 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. 2.4D). 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 (de Vanssay et al., 2012; Hermant et al., 2015; Le Thomas et al., 2014b). 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. 2.4E). 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 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 (data not shown). These results 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.

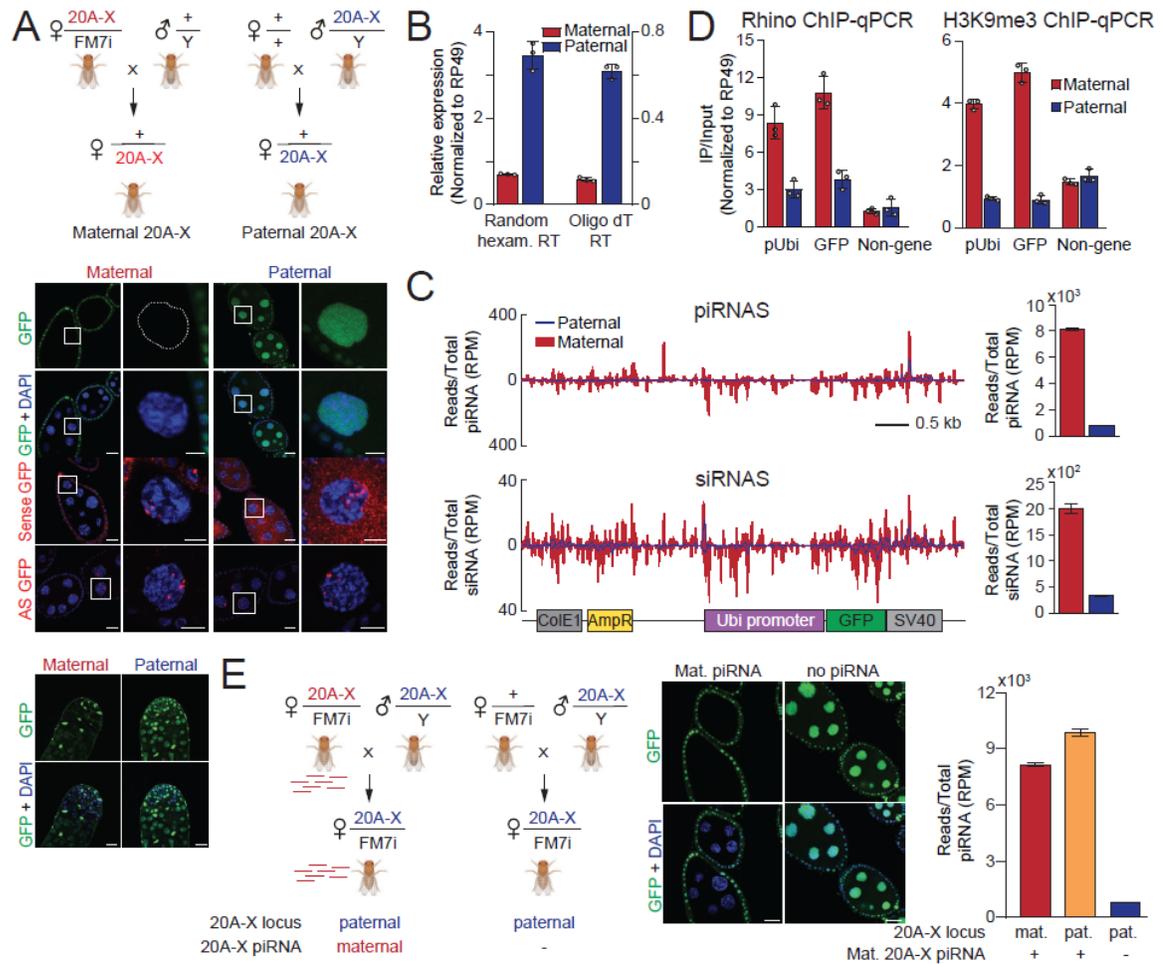


Figure 2.4 piRNA-induced repression of the 20A-X reporter depends on maternal transmission of cognate piRNAs.

(A) 20A-X repression is released after paternal transmission. Scheme of crosses to test effects of maternal and paternal transmission of 20A-X. Note that the genotype of the progeny from the two crosses are identical. Shown are GFP protein and sense and antisense GFP RNA expression in ovaries and GFP protein expression in testes of the progeny. Scale bar is 20 μ m and 2 μ m for egg chamber and single nurse cell nuclei, respectively, and 20 μ m for testis.

(B) GFP RNA expression increases after paternal transmission of 20A-X. RT-qPCR of GFP RNA in ovaries of progenies from the two crosses in (A) were performed with random hexamer and oligo dT primers and normalized to rp49 mRNA level. 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 ovaries of progeny from the two crosses shown in (A). Bar graphs on the right show number of piRNA and siRNA reads mapping to the reporter normalized to total piRNA and siRNA reads, respectively. 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 using primers against the pUbi and GFP regions as well as a control non-cluster region (chr 2L: 968,088 – 968,187, dm6) and normalized to the rp49 locus. H3K9me3 and Rhino levels drop after paternal transmission to levels similar to that of the control region. 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 cytoplasmic piRNA inheritance on repression of paternally transmitted 20A-X. Progeny of both crosses inherit the 20A-X locus from their fathers and are genetically identical. The only difference is cytoplasmic inheritance of small 20A-X-derived RNAs from their mothers, which is sufficient to restore both GFP repression and piRNA generation in the germline of the progeny. Number of piRNA reads mapping to the reporter was normalized to total piRNA reads mapping to the genome in each library. Error bars indicate standard deviation of two biological replicates. Scale bar is 20µm.

2.5 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. 2.5A). 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. 2.5A). Next, we analyzed small RNA profiles in ovaries of 20A-X flies 3, 11 and 21 months after establishment of the stock. piRNAs and siRNAs derived from 20A-X were already present at 3 months, but their abundance increased 4-fold and 3.7-fold, respectively, by 11 months (Fig. 2.5B). 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 3 months and 11 months were 1.0 and 4.2, respectively) (Fig. 2.5C, S2.5A). 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. S2.5C). We monitored whether 20A-X can recover its ability to generate piRNAs in future generations upon continuous maternal transmission (Fig. 2.5D). 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. 2.5D, S2.5B). 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. 2.5D, 2.5E, S2.5B). For example, in G2 almost equal fractions of flies showed normal expression, complete silencing or a mixed phenotype (Fig. 2.5E).

We profiled small RNAs 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 levels increased over generations. Importantly, the three groups of G3 ovaries with different levels of GFP repression had proportionally different levels of reporter piRNAs, indicating that repression correlates with piRNA abundance (Fig. 2.5F).

Finally, we determined enrichment of 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. 2.5G). 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.

