Université Pierre et Marie Curie

California Institute of Technology

Ecole doctorale Complexité du Vivant

Alexei Aravin Laboratory

Piwi function and piRNA cluster regulation

Drosophila melanogaster

Par Adrien Le Thomas

Thèse de doctorat de Biologie

Dirigée par Déborah Bourc'His / Alexei Aravin

Présentée et soutenue publiquement le 11 septembre 2014

Devant un jury composé de :

Stéphane Ronsseray, DR, HDR

Hervé Le Hir, DR, HDR

Hervé Seitz, CR, HDR

Alain Pélisson, DR, HDR, rapporteur

Olivier Voinnet, DR, HDR, rapporteur

Except where otherwise noted, this work is licensed under
http://creativecommons.org/licenses/by-nc-nd/3.0/

Acknowledgments

I would like to thank first my two thesis supervisors, Dr. Déborah Bourc'His and Dr. Alexei Aravin, who gave me the opportunity to achieve this PhD after three years of work. I thank Déborah for her support, her help, her advices, and her availability. I thank Alexei for welcoming me in his team at Caltech and for the trust he granted me from the beginning. He has been an invaluable mentor all along through, I quote, "the dark days of hard PhD work". Working at his side made me discover the daily work of scientific research and its outcomes, both exciting and extremely rich intellectually. He was also particularly great outside the scientific environment, whether it is to discover the nature around Los Angeles and the United-States, or to learn how to drink vodka shots like a Russian man.

I would like to thank all the members of my thesis committee and the defense jury who accepted to participate and took their time to evaluate my work. I am very honored by their presence and for the interest they give to my project.

I would like to address a special thanks to Dr. Katalin Fejes-Tóth for guiding me while taking my first steps into PhD and until the last. Kata's knowledge and guidance resulted in the publication of my first paper: her help was crucial and very precious. Besides, Kata, and Alexei, never left me alone while isolated from family during Thanksgiving, Christmas and Easter, I will never forget those moments.

A big thank to the now Doctors Ivan Olovnikov, for his vast knowledge and skills that he shared unconditionally, and Georgi Marinov, who taught me everything I know about computational analysis. Both were outstanding teachers.

I'd like to also heartily thank professors and teachers, who welcomed me in their team, over the years prior to my PhD: Pr. Bernard Dujon, Dr. Romain Koszul, Pr. Susan Gerbi, and Dr. Lionel Navarro. Internships I performed in their labs made me discover different research approaches that helped me determine my choices.

I would like to thank everyone in my lab for their assistance and constant advices: Dubi, Sergei, Alex, Evelyn, Ariel, Hamady, Junho, Ken, Svetlana, Alicia, Yicheng, Irina and Katya. You all had a tremendous impact on the success of my work.

I'd like to thank all my friends stayed in France, for their absolute moral support far from home. I am very grateful to have them around me.

Finally, I would like thank my parents, my sister, and my brother. My family has always been supporting throughout my life and I would never have made it that far without them. Despite the distance, new means of communication that internet provide, allowed their presence to be comforting and irreplaceable.

Contents

Introduction 3
Chapter 1 - The role of Piwi in Drosophila germline nuclei 11
Chapter 2 - piRNA cluster promoter and chromatin 12
Chapter 3 - Selection of piRNA cluster precursors 27
Part 1 – *Drosophila melanogaster* 28
Part 2 – *Drosophila virilis* 43
Discussion 52
Materials & Methods 59
References 65
Annexes – Supplementary Tables 71
Annexes – Supplementary Figures 77

Introduction

Throughout the eukaryotic lineage, to ensure cellular division and differentiation, various intracellular mechanisms exist to regulate gene expression and protect the genome against the deleterious influence of exogenous and endogenous selfish genetic elements. One of those regulatory mechanisms is RNA silencing, which provides highly specific RNA transcripts inhibition or cleavage, through complementary recognition of RNA targets by small non-coding RNA molecules[1]. Small silencing RNAs in metazoans can be divided into three classes: microRNAs (miRNAs), small interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs). They are defined by their short length (~20–34 nucleotides), and their association with members of the Argonaute (AGO) family of proteins, which guide these short RNAs to their targets. These three classes differ in their biogenesis, their modes of target regulation and in the biological pathways they are involved in[2]. The piRNA pathway was discovered and started to be characterized a few years ago[3-5]. piRNAs represent a distinct class of small non-coding RNAs whose processing mechanism is different from that of siRNAs and microRNAs. It was found that piRNAs function to repress transposable elements, specifically in the germline cells of animal gonads.

Oogenesis

Each new generation of a multicellular organism, undergoing sexual reproduction, requires the intervention of the female germline and the male germline. The male germline (sperm) fertilizes the female germline (egg), producing a multicellular embryo which becomes a new individual. Drosophila is an excellent model system to study germ cells development and especially the female gonads. The female reproductive organ, the ovary, produces eggs constantly over the span of its life, and all stages of the oogenesis are present in sub-structures called ovariole[6]. There are around 15 to 20 ovarioles per ovary, which comprise two major cell types: the germline cells surrounded by the somatic follicular cells supporting and maintaining the germline cells. These two cell types are dispatched over the different developmental stages in the ovariole from the germarium up to the final egg (12 stages total). Oogenesis debuts at the anterior tip of the ovariole, in the germline stem cell (GSCs) niche of the germarium. The GSC niche consists of necessary somatic cells-terminal filament cells, cap cells, escort cells, and

other cells which function is to maintain the GSCs. There are 2 GSCs in the GSC niche, which are directly attached to somatic cap cells and escort stem cells. The asymmetrical division of a GSC leads to the formation of two daughter cells: a new GSC and a differentiated cystoblast which then undergoes four rounds of incomplete mitosis until it eventually gives rise to a mature egg chamber consisting of 15 nurse cells, and one oocyte, the future egg. The nurse cells actively transcribe their genome and provide the oocyte with cytoplasmic components. In contrast, the oocyte enters meiosis, and is mostly inactive transcriptionally.

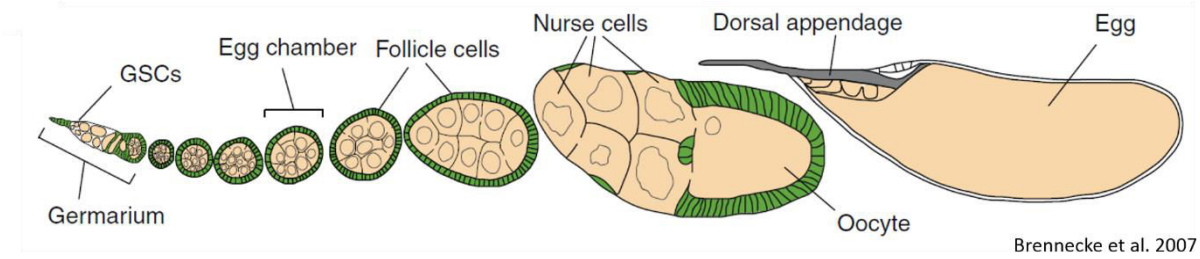


Figure: Cartoon depicting *Drosophila melanogaster* ovariole.

The germline cells' genome being the only one to be transmitted to the next generation, it is therefore of outmost importance to protect them from internal and external threats. Any flaws in the integrity of the germ cells genome can lead to sterility, or strong morphological defects and lethality of the progeny. It is essential for the oocyte to be able to transmit a correct genome, so the next generation can in turn ensure survival of the species.

One of the major internal threat to germline cells genome stability are transposable elements (TEs). These mobile genetic elements are repetitive parts of DNA that can move to insert at new locations in the genome[7,8]. They are very abundant in the *Drosophila melanogaster* genome, which consist of 15%–20% of TEs. Uncontrolled activity of TEs triggers DNA breaks causing cell cycle arrest, insertional mutagenesis causing gene disruption, and illegitimate recombination[9,10].

RNA interference

To protect themselves against such invasive genetic elements, Drosophila germline have established a very effective defense mechanism that involves one of the major small RNA interference pathway: the piRNAs pathway[11]. The piRNA pathway has a memory of past transposon invasions as well as an ability to mount acute responses to active transposons.

In animals, three major small RNA interference pathways have been identified: the microRNA (miRNA) pathway , the small interfering RNA (siRNA) pathway, and the Piwi interacting RNA (piRNA) pathway[2]. At the core of these three pathways lies a ribonucleoprotein complex consisting of a small RNA, responsible for the target recognition, and a member of the Argonaute protein family carrying the effector function. miRNAs and siRNAs are ubiquitously expressed and play important roles in post-transcriptional regulation of gene expression and defense against exogenous viral agents, respectively[12]. The microRNA pathway is believed to work mainly through prevention of translation of mRNAs which have complementarity to miRNAs. First encoded in the genome as primary miRNAs (pri-miRNA) and then cleaved first by Drosha inside the nucleus into a precursor miRNA (pre-miRNA). Secondly, after export to the cytoplasm, pre-miRNAs are cleaved into miRNA by Dicer (DCR-1), and the duplex is then loaded in AGO1 for further translational repression of target RNAs. The siRNA pathway responds to the presence of a double stranded RNA in the cell, which can be of either exogenous (e.g. viral) or endogenous (so called endo-siRNA) nature. It is active in most tissues and, besides being an object of extensive research, can be used as a powerful research tool: one can down-regulate expression of specific genes in a given tissue by expressing a dsRNA against these genes. In Drosophila, the siRNA pathway is first triggered by cleavage of a double strand (ds) RNA precursor (exogenous or endogenous origin) by a Dicer protein (DCR-2) into siRNA duplexes containing guide and passenger strands. They are then loaded into AGO2 to direct target cleavage.

miRNAs and siRNAs are processed from double-stranded or hairpin precursors by the type III ribonucleases Drosha and Dicer. In opposite, piRNAs are expressed exclusively in the gonads of animals, are processed from single-stranded precursors, and do not require the action of Dicer proteins. Three Argonaute effector proteins, the Piwi proteins, are active in the *Drosophila melanogaster* piRNA pathway, from which two, Aubergine (AUB) and Argonaute 3 (AGO3), are known to use their endonucleolytic activity to destroy transposon transcripts in the cytoplasm[4,13]. The third one, PIWI itself, is nuclear, but the molecular mechanism of its nuclear function remained unknown.

piRNA clusters

The biogenesis of piRNAs is far less elucidated than siRNAs and miRNAs. piRNAs are very diverse: hundreds of thousands of unique piRNAs are found that do not show any structure or sequence motif similarities[4]. In Drosophila, mapping of these piRNA sequences to the

genome revealed that piRNAs are coming from two types of genomic locations: the first and main source are discrete genomic loci, called piRNA clusters, while a much smaller fraction of piRNAs map to a hand-full of protein coding genes[14]. piRNA clusters are strongly enriched in repeats, transposon remnants, and are devoid of protein coding genes. Transposon remnants have accumulated enough mutations over time to lose their ability to move across the genome and represent a database of transposons that are to be recognized and silenced by the pathway: invasion of naïve host populations by a new transposon leads to uncontrolled jumping of the TE in the genome until a random integration occurs in a cluster leading to the generation of TE-derived piRNAs. Thus, the sequence of piRNA clusters is in constant flux to maintain up-to-date information about TE invasions[15]. Clusters can span up to 200kb and are located in pericentromeric and subtelomeric regions. Clusters are arranged in two constellations based on the orientation of their transcription: uni-directional clusters are transcribed in only one direction, while bi-directional clusters are transcribed convergently from two ends. Interestingly, the two cluster types differ in their expression pattern. Uni-directional clusters are mainly expressed in the somatic follicular cells of the Drosophila ovary, while bi-directional clusters are only transcribed in germline-derived nurse cells. P-element insertion at the beginning of the uni-directional *Flamenco* cluster disrupts piRNA expression up to 200kb downstream of the insertion sites, advocating for the presence of a single promoter unit responsible for the transcription of the whole cluster[4].

The second source of piRNAs, much lower in yield production, is protein coding genes. Some piRNAs are indeed mapping to the 3' untranslated region (UTR) of genes. The most prominent genic piRNAs come from the gene encoding for the transcription factor Traffic Jam (*tj*)[16].

Two distinct biogenesis pathways have been proposed for piRNAs: the primary piRNA processing pathway and the secondary pathway, or also called the ping-pong amplification loop.

The Primary pathway

As piRNAs can originate from single-stranded RNA precursors, their processing differs from that of siRNAs and miRNAs. It was proposed that long single-stranded precursor transcripts are cleaved by an endonuclease to generate the 5' end of the mature piRNAs[17]. Recently, a combination of genetic, biochemical and structural data identified Zucchini as a

candidate nuclease for piRNA 5' end processing[18,19]. Zucchini is a member of the phospholipase D family of phosphodiesterases and resides in the outer membrane of mitochondria. It cleaves the RNA leaving a 5' phosphate residue characteristic of piRNAs. Following this cleavage, which generates the 5' end of piRNAs, the piRNA precursors are thought to be loaded into Piwi proteins before their 3' end is trimmed by a yet unidentified factor that was named trimmer[20]. Interestingly, each Piwi protein is associated with piRNAs of a distinct size range[4,21]. It was proposed that trimmer stops when it reaches the region of the RNA protected by the footprint of the Piwi protein, generating piRNAs of different lengths specific to each Piwi protein[20]. After piRNAs are trimmed to the final size, the 3' end of the RNA is 2'-O-methyl-modified by Hen1, resulting in mature piRNAs[22]. piRNAs have a strong bias for a uridine residue as their 5' nucleotide. Currently, the origin of this bias is unknown: it can be caused by preferential cleavage by Zucchini or due to selective incorporation of piRNAs containing a 5' uridine into Piwi proteins. The latter model is supported by the observation that the 5' binding pocket of many Argonaute proteins can discriminate the first base[23,24].

The Ping-Pong loop

Many proteins that were implicated in piRNA biogenesis localize to *nuage* granules, cytoplasmic granular structures that are tightly associated with the outer nuclear membrane[25]. Accordingly, it is believed that many or all steps of piRNA biogenesis happen in *nuage* granules. In Drosophila, *nuage* granules contain two cytoplasmic Piwi proteins, AUB and AGO3. Computational analysis of piRNAs present in these two proteins revealed that Piwi proteins themselves can serve as nucleases that generate the 5' end of new piRNAs, in a process that was named the Ping-Pong amplification cycle[4,26]. AUB-loaded piRNAs recognize complementary transcripts (derived from active TEs or the opposite strand of the same piRNA cluster) and AUB cleaves them 10 nucleotides from the 5' end of the original piRNA. This forms the 5' end of a new piRNA, which is then incorporated into AGO3 and trimmed as described above. AGO3-associated piRNAs can in turn recognize complementary transcripts and cleave them, resulting in generation of new piRNAs that are identical in sequence to the initial piRNA that started the circle. Importantly, the Ping-Pong amplification loop is sensitive to target transposon expression and therefore might lead to amplification of piRNAs that target active elements.

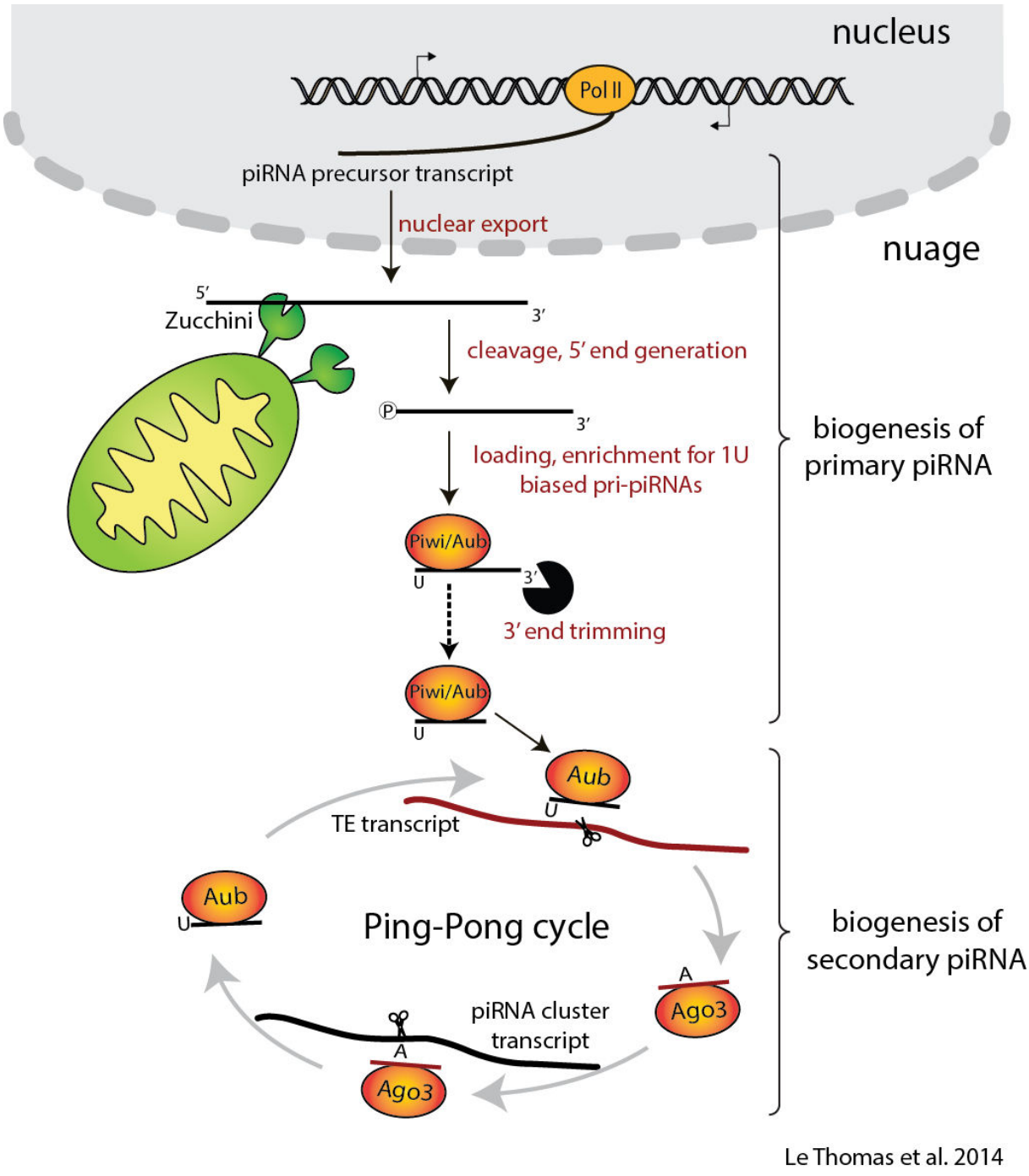


Figure: The two pathways involved in piRNA biogenesis.

PIWI function and piRNA cluster regulation

Piwi proteins and the associated piRNAs expressed in the maternal germline during oogenesis are deposited into the developing egg and are present in the early embryo before the start of zygotic transcription[27]. In Drosophila, such maternally inherited piRNAs are essential for effective piRNA-mediated silencing and fertility of the progeny. Recent studies of a long-known phenomenon called hybrid dysgenesis showed that the absence of piRNAs inherited from the previous generation causes a failure to silence TEs in the progeny[28].

Piwi proteins are essential to carry the effector function of the pathway, and are maternally transmitted. However, while a lot is known about the two cytoplasmic Piwis, AUB and AGO3, in post-transcriptional silencing, the mechanism of PIWI mediated transposons silencing in the nucleus is unknown. Loaded with piRNAs that are antisense to TE transcripts, defect in PIWI lead to up-regulation of TEs and sterility of the organism.

piRNA cluster expression is central for piRNA biogenesis, thus some elements should exist within that differentiate them from non-piRNA producing loci. Clusters precursor transcripts are channeled into the piRNA biogenesis pathway through an unknown mechanism. In comparison to miRNA or siRNA pathways, for which the biogenesis of small RNAs is clearly identified and described, it is currently not known, how piRNA cluster transcripts are selected for piRNA production.

The fact that only a very specific subset of transcripts is processed into piRNAs indicates that some features of either the genomic locus or the transcript itself must differentiate piRNA precursors from other cellular RNAs. Three hypotheses can be proposed on how piRNA-producing loci are identified. First, distinct sequence or structure motifs in precursor transcripts or DNA inside or around clusters could signal the biogenesis machinery to process the transcript into piRNAs. Second, the chromatin structure of piRNA clusters could be marking these genomic regions for differential processing of their transcripts. Finally, piRNAs that are inherited from the previous generation could be responsible for selection of regions for piRNA processing in the progeny. It is not excluded however that all these system work in concert.

PhD projects

Critical questions, regarding the role of the PIWI protein in the nucleus for TE silencing and the definition of piRNA producing loci (piRNA clusters), remain unanswered. Accordingly, the research projects I carried out during these three years of PhD aimed to characterize the mechanism of PIWI function in the nucleus, and aimed to uncover the specific features that make a piRNA cluster different from any other RNA producing locus. The work involved the use of several different techniques and different model organisms, *Drosophila melanogaster* and *Drosophila virilis*, in order to answer these central questions:

a) What is role of PIWI in the germline cells nuclei?
b) How piRNA clusters are transcribed and what is their chromatin structure?
c) How cluster transcripts are selected for piRNA biogenesis?

Chapter 1 - The role of Piwi in Drosophila germline nuclei

The Piwi protein is essential for the ovary integrity and was first discovered in a screen for factors affecting stem cell maintenance in Drosophila germ line[29]. Mutants have very small gonads, and were therefore named *P-element Induced WImpy testis*. Among all three Drosophila Piwis, PIWI itself, is the only one nuclear. Although its cleavage activity is not required to achieve silencing[30], in contrast to other Piwis AUB and AGO3, loss of PIWI results in transposons up-regulation, and a general arrest in the ovary development. PIWI function is therefore necessary for the survival of the species. However, how Piwi is carrying its role in silencing transposons in the nucleus, is not yet known.

Downloaded from genesdev.cshlp.org on April 10, 2013 - Published by Cold Spring Harbor Laboratory Press

Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state

Adrien Le Thomas,[1,2,3] Alicia K. Rogers,[1,3] Alexandre Webster,[1,3] Georgi K. Marinov,[1,3] Susan E. Liao,[1] Edward M. Perkins,[1] Junho K. Hur,[1] Alexei A. Aravin,[1,4] and Katalin Fejes Tóth[1,4]

[1]California Institute of Technology, Pasadena, California 91125, USA; [2]Université Pierre et Marie Curie, Ecole Doctorale Complexité du Vivant, 75005 Paris, France

In the metazoan germline, piwi proteins and associated piwi-interacting RNAs (piRNAs) provide a defense system against the expression of transposable elements. In the cytoplasm, piRNA sequences guide piwi complexes to destroy complementary transposon transcripts by endonucleolytic cleavage. However, some piwi family members are nuclear, raising the possibility of alternative pathways for piRNA-mediated regulation of gene expression. We found that *Drosophila* Piwi is recruited to chromatin, colocalizing with RNA polymerase II (Pol II) on polytene chromosomes. Knockdown of Piwi in the germline increases expression of transposable elements that are targeted by piRNAs, whereas protein-coding genes remain largely unaffected. Derepression of transposons upon Piwi depletion correlates with increased occupancy of Pol II on their promoters. Expression of piRNAs that target a reporter construct results in a decrease in Pol II occupancy and an increase in repressive H3K9me3 marks and heterochromatin protein 1 (HP1) on the reporter locus. Our results indicate that Piwi identifies targets complementary to the associated piRNA and induces transcriptional repression by establishing a repressive chromatin state when correct targets are found.

[*Keywords*: piRNA; Piwi; chromatin; RNA polymerase II; transcription; transposon]

Supplemental material is available for this article.

Received November 8, 2012; revised version accepted January 14, 2013.

Diverse small RNA pathways function in all kingdoms of life, from bacteria to higher eukaryotes. In eukaryotes, several classes of small RNA associate with members of the Argonaute protein family, forming effector complexes in which the RNA provides target recognition by sequence complementarity, and the Argonaute provides the repressive function. Argonaute–small RNA complexes have been shown to regulate gene expression both transcriptionally and post-transcriptionally. Post-transcriptional repression involves cleavage of target RNA through either the endonucleolytic activity of Argonautes or sequestering targets into cytoplasmic ribonucleoprotein (RNP) granules (Hutvagner and Simard 2008).

The mechanism of transcriptional repression by small RNAs has been extensively studied in fission yeast and plants. Several studies showed that Argonaute–small RNA complexes induce transcriptional repression by tethering chromatin modifiers to target loci. In fission yeast, the effector complex containing the Argonaute and the bound siRNA associates with the histone H3 Lys 9 (H3K9) methyltransferase Clr4 to install repressive H3K9-dimethyl marks at target sites (Nakayama et al. 2001; Maison and Almouzni 2004; Sugiyama et al. 2005; Grewal and Jia 2007). Methylation of histone H3K9 leads to recruitment of the heterochromatin protein 1 (HP1) homolog Swi6, enhancing silencing and further promoting interaction with the Argonaute complex. The initial association of Ago with chromatin, however, requires active transcription (Ameyar-Zazoua et al. 2012; Keller et al. 2012). Plants also use siRNAs to establish repressive chromatin at repetitive regions. Contrary to yeast, heterochromatin in plants is marked by DNA methylation, although repression also depends on histone methylation by a Clr4 homolog (Soppe et al. 2002; Onodera et al. 2005). Although siRNA-mediated gene silencing is predominant on repetitive sequences, it is not limited to these sites. Constitutive expression of dsRNA mapping to promoter regions results in production of corresponding siRNAs, de novo DNA methylation, and gene silencing (Mette et al. 2000; Matzke et al. 2004).

In metazoans, small RNA pathways are predominantly associated with post-transcriptional silencing. One class

[3]**These authors contributed equally to this work.**
[4]**Corresponding authors**
E-mail kft@caltech.edu
E-mail aravin@caltech.edu
Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.209841.112.

Downloaded from genesdev.cshlp.org on April 10, 2013 - Published by Cold Spring Harbor Laboratory Press

of small RNA, microRNA, regulates expression of a large fraction of protein-coding genes (Friedman et al. 2009). In *Drosophila*, siRNAs silence expression of transposable elements (TEs) in somatic cells (Chung et al. 2008; Ghildiyal et al. 2008) and target viral genes upon infection (Galiana-Arnoux et al. 2006; Wang et al. 2006; Zambon et al. 2006). Another class of small RNAs, Piwi-interacting RNAs (piRNAs), associates with the Piwi clade of Argonautes and acts to repress mobile genetic elements in the germline of both *Drosophila* and mammals (Siomi et al. 2011). Analysis of piRNA sequences in *Drosophila* revealed a very diverse population of small RNAs that primarily maps to transposon sequences and is derived from a number of heterochromatic loci called piRNA clusters, which serve as master regulators of transposon repression (Brennecke et al. 2007). Additionally, a small fraction of piRNAs seems to be processed from the mRNA of several host protein-coding genes (Robine et al. 2009; Saito et al. 2009). The *Drosophila* genome encodes three piwi proteins: Piwi, Aubergine (AUB), and Argonaute3 (AGO3). In the cytoplasm, AUB and AGO3 work together to repress transposons through cleavage of transposon transcripts, which are recognized through sequence complementarity by the associated piRNAs (Vagin et al. 2006; Agger et al. 2007; Brennecke et al. 2007; Gunawardane et al. 2007).

In both *Drosophila* and mammals, one member of the Piwi clade proteins localizes to the nucleus. Analogously to small RNA pathways in plants, the mouse piRNA pathway is required for de novo DNA methylation and silencing of TEs (Carmell et al. 2007; Aravin et al. 2008; Kuramochi-Miyagawa et al. 2008); however, the exact mechanism of this process is unknown. In *Drosophila*, DNA methylation is absent; however, several studies indicate that elimination of Piwi from the nucleus causes changes in histone marks on TEs (Klenov et al. 2011; Pöyhönen et al. 2012), yet a genome-wide analysis of Piwi's effect on chromatin marks and transcription is lacking.

Here we show that Piwi interacts with chromatin on polytene chromosomes in nurse cell nuclei. We found that Piwi exclusively represses loci that are targeted by piRNAs. We show that Piwi-mediated silencing occurs through repression of transcription and correlates with installment of repressive chromatin marks at targeted loci.

Results

To analyze the role of Piwi in the nucleus, we generated transgenic flies expressing a GFP-tagged Piwi protein (GFP-Piwi) under the control of its native regulatory region. GFP-Piwi was expressed in the ovary and testis in a pattern indistinguishable from the localization of native Piwi and was able to rescue the piwi-null phenotype as indicated by ovarian morphology, fertility, transposon expression, and piRNA levels. GFP-Piwi was deposited into the mature egg and localized to the pole plasm; however, contrary to a previous observation (Brower-Toland et al. 2007), we did not detect Piwi expression outside of the ovary and testis in third instar larvae or adult flies. We also did not observe the association of Piwi with polytene chromosomes in salivary gland cells of third instar larvae. In both follicular and germline cells of the *Drosophila* ovary, GFP-Piwi localized exclusively in the nucleus, with slightly higher concentrations apparent in regions enriched for DAPI, indicating a possible interaction with chromatin. To gain further insight into Piwi localization in the nucleus, we took advantage of the fact that nurse cell chromosomes are polytenized and can be visualized on the *otu* mutant background (Mal'ceva et al. 1997). Analysis of polytene chromosomes from nurse cells demonstrated that GFP-Piwi associates with chromatin in a specific banding pattern. Interestingly, coimmunostaining showed that a GFP-Piwi signal on polytene chromosomes generally overlaps with the RNA polymerase II (Pol II) signal, which marks sites of active transcription (Fig. 1A).

In order to identify factors that might be responsible for targeting Piwi to chromatin, we immunoprecipitated Piwi complexes from the *Drosophila* ovary and analyzed Piwi interaction partners by mass spectrometry. We purified Piwi complexes from ovaries of three different transgenic lines expressing GFP-Piwi, myc-Piwi, or Flag-Piwi using antibodies against each respective tag. As a control, we used flies expressing free GFP in the ovary.

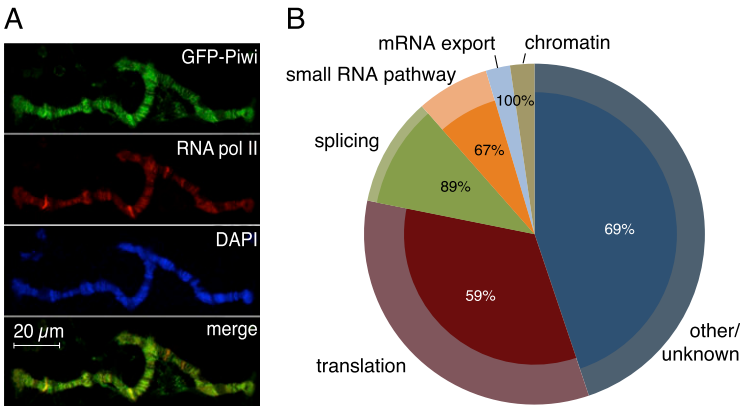


Figure 1. Piwi associates with chromatin and nuclear transcripts. (*A*) Polytene chromosomes from *Drosophila* nurse cells expressing GFP-Piwi on the *otu[7]/otu[11]* background. Piwi pattern on chromosomes correlates with Pol II staining. (*B*) Mass spectrometry analysis of Piwi interaction partners. Piwi complexes were precipitated in the presence and absence of RNase A. The outer circle represents classification of Piwi-associated proteins based on GO term analysis. The inner pies represent the fraction of each group whose association with Piwi depends on RNA (percentage indicated). Note that chromatin, splice, and mRNA export factors are virtually absent after RNase A treatment.

Downloaded from genesdev.cshlp.org on April 10, 2013 - Published by Cold Spring Harbor Laboratory Press

We identified >50 factors that showed significant enrichment in all three Piwi purifications but were absent in the control. We were unable to identify chromatin-associated factors that directly associate with Piwi but identified several RNA-binding proteins that associate with nascent transcripts, such as splicing (Rm62, Pep, Ref1, Yps, CG9684, CG31368, CG5728, and Mago) and nuclear export (Tho2 and Hpr1) factors (Fig. 1B). Upon RNase A treatment prior to immunoprecipitation, the presence of most of these RNA-binding proteins in purified Piwi complexes was eliminated.

Piwi proteins are believed to find their targets through sequence complementarity of the associated piRNA. In fact, it has been proposed that lack of the associated piRNA leads to destabilization of piwi proteins and to Piwi's inability to localize to the nucleus (Saito et al. 2009; Haase et al. 2010; Olivieri et al. 2010; Handler et al. 2011; Ishizu et al. 2011). On the other hand, Piwi has been proposed to have functions that are independent of its role in transposon control by regulating stem cell niche development (Cox et al. 1998; Klenov et al. 2011). To address the role of piRNA in translocation of Piwi into the nucleus and its function, we generated transgenic flies expressing a point mutant Piwi—referenced as Piwi-YK—that is deficient in piRNA binding due to a substitution of two conserved amino acid residues (Y551L and K555E) in the 5′ phosphate-binding pocket (Kiriakidou et al. 2007; Djuranovic et al. 2010). The Piwi-YK mutant was expressed in *Drosophila* follicular and germ cells at levels similar to that of wild-type Piwi but was completely devoid of associated piRNA (Fig. 2A). In contrast to wild-type Piwi, Piwi-YK could be found in the cytoplasm, supporting the existence of a quality control mechanism that prevents entrance of unloaded Piwi into the nucleus (Ishizu et al. 2011). Nevertheless, a significant amount of piRNA-deficient Piwi localized to the nucleus (Fig. 2B). Similar to wild-type Piwi, Piwi-YK seemed to associate with chromatin, as indicated by its localization in DAPI-stained regions of the nuclei, and this is consistent with fluorescence loss in photobleaching (FLIP) experiments that demonstrated reduced nuclear mobility compared with free diffusion (Supplemental Fig. S1). Based on sterility and ovarian morphology, the *piwi-YK* transgene was unable to rescue the piwi-null phenotype despite its nuclear localization (Fig. 2C), indicating that while piRNA binding is not absolutely essential for stability and nuclear localization of Piwi, it is required for Piwi function.

To directly test the function of Piwi in the nucleus, we analyzed the effect of Piwi deficiency on gene expression and chromatin state on a genome-wide scale. Piwi mutant females have atrophic ovaries caused by Piwi deficiency in somatic follicular cells (Lin and Spradling 1997; Cox et al. 1998), which precludes analysis of Piwi function in null mutants. Instead, we used RNAi knockdown to deplete Piwi in germ cells while leaving it functionally intact in somatic follicular cells. The Piwi knockdown flies did not exhibit gross morphological defects in the ovary; however, they showed drastic reduction in GFP-Piwi expression in germ cells and were sterile (Fig. 3A,B).

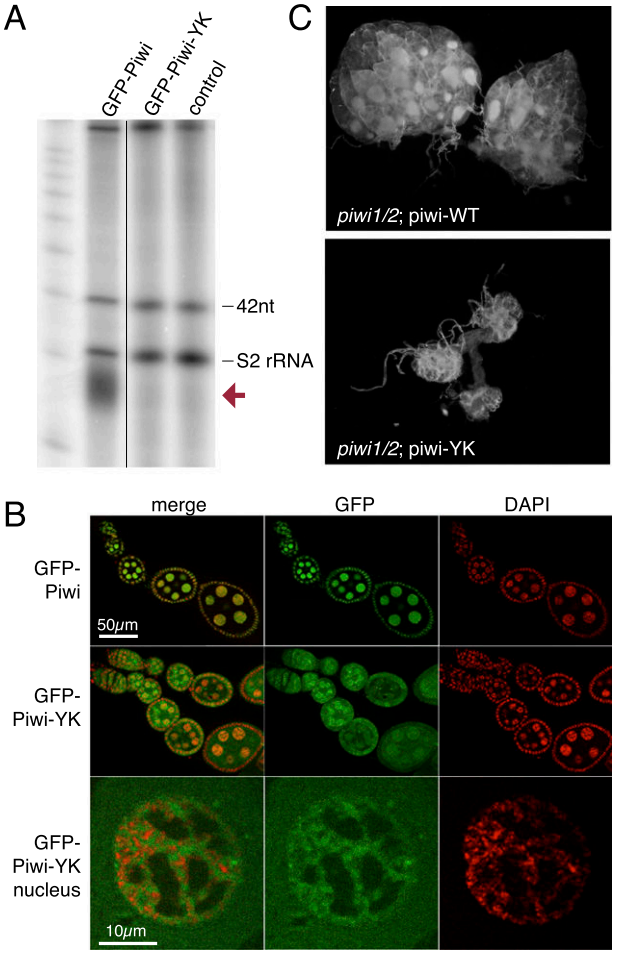


Figure 2. Piwi function, but not its nuclear localization, requires piRNA association. (*A*) The Piwi-YK mutant does not associate with piRNA. Immunoprecipitation of Piwi–piRNA complexes was performed with GFP antibody on ovaries from GFP-Piwi and GFP-Piwi-YK transgenic flies and a control strain. Small RNAs were isolated, 5′-labeled, and resolved on a denaturing gel. The same amount of 42-nucleotide RNA oligonucleotides was spiked into all samples prior to RNA isolation to control for loss of RNA during isolation and labeling. piRNAs (red arrow) are absent in the Piwi-YK complex. (*B*) GFP-Piwi-YK is present in the nuclei of nurse cells and colocalizes with chromatin (DAPI-stained areas). (*C*) The Piwi-YK mutant does not rescue the morphological changes caused by the piwi-null mutation. Dark-field images of ovaries where either the wild-type *piwi* or the *piwi-YK* transgene has been backcrossed onto the piwi-null background.

To analyze the effect of Piwi deficiency on the steady-state transcriptome as well as the transcription machinery, we performed RNA sequencing (RNA-seq) and Pol II chromatin immunoprecipitation (ChIP) combined with deep sequencing (ChIP-seq) experiments from Piwi knockdown and control flies.

In agreement with previous observations that implicated Piwi in transposon repression (Saito et al. 2006; Aravin et al. 2007; Brennecke et al. 2007), we found that steady-state transcript levels of several TEs were increased

Downloaded from genesdev.cshlp.org on April 10, 2013 - Published by Cold Spring Harbor Laboratory Press

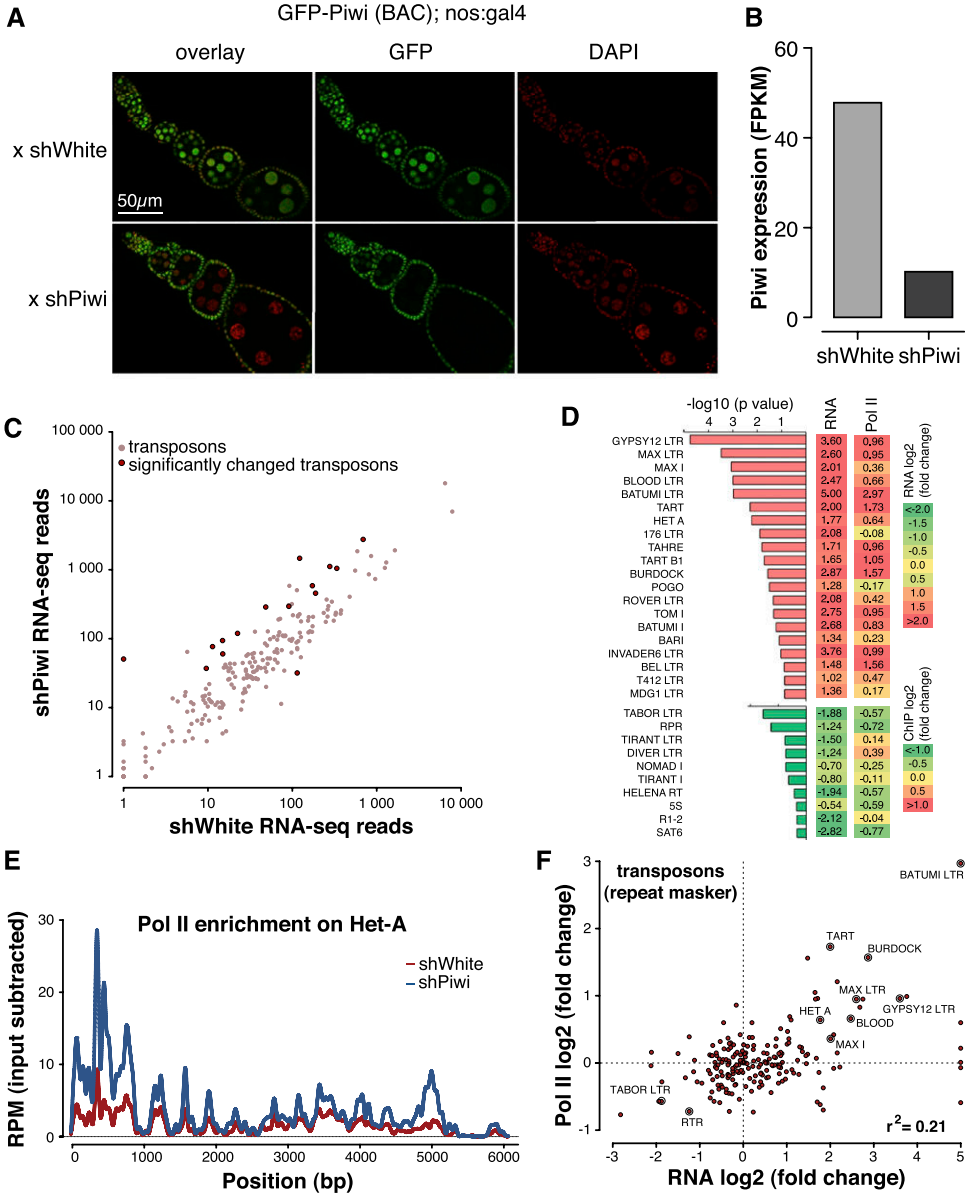


Figure 3. Piwi transcriptionally represses TEs. (*A*) Piwi knockdown is efficient and specific to ovarian germ cells as indicated by GFP-Piwi localization. GFP-Piwi; Nanos-Gal4-VP16 flies were crossed to control shRNA (shWhite) or shPiwi lines. Piwi is specifically depleted in germ cells and not in follicular cells, consistent with expression of the Nanos-Gal4-VP16 driver. (*B*) Piwi expression as measured by RNA-seq in the Piwi knockdown and control lines. Note that Piwi expression is unaffected in follicular cells, leading to relatively weak apparent knockdown in RNA-seq libraries from whole ovaries. (*C*) Effect of Piwi knockdown on the expression of TEs. Two biological replicate RNA-seq experiments were carried out, and differential expression was assessed using DESeq. Transposons that show significant change ($P < 0.05$) are indicated by dark-red circles. Out of 217 individual RepeatMasker-annotated TEs, 15 show a significant increase in expression upon Piwi knockdown. (*D*) The change in the levels of TE transcripts and Pol II occupancy on their promoters upon Piwi knockdown. Twenty up-regulated and 10 down-regulated transposons with the most significant changes in expression level are shown. Note the low statistical significance for down-regulated transposons. For a complete list of transposons, see Supplemental Figure S2. (*E*) Pol II signal over the Het-A retrotransposon in control flies (shWhite; red) and upon Piwi knockdown (shPiwi; blue). (*F*) Increased abundance of transposon transcripts upon Piwi depletion correlates with increased Pol II occupancy over their promoters ($r^2 = 0.21$). Note that the majority of elements do not show significant change in either RNA abundance or Pol II occupancy.

upon Piwi knockdown in germ cells (Fig. 3C,D; Supplemental Fig. S2). We found little to no change of RNA levels for transposons whose activity is restricted to follicular cells of the ovary, indicating that the observed changes are indeed due to loss of Piwi in the germline (Supplemental Fig. S2). The analysis of Pol II ChIP-seq showed that Pol II occupancy increased over promoters of multiple TEs (Fig. 3D–F; Supplemental Fig. S3). Indeed,

Downloaded from genesdev.cshlp.org on April 10, 2013 - Published by Cold Spring Harbor Laboratory Press

the change in steady-state levels of transposon transcripts upon Piwi depletion correlated with changes of Pol II occupancy (Fig. 3F). This result demonstrates that Piwi ensures low levels of transposon transcripts through a repressive effect on the transcription machinery.

To test whether Piwi-mediated transcriptional repression is accompanied by a corresponding change in chromatin state, we used ChIP-seq to analyze the genome-wide distribution of the repressive H3K9me3 mark in the ovary upon Piwi knockdown. We identified 705 genomic loci at which the level of H3K9me3 significantly decreased. More than 90% of the regions that show a decrease in the H3K9me3 mark upon Piwi depletion overlapped TE sequences, compared with the 33% that is expected from random genome sampling (Fig. 4A). Furthermore, these regions tend to be located in the heterochromatic portions of the genome that are not assembled on the main chromosomes (Fig. 4B). Only 20 of the identified regions localized to the euchromatic parts of the genome. Of these, 15 (75%) contained potentially active annotated copies of transposons. Taken together, our results indicate that Piwi is required for installment of repressive H3K9me3 chromatin marks on TE sequences of the genome.

While the vast majority of protein-coding host genes did not show significant changes in transcript level or Pol II occupancy upon Piwi knockdown, the expression of a small set of protein-coding genes (150 genes with a

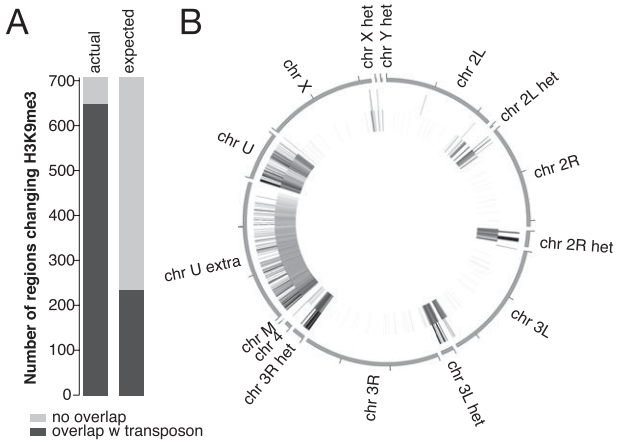


Figure 4. Piwi-induced transcriptional repression correlates with establishment of a repressive chromatin state. (*A*) Overlap between genomic regions of H3K9me3 depletion upon Piwi knockdown and TEs. Two replicates of H3K9me3 ChIP-seq experiments were carried out on control and Piwi-depleted ovaries, and enriched regions were identified using DESeq (see the Materials and Methods for details). A total of 705 regions show significant ($P < 0.05$) decrease in H3K9me3 occupancy upon Piwi knockdown, while only 30 regions showed a similarly significant increase. Out of the 705 regions that show a decrease in H3K9me3 marks upon Piwi knockdown, 91% (646) overlap with TE sequences compared with the 33% expected from random genome sampling. (*B*) Genomic positions of H3K9me3-depleted regions upon Piwi depletion (outer circle) and RepeatMasker-annotated transposons (inner circle). Note that almost all regions are localized in heterochromatic and repeat-rich portions of the genome (Het, chrU, and chrUExtra chromosomes).

P-value <0.05) was significantly increased (Fig. 5A; Supplemental Table 1). There are several possible explanations for Piwi's effect on host gene expression. First, failure in the piRNA pathway might cause up-regulation of several genes that generate piRNAs in wild-type ovaries (Robine et al. 2009; Saito et al. 2009). However, the genes up-regulated in Piwi-deficient ovaries were not enriched in piRNAs compared with other genes. Second, H3K9me3 marks installed on TE sequences in a Piwi-dependent manner might spread into neighboring host genes and repress their transcription, as was recently demonstrated in a follicular cell culture model (Sienski et al. 2012). To address this possibility, we analyzed genomic positions of the genes whose expression was increased upon Piwi knockdown relative to genomic regions that showed a decrease in H3K9me3 marks. We found that up-regulated genes did not show a significant change in the H3K9me3 mark (Fig. 5B; Supplemental Fig. S4). Furthermore, the few genes located close to the regions that show a decrease in H3K9me3 signal had unaltered expression levels upon Piwi knockdown. Next, we analyzed the functions of up-regulated genes using gene ontology (GO) term classifications and found significant enrichment for proteins involved in protein turnover and stress and DNA damage response pathways (Fig. 5C). Particularly, we found that 31 subunits of the proteasome complex were overexpressed. Therefore, our analysis indicates that up-regulation of specific host genes is likely a secondary response to elevated transposon levels and genomic damage.

In contrast to host genes, transcripts of TEs are targeted by piRNA. To directly address the role of piRNA in Piwi-mediated transcriptional silencing, we took advantage of a fly strain that expresses artificial piRNAs against the *lacZ* gene, which are loaded into Piwi complexes and are able to repress *lacZ* reporter expression in germ cells (Fig. 6A; Josse et al. 2007; Muerdter et al. 2012). Expression of piRNAs that are antisense to the reporter gene caused transcriptional silencing of the *lacZ* gene as measured by Pol II occupancy (Fig. 6B). Furthermore, we found that piRNA-induced silencing of the reporter gene was associated with an increase in the repressive H3K9me3 mark and HP1 occupancy and a decrease in the abundance of the active H3K4me2/3 marks at the reporter locus (Fig. 6C). This result is in good agreement with the genome-wide effect of Piwi depletion on distribution of the H3K9me3 mark and suggests that transcriptional silencing correlates with the establishment of a repressive chromatin structure and is mediated by piRNAs that match the target locus.