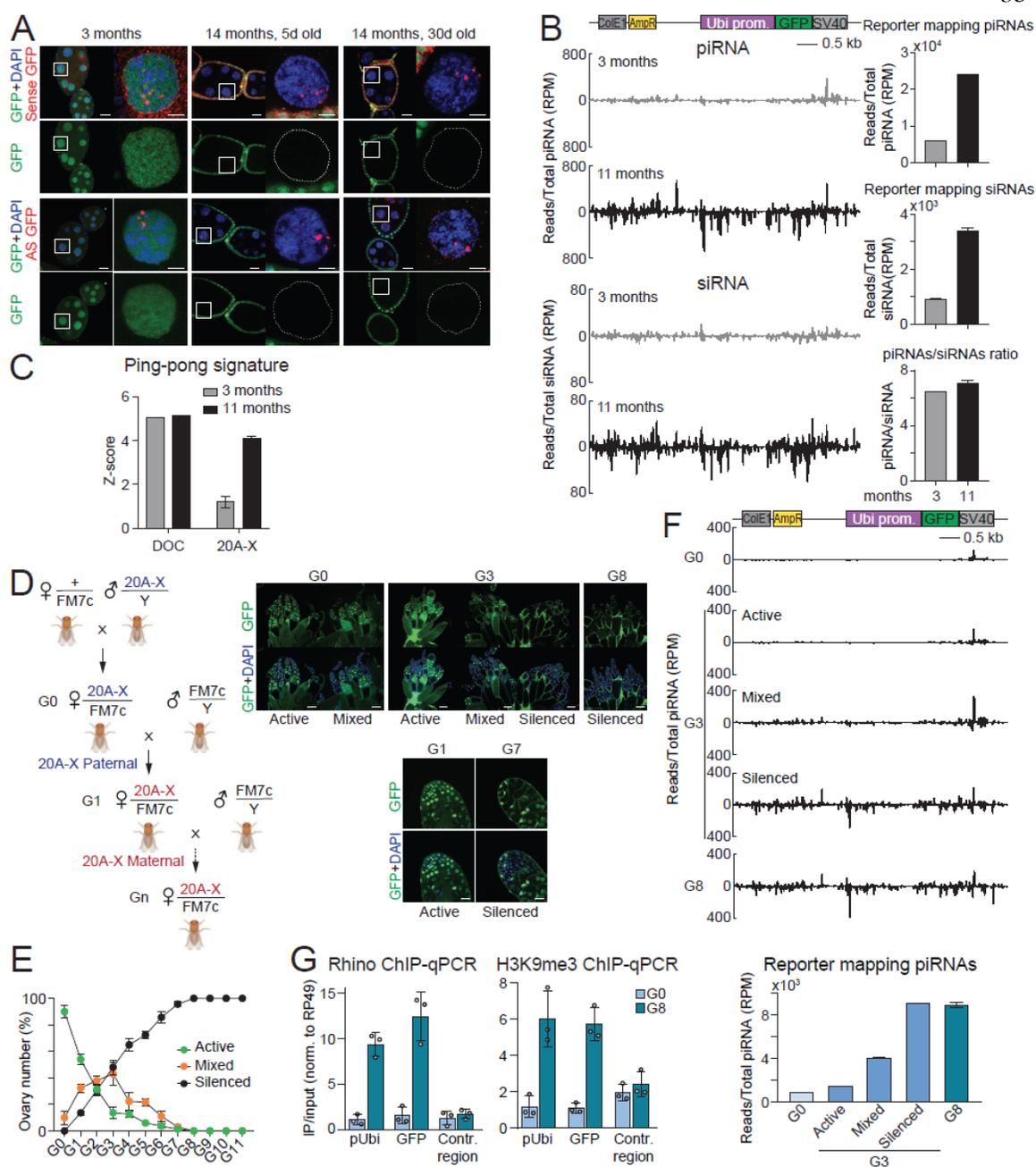


Figure 2.5 piRNA cluster is established over several generations.

(A) Establishment of 20A-X repression over several generations. Shown are GFP protein and RNA expression in ovaries of 20A-X flies 3 and 14 months after establishment of this line by RMCE. Young (5 daydays after hatching) and old (30 days after hatching) flies show no difference in expression. Scale bar is 20µm and 2µm for egg chamber and single nurse cell nuclei, respectively.

(B) 20A-X piRNA level increases over several generations. Shown are ovarian reporter-mapping piRNA and siRNA profiles 3 and 11 months after establishing the 20A-X line. Bar graph shows the levels of 20A-X piRNAs and siRNAs normalized to total number of piRNA and siRNA reads in each library. 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 (10 nt distance between 5' ends of complementary piRNAs) 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. Left: Scheme of crosses to monitor 20A-X after its paternal transmission. After paternal transmission in the first cross (G0), 20A-X is inherited maternally in each subsequent generation (G1-G8). Right: Expression of GFP in selected generations in ovaries and testes. Several generations (G1-G5) show variable GFP expression between individual flies and within each fly ovary. Scale bar is 100µm and 20µm for ovary and testis, respectively.

(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' indicate expression in the majority of germline nuclei, while 'mixed' indicate variable expression between individual egg chambers within the same ovary. Plotted is the fraction of ovaries with corresponding expression pattern in each generation from G0 to G11. The experiment was repeated three times and 100 ovaries were counted in each generation in each replica. 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 in different generations. In G3, ovaries were separated in three groups according to GFP expression as described in (E) and for each group independent small RNA libraries were prepared. Bar graph (bottom) shows 20A-X piRNA levels normalized to total piRNA reads in each library. Error bars indicate standard deviation of two biological replicates.

(G) 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 using primers against the pUbi and GFP regions as well as a control, non-cluster region (chr 2L: 968,088 – 968,187, dm6) and normalized to the rp49 locus. Error bars indicate the standard deviation of three biological replicates.

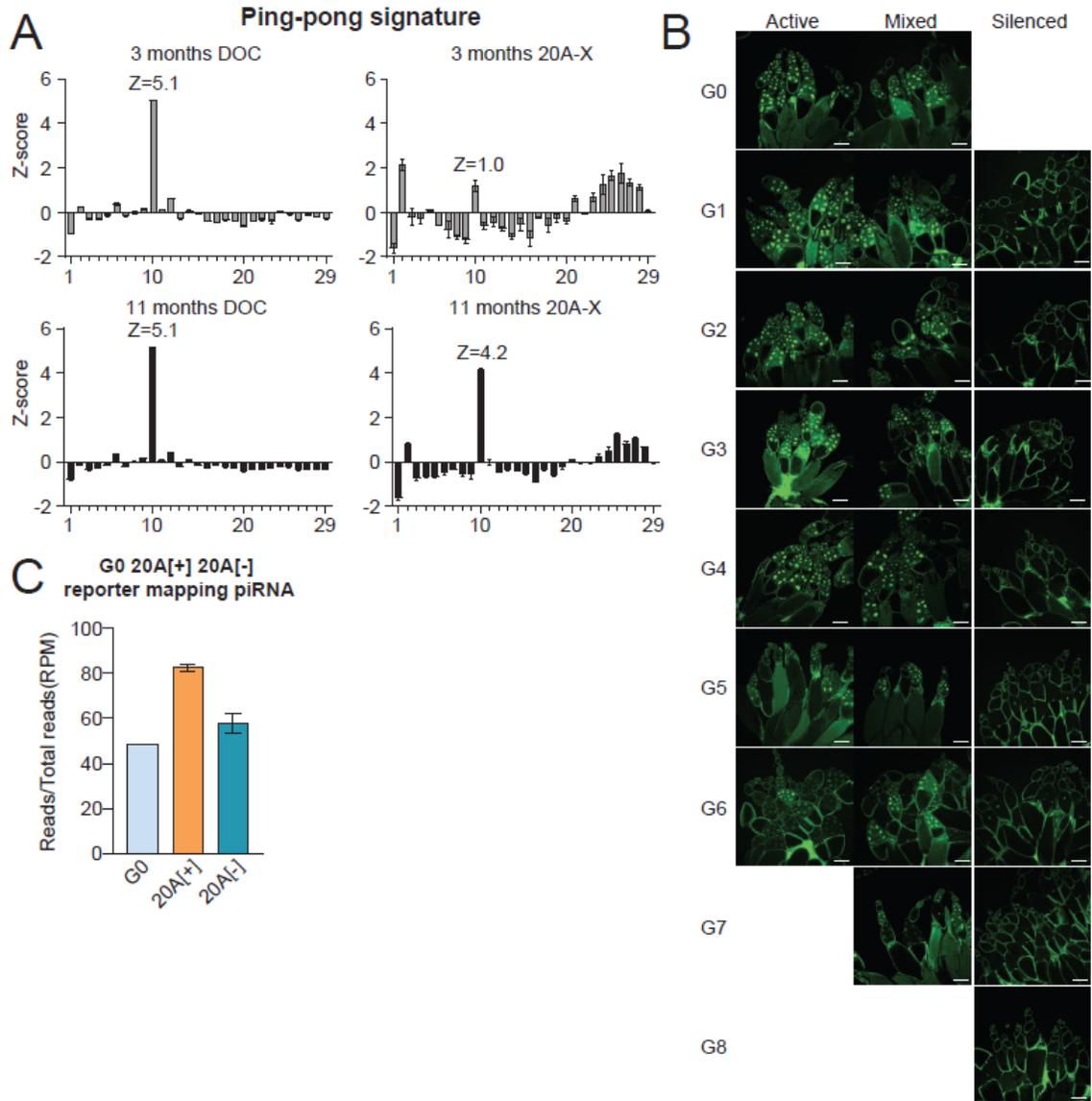


Figure S2.5 Analysis of 20A locus over generations.