Discussion

Little is known about the function of nuclear piwi proteins. The nuclear piwi in mice (Miwi2) affects DNA methylation of TEs (Carmell et al. 2007; Aravin et al. 2008; Kuramochi-Miyagawa et al. 2008). Several recent reports implicate *Drosophila* Piwi in regulation of chromatin marks on transposon sequences (Lin and Yin 2008; Klenov et al. 2011; Wang and Elgin 2011; Sienski et al. 2012). The mechanism of these processes is unknown in

Downloaded from genesdev.cshlp.org on April 10, 2013 - Published by Cold Spring Harbor Laboratory Press

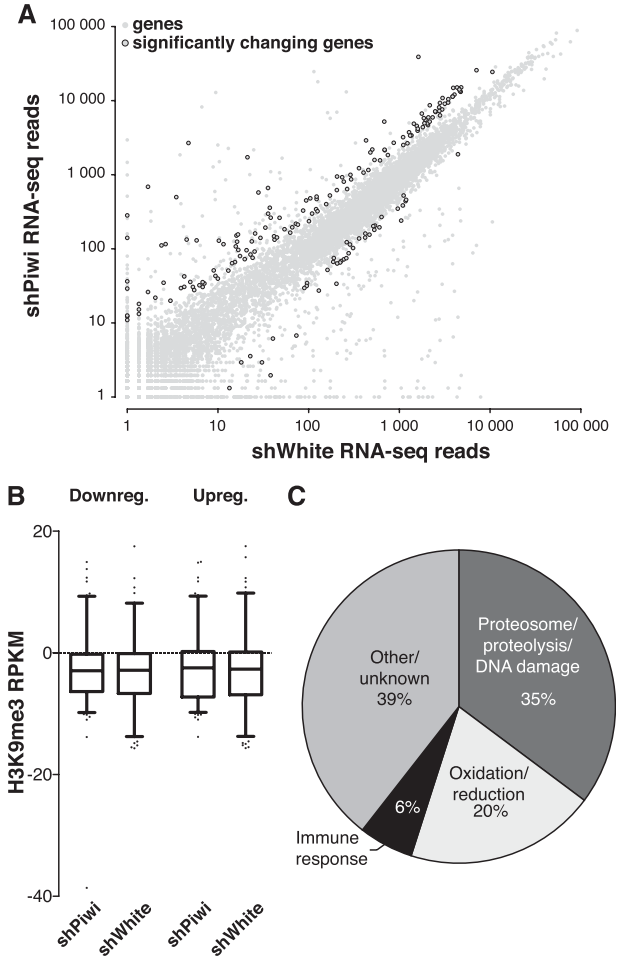


Figure 5. Piwi does not directly repress protein-coding genes. (*A*) Effect of Piwi knockdown on the expression of genes. Two replicate RNA-seq experiments were carried out, and differential expression was assessed using DESeq. Genes that show significant change ($P < 0.05$) are indicated by black circles. The vast majority of genes does not change significantly upon germline Piwi knockdown (shPiwi) compared with control (shWhite). (*B*) H3K9me3 mark density does not change over genes that show a significant change in expression upon Piwi knockdown (see Fig. 3C). Up-regulated and down-regulated genes are plotted separately. Signal indicated is after background subtraction. (*C*) Functional analysis of up-regulated genes by the Database for Annotation, Visualization, and Integrated Discovery (DAVID) reveals activation of the protein degradation and DNA damage response pathways. Percentages of all up-regulated genes are indicated.

both organisms. Previously, Piwi was shown to associate with polytene chromosomes in salivary gland cells and colocalize with HP1, a chromodomain protein that binds to heterochromatin and a few loci in euchromatin, suggesting that HP1 mediates Piwi's interaction with chromatin (Brower-Toland et al. 2007). However, recent results showed that the putative HP1-binding site on Piwi is dispensable for Piwi-mediated transposon silencing (Wang and Elgin 2011).

We did not detect Piwi expression outside of the ovary and testis, including in salivary gland cells, using a GFP-Piwi transgene expressed under native regulatory elements. We detected GFP-Piwi on polytene chromosomes in ovarian nurse cells that have a germline origin; however, it localizes in a pattern that largely does not overlap with HP1. FLIP experiments with GFP-Piwi indicated a relatively fast rate of fluorescence redistribution as compared with histone H2A (Supplemental Fig. S1), implying a transient interaction of Piwi with chromatin. Our proteomic analysis of Piwi complexes isolated from *Drosophila* ovaries did not identify chromatin-associated factors but revealed several RNA-binding proteins, such as splicing and nuclear export factors that bind nascent RNA transcripts (Fig. 1B). Importantly, the interaction of most of these RNA-binding proteins with Piwi was dependent on RNA, indicating that Piwi associates with nascent transcripts. As Piwi itself lacks DNA- and RNA-binding domains (beyond the piRNA-binding domain),

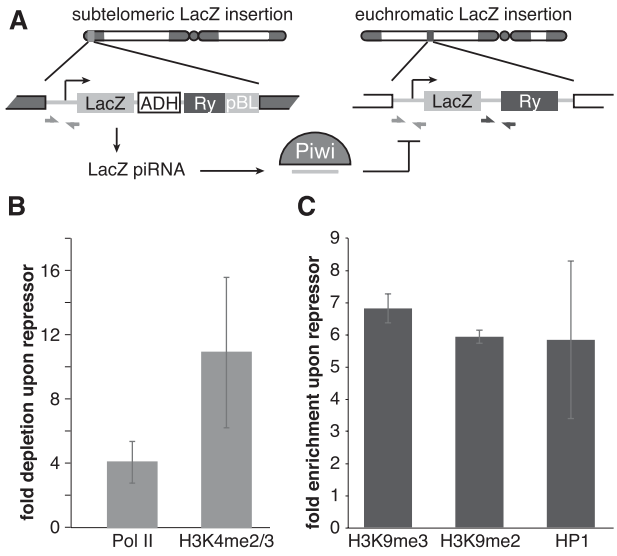


Figure 6. piRNA-dependent targeting of Piwi to a reporter locus leads to establishment of a repressive chromatin state and transcriptional silencing. (*A*) The mechanism of *trans*-silencing mediated by artificial piRNA and a schematic representation of the repressor and reporter *lacZ* constructs. The repressor construct is inserted in a subtelomeric piRNA cluster, leading to generation of piRNA from its sequence. Primers mapping to both constructs used for the Pol II and H3K4me2/3 ChIP-quantitative PCR (qPCR) are shown by light-gray arrows; primers specific to the reporter locus used for the H3K9me3, H3K9me2, and HP1 ChIP-qPCR are indicated by dark-gray arrows. (*B*) piRNAs induce transcriptional repression of the *lacZ* reporter. Pol II and H3K4me2/3 signals decreased on the *lacZ* promoter in the presence of artificial piRNAs as measured by ChIP-qPCR. Shown is the fold depletion of signal in flies that carry both repressor and reporter constructs compared with control flies that have only the reporter construct. The signal was normalized to RP49. (*C*) piRNAs induce an increase in H3K9me3 and H3K9me2 marks and HP1 binding as measured by ChIP-qPCR. Shown is the fold increase of corresponding ChIP signals downstream from the *lacZ* reporter in flies that carry both repressor and reporter constructs compared with control flies that have only reporter construct. The signal was normalized to RP49.

Downloaded from genesdev.cshlp.org on April 10, 2013 - Published by Cold Spring Harbor Laboratory Press

it is likely that the recruitment of Piwi to chromatin is through interactions with other RNA-binding proteins or sequence-specific interactions between Piwi-bound piRNA and nascent transcripts.

Using specific Piwi knockdown in germ cells of the *Drosophila* ovary, we analyzed the effect of Piwi depletion on gene expression, the transcription machinery, and H3K9me3 chromatin marks genome-wide. In agreement with previous results (Klenov et al. 2011), we found up-regulation of several TEs upon Piwi knockdown (Fig. 3C). The TEs that did not change their expression upon germline knockdown of Piwi might be expressed exclusively in somatic follicular cells of the ovary, such as the *gypsy* retrotransposon. Alternatively, some elements present in the genome might not have transcriptionally active copies, or the cytoplasmic AUB/AGO3 proteins may efficiently silence them at the post-transcriptional level.

The increase in steady-state levels of RNA upon Piwi depletion strongly correlates with an increase in Pol II occupancy on the promoters of transposons (Fig. 3D,F; Supplemental Fig S2). This result suggests that Piwi represses transposon expression at the transcriptional level, although we cannot completely exclude the possibility of an additional post-transcriptional effect. It was shown previously that depletion or mutation of Piwi leads to depletion of the repressive H3K9me3 mark and an increase in the active H3K4me2/3 marks on several transposon sequences (Klenov et al. 2011; Wang and Elgin 2011). Our ChIP-seq data extend these results to a genome-wide scale, proving that transposons are indeed the sole targets of Piwi, and demonstrate that changes in histone marks directly correlate with transcriptional repression.

Piwi depletion in the germline does not affect expression of the majority of host genes, although a small fraction of genes changes expression (Fig. 5A). One possible mechanism of the effect Piwi has on host genes is the spreading of repressive chromatin structure from transposon sequences to adjacent host genes. Indeed, such a spreading and the resulting repression of host gene transcription were observed in an ovarian somatic cell (OSC) culture model (Sienski et al. 2012). However, we did not find significant changes in the H3K9me3 mark for genes that are up-regulated upon germline depletion of Piwi, arguing against this mechanism playing a major role in host gene regulation. Instead, we found that the majority of host genes whose expression is increased as a result of Piwi depletion participate in protein turnover (e.g., proteasome subunits) and stress and DNA damage response pathways, indicating that they might be activated as a secondary response to cellular damage induced by transposon activation. The different effect of Piwi depletion on host gene expression in ovary and cultured cells might be explained by the fact that silencing of host genes due to transposon insertion would likely have a strong negative effect on the fitness of the organism but could be tolerated in cultured cells. Accordingly, new transposon insertions that cause repression of adjacent host genes should be eliminated from the fly population but can be detected in cultured cells. In agreement with this explanation, the majority of cases of repressive chromatin spreading in OSCs were observed for new transposon insertions that are absent in the sequenced *Drosophila* genome. Indeed, it was shown that the vast majority of new transposon insertions is present at a low frequency in the *Drosophila* population, likely due to strong negative selection (Petrov et al. 2003). Such selection was primarily attributed to the ability of TE sequences to cause recombination and genomic rearrangements. We propose that in addition to the effects on recombination, the selection against transposons can be driven by their negative impact on host gene expression in the germline linked to Piwi-mediated chromatin silencing.

How does Piwi discriminate its proper targets—transposons—from host genes? In the case of cytoplasmic Piwi proteins AUB and AGO3, recognition and post-transcriptional destruction of TE transcripts is guided by associated piRNAs. Our results indicate that piRNAs provide guidance for transcriptional silencing by the nuclear Piwi protein as well. First, in contrast to host genes that are not targeted by piRNAs, TE transcripts, which are regulated by Piwi, are recognized by antisense Piwi-bound piRNA (Brennecke et al. 2007). Second, a Piwi mutant that is unable to bind piRNA failed to rescue the piwi-null mutation despite its ability to enter the nucleus. Finally, expression of artificial piRNAs that target a reporter locus induced transcriptional silencing associated with an increase in repressive H3K9me3 and HP1 chromatin marks and a decrease in the active H3K4me2/3 marks (Fig. 6B,C). In contrast, the tethering of Piwi to chromatin in a piRNA-independent fashion by fusing Piwi with the lacI DNA-binding domain that recognizes lacO sequences inserted upstream of a reporter gene did not lead to silencing of the reporter (data not shown). Together, our results demonstrate that piRNAs are the essential guides of Piwi to recognize its targets for transcriptional repression.

It is tempting to propose that, similar to Argonautes in fission yeast, *Drosophila* Piwi directly recruits the enzymatic machinery that establishes the repressive H3K9me3 mark on its targets. Establishment of repressive marks can lead to stable chromatin-based transcriptional silencing that does not require further association of Piwi with target loci. This model explains why we found that Piwi is relatively mobile in the nucleus, indicative of only a transient interaction with chromatin. The Piwi-mediated transcriptional silencing has an interesting parallel in *Caenorhabditis elegans*, where the Piwi protein PRG-1 and associated 21U RNAs are able to induce stable transgenerational repression that correlates with formation of silencing chromatin marks on target loci. Interestingly, PRG-1 and 21U RNAs are necessary only for initial establishment of silencing, while continuing repression depends on siRNA and the WAGO group of Argonautes (Ashe et al. 2012; Bagijn et al. 2012; Buckley et al. 2012; Shirayama et al. 2012). Future studies should reveal the pathway that leads to transcriptional repression downstream from Piwi in *Drosophila* and the differences from and similarities to other species.

Downloaded from genesdev.cshlp.org on April 10, 2013 - Published by Cold Spring Harbor Laboratory Press

Materials and methods

Drosophila *stocks*

Nanos-Gal4-VP16 (BL4937), *UASp-shWhite* (BL33623), *UASp-shPiwi* (BL 33724), and Chr. I and II Balancer (BL7197) were purchased from the Bloomington Stock Center. GFP-Piwi-expressing flies (see below) were backcrossed onto the *piwi1*/*piwi2* (available from Bloomington Stock Center) background or the *otu7*/*otu11* (available from Bloomington Stock Center) background, respectively. *LacZ* reporter lines were a generous gift from S. Ronsseray.

Generation of transgenic fly lines

The GFP-Piwi, 3xFlag-HA-Piwi, and myc-Piwi constructs were generated using bacterial recombineering (Gene Bridges Counter Selection kit) to insert the respective tag after the start codon of the Piwi genomic region cloned in BAC clone BACN04M10. The KpnI–XbaI genomic fragment that contains the Piwi gene and flanking sequences was transferred to corresponding sites of the pCasper4 vector to create pCasper4/tagged Piwi.

The pCasper4/GFP-Piwi construct was used to generate pCasper4/GFP-Piwi-YK with two point mutations, Y551I and K555E. Mutations were introduced by PCR, amplifying products corresponding to a 3.1-kb upstream fragment and a 2.58-kb downstream fragment. The upstream fragment included a unique XbaI site at the 5′ end of the amplicon and overlapped 39 base pairs (bp) with the downstream fragment, which included a unique BamHI site at its 3′ end. The single XbaI–BamHI fragment was generated by overlap PCR with outside primers and cloned into corresponding sites of pCasper4/GFP-Piwi to replace the wild-type fragment. Transgenic flies were generated by P-element-mediated transformation (BestGene).

Immunoprecipitation of Piwi proteins and RNA gel of piRNA

Dissected ovaries were lysed in lysis buffer (20 mM HEPES at pH 7.0, 150 mM KCl, 2.5 mM MgCl, 0.5% Triton X-100, 0.5% Igepal, 100 U/mL RNasin [Promega], EDTA-free Complete Protease Inhibitor Cocktail [Roche]) and supernatant clarified by centrifugation. Supernatant was incubated with anti-eGFP polyclonal antibody (Covance) conjugated to Protein-G Dynabeads at 4°C. Beads were spiked with 5 pmol of synthesized 42-nucleotide RNA oligomer to assess purification efficiency, proteinase K-digested, and phenol-extracted. Isolated RNA was CIP-treated, radiolabeled using PNK and γ-P32-labeled ATP, and run on a 15% urea-PAGE gel. Western blots of ovary lysate and anti-eGFP immunoprecipitates were obtained from 8% SDS-PAGE gels and probed with polyclonal rabbit anti-eGFP antibody to confirm expression of the full-length transgene.

Mass spectrometric analysis of Piwi interaction partners

Lysis and clarification of ovary samples were performed as described above using lysis buffer with reduced detergent (0.1% Triton X-100, 0.1% Igepal). Piwi proteins with Flag, Myc, or GFP tag were purified from *Drosophila* ovaries using corresponding antibodies covalently coupled to M-270 epoxy Dynabeads (Invitrogen) (Cristea et al. 2005). Immunoprecipitation of free GFP from GFP-expressing ovaries was used as a negative control. Immunoprecipitations were performed in the presence or absence of RNase A (100 μg/mL; 30 min at 25C). Piwi and copurified interacting proteins were resolved on NuPAGE Novex 4%–12% Bis-Tris gels and stained with colloidal Coomassie blue. Gel fragments that contained protein bands were excised and in-gel-trypsinized, and the peptides were extracted following the standard protocol of the Proteome Exploration Laboratory at California Institute of Technology. Peptide analyses were performed on an LTQ-FT Ultra (Thermo Fisher Scientific) equipped with a nanoelectrospray ion source (Thermo Fisher Scientific) connected to an EASY-nLC. Fractionation of peptides was performed on a 15-cm reversed-phase analytical column (75-μm internal diameter) in-house-packed with 3-μm C18 beads (ReproSil-Pur C18-AQ medium; Dr. Maisch GmbH). Acquired spectra were searched against the *Drosophila melanogaster* proteome using the search engine Mascot (Matrix Science, version 2.2.06), and protein inferences were performed using Scaffold (Proteome Software, version 3). For an Excel file of Piwi interaction partners, see the Supplemental Material.

ChIP, ChIP-seq, and RNA-seq

ChIP was carried out using standard protocols (Moshkovich and Lei 2010). ChIP-seq and RNA-seq library construction and sequencing were carried out using standard protocols following the general principles described by Johnson et al. (2007) and Mortazavi et al. (2008), respectively. Data analysis was carried out using a combination of publicly available software tools and custom-written python scripts. Additional details regarding high-throughput data analysis are described in the Supplemental Material. For quantitative PCR (qPCR) primers, see Supplemental Table 2. GO term analysis of genes up-regulated upon Piwi knockdown was performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Huang et al. 2009a,b) and FlyBase for additional assignment of GO terms. Sequencing data is available through Gene Expression Omnibus (accession no. GSE43829).

Antibodies

eGFP antibody (rabbit polyclonal serum; Covance) was affinity-purified in our laboratory. Anti-myc (Millipore), anti-Flag (Sigma), Pol II (ab5408), and Pol II pSer5 (ab5131) are commercially available.

Imaging of ovaries

Ovaries were fixed in 4% PFA in PBS for 20 min, permeabilized in 1% Triton X-100 in PBS, DAPI-stained (Sigma-Aldrich), washed, and mounted in 50% glycerol/PBS. Images were captured using an AxioImager microscope; an Apotome structured illumination system was used for optical sections (Carl Zeiss).

FLIP

FLIP time series were captured on an LSM510 confocal microscope equipped with a 40×/0.9 NA Imm Corr multi-immersion objective. Ovaries were dissected into halocarbon 700 oil (Sigma) and mounted under a 0.17-mm coverslip (Carl Zeiss) immediately before imaging. Two initial baseline images were captured, followed by 80–100 iterations consisting of two bleach iterations at 100% laser power (488 nm or 543 nm for GFP- and RFP-tagged proteins, respectively), followed by two images with reduced illumination intensity. FLIP series were cropped and median-filtered with a 2-pixel radius to reduce noise using FIJI (Schindelin et al. 2012) and the "Rigid Body" function of the StackReg plugin (Thévenaz et al. 1998) to correct drift when needed. Using Matlab software (The Mathworks), images were background-subtracted and corrected for acquisition bleaching. A value representing the true loss of intensity relative to the initial prebleach images, where 0 indicates no change in

Downloaded from genesdev.cshlp.org on April 10, 2013 - Published by Cold Spring Harbor Laboratory Press

intensity and 1 represents complete photobleaching, was calculated for each pixel and each bleach/capture cycle and plotted with a color lookup table and calibration bar. Scale bars and annotations were made in Inkscape (http://inkscape.org).

Preparation of polytene squashes for immunofluorescence

Flies carrying the GFP-Piwi BAC construct were backcrossed onto the *otu[7]* and *otu[11]* background. Progeny from the cross of the two lines were grown at 18°C. Stage 7–12 egg chambers were separated and transferred to a polylysine-coated microscopic slide into PBST. From here, the "smush" protocol was followed (Johansen et al. 2009), but PFA cross-linking was reduced to 10 min. Slides were imaged using an AxioImager microscope and a 63× oil immersion objective (Carl Zeiss).

Acknowledgments

We are grateful to Evelyn Stuwe from the Aravin laboratory for purifying the GFP antibody; I. Antoshechkin of the Millard and Muriel Jacobs Genetics and Genomics Laboratory for sequencing; D. Trout, H. Amrhein, and S. Upchurch for computational assistance; the Bloomington Stock Center for fly stocks; and S. Hess, B. Graham, and M Sweredoski from the Proteome Exploration Laboratory at the Beckmann Institute, California Institute of Technology, for assistance with the mass spectrometry experiments. We thank members of the Aravin laboratory for critical comments on the manuscript. We thank Barbara Wold and members of the Wold laboratory for helpful discussions on ChIP protocols and analysis. A.K.R. and E.M.P. are supported by the Institutional Training Grant NIH/NRSA 5T32 GM07616, and E.M.P. is additionally supported by the Gordon Ross Medical Foundation. G.K.M. is supported by The Beckman Foundation, the Donald Bren Endowment, and NIH grant U54 HG004576. This work was supported by grants from the National Institutes of Health (R01 GM097363, R00 HD057233, and DP2 OD007371A to A.A.A.), the Searle Scholar Award (to A.A.A.), and the Ellison Medical Foundation New Scholar in Aging Award (to K.F.T.).

References

Agger K, Cloos P, Christensen J, Pasini D, Rose S, Rappsilber J, Issaeva I, Canaani E, Salcini A, Helin K. 2007. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449:** 731–734.

Ameyar-Zazoua M, Rachez C, Souidi M, Robin P, Fritsch L, Young R, Morozova N, Fenouil R, Descostes N, Andrau J-C et al. 2012. Argonaute proteins couple chromatin silencing to alternative splicing. *Nat Struct Mol Biol* **19:** 998–1004.

Aravin A, Hannon G, Brennecke J. 2007. The Piwi–piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318:** 761–764.

Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, Bestor T, Hannon GJ. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* **31:** 785–799.

Ashe A, Sapetschnig A, Weick E-M, Mitchell J, Bagijn M, Cording A, Doebley A-L, Goldstein L, Lehrbach N, Le Pen J et al. 2012. piRNAs can trigger a multigenerational epigenetic memory in the germline of *C. elegans*. *Cell* **150:** 88–99.

Bagijn M, Goldstein L, Sapetschnig A, Weick E-M, Bouasker S, Lehrbach N, Simard M, Miska E. 2012. Function, targets, and evolution of *Caenorhabditis elegans* piRNAs. *Science* **337:** 574–578.

Brennecke J, Aravin A, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon G. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128:** 1089–1103.

Brower-Toland B, Findley S, Jiang L, Liu L, Yin H, Dus M, Zhou P, Elgin S, Lin H. 2007. *Drosophila* PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev* **21:** 2300–2311.

Buckley B, Burkhart K, Gu S, Spracklin G, Kershner A, Fritz H, Kimble J, Fire A, Kennedy S. 2012. A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. *Nature* **489:** 447–451.

Carmell M, Girard A, van de Kant H, Bourc'his D, Bestor T, de Rooij D, Hannon G. 2007. MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* **12:** 503–514.

Chung W-J, Okamura K, Martin R, Lai E. 2008. Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. *Curr Biol* **18:** 795–802.

Cox D, Chao A, Baker J, Chang L, Qiao D, Lin H. 1998. A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev* **12:** 3715–3727.

Cristea IM, Williams R, Chait BT, Rout MP. 2005. Fluorescent proteins as proteomic probes. *Mol Cell Proteomics* **4:** 1933–1941.

Djuranovic S, Zinchenko M, Hur J, Nahvi A, Brunelle J, Rogers E, Green R. 2010. Allosteric regulation of Argonaute proteins by miRNAs. *Nat Struct Mol Biol* **17:** 144–150.

Friedman R, Farh K, Burge C, Bartel D. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19:** 92–105.

Galiana-Arnoux D, Dostert C, Schneemann A, Hoffmann J, Imler J-L. 2006. Essential function in vivo for Dicer-2 in host defense against RNA viruses in *Drosophila*. *Nat Immunol* **7:** 590–597.

Ghildiyal M, Seitz H, Horwich M, Li C, Du T, Lee S, Xu J, Kittler E, Zapp M, Weng Z et al. 2008. Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* **320:** 1077–1081.

Grewal S, Jia S. 2007. Heterochromatin revisited. *Nat Rev Genet* **8:** 35–46.

Gunawardane L, Saito K, Nishida K, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi M. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. *Science* **315:** 1587–1590.

Haase A, Fenoglio S, Muerdter F, Guzzardo P, Czech B, Pappin D, Chen C, Gordon A, Hannon G. 2010. Probing the initiation and effector phases of the somatic piRNA pathway in *Drosophila*. *Genes Dev* **24:** 2499–2504.

Handler D, Olivieri D, Novatchkova M, Gruber F, Meixner K, Mechtler K, Stark A, Sachidanandam R, Brennecke J. 2011. A systematic analysis of *Drosophila* TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. *EMBO J* **30:** 3977–3993.

Huang DW, Sherman B, Lempicki R. 2009a. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37:** 1–13.

Huang DW, Sherman B, Lempicki R. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4:** 44–57.

Hutvagner G, Simard M. 2008. Argonaute proteins: Key players in RNA silencing. *Nat Rev Mol Cell Biol* **9:** 22–32.

Ishizu H, Nagao A, Siomi H. 2011. Gatekeepers for Piwi–piRNA complexes to enter the nucleus. *Curr Opin Genetic Dev* **21:** 484–490.

Downloaded from genesdev.cshlp.org on April 10, 2013 - Published by Cold Spring Harbor Laboratory Press

Johansen K, Cai W, Deng H, Bao X, Zhang W, Girton J, Johansen J. 2009. Polytene chromosome squash methods for studying transcription and epigenetic chromatin modification in *Drosophila* using antibodies. *Methods* **48:** 387–397.

Johnson D, Mortazavi A, Myers R, Wold B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316:** 1497–1502.

Josse T, Teysset L, Todeschini A-L, Sidor C, Anxolabéhère D, Ronsseray S. 2007. Telomeric *trans*-silencing: An epigenetic repression combining RNA silencing and heterochromatin formation. *PLoS Genetic* **3:** 1633–1643.

Keller C, Adaixo R, Stunnenberg R, Woolcock KJ, Hiller S, Buhler M. 2012. HP1(Swi6) mediates the recognition and destruction of heterochromatic RNA transcripts. *Mol Cell* **47:** 215–227.

Kiriakidou M, Tan G, Lamprinaki S, De Planell-Saguer M, Nelson P, Mourelatos Z. 2007. An mRNA m7G cap binding-like motif within human Ago2 represses translation. *Cell* **129:** 1141–1151.

Klenov M, Sokolova O, Yakushev E, Stolyarenko A, Mikhaleva E, Lavrov S, Gvozdev V. 2011. Separation of stem cell maintenance and transposon silencing functions of Piwi protein. *Proc Natl Acad Sci* **108:** 18760–18765.

Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri T et al. 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* **22:** 908–917.

Lin H, Spradling A. 1997. A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. *Development* **124:** 2463–2476.

Lin H, Yin H. 2008. A novel epigenetic mechanism in *Drosophila* somatic cells mediated by Piwi and piRNAs. *Cold Spring Harb Symp Quant Biol* **73:** 273–281.

Maison C, Almouzni G. 2004. HP1 and the dynamics of heterochromatin maintenance. *Nat Rev Mol Cell Biol* **5:** 296–304.

Mal'ceva N, Belyaeva E, King R, Zhimulev I. 1997. Nurse cell polytene chromosomes of *Drosophila melanogaster* otu mutants: Morphological changes accompanying interallelic complementation and position effect variegation. *Dev Genetic* **20:** 163–174.

Matzke M, Aufsatz W, Kanno T, Daxinger L, Papp I, Mette M, Matzke A. 2004. Genetic analysis of RNA-mediated transcriptional gene silencing. *Biochim Biophys Acta* **1677:** 129–141.

Mette M, Aufsatz W, van der Winden J, Matzke M, Matzke A. 2000. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J* **19:** 5194–5201.

Mortazavi A, Williams B, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat Methods* **5:** 621–628.

Moshkovich N, Lei E. 2010. HP1 recruitment in the absence of argonaute proteins in *Drosophila*. *PLoS Genetics* **6:** e1000880.

Muerdter F, Olovnikov I, Molaro A, Rozhkov N, Czech B, Gordon A, Hannon G, Aravin A. 2012. Production of artificial piRNAs in flies and mice. *RNA* **18:** 42–52.

Nakayama J, Rice J, Strahl B, Allis C, Grewal S. 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292:** 110–113.

Olivieri D, Sykora M, Sachidanandam R, Mechtler K, Brennecke J. 2010. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *EMBO J* **29:** 3301–3317.

Onodera Y, Haag J, Ream T, Costa Nunes P, Pontes O, Pikaard C. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120:** 613–622.

Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE. 2003. Size matters: Non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Mol Biol Evol* **20:** 880–892.

Pöyhönen M, de Vanssay A, Delmarre V, Hermant C, Todeschini A, Teysset L, Ronsseray S. 2012. Homology-dependent silencing by an exogenous sequence in the *Drosophila* germline. *G3 (Bethesda)* **2:** 331–338.

Robine N, Lau N, Balla S, Jin Z, Okamura K, Kuramochi-Miyagawa S, Blower M, Lai E. 2009. A broadly conserved pathway generates 3′UTR-directed primary piRNAs. *Curr Biol* **19:** 2066–2076.

Saito K, Nishida K, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi M. 2006. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev* **20:** 2214–2222.

Saito K, Inagaki S, Mituyama T, Kawamura Y, Ono Y, Sakota E, Kotani H, Asai K, Siomi H, Siomi M. 2009. A regulatory circuit for piwi by the large Maf gene traffic jam in *Drosophila*. *Nature* **461:** 1296–1299.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B et al. 2012. Fiji: An open-source platform for biological-image analysis. *Nat Methods* **9:** 676–682.

Shirayama M, Seth M, Lee H-C, Gu W, Ishidate T, Conte D, Mello C. 2012. piRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. *Cell* **150:** 65–77.

Sienski G, Dönertas D, Brennecke J. 2012. Transcriptional silencing of transposons by piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151:** 964–980.

Siomi M, Sato K, Pezic D, Aravin A. 2011. PIWI-interacting small RNAs: The vanguard of genome defence. *Nat Rev Mol Cell Biol* **12:** 246–258.

Soppe WJ, Jasencakova Z, Houben A, Kakutani T, Meister A, Huang M, Jacobsen S, Schubert I, Fransz P. 2002. DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in *Arabidopsis*. *EMBO J* **21:** 6549–6559.

Sugiyama T, Cam H, Verdel A, Moazed D, Grewal S. 2005. RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. *Proc Natl Acad Sci* **102:** 152–157.

Thévenaz P, Ruttimann U, Unser M. 1998. A pyramid approach to subpixel registration based on intensity. *IEEE Trans Image Process* **7:** 27–41.

Vagin V, Sigova A, Li C, Seitz H, Gvozdev V, Zamore P. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313:** 320–324.

Wang S, Elgin S. 2011. *Drosophila* Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proc Natl Acad Sci* **108:** 21164–21169.

Wang X-H, Aliyari R, Li W-X, Li H-W, Kim K, Carthew R, Atkinson P, Ding S-W. 2006. RNA interference directs innate immunity against viruses in adult *Drosophila*. *Science* **312:** 452–454.

Zambon RA, Vakharia VN, Wu LP. 2006. RNAi is an antiviral immune response against a dsRNA virus in *Drosophila melanogaster*. *Cell Microbiol* **8:** 880–889.

Chapter 2 - piRNA cluster promoter and chromatin

piRNAs originate from peri-centromeric and subtelomeric loci, called piRNA clusters, which are transcribed as continuous units producing long RNA precursors later processed into piRNAs. Previous studies demonstrate that the heterochromatic nature of germline piRNA clusters is very important for piRNA production[31-33].

It was, however, never precisely described how different the chromatin landscape is between the two piRNA cluster types (bi-directional and uni-directional), and how does that compare with the rest of the genome. In addition, how their transcription is starting is also not known. Here, we try to bring light into those two questions.

Germline bi-directional and uni-directional piRNA clusters have a distinct chromatin structure.

In order to understand the chromatin status of piRNA clusters we performed ChIP-seq for a variety of active and repressive histone marks[34] together with the HP1a protein in *Drosophila melanogaster* ovaries. The profiling of these marks along bi-directional piRNA clusters (Fig. 1), expressed in the germline, shows an evident depletion for active marks, such as H3K36me1, H3K36me3, and also H3K4me2/3, whereas a strong enrichment for repressive marks is observed, as seen for HP1a, H3K9me2, and more specifically H3K9me3 and Rhino.

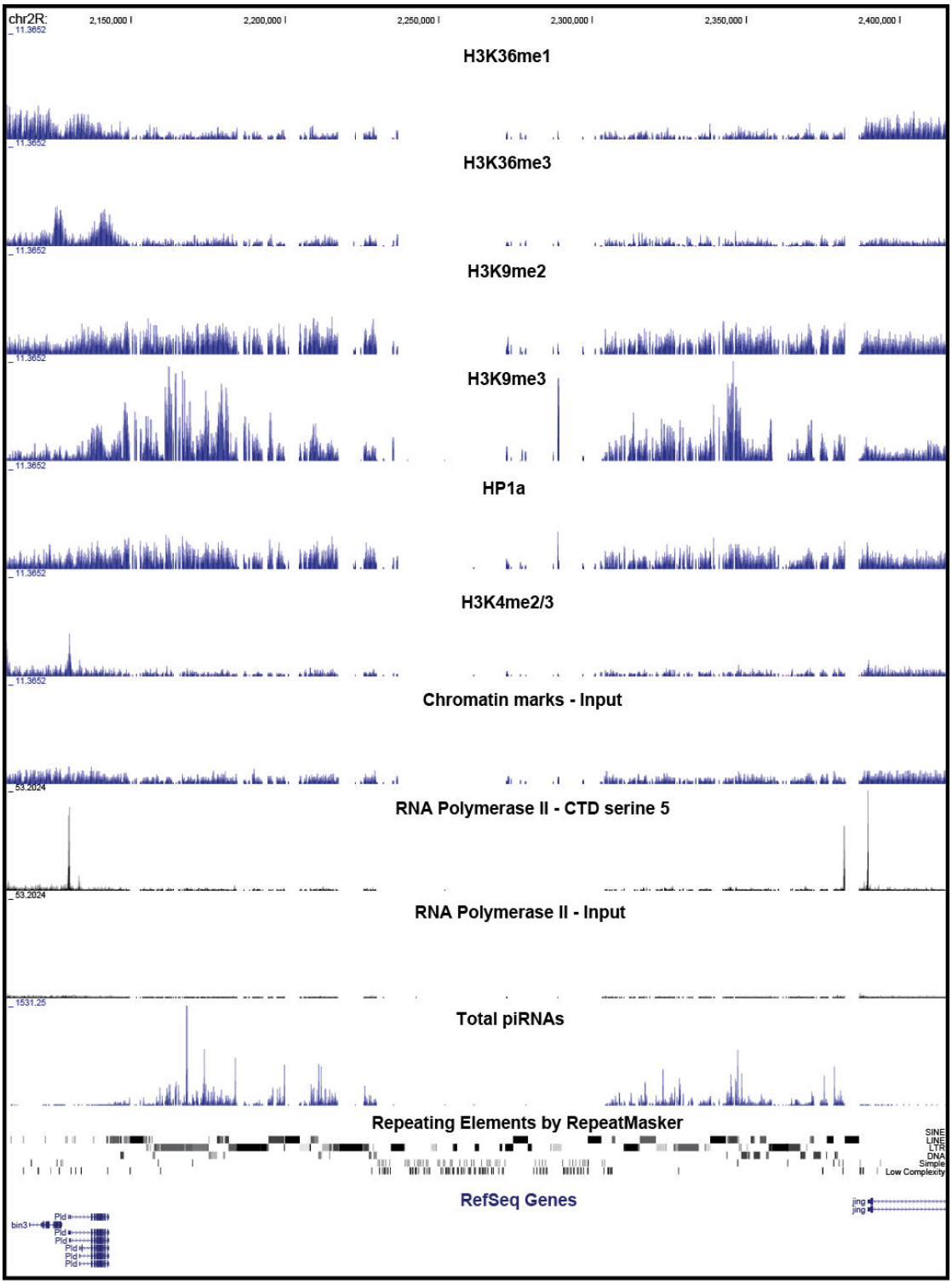


Figure 1. Germline Bi-directional piRNA clusters chromatin structure.
Chromatin marks and RNA Polymerase II densities at the bi-directional piRNA cluster 42AB (delimited by the piRNA density--Total piRNA track) in *Drosophila melanogaster*. An enrichment for repressive chromatin marks (H3K9me2, H3K9me3, and HP1) is observed along the cluster. Active marks, H3K4me2/3, H3K36me1, and H3K36me3, are depleted along the cluster. Repressive marks tend to spread in the vicinity of the cluster at immediate flanking regions, especially overlapping with the last exons of the Pld gene that are signaled by the active H3K36me3 chromatin mark.

To quantitatively compare the chromatin of piRNA clusters with other genomic regions, we defined enrichment of H3K9me3 (repressive) and H3K36me3 (active) chromatin marks (defined as a ratio of depth-normalized read number in ChIP to input) for prominent piRNA clusters in the genome, compared to their immediate genomic environment (Fig. 2). All bi-directional clusters, including the most prominent cluster at region 42AB, show a significantly higher H3K9me3 mark enrichment and lower H3K36me3 mark enrichment than their neighboring regions. The only uni-directional cluster expressed in the germline, X upstream, in contrary shows a totally opposite profile. These results indicate that high level of H3K9me3 mark is an important characteristic of piRNA clusters.

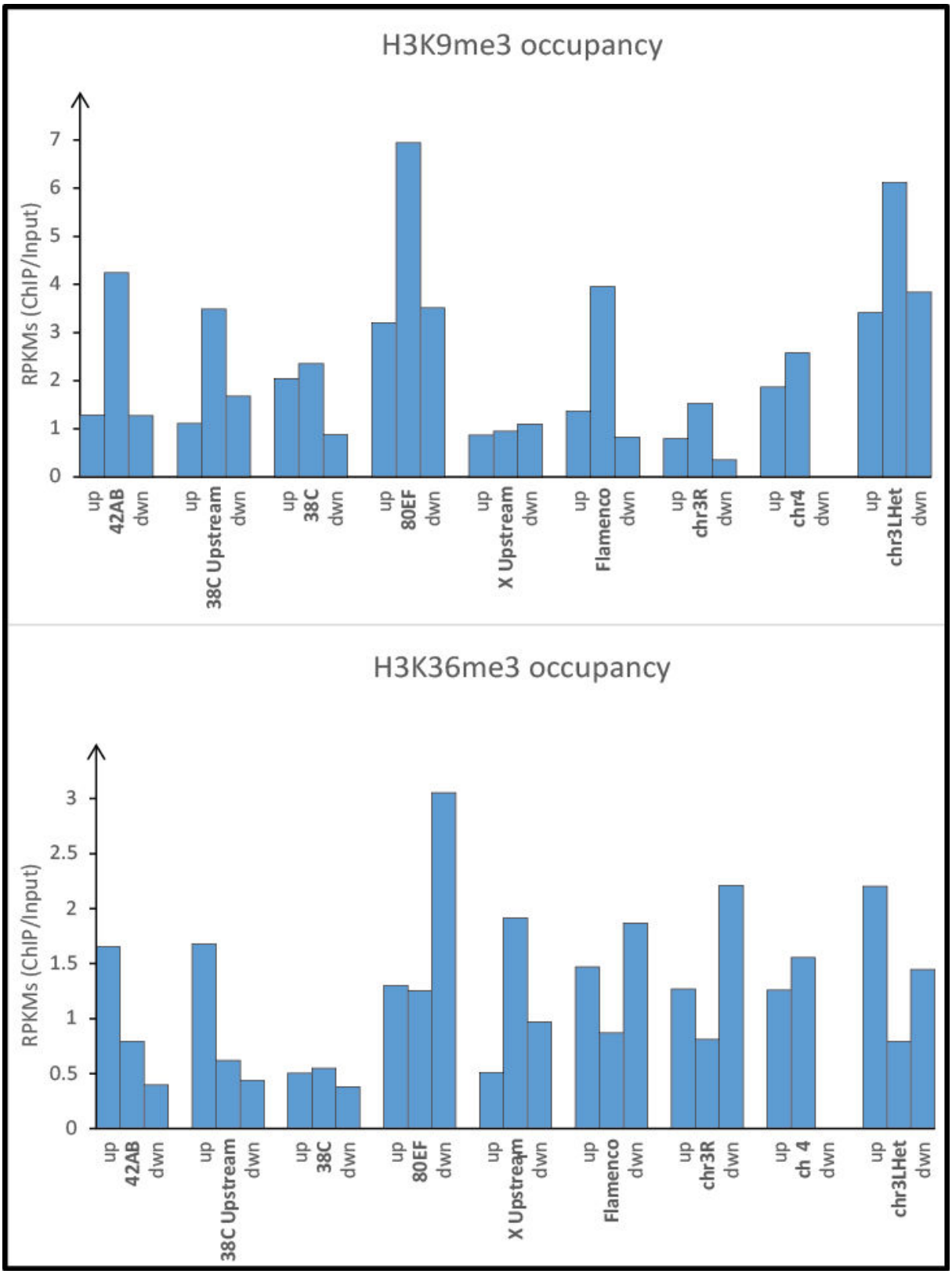


Figure 2. Germline piRNA clusters chromatin environment.
Chromatin marks changes around known piRNA clusters. Taking the list of major piRNA clusters[4], RPKMs values (IP/Input) were calculated for a 50kb region flanking each cluster upstream (up) and downstream (dwn), and for the cluster itself. The repressive mark H3K9me3 is depleted around cluster whereas the active mark H3K36me3 is regained.

In addition, we used all the marks we profiled to group genomic regions using Self-Organized Maps (SOM) (Fig. 3). SOM allows for unbiased clustering of genomic regions that possess similar chromatin features. Subsequently, the chromatin characteristics can be extracted for any group of genomic features (such as promoters, exons, transposons or clusters, and compared to the chromatin of other genomic entities). As expected, promoters of protein coding genes are characterized by enrichment of the H3K4me2/3 mark and occupancy of Serine 5 phosphorylated RNA Pol II. Genomic sequences occupied by protein-coding genes group together and are defined by high level of H3K36me3 mark found on exons. In contrast, piRNA clusters occupy a different territory on SOM maps that overlap with regions characterized by high level of H3K9me3 mark and HP1. As the vast majority of protein-coding genes are located in euchromatin, we decided to compare the chromatin of piRNA clusters with that of expressed protein-coding genes located in heterochromatin. We defined heterochromatic genes by a high fraction of repetitive sequences in their environment (both up- and downstream of the gene and in their introns). Heterochromatic genes did not show a significant enrichment for the H3K9me3 mark and HP1a. Therefore, the chromatin signature of clusters is clearly distinct from that of the majority of protein-coding genes.

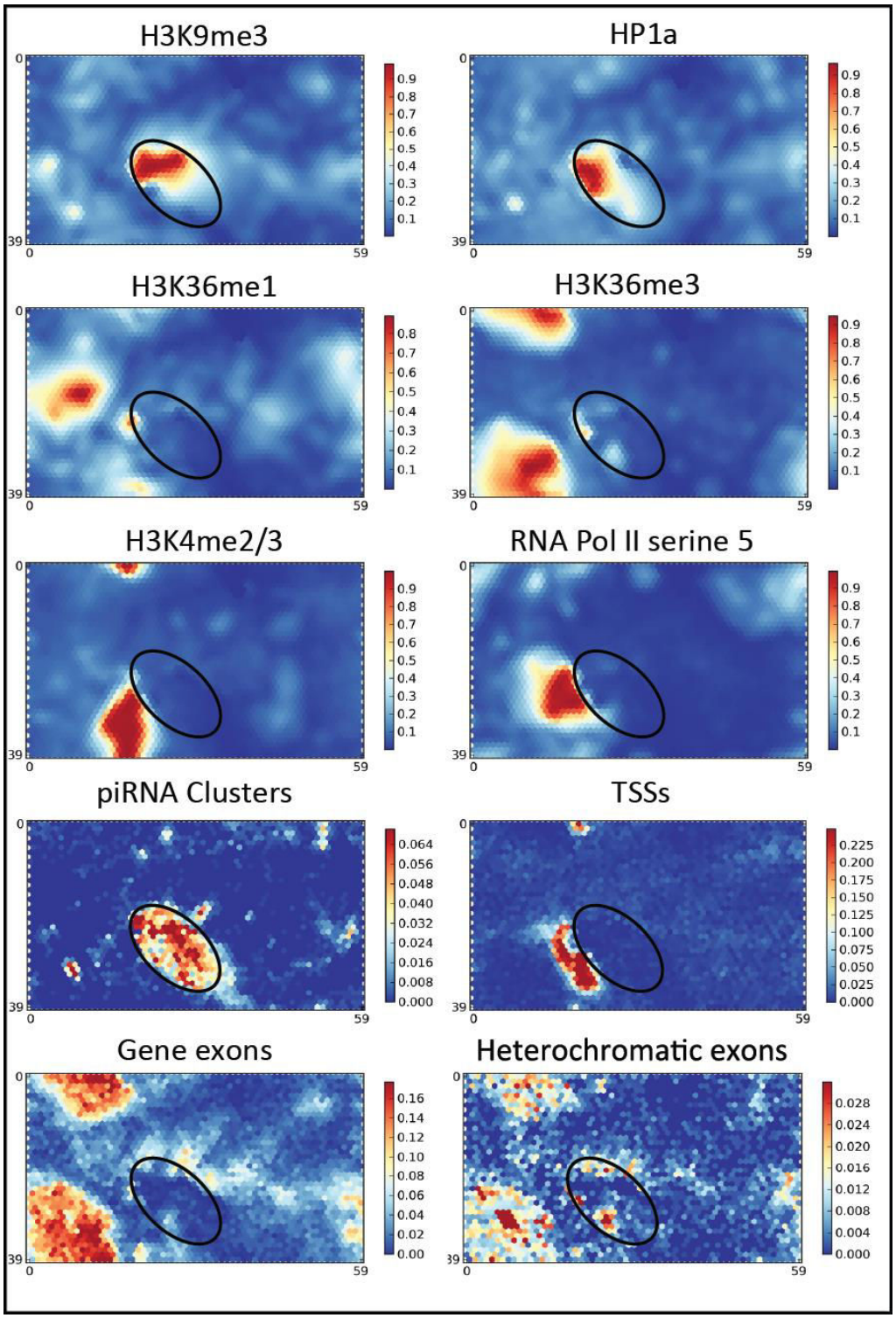


Figure 3. Chromatin marks and RNA Polymerase II densities clustered genome-wide: Self-Organizing Maps (SOMs).

The first six represent the enrichment of each individual chromatin mark at different locations on the same background map covering the whole genome. Each dots on the maps represent an ensemble of 100bp genome bins that have the same densities for each chromatin mark. The more similar the dots are from each other, the closer they are. Clusterization was performed 10 times, only the best score was kept to generate the final SOM. Signal density is calculated by RPKMs normalized to mappability of 50bp reads. In addition, were mapped on this SOM: piRNAs clusters, TSSs, gene exons, and heterochromatic gene exons (last 4 maps). The black ellipse represent the location of piRNA cluster bins for easy comparisons in maps.

In contrast, the only germline expressed uni-directional piRNA cluster, X upstream, displays a dramatically different chromatin pattern (Fig. 4). Active marks H3K4me2/3 and H3K36me3 are enriched respectively at the putative promoter region and along the body of the cluster. On the other hand, repressive marks are all depleted along the cluster. We also did RNA-seq where RNA transcripts were Poly-A selected prior to library preparation and sequencing. We were surprised to see that the X upstream cluster appears to be the only piRNA cluster that has a strong RNA-seq signal all along its body. This result indicates that the precursor transcript arising from this piRNA cluster has a poly-A tail, and a poly-A site must be present in its genomic sequence (Fig. 4). This feature along, with the presence of active chromatin marks, suggests that, unlike any other piRNA cluster, the X upstream genomic region and precursor mostly behave like a regular gene. The X upstream cluster has the singularity to be expressed, not only in germ cells, but also in somatic follicular cells of the ovary. Therefore, it is conceivable that these particular features are necessary to ensure proper expression of the cluster in both cell types. However, other characteristics, absent to genic regions, must be present at the X upstream cluster to channel the precursor transcript into the piRNA pathway.