(A) Analysis of ping-pong signature (overlap between 5'ends of piRNA mapping in opposite orientation) of piRNAs mapping to 20A-X and DOC piRNAs. Error bars indicate standard deviation of two biological replicates.

(B) GFP expression in ovary of each generation (G0 to G8) from the crosses shown on Fig. 2.5D. Scale bar is 100 μ m.

(C) Comparison of the number of reporter-mapping piRNAs in small RNA libraries from ovaries of 20A-X flies after paternal transmission (G0) (after normalized to copy number based on GFP qPCR results) and in small RNA libraries from ovaries of 20A[+] and 20A[-] flies.

2.6 Maternal siRNA triggers activation of piRNA cluster in the progeny

How loci like 20A-X start to function as dual-strand cluster and generate piRNAs remains unclear. In addition to piRNAs, all studied dual-strand clusters generate siRNA. Although we found that transgenes containing simple inverted repeats produce exclusively siRNAs and no piRNAs, it is still possible that the presence of cognate siRNAs might provide the initial trigger to activate piRNA production.

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. 2.6A). As seen before, paternal transmission of 20A-X led to release of GFP repression in the germline of the progeny (Fig. 2.6B). 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. 2.6B).

To further explore the role of siRNAs, we abrogated the siRNA pathway either in the mothers or in the progeny using Ago2 mutation (Fig. 2.6C). In flies, Ago2 is required for the stability and function of siRNAs and its mutation completely disrupts the siRNA pathway (Okamura et al., 2004). GFP repression was strongly disrupted in the progeny of Ago2-deficient mothers, confirming that it requires trans-generational cytoplasmic transmission of siRNA (Fig. 2.6D). In contrast, Ago2-deficient progeny that inherited siRNA from their heterozygous mothers show strong GFP repression. 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. 2.6E). 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. 2.6E). 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. Furthermore, 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. 2.6F). These results suggest that siRNAs are able to provide the initial trigger that converts the 20A-X locus into a dual-strand piRNA cluster.

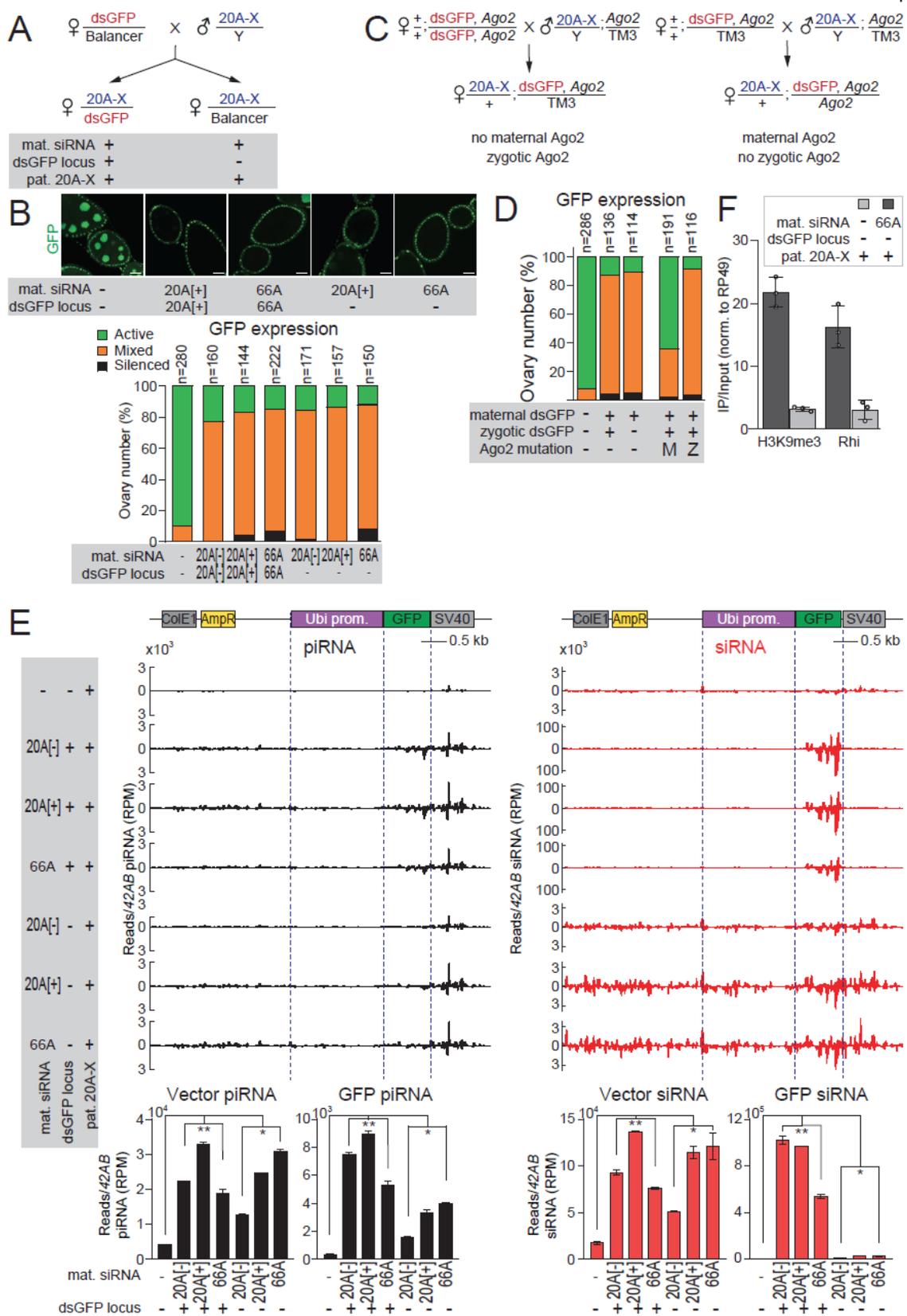


Figure 2.6 Cytoplasmic inheritance of siRNAs activates piRNA biogenesis in the progeny.

(A) Crossing scheme to test the role of siRNAs in triggering 20A-X repression. The 20A-X reporter is inherited from the father, while mothers harbor an inverted repeat dsGFP construct that generates siRNA. 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). Two independent dsGFP loci in 20A and in a locus in band 66A were analyzed. Maternal siRNA triggered repression of 20A-X in the progeny independently of inheritance of siRNA-generating locus or the genomic position of the siRNA-generating locus. Scale bar is 20 μ m. N indicates the number of ovaries analyzed for each genotype. GFP expression in progenies of each genotype was assessed by fluorescent microscopy.

(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 mapping to the reporter in ovaries of progenies of crosses shown in (A). Bar graph shows 20A-X piRNA and siRNA levels normalized to those generated from the 42AB cluster. 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 level on chromatin of 20A-X was measured by ChIP-qPCR in progenies of crosses shown in (A). ChIP signal in 20A-X is normalized to the rp49 locus. Error bars indicate the standard deviation of three biological replicates.

Chapter 3

DISCUSSION

3.1 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 target new invader elements. Previous studies have revealed mechanisms to store information about genome invaders in piRNA clusters and to maintain piRNA biogenesis through a feed-forward loop that involves trans-generational cytoplasmic transmission of piRNAs. These inherited piRNAs guide deposition of the RDC chromatin complex on clusters, which is required for efficient piRNA expression in the progeny. 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 integration of new transposons into pre-existing piRNA clusters, which would lead to generation of novel piRNAs, a process that has been modeled experimentally (Le Thomas et al., 2014b; Muerdter et al., 2012) and observed naturally (Khurana et al., 2011; Zhang et al., 2020). However, other studies suggest that entire new piRNA-generating regions can arise in evolution, providing another mechanism for acquiring immunity against novel elements. Indeed, some piRNA clusters are active only in one but not other *D. virilis* strains (Gebert et

al., 2021; Le Thomas et al., 2014a), suggesting their recent formation. Recent comprehensive evolutionary analysis showed that piRNA cluster regions are extremely labile in *Drosophila* evolution, suggesting frequent acquisition and loss of piRNA clusters (Gebert et al., 2021). Finally, spontaneous formation of novel piRNA clusters from transgenic sequences has been observed (de Vanssay et al., 2012). Though these findings suggested that new piRNA-producing genomic regions that contain no sequence homology to pre-existing piRNAs can arise, the conceptual framework to explain this process was lacking.