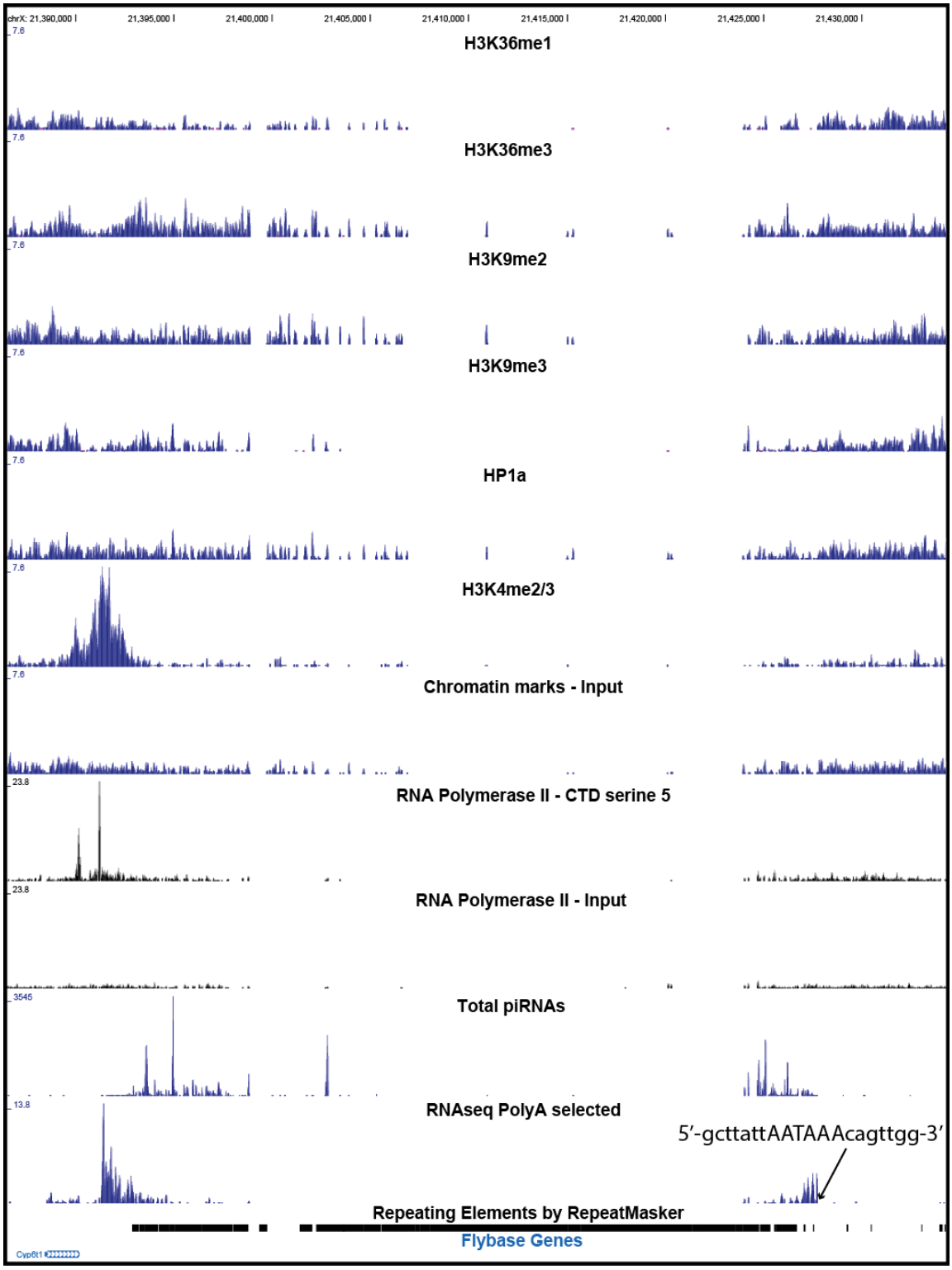
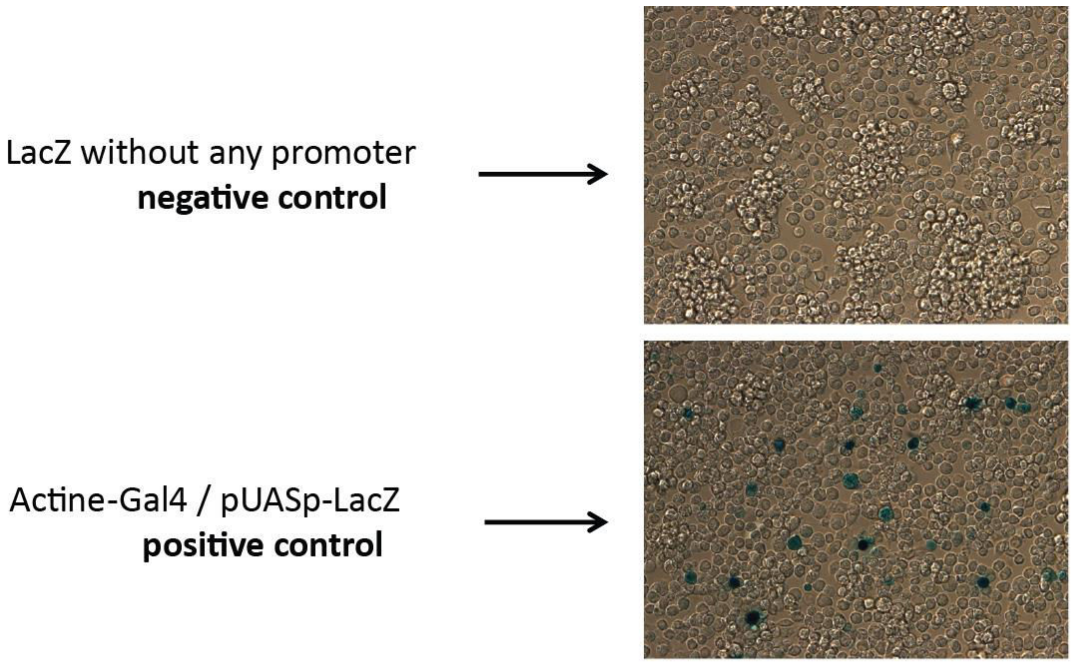


Figure 4. Germline Uni-directional piRNA cluster chromatin structure.
Chromatin marks and RNA Polymerase II densities at the uni-directional piRNA cluster X upstream (delimited by the piRNA density--Total piRNA track) in *Drosophila melanogaster*. In opposite to bi-directional clusters, the active mark specific for promoter, H3K4me2/3, is strongly enriched at the beginning of the cluster, and the active mark H3K36me3 also show a slight enrichment along the cluster. The H3K4me2/3 signal overlap with the RNA Pol II signal. Accordingly, repressive marks are depleted from the body of the cluster. One can notice the strong enrichment for mRNA polyA selected signal along the cluster (absent at bi-directional clusters) correlating with the piRNA density.

piRNA cluster promoters are necessary for piRNA expression

Along with chromatin marks, we also ChIPed for the RNA polymerase II phosphorylated on the serine 5 of its CTD domain to identify putative promoter regions of every piRNA cluster (Fig. 1,2). Doing so, we were able to find distinct RNA pol II signals around bi-directional clusters, and a unique signal upstream of uni-stranded clusters X upstream and Flamenco.

In order to determine if the regions corresponding to the RNA Pol II ChIP signal are really promoting transcription of piRNA clusters, we performed two types of experiment: first, we tested the ability of putative promoter cluster sequences to drive transcription by cloning them in front of a LacZ reporter gene and transfecting resulting constructs into embryonic S2 culture cells. The transcription output was assessed by measuring the level of LacZ expression in the cells (Fig. 5). Each putative promoter tested was able to drive expression of LacZ, indicating that those regions are indeed promoters, and do not need the presence of an active piRNA pathway to perform their function.

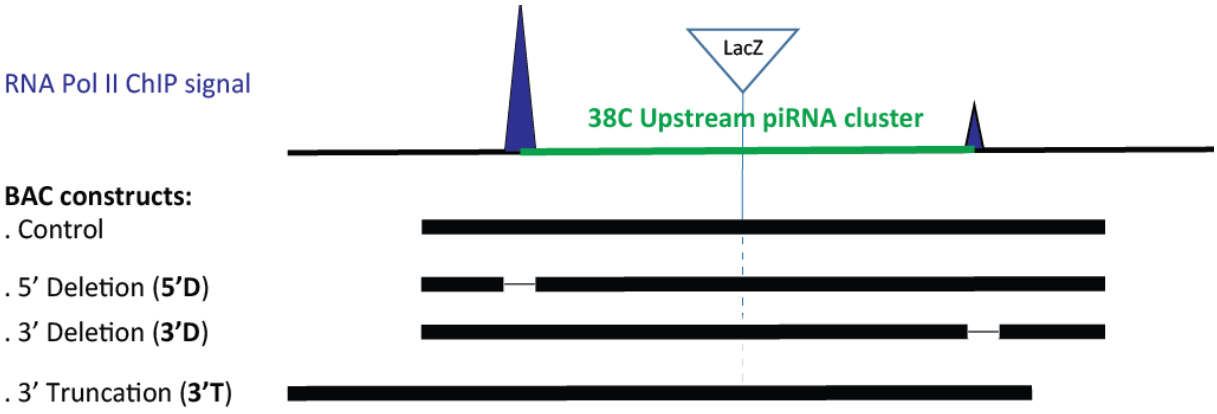


βGal quantitative assay (A420nm):

38C_up+	X_Up	Flamenco	42AB-	- Control	+ Control
1970	1237	514	418	1	2433

Figure 5. Identification of diverse piRNA cluster promoter regions. Qualitative and quantitative LacZ staining assays using piRNA cluster putative promoter regions. Quantitative assay values are normalized to Luciferase expression (co-transfected plasmid).

Secondly, we tested if the absence of these promoter regions would affect the production of piRNAs. Using a BAC construct containing a full length piRNA cluster, 38C upstream, in which we inserted inverted sequences of LacZ fragments (non-coding), we deleted the putative promoter regions (corresponding to the RNA pol II peak signals at either the 5' or the 3' end). Those constructs were injected into flies and piRNAs were sequenced (Fig. 6). We also sequenced piRNAs from another publicly available BAC construct that present large truncation of the 3' end, and in which LacZ fragments were also inserted at the exact same position. However, it should be noted that the 3' end truncation does not eliminate the 3' RNA Pol II ChIPseq peak signal.



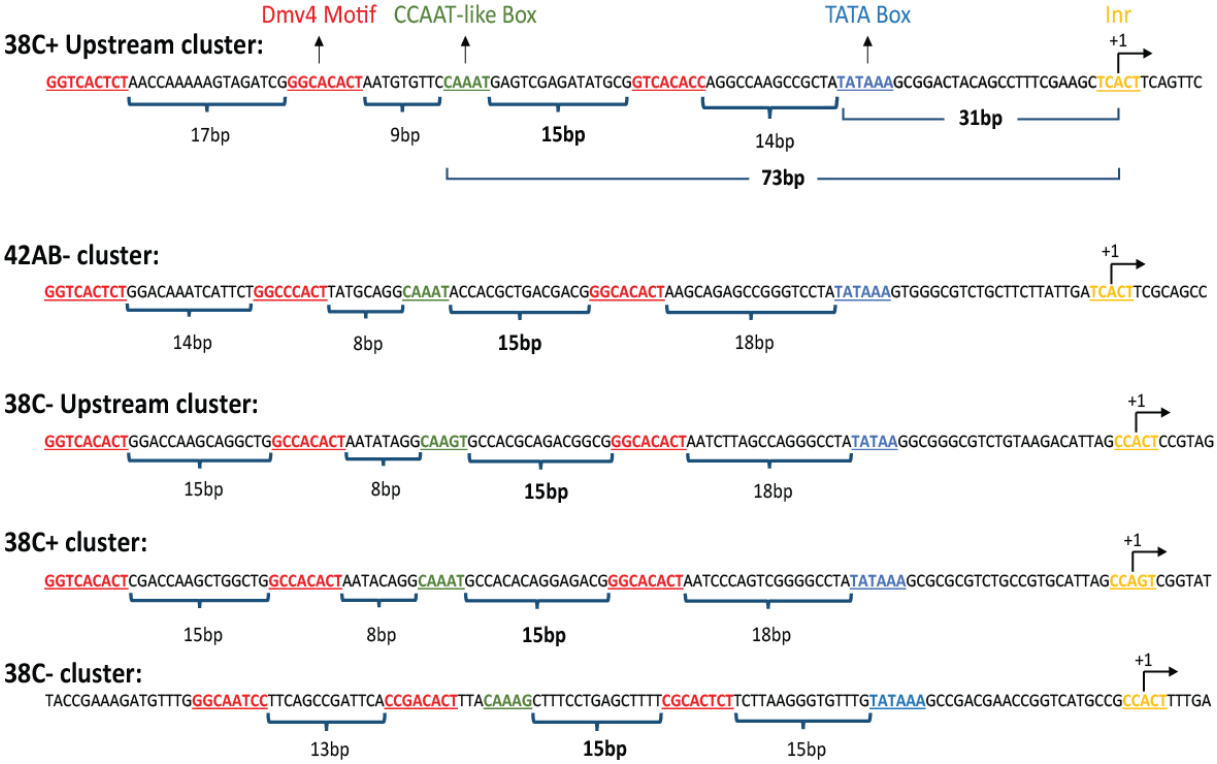
24-29nt piRNAs:	Control	5'D	3'D	3'T
read mapping both LacZ strand	1248	539	3479	2636
read mapping LacZ positive strand	1006	276	3089	2097
read mapping LacZ negative strand	242	263	390	539
Percentage + reads	**81%**	**51%**	**89%**	**80%**
Percentage - reads	**19%**	**49%**	**11%**	**20%**

Figures 6. piRNA production depends on active promoter transcription.
Deletion of regions corresponding to the RNA Pol II ChIP signal flanking the bi-directional piRNA cluster 38C upstream disrupt piRNA production from that cluster. After insertion of fragments a of the LacZ sequence into a BAC construct encompassing the piRNA cluster full length, precise deletion of the 5' or 3' RNA Pol II site were done. The constructs were injected into *D. melanogaster* and finally piRNAs were sequenced. In addition, the LacZ sequence was also inserted, at the same location, into another BAC constructs showing a truncation of 3' region. The 3' truncation does not include the 3' RNA Pol II site. LacZ piRNAs mapping to the positive strand strongly decrease when the 5' RNA Pol II signal is lost (read numbers and ratios). However, 3' deletion only show a small decrease in piRNAs levels.

The amount of piRNAs mapping to the LacZ fragment of the transgenic cluster in the control BAC line correlates with what is expected from the endogenous 38C upstream cluster: around 80% map the positive strand, and 20% map to the negative strand. piRNAs coming from antisense transcription are particularly low in read numbers, so piRNAs, here, mostly arise from the sense transcript of the cluster. In the fly line that was injected with a transgenic 38C upstream cluster that has its 5' promoter deleted, piRNAs mapping to positive strand decrease to the level of the negative strand (almost 50% each). 3' deletion of the promoter also reduce the percentage of piRNA mapping to the antisense strand (11% vs 20% originally), however, no change was seen in the case of the 3' truncation (RNA Pol II signal still present). It confirms that even if for a low transcriptional level, this region is needed as an antisense promoter. The same experiment was performed, on the uni-directional X upstream cluster, by another student in the lab. The putative promoter region was removed and piRNAs sequenced. In addition, another construct was made where the putative promoter was replaced by an inducible UASp promoter. The interesting result is that more piRNAs seemed to be produced when the cluster is driven by the inducible UASp promoter (data not shown here). However, as for the 38C upstream cluster, piRNAs are lost upon promoter deletion. Overall, these results indicates that we identified promoters that drive cluster expression.

Cluster promoter motifs, and effect on transcription

Once we identified promoter sequences, we decided to align together those of major bi-directional piRNA clusters (5 peaks). We quickly noticed a 150-200 bp region that is highly similar in every promoter sequence. Inside this homologous region, we can identify several motifs that have already been described previously for promoters: a TATA box, an Initiation motif (Inr), and a CCAAT-box like[35,36] motif (Fig.7). These motifs all seem to be relatively well conserved between different Drosophila species (see UCSC Genome Browser). In addition to these known motifs, we can also find another one that appear three times in that same region: G[G,T,C]C[A,C]C[A,T]C[C,T]. It is an eight nucleotides motif that was only identified previously in large-scale computational studies of Drosophila gene promoters. It was shown to be present mostly at genes expressed in Drosophila ovaries[37,38].



Figures 7. Similarities between bi-directional piRNA cluster promoters.
Here is represented part of the homologous sequence present among the five RNA Pol II peak signal at the beginning of double-stranded piRNA clusters. + and – signify the orientation of the transcription by the RNA Pol II (bi-directional clusters). We can identify the three DMv4 motifs (in red), the CCAAT-like box (in green), the TATA box (in blue), and the transcription initiation motif (Inr, in yellow). – sequences had to be reverse-complemented in order to detect similarities.

Following these studies, we refer to this motif as the DMv4 motif. Unfortunately, we were not able to find more information to date for this motif in the literature: nothing is known about its role or what transcription factor can bind to it. It is especially worth noting that this motif was never identified to occur three times in the same genic promoter in those previous studies, but just once (also confirmed by ourselves computationally). Therefore this seem to be a peculiarity of piRNA cluster promoters.

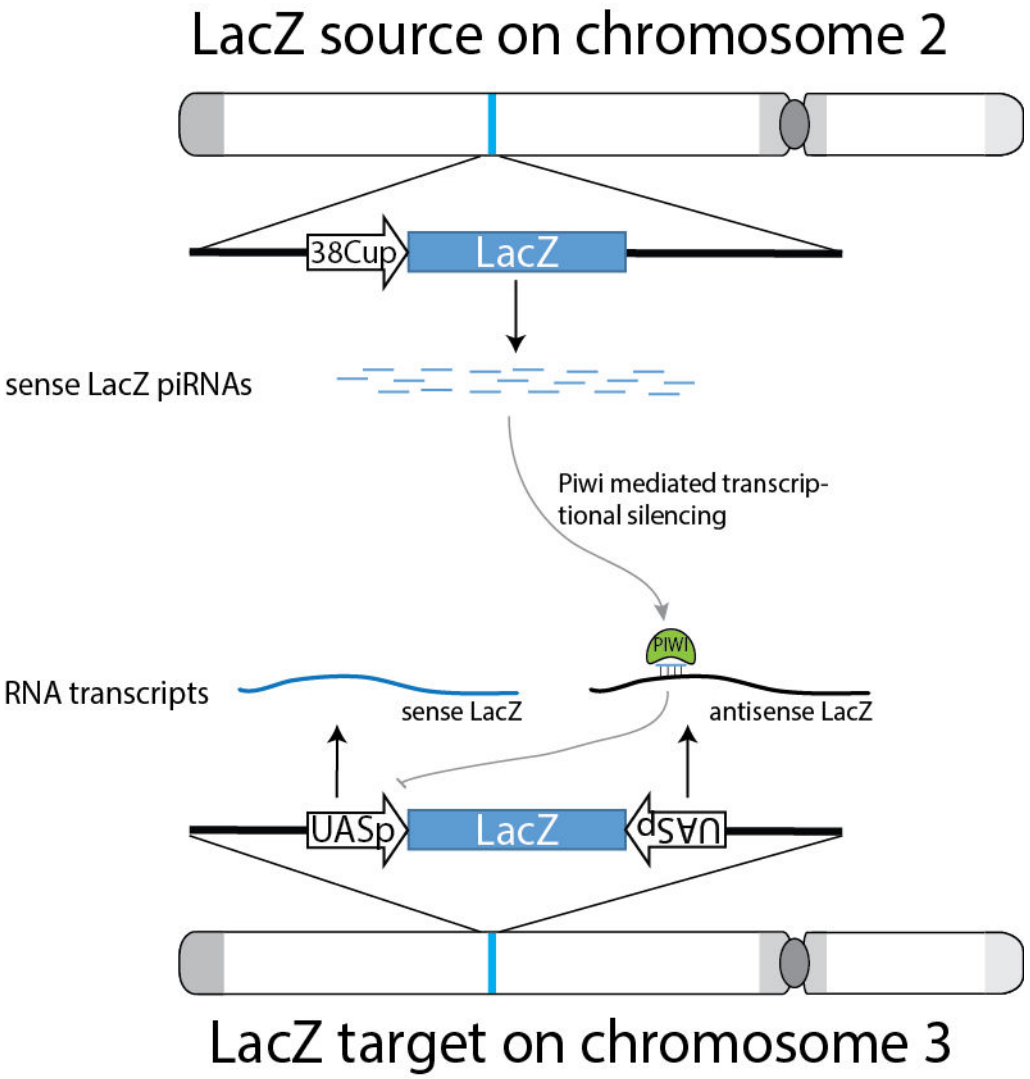
We tested the importance of these motifs in transcription efficiency. We cloned the promoter region of the 38C upstream cluster with, beside an untouched version of the promoter (WT), either all the DMv4 motifs (DMv4), or only the CCAAT-box like motif (CAAAT) scrambled, in front of a LacZ coding sequence and transfected the constructs into S2 cultured cells. Interestingly, LacZ staining and OPNG assay experiments, show that LacZ expression is lost, when the DMv4 motifs are being scrambled, but not for the CCAAT-box motif (table

below, values are normalized to Luciferase expression). The DMv4 motif is therefore essential for transcription.

βGal quantitative assay (A420nm):

WT	Dmv4	CAAAT	- Control
537	88	367	15

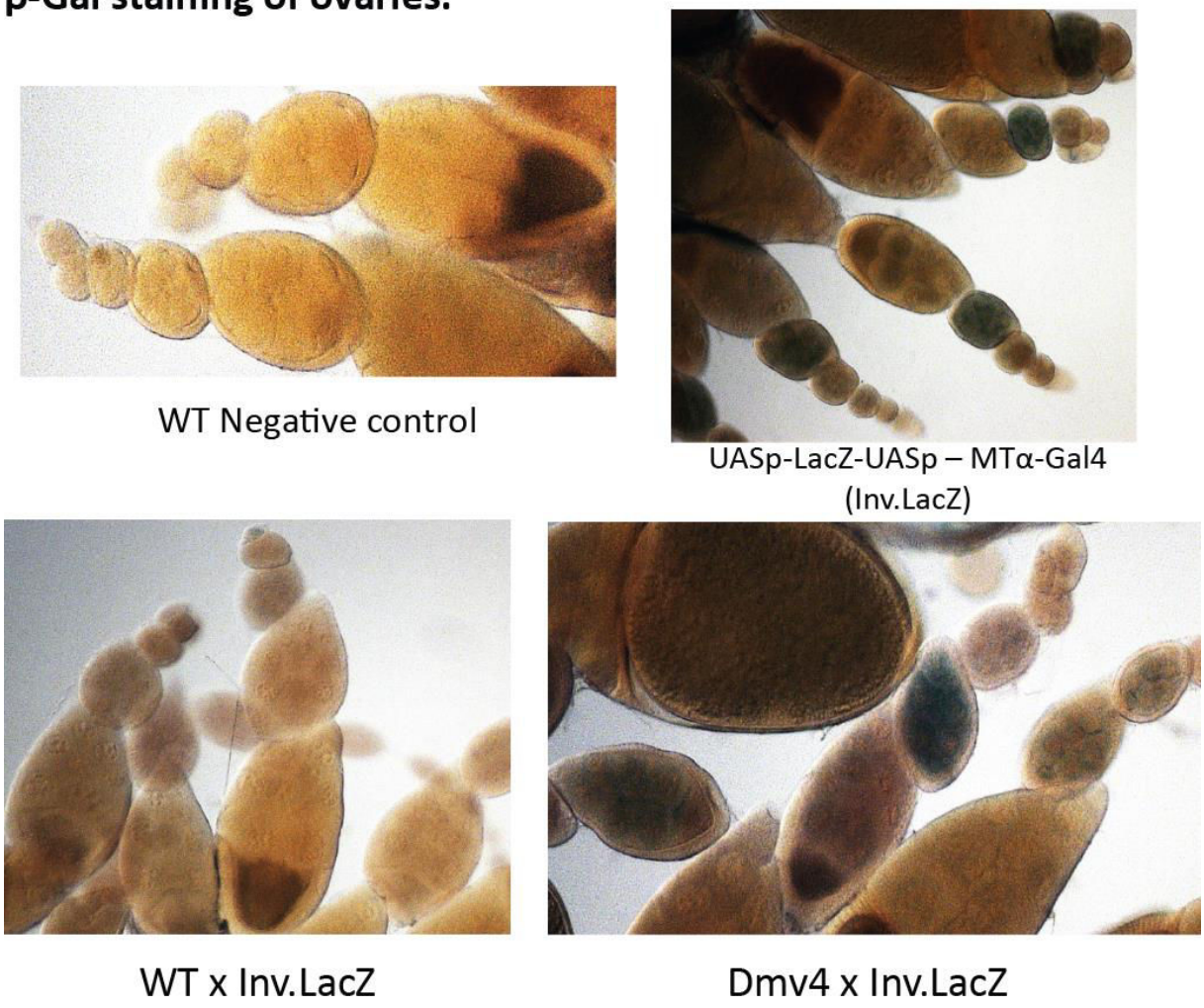
We next injected the same constructs into flies and tested their expression to see when exactly piRNA clusters might be expressed in the ovaries. No LacZ staining was visible in ovaries where the wild-type (WT) version of the promoter was cloned in front of LacZ. However, the WT and CAAAT constructs do produce a detectable RNA transcript in ovaries in comparison to the DMv4 construct and carcasses of the same flies (RT-PCR). In contrary to S2 cells, ovarian germline cells have a fully active piRNA pathway (all factors expressed). It is possible that the construct is being silenced by the pathway or that cluster promoters are sufficient to induce piRNA production from the LacZ reporter sequence. Indeed, when knocking-down the primary piRNA processing factor Zucchini, responsible for 5' end processing of piRNA cluster precursor transcripts, we can retrieve LacZ expression from the WT construct of the ovary. Next, we tested if this WT construct (called now, LacZ source) was able to produce piRNAs that can silence, in *trans*, a LacZ reporter construct (called here LacZ target). However, the main problem in this scenario is that, if the LacZ source transgene does produce piRNAs, those are sense to the LacZ source and target transcripts, and therefore unable to mediate target recognition by complementarity with the LacZ reporter construct. Therefore, in order for the newly made LacZ piRNAs to silence in *trans*, it needs to target an antisense LacZ RNA transcript. Fortunately, we have in our lab a fly line containing a LacZ coding sequence surrounded on both side by UASp promoters: one sense upstream, and another one, inverted, anti-sense downstream. This produce two types of transcripts, one sense that results in LacZ expression in ovaries when driven by the Nanos-Gal4 driver (germline specific), and one antisense that can now be targeted by piRNAs from the LacZ source transgene (Fig. 8). According to our previous results in Piwi transcriptional silencing (see Chapter 1), we can expect to see silencing of the LacZ target if proper piRNAs are being produce from the source: deposition of H3K9me3 over the target region and RNA Pol II repression.



Figures 8. LacZ *trans*-silencing scheme.
Recognition, by sense piRNAs coming from the LacZ source transgene, of the antisense LacZ target RNA transcript should induce co-transcriptional repression of the locus, through deposition of repressive chromatin marks, as shown in Chapter 1, over the genomic region.

And indeed, we were able to observe loss of LacZ staining when LacZ source and target transgenes are combined in the same fly line (Fig. 9). However, although very promising, this results are also very preliminary and need more attention. Beside ChIPing for H3K9me3 at the target locus, piRNAs obviously need to be sequenced, but in absence of an active endogenous 38C upstream cluster, to rule out any involvement of paramutation (see next chapter) effects (possible, and currently in process, through the help of the newly developed CRISPR system[39]).

Overall, these results confirm the necessity for a dedicated promoter for proper piRNA cluster expression and subsequently piRNA production.



Figures 9. piRNA cluster promoter induces mature piRNA production.
LacZ staining of the ovaries that seem to reveal the ability of the 38C upstream cluster promoter, cloned in front of a LacZ sequence, to produce mature LacZ piRNAs by itself. These LacZ piRNAs can, in turn, silence in *trans* another LacZ reporter construct present in the genome. The first picture represent staining in the fly line possessing only the LacZ source construct (WT= 38Cup-LacZ): no staining is observed. The second picture represent the staining in fly line possessing only the LacZ target (Inv. LacZ) construct, driven by the Maternal-alpha tubulin (MTα) driver: LacZ staining can be observed in later stage of the ovaries, which correspond to the driver activation pattern. The third picture represent the staining in the progeny of the cross of the first two fly lines: there is loss of the LacZ staining in the ovarioles. The fourth and last picture represent staining in the progeny of the cross between a special version of LacZ source fly line, where all DMv4 motifs have been scrambled the 38Cup promoter, and the fly line possessing the LacZ target (Inv. LacZ) construct: LacZ staining is still present.

Chapter 3 - Selection of piRNA cluster precursors

A recent study of de Vanssay and colleagues showed that a transgenic locus that generates piRNAs is able to induce piRNA production from another homologous locus that was originally incompetent for piRNA generation[40]. Importantly, the study revealed that deposition of a cytoplasmic factor from the mother, which expresses the active locus, into the embryo is sufficient to activate piRNA generation from the recipient locus without the inheritance of the active locus itself. The authors proposed that the epigenetic signal that activates piRNA biogenesis in the next generation is a pool of piRNAs that are generated from the active locus and inherited by the progeny. Indeed, Piwi proteins and the associated piRNAs that are expressed in the maternal germline during oogenesis are deposited into the developing egg and are present in the early embryo before the start of zygotic transcription[4,27,41]. Furthermore, maternally inherited piRNAs are essential for effective piRNA-mediated silencing and the fertility of the progeny[28]. Trans-generational inheritance of piRNAs that initiate piRNA biogenesis from homologous genomic regions in the progeny provides an attractive solution to the specification of piRNA precursors, however, the molecular mechanism for this process has remained unknown.

Here, we use several genetic systems and two different Drosophila species, *D. melanogaster* and *D. virilis*, to dissect requirements for piRNA biogenesis.

Part 1 - Trans-generationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing.

Activity of transgenic piRNA clusters correlates with the level of H3K9me3 mark and with the presence of trans-generationally inherited piRNAs

To identify the features that discriminate piRNA-generating regions from other genomic loci, we searched for examples when the same locus is active in piRNA generation in one Drosophila strain but inactive in another. Two *D. melanogaster* strains, T1 and BX2, contain an identical number of tandem repeats of the P-lacZ transgene inserted in the same locus in the middle of chromosome arm 2R, a genomic position that does not give rise to piRNAs in other strains[42,43]. In the T1 strain the transgene gives rise to abundant piRNAs in germ cells, and these piRNAs are able to silence expression of another lacZ transgene in *trans*. In contrast, no piRNAs are generated from the same transgene in the BX2 strain[40] (Fig. 1A). As the sequence of the T1 and the BX2 transgenes is identical, the reason for their differential ability to generate piRNAs is unknown. Recent studies revealed that two chromatin factors, SetDB1 (a methyltransferase responsible for installation of the H3K9me3 mark) and the HP1 homologue Rhino, are required for piRNA biogenesis[32,33]. Based on this, we hypothesized that a difference in the chromatin state of the T1 and the BX2 transgenes explains their differential ability to produce piRNAs. To test this we profiled the histone 3 lysine 9 trimethylation (H3K9me3) mark on the transgene sequences in ovaries of flies from both strains using ChIP-qPCR. The H3K9me3 mark was enriched by ~4 fold over the transgene in the 'active' T1 strain compared to the 'inactive' BX2 strain (Fig. 1B). To rule out the possibility that the differences in H3K9me3 mark between T1 and BX2 are caused by changes in nucleosome occupancy, we performed total H3 ChIP that showed no difference between the two strains (Fig. S1). This result indicates that the ability of a genomic locus to generate piRNAs indeed correlates with its chromatin structure and in particular with a high level of the H3K9me3 mark.

In Drosophila, piRNAs expressed in germ cells during oogenesis are deposited into the embryo and are important for transposon silencing in the next generation[28]. It was previously found that the BX2 locus, which is deficient in piRNA production, can be converted into an active locus (designated as BX2*) by exposure to maternally inherited cytoplasm, carrying piRNAs homologous to the lacZ transgene[40] (Fig. 1A). To determine whether activation of the

BX2 locus caused by trans-generationally inherited piRNAs was coupled with a change in the chromatin state of the locus, we analyzed this region for the presence of the H3K9me3 mark before and after conversion. Conversion of BX2 to BX2* was accompanied by an increase in the H3K9me3 mark on the transgene (Fig. 1C). The level of the H3K9me3 mark on the activated BX2* transgene was similar to the signal observed on the locus in the T1 strain. Therefore, exposure of an inactive locus to homologous piRNAs inherited from the previous generation leads to installment of the H3K9me3 mark and conversion of the locus to an active piRNA cluster.

The ability of the T1 and the activated BX2* loci to generate piRNAs was reported to be stable over many generations if the locus was inherited from the mother[40] (Fig. 1A). However, the transgenic loci lose their capacity to generate piRNAs after paternal transmission, suggesting that the presence of piRNAs is necessary to maintain the locus in an active state. We measured the enrichment of the H3K9me3 mark on the T1 transgene after maternal and paternal transmission. The paternal transmission reduced the level of the H3K9me3 mark ~2 fold, although levels were slightly higher than those on the non-activated BX2 locus (Fig. 1D). Thus, lack of trans-generationally inherited piRNAs homologous to the transgene correlates with decreased levels of the H3K9me3 mark.

Taken together, our experiments demonstrate that the ability of a transgenic locus to generate piRNAs correlates with a high level of H3K9me3 over the transgene. Reciprocally, the high level of H3K9me3 mark on the locus requires the presence of trans-generationally inherited piRNAs homologous to the locus. Specifically, an active locus loses its high H3K9me3 signal and becomes inactive if it is not exposed to homologous piRNAs, whereas an inactive locus with low H3K9me3 signal acquires both a higher level of the H3K9me3 chromatin mark and the ability to generate piRNAs after exposure to homologous piRNAs. Together these observations suggest that piRNAs inherited from the previous generation induce a change in the chromatin state of homologous genomic regions in the progeny and this change in chromatin is required to produce piRNAs from these regions.

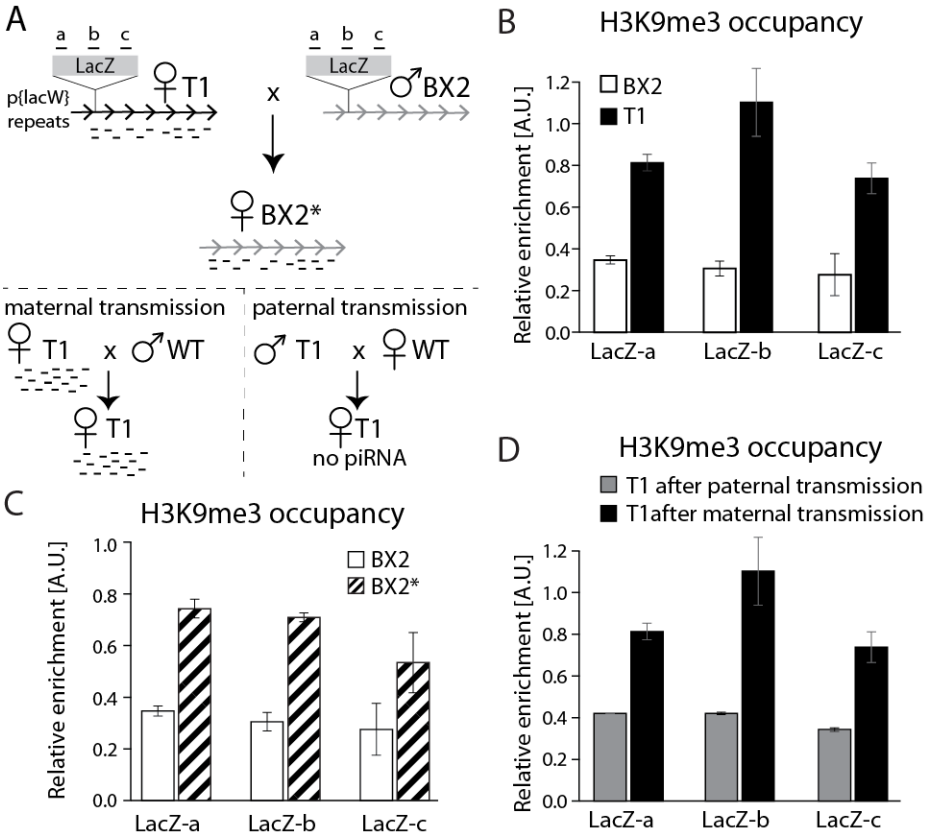


Figure 1. Activity of transgenic piRNA clusters correlates with the level of H3K9me3 mark and with the presence of trans-generationally inherited piRNAs

(A) Transgenic *D. melanogaster* strains used for the study. Both T1 and BX2 strains contain seven identical tandem copies of the P{LacW} transgene inserted at the cytological site 50C. Strain T1 was derived from BX2 by X-ray irradiation and carries large-scale chromosome rearrangements. Transgene-derived piRNAs are generated in T1 but not in BX2 females. Trans-generational inheritance of piRNAs from T1 females leads to activation of the BX2 locus in the progeny (indicated as BX2*) resulting in piRNA production. Note that inheritance of the T1 locus is not required for the activation of BX2. The ability to generate piRNAs from the T1 transgene is inherited through the female, but not the male, lineage (bottom). The locations of regions A, B and C along the *LacZ* sequence, analyzed in the ChIP-qPCR experiments, are indicated.

(B) piRNA production from transgenes correlates with enrichment of the H3K9me3 mark over the transgene. H3K9me3 signal on T1 and BX2 as measured by ChIP-qPCR on three different regions of the *lacZ* sequence shown on (A). Signal was normalized to input and compared to signal at a control genomic region (chr2R: 2336913-2337023) highly enriched in the H3K9me3 mark. The ChIPs were performed on two independent biological samples followed by duplicate qPCRs on each sample. Two-sided t-test was used to calculate significance of the differences ($p<0.05$).

(C) Conversion of the inactive BX2 locus to a piRNA producing BX2* locus correlates with an increase in the H3K9me3 signal, as measured by ChIP-qPCR on three different regions of the *lacZ* sequence. The ChIPs were performed on two independent biological samples followed by duplicate qPCRs on each sample. Two-sided t-test was used to calculate significance ($p<0.05$).

(D) Paternal transmission of the T1 transgene is accompanied by a loss of H3K9me3 signal, as measured by ChIP-qPCR. The ChIPs were performed on two independent biological samples followed by duplicate qPCRs on each sample. Two-sided t-test was used to calculate significance ($p<0.05$).

Trans-generationally inherited piRNAs enhance the ping-pong processing of piRNAs

To test if trans-generational inheritance of piRNAs is required for piRNA biogenesis from other loci we used another transgenic system that contains single-copy unique sequences. The insertion of the P{lArB} construct into a subtelomeric piRNA cluster in strain P1152 leads to generation of abundant piRNAs from this transgene[44,45]. Importantly, piRNAs generated from the lacZ sequence in P1152 are able to silence other lacZ transgenes located in euchromatic sites, such as the P{A92} in strain BC69, in *trans*[40,44,46].

We analyzed the generation of piRNAs from the P1152 locus in the parental strain and two reciprocal progenies of crosses between P1152 and BC69 stocks. Maternal Deposition (MD) progeny inherit transgene-derived piRNAs from their P1152 mothers, while genetically identical No Maternal Deposition (NMD) progeny of the reciprocal cross do not inherit piRNAs (Fig. 2A). In addition to profiling total piRNA populations in the ovaries of each genotype, we immunoprecipitated the PIWI, AUB and AGO3 proteins and defined piRNA populations associated with each protein. As reported before[40,45] we detected a large number of piRNAs mapping to the P1152 transgene in the parental P1152 strain. Transgenic piRNAs are found in all three Piwi complexes (Fig. 2B). In contrast, practically no piRNAs were detected in the BC69 strain. The level of piRNAs mapped to the transgene drops two-fold in the MD progeny compared to the parental P1152 flies reflecting the fact that only one allele of the transgene is present in MD flies compared to two copies in the parental P1152 strain. The level of piRNAs decreases ~10-fold in the NMD progeny compared to the genetically identical MD flies, indicating that inherited piRNAs are essential for piRNA generation (Fig. 2B).

One mechanism by which trans-generationally inherited piRNAs might boost production of new piRNAs in the progeny is processing of complementary transcripts through the ping-pong loop. In the ping-pong loop, piRNA guide endonucleolytic cleavage of complementary transcript leading to the generation of new piRNA from the cleaved product[4,13,47]. Indeed, piRNA clusters that are expressed in germ cells, including telomeric cluster containing the P1152 locus, are transcribed from both genomic strands, therefore, trans-generationally inherited piRNAs can target transcripts from these regions to initiate their processing. To study the impact of inherited piRNAs on the ping-pong processing, we analyzed the characteristic feature of ping-pong, the number of complementary piRNA pairs with a 10-nt overlap between their 5' ends (so called ping-pong pairs). To account for different numbers

of sequences in each library we sampled the same number of reads from each library 1000 times and calculated the fraction of piRNAs in ping-pong pairs. A large number of piRNAs derived from the transgene in the P1152 strain was generated by ping-pong processing (Fig. 2C, S2). In agreement with previous studies[4,48], the strongest ping-pong interaction was observed between AUB and AGO3, while PIWI participation in the ping-pong is negligible. The fraction of the ping-pong pairs in the total piRNA populations as well as in AUB/AGO3 pairs is similar in the parental P1152 strain and MD progeny. However, the ping-pong processing dramatically decreased in the NMD progeny: the fraction of piRNAs in ping-pong pairs decreases 4 to 7 fold in the most prominent partners, AUB and AGO3 (Fig. 2C, S2). This result shows that inheritance of homologous piRNAs indeed boosts ping-pong processing in the progeny.

The level of PIWI-bound piRNAs mapped to P1152 transgene is decreased 8.3-fold in the absence of inherited piRNAs in NMD progeny (Fig. 2B). However, analysis of piRNA sequences associated with PIWI showed that they, in contrast to AUB and AGO3 piRNAs, do not form significant ping-pong pairs (Fig. 2C, S2) supporting previous studies that indicate that PIWI-bound piRNA are generated by other biogenesis mechanism, so called primary processing. Therefore, the boost in ping-pong processing is not sufficient to explain the whole impact of inherited piRNAs that also enhance primary piRNA biogenesis.

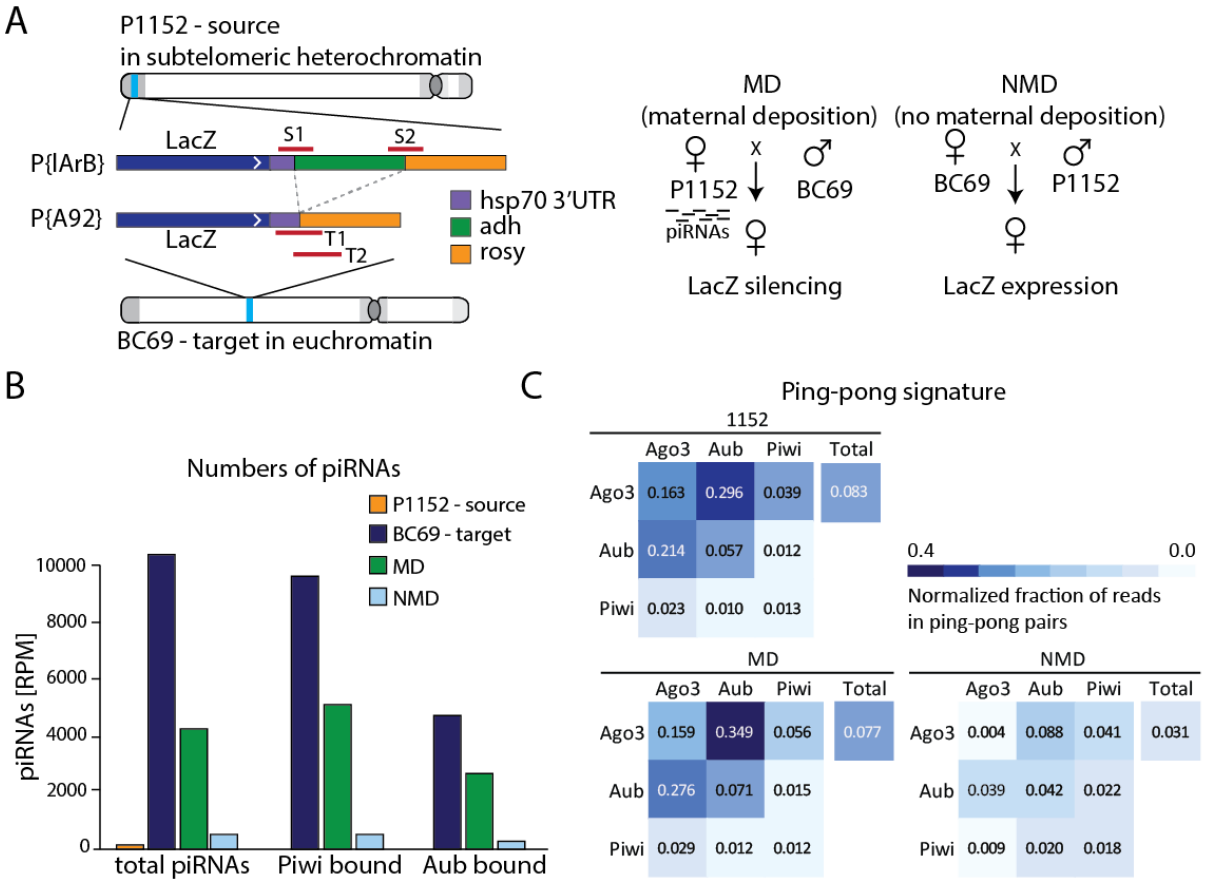


Figure 2. Inherited piRNAs activate the ping-pong cycle

(A) Transgenic D. melanogaster strains used for the study. The strain P1152 contains a single copy of the P{lArB} insertion at the telomeric X-TAS piRNA cluster on the X chromosome, which generates abundant piRNAs from both genomic strands (Josse et al. 2007; Muerdter et al. 2012). Strain BC69 contain a single copy of the P{A92} insertion at the euchromatic site 35B-C of chromosome 2 that does not generate piRNAs.The P{lArB} and P{A92} transgenes are identical in sequence, except of the Adh region that is specific to the P{lArB} construct in P1152. The bars (S1-2 and T1-2) indicate sequences used for ChIP-qPCR to measure chromatin the structure of the P1152 (source) and the BC69 (target) transgenes. BC69 females express lacZ from the transgene in the ovary, however, lacZ is repressed by lacZ piRNAs generated from the P1152 transgene if both transgenes are combined. The cross between P1152 and BC69 strains was performed in two directions: the progeny designated as Maternal Deposition or MD inherited P1152-derived piRNAs from their mothers. In contrast the progeny of the reciprocal cross, designated as No Maternal Deposition or NMD, did not inherit these piRNAs.

(B) piRNA production from the P1152 and the BC69 transgenes in the parental strains and the MD and NMD progenies. Shown are the numbers of piRNA reads normalized to library depth (RPM) mapped to the P{lArB} transgene in total RNA and in RNA purified from the PIWI and AUB complexes. Abundant piRNAs are present in the P1152 (piRNA source), but not in the BC69 (piRNA target) parental strain. piRNA levels in the MD progeny are 2-fold lower than in the P1152 strain reflecting the heterozygous presence of the source locus in the progeny. The piRNA level is ~10-fold lower in the NMD compared to the MD progeny in the total cellular piRNA population and in the PIWI and AUB complexes.

(C) Analysis of ping-pong processing. An equal number of piRNA reads mapped to the P{lArB} transgene in each library was sampled 1000-times to calculate the fraction of reads in ping-pong pairs. Shown are the mean values of the normalized fraction of piRNAs that are in ping-pong pairs; the distributions of values from the 1000-times sampling are shown on box-plots in Fig. S1. Both the P1152 parental strain and the MD progeny have high level of AUB- and AGO3-bound piRNAs generated by ping-pong processing. The fraction of piRNA generated by ping-pong processing drops 4 to 7-fold in NMD progeny. PIWI-bound piRNAs are generated by primary biogenesis mechanism and accordingly do not have a strong-ping-pong signature.

Inherited piRNAs trigger the deposition of the H3K9me3 mark and initiation of primary piRNA biogenesis from the same genomic sequence

We analyzed chromatin states of both source (P1152) and target (BC69) loci in ovaries of MD and NMD progenies (Fig. 2A). Previously, we found that the presence of piRNAs generated from the P1152 source locus leads to an increase in the H3K9me3 mark and a decrease in the H3K4me mark on the BC69 target locus[49]. We confirmed accumulation of the H3K9me3 mark on the BC69 target locus and found that it occurs only in MD progeny that inherited homologous piRNAs from their mothers (Fig. 3A). Importantly, we also observe that in the absence of inherited piRNAs, in NMD progeny, the level of H3K9me3 decreases on the P1152 source construct (Fig. 3A). These results showed that high level of H3K9me3 on both the source and the target loci requires trans-generational piRNA inheritance.

To test if the deposition of H3K9me3 on the BC69 target also triggers initiation of piRNA biogenesis we mapped piRNAs to the junction between the *hsp70* and *rosy* sequence segments that is specific to this construct (Fig. 2A, 3B). As expected no target-derived piRNAs were found in piRNA populations from the parental P1152 and BC69 strains. However, 9 distinct sequences were detected in the MD progeny, while none was detected in the NMD progeny (Fig. 3B, S3). This result shows that the same single-copy sequence acquires the H3K9me3 mark and becomes a source of new piRNA at the same time. Importantly, the target-derived piRNAs were detected in the MD progeny in the total piRNA population (2 piRNA reads) and in complex with PIWI (7 reads), while they were absent in AUB and AGO3. This indicates that the new target-derived piRNAs are generated by primary processing and not by ping-pong. Overall, our data show that exposure to homologous piRNAs leads to an increase in the H3K9me3 mark and initiation of primary piRNA biogenesis on the target locus.