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-in form of active piRNA clusters-to recognize its targets, the siRNA pathway uses a simple rule for self/non-self discrimination. Unlike normal genes, transposons often generate both sense and antisense transcripts that form dsRNAs, which are recognized by Dicer and processed into siRNAs. Indeed, in addition to sense transcription from their own promoters, transposons are often transcribed in antisense orientation from host genes' promoters, for example if a TE is inserted into an intron or the 3'UTR in opposite orientation than the host gene orientation. In addition, recursive insertion of transposons into each other creates inverted repeats that

generate hairpin RNAs. The presence of sense and antisense transcripts leads to dsRNA and siRNA formation, 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, especially in TE-rich plant genomes. In contrast, the piRNA pathway is restricted to Metazoa, suggesting that it is a more recent evolutionary innovation. As both pathways target foreign genetic elements, siRNAs provide an ideal signal to activate piRNA biogenesis against novel invaders. Activation of piRNA response by siRNAs can be compared to stimulation of the robust and long-lasting adaptive immune response by the first-line innate immune systems (Medzhitov, 2007).

Our results indicate that siRNAs are important to jump-start piRNA cluster activity, but are dispensable later on. Indeed, cytoplasmically-inherited siRNAs are sufficient to trigger piRNA biogenesis in the progeny, while the zygotic siRNA pathway is completely dispensable for this process (Fig. 2.6). Consistent with siRNAs being dispensable for maintenance of piRNA biogenesis, a previous study showed that the siRNAs are not required for the activity of an artificial piRNA cluster (T1/BX2)(de Vanssay et al., 2012) . On the other hand, increase in the level of antisense transcripts was proposed to be linked to spontaneous activation of BX2 (Casier et al., 2019), suggesting a general role of antisense

RNA (and hence the siRNA pathway) in activation of piRNA clusters. Overall, our results and the previous findings suggest a two-step model of cluster activation (Fig. 3.1): 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 for cluster maintenance.

The precise molecular mechanism by which siRNAs trigger piRNA biogenesis remains to be understood. In yeast, siRNAs guide a methyltransferase complex to their genomic targets leading to H3K9me3 deposition on chromatin (Verdel et al., 2004). In *Drosophila*, this modification is required for deposition of the RDC complex and thus robust piRNA biogenesis (Klattenhoff et al., 2009; Le Thomas et al., 2014b; Mohn et al., 2014; Yu et al., 2015a; Zhang et al., 2014). However, in flies siRNA repression seems to be restricted to target RNA cleavage in the cytoplasm, suggesting that they induce piRNA biogenesis through a different mechanism. The cleavage of complementary transcripts by siRNAs creates aberrant RNAs with 5'-monophosphorylated ends, which are good substrates for the cytoplasmic piRNA processing machinery. Thus, we propose 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 (Huang et al., 2017; Pandey et al., 2017; Rogers et al.,

2017). As the nuclear Piwi/piRNA complex guides establishment of the H3K9me3 mark, this model explains how cytoplasmic siRNAs are capable of inducing chromatin changes that are associated with piRNA cluster activation.

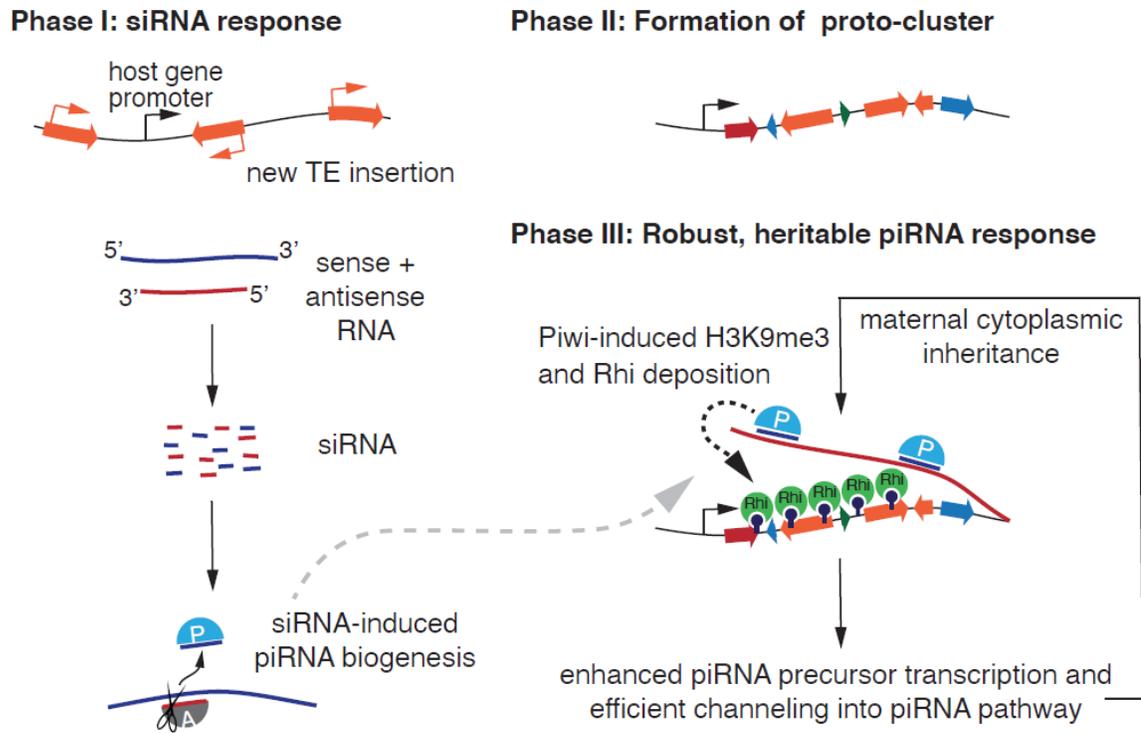


Figure 3.1 The model for siRNA-triggered activation of piRNA immunity.

3.2 Genomic requirements for piRNA cluster function

While our results indicate that siRNAs can trigger piRNA biogenesis from cognate genomic regions, we and others also found that simple inverted repeats generate exclusively siRNAs and not piRNAs. Thus, 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. Although 20A-X is composed of the same sequences as the simple inverted repeat construct, it contains ~ 10 copies of the transgenic sequence. The only other artificial piRNA clusters described, T1/BX2, also contain multiple tandem sequences. Thus, the extended length and repetitive nature seems to be important for de novo establishment of clusters. These features might be linked to the important role that chromatin organization plays in piRNA cluster function. Specifically, the extended repetitive organization might be required for maintenance of the RDC complex chromatin compartment, which is essential for transcription and post-transcriptional processing of piRNA precursors (Andersen et al., 2017; Chen et al., 2016; ElMaghraby et al., 2019; Hur et al., 2016; Klattenhoff et al., 2009; Mohn et al., 2014; Zhang et al., 2014). piRNA clusters are enriched in the heterochromatic H3K9me3 mark and the RDC complex that binds this mark (Klattenhoff et al., 2009; Le Thomas et al., 2014b; Mohn et al., 2014; Yu et al., 2015a; Zhang et al., 2014). Rhi is a paralog of HP1 that forms 'classic' heterochromatin domains, which depend on cooperative interactions between multiple HP1 molecules and associated proteins (through interactions between HP1 dimers bound to neighboring nucleosomes) (Le Thomas et al., 2014b; Yu et al., 2015a; Yu et al., 2018). Accordingly, extended length might be required for formation of a stable HP1 chromatin compartment. Tandem repeats seem to be particularly prone to formation of heterochromatin, though the underlying mechanisms is not

completely clear (Dorer and Henikoff, 1994). Similar to HP1, Rhi is capable of self-interactions through its chromo and chromo-shadow domains and these interactions are required for formation of RDC complex compartments and the function of piRNA clusters (Le Thomas et al., 2014b; Yu et al., 2015a). Thus, extended length and tandem repeats might be necessary to form a stable RDC complex chromatin compartment in the nucleus. 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. 3.1).

*Chapter 4***MATERIALS AND METHODS*****Drosophila* stocks**

All flies were raised at 25°C. The shWhite (BDSC #33623) and shRhino (BDSC #35171) stocks were obtained from Bloomington, shZucchini (#313693) stock was obtained from the Vienna *Drosophila* Resource Center, *Ago2*[414] (#109027) stock was obtained from Kyoto Stock Center. shRNAs were driven by the nos-GAL4 driver (BDSC #4937).

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 (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 (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 (Port and Bullock, 2016). The DNA oligo sequences are shown below:

20A-X deletion gRNA sequence

Forward:

GCGGCCCGGGTTCGATTCCCGGCCGATGCATTGAAGCTCCCACGAAGTTAGTT

TTAGAGCTAGAAATAGCAAG

Reverse:

ATTTAACTTGCTATTTCTAGCTCTAAACTAGTTGACGAGTGTCGCTTTGCA

CCAGCCGGGAATCGAACCC

20A promoter gRNA sequence

Forward:

GCGGCCCGGGTTCGATTCCCGGCCGATGCAACTACGTTACTAAGCATTGGTT
TTAGAGCTAGAAATAGCAAG

Reverse:

ATTTTAACTTGCTATTTCTAGCTCTAAAACGATGTCCAACTTGCAATTTTGCA
CCAGCCGGGAATCGAACCC

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 progeny were screened to verify the promoter deletion by genomic PCR followed by sanger sequencing. Transgenic flies used in this study are listed in Supplementary Table 1.

RNA HCR-FISH

The HCR-FISH RNA protocol was adapted from a previous protocol (Choi et al., 2018; Luo et al., 2020). Fly ovaries were dissected in cold PBS and fixed in 300 μ l fixation solution (4% paraformaldehyde, 0.15% Triton X-100 in PBS) at room temperature followed by three washes with PBX (PBS, 0.3% Triton X-100) for 5 min each at room temperature. Samples were dehydrated in 500 μ l 70% ethanol and permeabilized overnight at 4°C on a nutator. Samples were rehydrated in 500 μ l wash buffer (2 \times SSC, 10% [v/v] formamide) for 5 min at room temperature and pre-hybridized in 500 μ l hybridization buffer (50% formamide, 5x SSC, 9 mM citric acid pH 6.0, 0.1% Tween 20, 50 μ g/ml heparin, 1x Denhardt's solution, 10% dextran sulfate) for 30 min at 37°C. Following pre-hybridization, the hybridization solution containing 2 pmol of each probe was added and samples were incubated 12–16 hr at 37°C. Samples were washed four times with 500 μ l probe wash buffer (50% formamide, 5x SSC, 9 mM citric acid pH 6.0, 0.1% Tween 20, 50 μ g/ml heparin) for 15 min each at 37°C and three times with 5 x SSCT for 5 min each at room temperature. Samples were incubated in 500 μ l amplification buffer (5x SSC, 0.1% Tween 20, 10% dextran sulfate) for 30 min at room temperature. 30 pmol each of hairpin H1 hairpin H2 were prepared separately by incubating at 95°C for 90 s and letting them cool to room temperature in the dark for 30 min. Samples were incubated with the hairpin solution for 12–16hr in the dark at room

temperature followed by washing with 500 μ l 5x SSCT at room temperature in following order: 2×5 min, 2×30 min and 1×5 min. Samples were preserved on glass slides with mounting medium and imaged using a ZEISS LSM880 microscope. FISH signal was analyzed by using Fiji software. The probes were designed and synthesized by Molecular Technologies and Alexa594 was used for probe detection. Probe sequences are listed in Table 2.

ChIP-seq and ChIP-qPCR

All ChIP experiments were performed based on previous description(Chen et al., 2016) with some modifications. About 100 ovaries were dissected in cold PBS, transferred ovaries into 1.5mL Eppendorf tube. Ovaries were fixed by adding 500 μ L 1% formaldehyde in PBS for 10min at room temperature. Quenching the fixation by adding Glycine to a final 200mM concentration solution, incubating 5min at room temperature. Rinsing ovaries 3 times with PBS for 5min per each at room temperature. Washing fixed ovaries twice in 1mL Farnham buffer (5mM HEPES pH 8.0, 85mM KCl, 0.5% NP-40, 10mM NaF, 0.2mM Na₃VO₄, protease inhibitor cocktail) and breaking the ovaries with pipette up and down rigorously. Spinning down ovaries and discarding supernatant. Ovaries were transferred into tighter

dounce homogenizer with 350 μ L RIPA IP buffer (20mM Tris pH 7.4, 150mM NaCl, 1% NP-40, 0.5% Sodium Deoxycholate, 0.35% SDS, 10mM NaF, 0.2mM Na₃VO₄, protease inhibitor cocktail), dounced around 30 strokes till solution show milky. Transferring ovary lysate into sonication tubes no more than 300 μ L, sonicating ovary lysate at 30 cycles, 30sec on and 30sec off, high intensity, 4°C water bath. After sonication, adding 700 μ L RIPA IP dilute buffer (20mM Tris pH 7.4, 150mM NaCl, 1% NP-40, 0.5% Sodium Deoxycholate, 10mM NaF, 0.2mM Na₃VO₄, protease inhibitor cocktail), then lysate was centrifuged at 14000rpm for 15min at 4°C, carefully transfer supernatant to a new Eppendorf tube. 50 μ L per each reaction Dynabeads protein G (Invitrogen) was cleaned by adding 1mL 5mg/mL BSA and 1 μ L 10 mg/mL Salmon Sperm DNA Solution, rotated at room temperature for 1hr. Washing beads 3 times with buffer (20mM Tris pH 7.4, 150mM NaCl, 1% NP-40, 0.5% Sodium Deoxycholate, 0.1% SDS, 10mM NaF, 0.2mM Na₃VO₄, protease inhibitor cocktail). Before added cleaned beads into ovary lysate, saving 50 μ L ovary lysate as input, incubate beads and ovary lysate at 4°C for overnight. Washing beads 5 times with LiCl IP wash buffer (10mM Tris pH 7.5, 500mM LiCl, 1% NP-40, 1% Sodium Deoxycholate), at last wash, transferring beads into a new Eppendorf tube. After completely removing wash buffer, resuspend beads by adding 300 μ L PK buffer (200mM Tris pH 7.4, 25mM EDTA, 300mM NaCl, 2% SDS) and 7 μ L Proteinase K (800 units/mL), input samples were also added

250 μ L PK buffer and 7 μ L Proteinase K (800 units/mL). Input and IP samples were incubated on ThermoMixer for 2.5hr at 55°C then overnight at 65°C to reverse crosslink. Adding equal volume of Phenol-Chloroform pH 8.0 to the solution and vortex strongly 20sec, then immediately spin down 5min at maximum speed. Transferring aqueous (upper) phase to a new Eppendorf tube, follow with adding equal volume of Chloroform, vortex strongly 20sec, and then spin down 5min at maximum speed. Transferring aqueous (upper) phase to a new 1.5 mL Eppendorf tube and adding 2 μ L of Glycogen blue and 2.5 volumes of ice-cold 100% ethanol, mix by inverting tube a few times and place tubes at -20°C for overnight. Next day, spinning down tubes in cold centrifuge for 30min at maximum speed, the pellets were washed with 70% ethanol. Completely removing ethanol and air dry pellets at room temperature. The pellets were dissolved with 20 μ L ddH₂O and measured the concentration by using Qubit (Invitrogen).

The antibodies were used for ChIP are anti-H3K9me3 antibody from Abcam (ab8898) and anti-Rhino antibody obtained from the Brennecke lab. 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 Table 2. 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). 4 μ g total RNA was loaded onto a 15% polyacrylamide gel and small RNA between 19 and 29 nt in length was excised and isolated. 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. DNAase I treatment and reverse transcription was performed from 1 μ g 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 Table 2.