In addition to piRNAs that target the lacZ sequence, the P{lArB} transgene in the P1152 strain generates piRNAs that can potentially target three endogenous loci in the Drosophila genome, *Adh*, *hsp70* and *rosy*. We tested if exposure to homologous piRNAs initiated piRNA biogenesis at these genomic regions. Few piRNAs are generated from *Adh*, *hsp70* and *rosy* loci in NMD progeny. In contrast, abundant piRNAs are generated from endogenous *Adh* and *hsp70* sequences that flank the region targeted by transgene-derived piRNAs in MD progeny (Fig. 3C and data not shown). Importantly, the *Adh* sequences that generate piRNAs in MD flies are not directly targeted by transgene-derived piRNAs. This eliminates ping-pong processing initiated by transgene-derived piRNAs as a possible mechanism for biogenesis of piRNAs from

endogenous *Adh* locus. Instead this result indicates that exposure to inherited piRNAs activates primary piRNA biogenesis at previously naïve genomic loci (Fig. 3D).

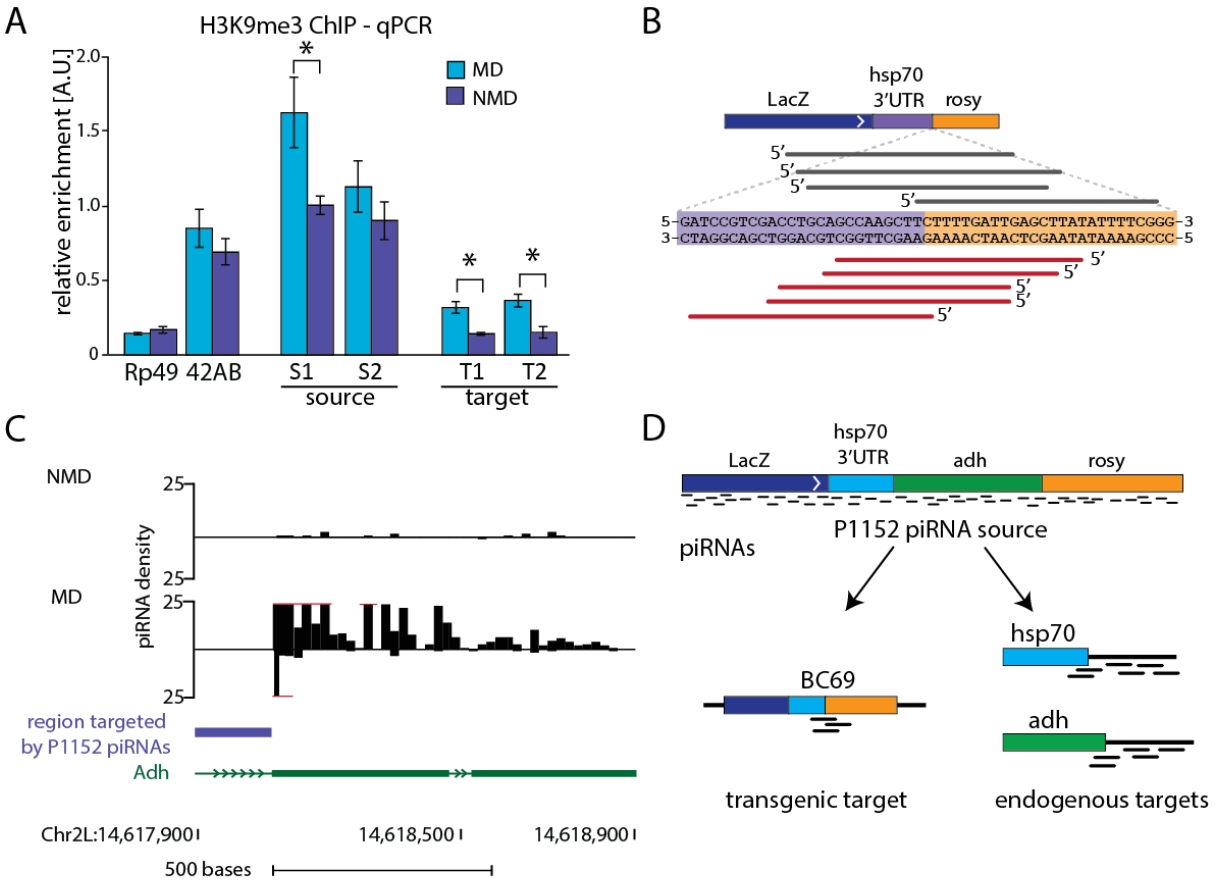


Figure 3. Inherited piRNAs activate piRNA biogenesis on target loci

(A) Repression of the BC69 target transgene by P1152 correlates with a 2-fold higher H3K9me3 signal in the MD compared to the NMD progeny, as measured by ChIP-qPCR. The H3K9me3 mark also decreases at the P1152 locus in the NMD progeny, where cognate piRNAs are not inherited. Therefore maternally deposited piRNAs are necessary to install the H3K9me3 mark on the piRNA target (BC69) and to maintain the high level of the mark on the source of piRNAs (P1152). For this experiment, MD progeny was obtained by crossing P1152; BC69 females to wild-type males and NMD progeny was obtained from the reciprocal cross. Signal was normalized to input and compared to signal at a control genomic region (chr2R: 2336913-2337023) highly enriched in the H3K9me3 mark. No H3K9me3 change is observed on the rp49 gene and the 42AB piRNA cluster. The ChIPs were performed on two independent biological samples followed by triplicate qPCRs on each sample. P values were calculated with a two-sided t-test, * indicates $p<0.05$.

(B) Primary piRNA processing is initiated on the BC69 target after exposure to homologous piRNAs in the MD progenies. Shown are piRNA sequences that mapped specifically to the BC69 target locus and not to P1152 (the junction between the hsp70 and the rosy segment, which is present in BC69, but not in P1152). Out of 9 distinct piRNAs, 2 were detected in the total cellular piRNA library and 7 in the Piwi-bound population in the MD progeny. None were detected in the parental strains or in the NMD progeny. BC69-derived piRNAs were also absent in the AUB and AGO3 complexes in the MD progeny indicating that the target-derived piRNAs were generated by primary processing and not by ping-pong. The density of BC69 target-derived piRNAs in MD progeny seems to be ~10-fold lower than that of P1152-derived piRNAs.

(C) piRNAs are produced from endogenous Adh locus targeted by the transgenic P1152 piRNAs in the MD progeny. The region targeted by the transgenic P1152 piRNAs is indicated by blue bar. Shown are only piRNAs that uniquely mapped to the endogenous Adh locus.

(D) Exposure to homologous piRNAs triggers piRNA generation from previously naïve loci. Note that piRNAs are generated from the regions that are not directly targeted by transgenic P1152 piRNA indicating the spreading around the originally targeted sequence.

The H3K9me3 mark on piRNA clusters is recognized by the HP1 homolog Rhino

One possible reader of the H3K9me3 mark on piRNA cluster chromatin is Rhino, a germline specific HP1 homolog, which was previously shown to be enriched on double-stranded piRNA clusters and to be required for piRNA biogenesis[32]. Rhino harbors a conserved chromodomain predicted to recognize methylated lysine residues (Fig. S4). To get insights into the genome-wide distribution of Rhino on chromatin, we performed ChIP-Seq on ovaries of flies that expressed a BioTAP-tagged Rhino protein. Genome-wide analysis showed that Rhino is enriched on the regions that have high level of H3K9me3 (Fig. S5). Analysis of the ChIP-Seq data confirmed the previous observations of Rhino enrichment on chromatin of double-stranded, but not single-stranded piRNA clusters (Fig. 4A).

Next, we tested if the high level of the H3K9me3 mark on the transgenic T1 locus that was induced by trans-generationally inherited piRNAs leads to binding of Rhino protein. ChIP-qPCR showed that Rhino is enriched on the active T1, but not the inactive BX2 locus (Fig. 4B). Furthermore, Rhino localization on T1 chromatin is lost together with the H3K9me3 mark after paternal transmission of the locus.

To test if the Rhino chromodomain directly recognizes the H3K9me3 mark, we monitored the binding affinity of the Rhino chromodomain as a function of the methylation state of the H3(1-15)K9me peptide using isothermal titration calorimetry (ITC). The H3K9me3 and H3K9me2 peptides showed similar binding affinity of 21.8 μM and 28.4 μM, respectively. By contrast, the H3K9me1 and unmodified H3 peptide showed no detectable binding affinity, establishing that recognition of the H3 tail by Rhino prefers high methylation states of lysine 9 (Fig. 4C).

To further investigate the molecular mechanism of recognition of H3K9me3 by Rhino, we determined the crystal structures of the Drosophila Rhino chromodomain in the free form and in complex with the H3(1-15)K9me3 peptide at 1.5 Å and 2.5 Å resolution, respectively. The overall structure of the Rhino chromodomain (Fig. 4D) resembles that of the classic HP1 chromodomain, which has an N-terminal three-stranded twisted β-sheet and a C-terminal short α-helix, and forms a homodimer. The Rhino-H3(1-15)K9me3 complex also adopts a dimeric arrangement, similar to its free form (Fig. 4E, S6A). We can trace and build the H3(1-15)K9me3 peptide from Thr3 to Ser10. The interaction between the peptide and the Rhino chromodomain consists of specific recognition of H3K9me3 by three conserved aromatic

residues Phe48, Try24 and Trp45, like that observed in the canonical recognition of H3K9me3 by the HP1 protein, complemented by hydrogen bonding interactions mainly to the peptide backbone together with a few side chains[50] (Fig. 4F, S6B). The dimer interface in the complex is stabilized by a network of hydrogen bonding and hydrophobic interactions (Fig. S7). Overall, our data suggest that the H3K9me3 mark installed on piRNA clusters in a manner that depends on trans-generationally inherited piRNAs is read by the Rhino chromodomain protein.

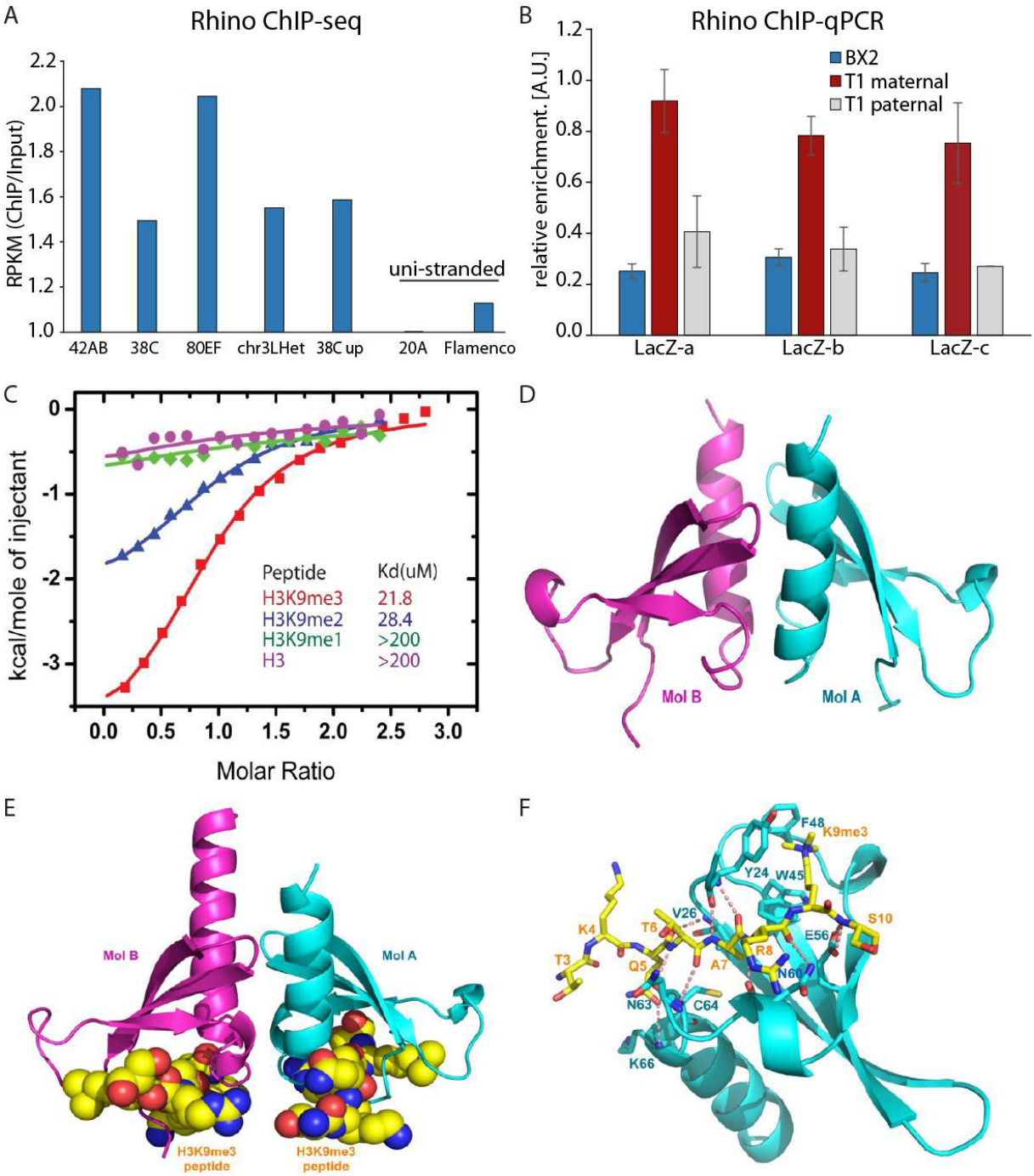


Figure 4. Chromodomain protein Rhino binds the H3K9me3 mark and is enriched on chromatin of piRNA clusters.
(A) Rhino ChIP-seq signal is enriched at double-stranded piRNA clusters. No enrichment is seen on flamenco, a uni-stranded cluster that is active only in follicular cells and on cluster 20A, a uni-stranded cluster that is expressed in germ cells. Shown are enrichments of ChIP-Seq signal relative to input. Read counts are normalized to sequencing depth and region length (RPKMs).
(B) Activity and H3K9me3 signal on the T1 and the BX2 transgenes correlates with Rhino binding. Rhino signal on T1 and BX2 was measured by ChIP-qPCR on three regions of *lacZ* (n=2). Signal was normalized to input and compared to signal at a control genomic region (chr2R: 2336913-2337023). Error bars represent the standard error of the mean between two independent biological replicates. Two-sided t-test was used to calculate significance ($p<0.05$).
(C) Binding of Rhino chromodomain to different methylated states of the H3(1-15)K9me peptide measured by isothermal titration calorimetry (ITC).
(D) Crystal structure of *D. melanogaster* Rhino chromodomain in the free state. The Rhino chromodomain forms a homodimer with each monomer labeled as Mol A and Mol B.
(E) Structure of the *D. melanogaster* Rhino chromodomain in complex with the H3(1-15)K9me3 peptide. Each monomer contains a bound H3K9me3 peptide.
(F) Specific recognition of the H3K9me3 peptide by the Rhino chromodomain.
Intermolecular hydrogen-bonding interactions are designated by dashed red lines.

Rhino recruits the piRNA biogenesis factor Cutoff and promotes efficient transcription of piRNA clusters

In the nucleus Rhino co-localizes with Cutoff, a protein that plays an important, but not well defined role in piRNA biogenesis[51]. Cutoff is homologous to proteins that participate in RNA quality control and transcriptional termination in yeast and mammals[52-55]. Localization of Cutoff in specific nuclear foci was reported to be dependent on Rhino protein suggesting that Cutoff might be recruited to chromatin through Rhino[51]. We immunoprecipitated tagged Rhino and Cutoff proteins from Drosophila ovaries and found that they form a complex (Fig. S8). Next, we used genome-wide ChIP-Seq to determine Cutoff localization on chromatin. In agreement with previous ChIP-PCR observations[51] we found that Cutoff localizes on double-stranded, but not single-stranded piRNA clusters in perfect correlation with Rhino localization (Fig. 5A). The Cutoff profile at the border of piRNA clusters closely correlates with the H3K9me3 profile (Fig. 5B), suggesting that Cutoff is recruited to piRNA cluster chromatin through the interaction with Rhino that binds this mark. Interestingly, enrichment of Cutoff on piRNA cluster chromatin was abolished if during the ChIP experiment the crosslinked chromatin was treated with RNase prior to immunoprecipitation, indicating that Cutoff also interacts with nascent transcripts (Fig. S9).

We tested if the high level of the H3K9me3 mark and Rhino on the transgenic T1 locus leads to binding of the Cutoff protein. ChIP-qPCR showed that Cutoff is more enriched on the active T1 compared to the inactive BX2 locus (Fig. 5C). Furthermore, Cutoff localization on T1 chromatin is lost together with the H3K9me3 mark and Rhino after paternal transmission.

To study the functional importance of Cutoff localization on piRNA cluster chromatin we profiled small and long RNA from ovaries of Cutoff mutants and control flies. In agreement with previous observations[51], we found that piRNAs from double-stranded piRNA clusters are almost entirely eliminated in Cutoff mutants (Fig. 6A, B). In contrast, Cutoff mutation did not affect piRNA generation from *flamenco*, a uni-stranded cluster that is active only in follicular cells, and from 20A, a uni-stranded cluster that is expressed in germ cells, but is not enriched in Rhino and Cutoff proteins. Double-stranded piRNA clusters also give rise to a significant fraction of the 21 nt siRNAs that are generated from double-stranded RNAs by Dicer[56] (Fig. S10). Importantly, siRNAs from double-stranded clusters were also eliminated in the Cutoff mutant (Fig. 6A, B) indicating that the function of Cutoff is broader than just targeting nascent transcripts to the piRNA processing machinery.

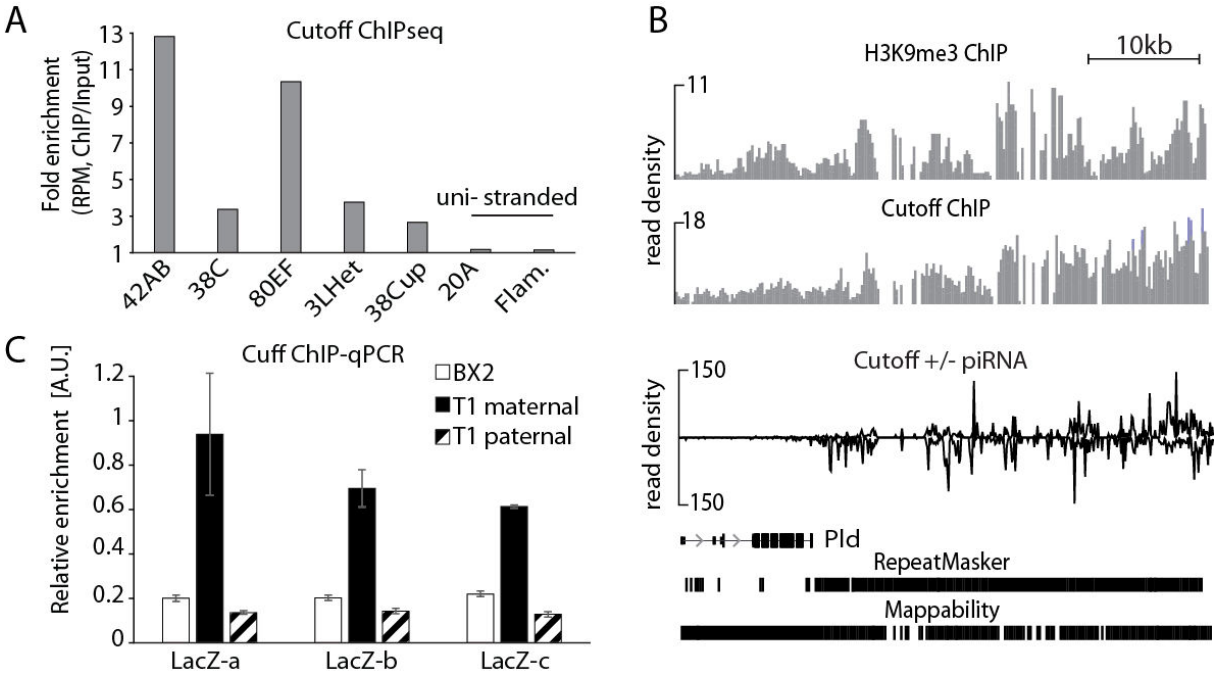


Figure 5. Rhino recruits the piRNA biogenesis factor Cutoff to chromatin of piRNA clusters.

(A) Cutoff ChIP-Seq signal is enriched at double-stranded piRNA clusters. No enrichment is seen on uni-stranded flamenco and 20A clusters**.** Shown are enrichments of ChIP-Seq signal relative to input. Read counts are normalized to sequencing depth and region length (RPKMs). **(B)** Profiles of H3K9me3, Cutoff, and piRNAs along a fragment of the 42AB cluster (chromosome 2R: 2129920 - 2181079). The Cutoff ChIP-Seq signal correlates with the H3K9me3 and piRNA profiles and progressively increases along the beginning of the 42AB cluster.

(C) Cutoff is enriched on chromatin of the active T1, but not the inactive BX2 or the paternally transmitted T1 transgene. Cutoff signal on T1 and BX2 as measured by ChIP-qPCR on three regions of *lacZ* (n=2). Signal was normalized to input and compared to signal at a control genomic region (chr2R: 2336913-2337023). Error bars represent the standard error of the mean between two technical replicates.

Instead this result suggests that Cutoff is required to generate precursor molecules for both si- and piRNA processing. Indeed, RNA-Seq of long rRNA-depleted RNA showed a decrease of transcripts from dual-stranded clusters upon depletion of Cutoff and Rhino (Fig. 6A, B). To study if Rhino/Cutoff complex is directly involved in transcription of piRNA precursor transcripts we performed nuclear run-on/GRO-Seq experiments on ovaries of control and Rhino-depleted flies. Quantification of nascent transcripts in two independent run-on experiments using two different methods, RT-qPCR and deep sequencing, showed that Rhino is indeed required for efficient transcription of dual-stranded piRNA clusters (Fig. 6C, S11). Together our data indicate that the H3K9me3 mark on piRNA clusters leads to recruitment of the Rhino/Cutoff complex, which is critical for transcription of piRNA precursors.

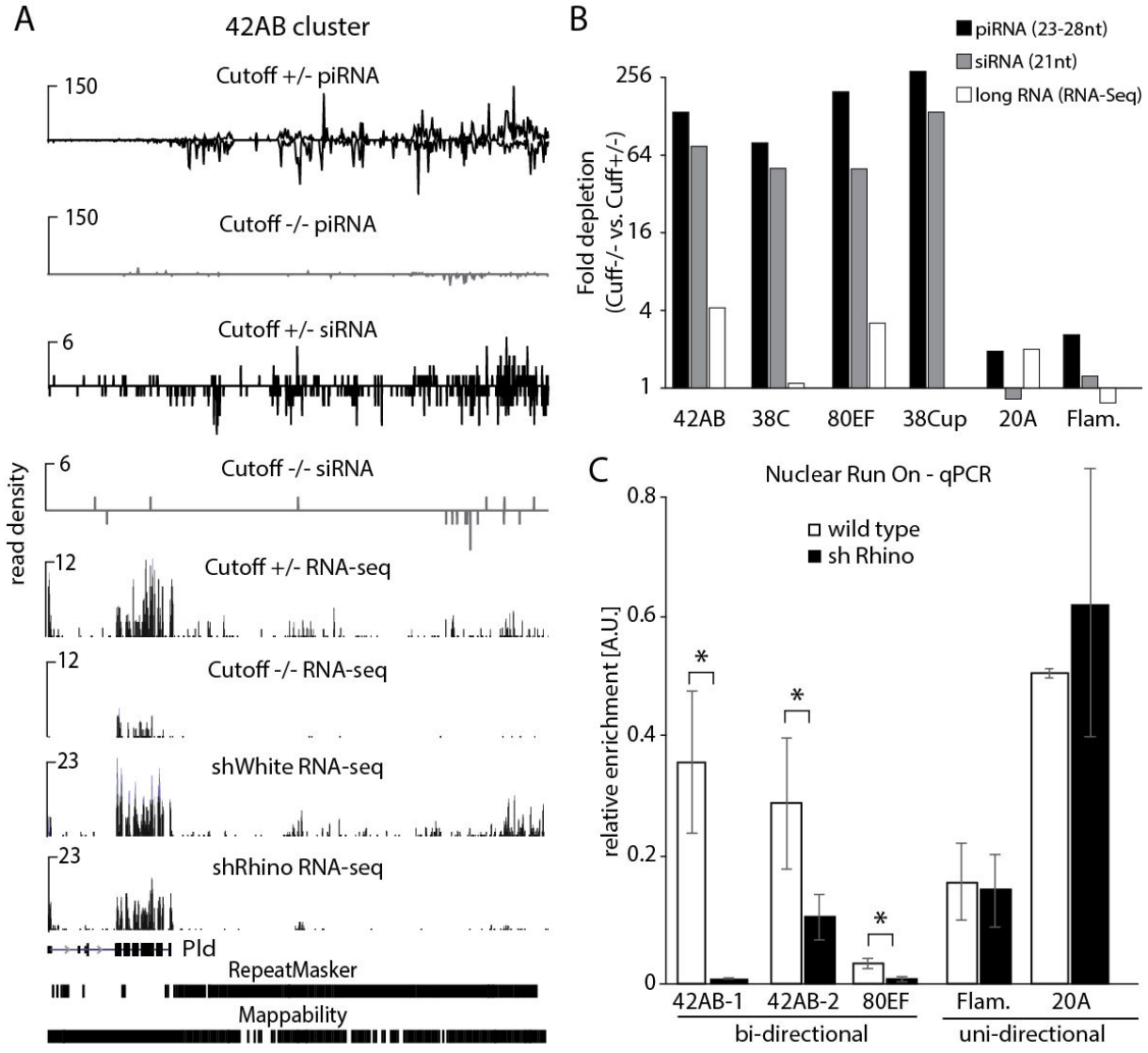


Figure 6. Transcriptional output of piRNA clusters is decreased in absence of Rhino and Cutoff.