DNA FISH

Polytene chromosomes DNA FISH was performed as previously described(Cai et al., 2010; Lavrov et al., 2004) 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) (Bolger et al., 2014) and cutadapt (version 1.15)(Martin, 2011) 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(Langmead et al., 2009) (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 (Ramirez et al., 2016). Regions blacklisted by ENCODE (Amemiya et al., 2019) were excluded from enrichment analysis. Read counts over equal-sized bins were calculated using deepTools2 and BEDOPS (Neph et al., 2012), and figures were made using Matlab. All the scripts we used can be found on GitHub (https://github.com/brianpenghe/Luo_2021_piRNA/blob/main/ChIP-seq.md).

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. The scripts are available on GitHub (https://github.com/brianpenghe/Luo_2021_piRNA/blob/main/FusionReads.md).

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 (Antoniewski, 2014). Reporter coverage was calculated based on 10 nt bin size, and figures were made using Matlab. The mapping scripts are available on GitHub (https://github.com/brianpenghe/Luo_2021_piRNA/blob/main/piRNA-seq.md).

Data availability

Libraries generated from this study are deposited in GEO under accession codes GSE193091.

The scripts are available on GitHub: https://github.com/brianpenghe/Luo_2021_piRNA
Pol II ChIP-seq data analyzed in this study were from GSE43829 (Le Thomas et al., 2013) and GSE97719 (Andersen et al., 2017).

Table 1. *Drosophila melanogaster* stocks.

Stock	Source	Identifier
pUbi > eGFP-NLS 20A[-]	This study	N/A
pUbi > eGFP-NLS 20A[+]	This study	N/A
pUbi > eGFP-NLS 20A-X	This study	N/A
pUbi > eGFP-NLS 42AB[-]	This study	N/A
pUbi > eGFP-NLS 66A6	This study	N/A
pUbi > eGFP(sense)-NLS-GFP(antisense) 20A[-]	This study	N/A
pUbi > eGFP(sense)-NLS-GFP(antisense) 20A[+]	This study	N/A
pUbi > eGFP(sense)-NLS-GFP(antisense) 66A6	This study	N/A
MI07308	Bloomington stock	BDSC #43121
MI08972	Bloomington stock	BDSC #50496
UASp > small hairpin white	Bloomington stock	BDSC #33623
UASp > small hairpin Rhino	Bloomington stock	BDSC #35171
UASp > small hairpin Zucchini	VDRC	#313693
20A-X deletion gRNA	This study	N/A
Δ 20A-X	This study	N/A
20A promoter gRNA	This study	N/A
Δ 20A promoter	This study	N/A
Δ 20A promoter 20A-X	This study	N/A
attP40[nos-Cas9]	NIG-FLY	CAS-0001
Maternal alpha-tubulin67C > Gal4	Bloomington stock	BDSC #7063
pNanos > Gal4	Bloomington stock	BDSC #4937
w[*]; Ago2[414]	Kyoto Stock Center	#109027

Table 2. Primers and probes.

Primers	Sequence (5' to 3')
20A-X upstream qPCR F	GCCAGTAGTCGTCTCTCATTTATGC
20A-X upstream qPCR R	GCTGAAGCACTTGATTGCCAAC
20A-X downstream qPCR F	GCTTCCCATAAACCTCCCATGTG
20A-X downstream qPCR R	TCGTGGGAGCTTCAAGAGTATTGG
20A qPCR F	GCCTACGCAGAGGCCTAAGT
20A qPCR R	CAGATGTGGTCCAGTTGTGC
20A-X genomic A	ATGAGTTCAATTCGCTACTGCGAG
20A-X genomic B	ACTTCAACAGGAGCATAACCGCTAC
20A-X genomic C	GTGCTTCCCCGTGTGTGG
20A-X genomic D	CCGACAACCACTACCTGAGC
20A-X deletion A	GCCAGTAGTCGTCTCTCATTTATGC (Same as 20A-X upstream qPCR F)
20A-X deletion B	GACCGTTCCAGATTCGCTGC
pUbi qPCR F	TGCATTTCAAGGTCTTTGTTCGG
pUbi qPCR R	GCGAAAATCAACACGCAAGTTTT
GFP qPCR F	TACAACAGCCACAAGGTCTATATCA
GFP qPCR R	GGTGTCTGCTGGTAGTGGTC
SV40 qPCR F	TGGTGGGAATGCCTTTAATGAGGA
SV40 qPCR R	CCTTGGGGTCTTCTACCTTCTC
ColE qPCR F	AACTATCGTCTTGAGTCCAACCC
ColE qPCR R	GTAGTTAGGCCACCACTTCAAGA
AmpR qPCR F	ACGATCAAGGCGAGTTACATGAT
AmpR qPCR R	TACGGATGGCATGACAGTAAGAG
Non-gene qPCR F	CCCCATTTCCAGACGAGTCC
Non-gene qPCR R	TGACGGCAATAAGGATGCGA
rp49 qPCR F	CCGCTTCAAGGGACAGTATCTG
rp49 qPCR R	ATCTCGCCGCAGTAAACGC
DNA targets for HCR FISH	Identifier
GFP sense probes	3066/A160
GFP antisense probes	4017/D577

BIBLIOGRAPHY

- 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.
- Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE blacklist: identification of problematic regions of the genome. *Sci Rep* 9, 9354.
- Andersen, P.R., Tirian, L., Vunjak, M., and Brennecke, J. (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature* 549, 54-59.
- Antoniewski, C. (2014). Computing siRNA and piRNA overlap signatures. *Methods Mol Biol* 1173, 135-146.
- Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., 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.
- Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* 31, 785-799.
- Batki, J., Schnabl, J., Wang, J., Handler, D., Andreev, V.I., Stieger, C.E., Novatchkova, M., Lampersberger, L., Kauneckaitė, K., Xie, W., *et al.* (2019). The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation. *Nat Struct Mol Biol* 26, 720-731.
- Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., and Moran, J.V. (2010). LINE-1 retrotransposition activity in human genomes. *Cell* 141, 1159-1170.
- Bejerano, G., Lowe, C.B., Ahituv, N., King, B., Siepel, A., Salama, S.R., Rubin, E.M., Kent, W.J., and Haussler, D. (2006). A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* 441, 87-90.
- Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409, 363-366.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120.

- Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H., *et al.* (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* *18*, 1752-1762.
- Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* *128*, 1089-1103.
- Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., and Kazazian, H.H., Jr. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A* *100*, 5280-5285.
- Burns, K.H. (2017). Transposable elements in cancer. *Nat Rev Cancer* *17*, 415-424.
- Cai, W., Jin, Y., Girton, J., Johansen, J., and Johansen, K.M. (2010). Preparation of *Drosophila* polytene chromosome squashes for antibody labeling. *J Vis Exp* *36* 1748.
- Casier, K., Delmarre, V., Gueguen, N., Hermant, C., Viode, E., Vaury, C., Ronsseray, S., Brasset, E., Teyssset, L., and Boivin, A. (2019). Environmentally-induced epigenetic conversion of a piRNA cluster. *Elife* *8*. e39842.
- Chen, J.M., Ferec, C., and Cooper, D.N. (2006). LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease: Mutation detection bias and multiple mechanisms of target gene disruption. *J Biomed Biotechnol* *2006*, 56182.
- Chen, M., Jia, L., Zheng, X., Han, M., Li, L., and Zhang, L. (2022). Ancient human endogenous retroviruses contribute to genetic evolution and regulate cancer cell type-specific gene expression. *Cancer Res* *82*, 3457-3473.
- Chen, Y.A., Stuwe, E., Luo, Y., Ninova, M., Le Thomas, A., Rozhavskaia, E., Li, S., Vempati, S., Laver, J.D., Patel, D.J., *et al.* (2016). Cutoff suppresses RNA polymerase II termination to ensure expression of piRNA precursors. *Mol Cell* *63*, 97-109.
- Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation in situ hybridization chain reaction: Multiplexed, quantitative, sensitive, versatile, robust. *Development* *145*, dev165753.
- Chuong, E.B., Rumi, M.A., Soares, M.J., and Baker, J.C. (2013). Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet* *45*, 325-329.
- Cordaux, R., and Batzer, M.A. (2009). The impact of retrotransposons on human genome evolution. *Nat Rev Genet* *10*, 691-703.
- Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., *et al.* (2008). An endogenous small interfering RNA pathway in *Drosophila*. *Nature* *453*, 798-802.

- Daniel, C., Behm, M., and Ohman, M. (2015). The role of Alu elements in the cis-regulation of RNA processing. *Cell Mol Life Sci* 72, 4063-4076.
- de Vanssay, A., Bouge, A.L., Boivin, A., Hermant, C., Teyssset, L., Delmarre, V., Antoniewski, C., and Ronsseray, S. (2012). Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus. *Nature* 490, 112-115.
- Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* 35, 41-48.
- Dorer, D.R., and Henikoff, S. (1994). Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. *Cell* 77, 993-1002.
- Elbarbary, R.A., Lucas, B.A., and Maquat, L.E. (2016). Retrotransposons as regulators of gene expression. *Science* 351, aac7247.
- ElMaghraby, M.F., Andersen, P.R., 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.
- Ewing, A.D., and Kazazian, H.H., Jr. (2010). High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res* 20, 1262-1270.
- Fabry, M.H., Ciabrelli, F., Munafo, M., Eastwood, E.L., Kneuss, E., Falciatori, I., Falconio, F.A., Hannon, G.J., and Czech, B. (2019). piRNA-guided co-transcriptional silencing coopts nuclear export factors. *Elife* 8.
- Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 9, 397-405.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806-811.
- Gammon, D.B., and Mello, C.C. (2015). RNA interference-mediated antiviral defense in insects. *Curr Opin Insect Sci* 8, 111-120.
- Gebert, D., Neubert, L.K., Lloyd, C., Gui, J., Lehmann, R., and Teixeira, F.K. (2021). Large *Drosophila* germline piRNA clusters are evolutionarily labile and dispensable for transposon regulation. *Mol Cell* 81, 3965-3978 e3965.
- Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z., *et al.* (2008). Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* 320, 1077-1081.

- Gitlin, L., Karelsky, S., and Andino, R. (2002). Short interfering RNA confers intracellular antiviral immunity in human cells. *Nature* *418*, 430-434.
- 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.
- Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* *315*, 1587-1590.
- Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* *286*, 950-952.
- Hancks, D.C., and Kazazian, H.H., Jr. (2016). Roles for retrotransposon insertions in human disease. *Mob DNA* *7*, 9.
- Hedges, D.J., and Deininger, P.L. (2007). Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat Res* *616*, 46-59.
- Helman, E., Lawrence, M.S., Stewart, C., Sougnez, C., Getz, G., and Meyerson, M. (2014). Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res* *24*, 1053-1063.
- Hermant, C., Boivin, A., Teyssset, 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.
- Huang, X., Fejes Toth, K., and Aravin, A.A. (2017). piRNA Biogenesis in *Drosophila melanogaster*. *Trends Genet* *33*, 882-894.
- Hur, J.K., Luo, Y., Moon, S., Ninova, M., Marinov, G.K., Chung, Y.D., and Aravin, A.A. (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.
- Ipsaro, J.J., Haase, A.D., Knott, S.R., Joshua-Tor, L., and Hannon, G.J. (2012). The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature* *491*, 279-283.
- Ito, J., Sugimoto, R., Nakaoka, H., Yamada, S., Kimura, T., Hayano, T., and Inoue, I. (2017). Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. *PLoS Genet* *13*, e1006883.
- Iwasaki, Y.W., Sriswasdi, S., Kinugasa, Y., Adachi, J., Horikoshi, Y., Shibuya, A., Iwasaki, W., Tashiro, S., Tomonaga, T., and Siomi, H. (2021). Piwi-piRNA complexes induce stepwise changes in nuclear architecture at target loci. *EMBO J* *40*, e108345.

- Kassiotis, G., and Stoye, J.P. (2016). Immune responses to endogenous retroelements: Taking the bad with the good. *Nat Rev Immunol* 16, 207-219.
- Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T.N., Siomi, M.C., and Siomi, H. (2008). *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* 453, 793-797.
- Khodosevich, K., Lebedev, Y., and Sverdlov, E. (2002). Endogenous retroviruses and human evolution. *Comp Funct Genomics* 3, 494-498.
- Khurana, J.S., Wang, J., Xu, J., Koppetsch, B.S., Thomson, T.C., Nowosielska, A., Li, C., Zamore, P.D., Weng, Z., and Theurkauf, W.E. (2011). Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell* 147, 1551-1563.
- Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., 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.
- Kneuss, E., Munafo, M., Eastwood, E.L., Deumer, U.S., Preall, J.B., Hannon, G.J., and Czech, B. (2019). Specialization of the *Drosophila* nuclear export family protein Nxf3 for piRNA precursor export. *Genes Dev* 33, 1208-1220.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25.
- Lavrov, S., Dejardin, J., and Cavalli, G. (2004). Combined immunostaining and FISH analysis of polytene chromosomes. *Methods Mol Biol* 247, 289-303.
- Le Thomas, A., Marinov, G.K., and Aravin, A.A. (2014a). A transgenerational process defines piRNA biogenesis in *Drosophila virilis*. *Cell Rep* 8, 1617-1623.
- Le Thomas, A., Rogers, A.K., Webster, A., Marinov, G.K., Liao, S.E., Perkins, E.M., Hur, J.K., Aravin, A.A., and Toth, K.F. (2013). Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev* 27, 390-399.
- Le Thomas, A., Stuwe, E., Li, S., Du, J., Marinov, G., Rozhkov, N., Chen, Y.C., Luo, Y., Sachidanandam, R., Toth, K.F., *et al.* (2014b). Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev* 28, 1667-1680.

- Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L.J., 3rd, Lohr, J.G., Harris, C.C., Ding, L., Wilson, R.K., *et al.* (2012). Landscape of somatic retrotransposition in human cancers. *Science* 337, 967-971.
- Li, Y., Zhang, G., and Cui, J. (2022). Origin and deep evolution of human endogenous retroviruses in pan-primates. *Viruses* 14, 1370.
- Lindbo, J.A., Silva-Rosales, L., Proebsting, W.M., and Dougherty, W.G. (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.
- Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., *et al.* (2004). Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430, 471-476.
- Lunyak, V.V., Prefontaine, G.G., Nunez, E., Cramer, T., Ju, B.G., Ohgi, K.A., Hutt, K., Roy, R., Garcia-Diaz, A., Zhu, X., *et al.* (2007). Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* 317, 248-251.
- Luo, Y., Fefelova, E., Ninova, M., Chen, Y.A., and Aravin, A.A. (2020). Repression of interrupted and intact rDNA by the SUMO pathway in *Drosophila melanogaster*. *Elife* 9.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011 17, 3.
- Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* 123, 607-620.
- Mc, C.B. (1950). The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A* 36, 344-355.
- Medzhitov, R. (2007). Recognition of microorganisms and activation of the immune response. *Nature* 449, 819-826.
- Mills, R.E., Bennett, E.A., Iskow, R.C., Luttig, C.T., Tsui, C., Pittard, W.S., and Devine, S.E. (2006). Recently mobilized transposons in the human and chimpanzee genomes. *Am J Hum Genet* 78, 671-679.
- 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.
- Muerdter, F., Olovnikov, I., Molaro, A., Rozhkov, N.V., Czech, B., Gordon, A., Hannon, G.J., and Aravin, A.A. (2012). Production of artificial piRNAs in flies and mice. *RNA* 18, 42-52.

- Neph, S., Kuehn, M.S., Reynolds, A.P., Haugen, E., Thurman, R.E., Johnson, A.K., Rynes, E., Maurano, M.T., Vierstra, J., Thomas, S., *et al.* (2012). BEDOPS: High-performance genomic feature operations. *Bioinformatics* 28, 1919-1920.
- Nguyen, T.H.M., Carreira, P.E., Sanchez-Luque, F.J., Schauer, S.N., Fagg, A.C., Richardson, S.R., Davies, C.M., Jesuadian, J.S., Kempen, M.H.C., Troskie, R.L., *et al.* (2018). L1 retrotransposon heterogeneity in ovarian tumor cell evolution. *Cell Rep* 23, 3730-3740.
- Ninova, M., Chen, Y.A., Godneeva, B., Rogers, A.K., Luo, Y., Fejes Toth, K., and Aravin, A.A. (2020a). Su(var)2-10 and the SUMO pathway link piRNA-guided target recognition to chromatin silencing. *Mol Cell* 77, 556-570 e556.
- Ninova, M., Godneeva, B., Chen, Y.A., Luo, Y., Prakash, S.J., Jankovics, F., Erdelyi, M., Aravin, A.A., and Fejes Toth, K. (2020b). The SUMO Ligase Su(var)2-10 controls hetero- and euchromatic gene expression via establishing H3K9 trimethylation and negative feedback regulation. *Mol Cell* 77, 571-585 e574.
- Nishimasu, H., Ishizu, H., Saito, K., Fukuhara, S., Kamatani, M.K., 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.
- Okamura, K., Ishizuka, A., Siomi, H., and Siomi, M.C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* 18, 1655-1666.
- Pandey, R.R., Homolka, D., Chen, K.M., Sachidanandam, R., Fauvarque, M.O., and Pillai, R.S. (2017). Recruitment of Armitage and Yb to a transcript triggers its phased processing into primary piRNAs in *Drosophila* ovaries. *PLoS Genet* 13, e1006956.
- Pelisson, A., Sarot, E., Payen-Groschene, G., and Bucheton, A. (2007). A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the *Drosophila* ovary. *J Virol* 81, 1951-1960.
- Port, F., and Bullock, S.L. (2016). Augmenting CRISPR applications in *Drosophila* with tRNA-flanked sgRNAs. *Nat Methods* 13, 852-854.
- Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dunder, F., and Manke, T. (2016). deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44, W160-165.
- Rand, T.A., Petersen, S., Du, F., and Wang, X. (2005). Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* 123, 621-629.

- Rebollo, R., Karimi, M.M., Bilenky, M., Gagnier, L., Miceli-Royer, K., Zhang, Y., Goyal, P., Keane, T.M., Jones, S., Hirst, M., *et al.* (2011). Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genet* 7, e1002301.
- Rodic, N., Steranka, J.P., Makohon-Moore, A., Moyer, A., Shen, P., Sharma, R., Kohutek, Z.A., Huang, C.R., Ahn, D., Mita, P., *et al.* (2015). Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. *Nat Med* 21, 1060-1064.
- Rogers, A.K., Situ, K., Perkins, E.M., and Toth, K.F. (2017). Zucchini-dependent piRNA processing is triggered by recruitment to the cytoplasmic processing machinery. *Genes Dev* 31, 1858-1869.
- Roulois, D., Loo Yau, H., Singhania, R., Wang, Y., Danesh, A., Shen, S.Y., Han, H., Liang, G., Jones, P.A., Pugh, T.J., *et al.* (2015). DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. *Cell* 162, 961-973.
- Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M.C. (2006). Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev* 20, 2214-2222.
- Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Goncalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* 148, 335-348.
- Scott, E.C., Gardner, E.J., Masood, A., Chuang, N.T., Vertino, P.M., and Devine, S.E. (2016). A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Res* 26, 745-755.
- Shukla, R., Upton, K.R., Munoz-Lopez, M., Gerhardt, D.J., Fisher, M.E., Nguyen, T., Brennan, P.M., Baillie, J.K., Collino, A., Ghisletti, S., *et al.* (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* 153, 101-111.
- Sienski, G., Batki, J., Senti, K.A., 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.
- Sienski, G., Donertas, D., and Brennecke, J. (2012). Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* 151, 964-980.
- Sun, X., Wang, X., Tang, Z., Grivainis, M., Kahler, D., Yun, C., Mita, P., Fenyó, D., and Boeke, J.D. (2018). Transcription factor profiling reveals molecular choreography and key

- regulators of human retrotransposon expression. *Proc Natl Acad Sci U S A* *115*, E5526-E5535.
- Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M.P., and Wang, T. (2014). Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res* *24*, 1963-1976.
- Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M., *et al.* (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* *453*, 534-538.
- Tang, G., Reinhart, B.J., Bartel, D.P., and Zamore, P.D. (2003). A biochemical framework for RNA silencing in plants. *Genes Dev* *17*, 49-63.
- Tang, Z., Steranka, J.P., Ma, S., Grivainis, M., Rodic, N., Huang, C.R., Shih, I.M., Wang, T.L., Boeke, J.D., Fenyo, D., *et al.* (2017). Human transposon insertion profiling: Analysis, visualization and identification of somatic LINE-1 insertions in ovarian cancer. *Proc Natl Acad Sci U S A* *114*, E733-E740.
- Trizzino, M., Park, Y., Holsbach-Beltrame, M., Aracena, K., Mika, K., Caliskan, M., Perry, G.H., Lynch, V.J., and Brown, C.D. (2017). Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res* *27*, 1623-1633.
- Ullu, E., and Tschudi, C. (1984). Alu sequences are processed 7SL RNA genes. *Nature* *312*, 171-172.
- Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* *313*, 320-324.
- Vance, V., and Vaucheret, H. (2001). RNA silencing in plants--defense and counterdefense. *Science* *292*, 2277-2280.
- Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* *303*, 672-676.
- Voinnet, O., Pinto, Y.M., and Baulcombe, D.C. (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.
- Wang, J., Vicente-Garcia, C., Seruggia, D., Molto, E., Fernandez-Minan, A., Neto, A., Lee, E., Gomez-Skarmeta, J.L., Montoliu, L., Lunyak, V.V., *et al.* (2015). MIR retrotransposon sequences provide insulators to the human genome. *Proc Natl Acad Sci U S A* *112*, E4428-4437.

- Warren, W.C., Hillier, L.W., Marshall Graves, J.A., Birney, E., Ponting, C.P., Grutzner, F., Belov, K., Miller, W., Clarke, L., Chinwalla, A.T., *et al.* (2008). Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453, 175-183.
- Xing, J., Zhang, Y., Han, K., Salem, A.H., Sen, S.K., Huff, C.D., Zhou, Q., Kirkness, E.F., Levy, S., Batzer, M.A., *et al.* (2009). Mobile elements create structural variation: Analysis of a complete human genome. *Genome Res* 19, 1516-1526.
- Yu, B., Cassani, M., Wang, M., Liu, M., Ma, J., Li, G., Zhang, Z., and Huang, Y. (2015a). Structural insights into Rhino-mediated germline piRNA cluster formation. *Cell Res* 25, 525-528.
- Yu, B., Lin, Y.A., Parhad, S.S., Jin, Z., Ma, J., Theurkauf, W.E., Zhang, Z.Z., and Huang, Y. (2018). Structural insights into Rhino-Deadlock complex for germline piRNA cluster specification. *EMBO Rep* 19.
- Yu, Y., Gu, J., Jin, Y., Luo, Y., Preall, J.B., Ma, J., Czech, B., and Hannon, G.J. (2015b). Panoramix enforces piRNA-dependent cotranscriptional silencing. *Science* 350, 339-342.
- Zhang, G., Tu, S., Yu, T., Zhang, X.O., Parhad, S.S., Weng, Z., and Theurkauf, W.E. (2018). Co-dependent assembly of *Drosophila* piRNA precursor complexes and piRNA cluster heterochromatin. *Cell Rep* 24, 3413-3422 e3414.
- Zhang, S., Pointer, B., and Kelleher, E.S. (2020). Rapid evolution of piRNA-mediated silencing of an invading transposable element was driven by abundant de novo mutations. *Genome Res* 30, 566-575.
- Zhang, Z., Wang, J., Schultz, N., Zhang, F., Parhad, S.S., Tu, S., Vreven, T., Zamore, P.D., Weng, Z., and Theurkauf, W.E. (2014). The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell* 157, 1353-1363.
- Zhao, K., Cheng, S., Miao, N., Xu, P., Lu, X., Zhang, Y., Wang, M., Ouyang, X., Yuan, X., Liu, W., *et al.* (2019). A Pandas complex adapted for piRNA-guided transcriptional silencing and heterochromatin formation. *Nat Cell Biol* 21, 1261-1272.