(A) Profiles of Cutoff, piRNA, siRNA and long RNA (rRNA-minus RNA-Seq) along a fragment of the 42AB cluster (chromosome 2R: 2129920 - 2181079). Both piRNAs and siRNAs as well as long RNA from the 42AB locus are almost entirely eliminated in Cutoff mutant flies. Similar decrease in long RNA is seen in Rhino-depleted flies (shRhino). RNA-seq signal on the left correspond to exons of the Pld gene.

(B) The levels of piRNAs, siRNAs and long RNAs (rRNA-minus RNA-Seq) from double-stranded piRNA clusters are decreased in the Cutoff mutant. Shown are fold-decrease in the normalized number of piRNA, siRNA and long RNA reads generated from each piRNA cluster in the ovaries of Cutoff homozygous mutant compared to control heterozygous animals.

(C) Transcriptional output from double-stranded piRNA clusters is decreased upon Rhino depletion. Shown are RT-qPCR quantification of nascent transcripts from nuclear run-on performed on wild type and Rhino knockdown ovaries. The experiment was performed on two biological samples with three technical replicates each; p-values were calculated with a two-sided t-test, * indicates $p<0.05$.

Part 2 - A trans-generational epigenetic process defines piRNA biogenesis in *Drosophila virilis*.

To understand the determinants of piRNA cluster activity and identify the features that discriminate piRNA-generating regions from other genomic loci we took advantage of previous finding of piRNA clusters that are differentially active in two *D. virilis* strains[57]. Three genomic regions designated as clusters #1, #2, and #3 are present in the genomes of both strains 9 and 160, yet piRNAs are only generated from these clusters in strain 160 (we will refer to strain 160 as the Active or A strain) (Fig. 1, 2A). Strain 9 (the Inactive or I strain) does not have general defects in the piRNA pathway as piRNAs are generated from other genomic regions in this strain. The reason for differential ability to generate piRNAs from certain genomic regions in the two strains is unknown.

Following our previous results, we hypothesized that a difference in the chromatin state of the genomic regions in strains A and I might explain their differential ability to produce piRNAs. To determine whether the chromatin state plays a role in the differential activity of the piRNA clusters, we profiled the H3K9me3 mark in ovaries from flies of both strains. For each strain, ChIP-Seq libraries from two independent biological samples were generated, sequenced and analyzed. Genomic regions that generate piRNAs exclusively in strain A (clusters #1-3) had significantly higher levels of H3K9me3 in this strain compared to strain I (Fig. 1, 2A, B, S1A).

In contrast, cluster #4, which produced a similar amount of piRNAs in both strains, had comparable levels of the H3K9me3 mark (Fig. 2A, B, S1A). The H3K9me3 profile on piRNA clusters in strain A closely parallels the profile of piRNAs generation: both signals drop at the same genomic position (Fig. 1). Therefore, the activity of native piRNA-generating regions in *D. virilis* correlates with high levels of the H3K9me3 mark for all three differentially expressed clusters.

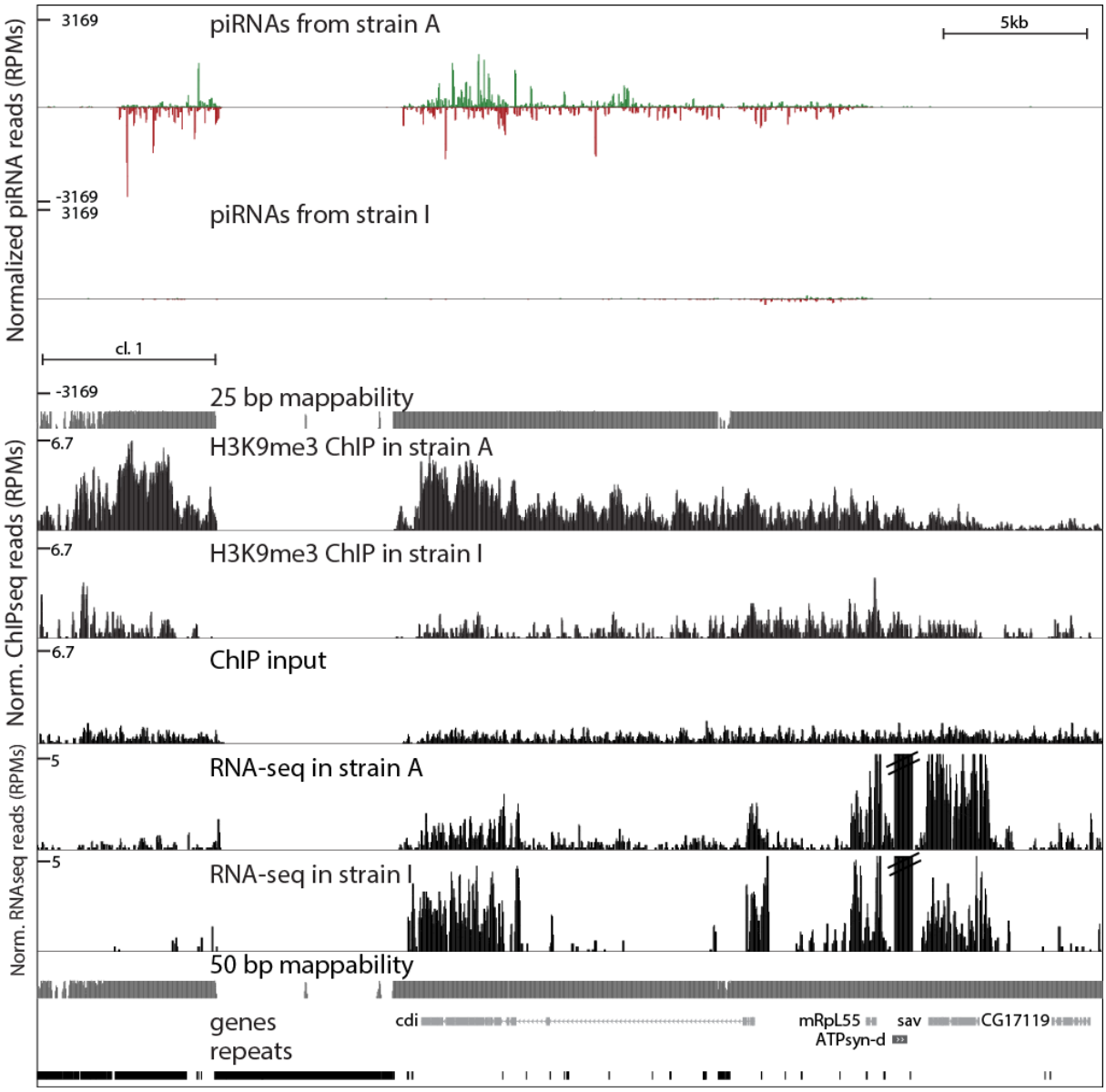


Figure 1. Profiles of the H3K9me3 mark, piRNAs and long RNAs on differentially expressed piRNA clusters.
piRNA, H3K9me3, and long RNA-seq (rRNA-depleted total RNA) densities on cluster #1 in the two *D. virilis* strains. High H3K9me3 signal is present over the cluster sequence in strain A that generates abundant piRNAs from this region; low H3K9me3 signal and no piRNAs are present in strain I. Only reads that mapped to the genome uniquely at this position are shown, resulting in the gap in read-density over a recent and non-divergent repetitive element (LTR retrotransposon) insertion. Unique-mappability tracks for 25-bp and 50-bp long reads are shown in gray. piRNAs mapping to the + and - strand are shown in green and red, respectively.

To understand how the H3K9me3 mark affects the activity of piRNA-generating loci, we used ChIP-seq to profile genome-wide RNA polymerase II occupancy and RNA-seq to measure transcript levels in ovaries from strains A and I. We found that high H3K9me3 levels on differentially expressed clusters in strain A correlated with high occupancy of RNA polymerase II on these regions (Fig. 2C, S1B). Furthermore, the piRNA clusters gave rise to more precursor transcripts in strain A than in strain I, indicating that the enrichment of Pol II at these sites represents polymerase engaged in active transcription (Fig. 2D, S2). Together, these results suggest that high level of H3K9me3 on piRNA clusters correlates with the generation of precursor transcript for piRNA processing.

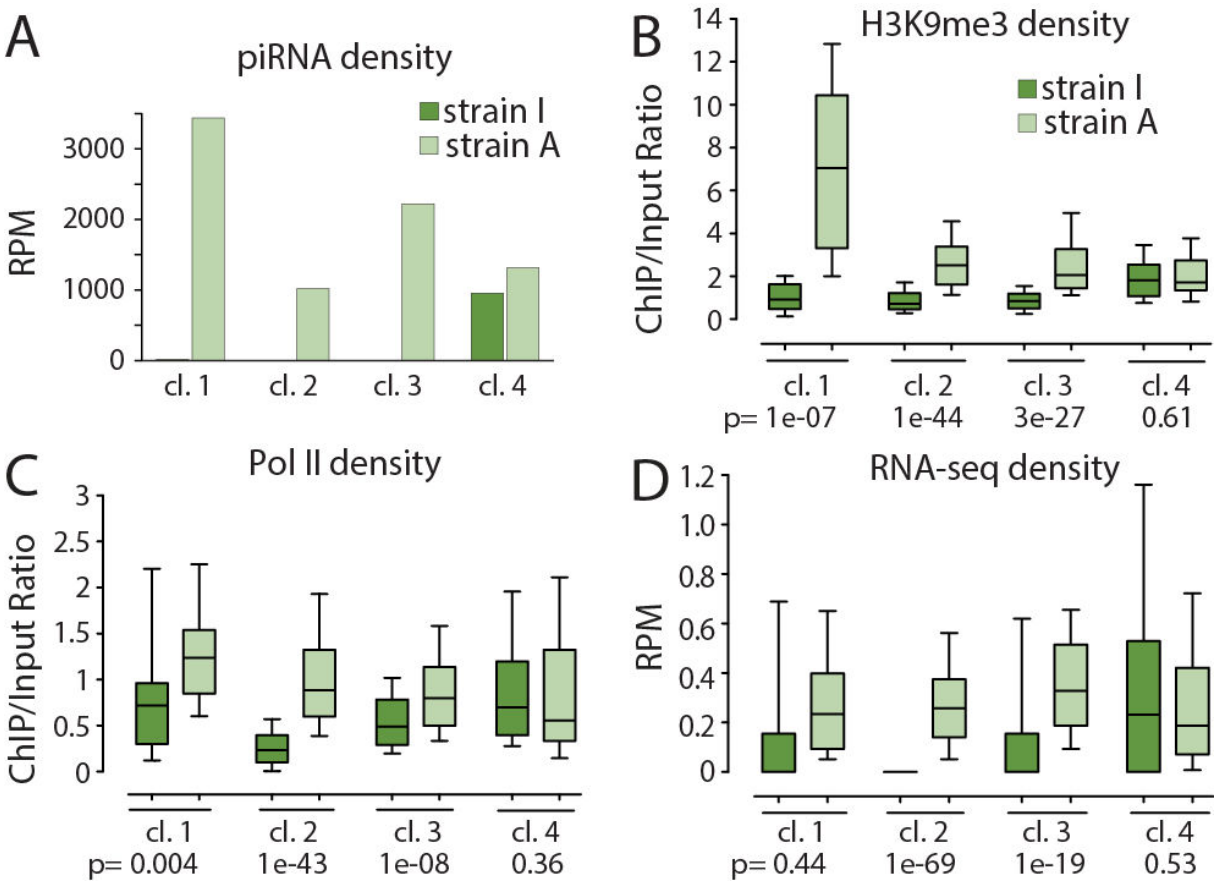


Figure 2. Enrichment of H3K9me3 on piRNA clusters correlates with increased transcription and piRNA generation.

(A) Differential expression of piRNA clusters in the two *D. virilis* strains. Clusters #1, #2, and #3 generate large numbers of piRNAs in strain A, but are inactive in strain I. Cluster #4 shows similar piRNA levels in both strains. Only the 5' portion of cluster 1 that does not contain protein-coding sequences (marked on Fig.1) was used for quantification on this and other panels.

(B) Production of piRNAs correlates with the enrichment of H3K9me3 over the genomic sequence. Clusters #1, #2, and #3 have high levels of H3K9me3 in strain A and low levels in strain I; cluster #4 has similar levels of H3K9me3 in both strains as measured by ChIP-seq. Cluster genomic regions were divided into 200 bp windows and a sequencing-depth normalized RPM (Reads Per Million) score was calculated for each window. The Mann-Whitney U test

was used to calculate the p-values and evaluate the significance of the observed differences. An independent biological replicate for this experiment is shown in Figure S1A.
(C) RNA polymerase II density over piRNA clusters in the two strains as measured by ChIP-seq. Clusters #1, #2, and #3 have higher RNA Pol II density in strain A, while cluster #4 shows similar levels of RNA Pol II signal in both strains. Cluster #1 contains a protein-coding gene; only sequences coming from the regions not overlapping with the gene were considered. Cluster genomic regions were divided into 200 bp windows and a sequencing-depth normalized RPM (Reads Per Million) score was calculated for each window. The Mann-Whitney U test was used to calculate the p-values and evaluate the significance of the observed differences. An independent biological replicate for this experiment is shown in Figure S1B.
(D) Long RNA expression from piRNA clusters in the two strains as measured by RNA-seq on rRNA-depleted total RNA. Clusters #1, #2, and #3 generate more transcripts in strain A than in strain I; levels were similar in the two strains for cluster #4. Cluster genomic regions were divided into 200 bp windows and a sequencing-depth normalized RPM (Reads Per Million) score was calculated for each window. The Mann-Whitney U test was used to calculate the p-values and evaluate the significance of the observed differences.

To further understand the properties of genomic regions that generate piRNAs we studied the activity of piRNA clusters in the progeny of the cross between strains A and I. The cross between the two strains can be performed in two different directions: the progeny of the cross between strain A females and strain I males was designated as MD (for Maternal Deposition) as the A chromosomes are inherited from the mothers (Fig. 3A). In the progeny of the opposite cross, designated as NMD (for No Maternal Deposition), the A chromosomes are inherited from the fathers. Importantly, the genotypes of MD and NMD progeny are absolutely identical.

Profiling of piRNAs in ovaries of MD and NMD flies showed that similar amounts of piRNAs were generated from cluster #4, which is equally expressed in both parental strains (Fig. 3B). Surprisingly, ~3-fold more piRNAs were generated from the three differentially active clusters in MD compared to NMD progeny. As the genotypes of MD and NMD progeny are identical, this result indicates that the trans-generational inheritance of an epigenetic signal that is transmitted from the mothers to their progeny enhances the ability to generate piRNAs.

One mechanism that can boost generation of piRNAs is processing of complementary transcripts through the ping-pong amplification loop. We analyzed the efficiency of ping-pong processing by measuring its characteristic feature, the number of complementary piRNA pairs with a 10-nt overlap between their 5' ends (so called ping-pong pairs). In the MD progeny, the majority of piRNA sequences coming from clusters 1-3 can form pairs with a 10-nt overlap between their 5' ends (Fig. 3C). In contrast, only a small fraction of piRNAs form ping-pong pairs in the NMD progeny. To account for the different number of sequences derived from differentially-expressed clusters in MD and NMD progenies, we sampled the same number of reads from each library and calculated the fraction of reads in ping-pong pairs a thousand times. A similar fraction of the piRNAs formed ping-pong pairs in strain A and in the MD progeny; however, the fraction of piRNAs generated through the ping-pong mechanism from clusters #1, #2, and #3 was dramatically reduced in the NMD progeny, compared to the fraction in the MD progeny (Fig. 3D). A similar fraction of piRNAs mapping to cluster #4 formed ping-pong pairs in strain A and in both the MD and the NMD progeny, indicating that the ping-pong machinery was intact in both progenies. These data suggest that an epigenetic signal inherited by MD progeny from their mothers boosts the ping-pong processing of cluster transcripts into mature piRNAs.

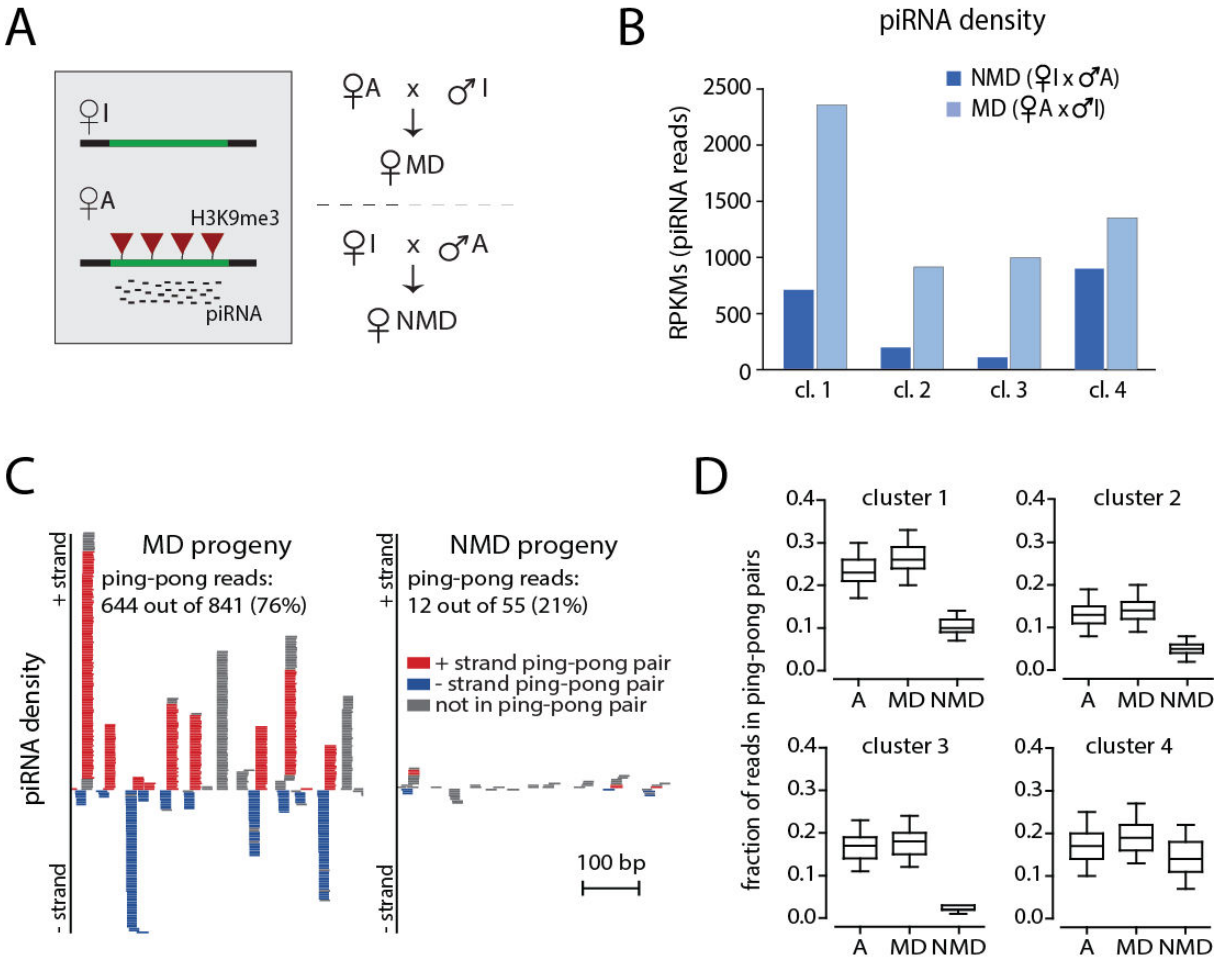


Figure 3. The effect of trans-generational epigenetic signal on the ping-pong piRNA processing.

(A) Scheme of the cross between A and I strains along with the chromatin status and piRNA levels at differentially expressed clusters in the parental strains. The level of H3K9me3 mark (red triangle) and piRNA abundance for the parental strains are shown in the gray box.

(B) Trans-generational inheritance of piRNAs from the mother leads to increased piRNA generation in the progeny. Expression of piRNAs in MD progeny (female A crossed to male I) that inherited piRNAs derived from clusters #1, #2, and #3 from their mothers and NMD progeny (female I crossed to male A) that did not inherit these piRNAs.

(C) piRNA density on a 500-bp fragment of cluster #1 shows higher numbers of ping-pong pairs in the MD progeny than in the NMD progeny. Shown are individual piRNA reads. Reads that have a ping-pong pair are shown in red and blue.

(D) Trans-generational epigenetic signal enhances the ping-pong amplification cycle in the progeny as indicated by the higher abundance of ping-pong pairs in MD compared to NMD progeny. To determine the fraction of piRNAs that participate in ping-pong pairs uniquely mapped piRNA reads that mapped to opposite strands of each other and have a 10 bp distance between their 5' ends were counted. To account for variable sequencing depth, we carried out the analysis by sampling reads to ~25% of the read counts in the library with the fewest reads in a region and repeating this 1000-times (see Material & Methods for details).

To test if increase in ping-pong biogenesis in MD progeny might be a consequence of increased production of piRNA precursors we analyzed the correlation between piRNA cluster expression and ping-pong biogenesis. If the increase in the fraction of piRNA participating in ping-pong pairs is a simple consequence of increased cluster expression then the fraction of such reads should correlate with cluster expression levels. We analyzed piRNAs generated by all double-stranded piRNA clusters and found no correlation between cluster expression and the fraction of reads in ping-pong pairs (Fig. S3). This result suggests that enhanced ping-pong processing in MD progeny is likely caused by a mechanism that is independent of the increase in the level of piRNA precursors.

Next, we studied the chromatin state of differentially active clusters in MD and NMD progenies. We compared the levels of H3K9me3 using separate ChIP-qPCR and ChIP-seq experiments performed on two independent biological samples of MD and NMD ovaries (Fig. 4A, S4). In both progenies H3K9me3 levels over the clusters were intermediate between the strong and weak enrichment seen in strain A and strain I, respectively. Importantly, both methods show that H3K9me3 enrichment was higher in ovaries of MD compared to NMD flies indicating that the maternally-supplied epigenetic signal affects chromatin of piRNA clusters in the progeny (Fig. 4A, S4).

The identical genomes of the MD and the NMD progenies each contain two alleles of differentially active piRNA clusters that have different H3K9me3 levels in their parents, the A and the I strains. To understand the impact of the maternally-supplied epigenetic signal on individual alleles, it is critical to differentiate sequences derived from the A and I chromosomes. We therefore carried out whole-genome sequencing of the genomes of both strains and generated 82x and 71x coverage for strains A and I, respectively. Next, we identified single nucleotide polymorphisms (SNPs) for each strain relative to the *D. virilis* reference genome (see the Methods section for details). We found 326,026 and 1,086,963 SNPs in genomes of A and I respectively, of which, 169,192 SNPs were shared. A total of 1,074,605 SNPs differed between the genomes of the two strains (average density of 5.19 such SNPs per Kb) allowing us to determine the allelic origin of piRNAs and ChIP-seq reads derived from polymorphic genomic regions. Importantly, we identified such SNPs in each of the differentially expressed piRNA clusters (Table S1). This allowed us to unambiguously map a fraction of the ChIP-seq and piRNA reads to each individual allele (Fig. 4B-D, S5).

We found that the two alleles of differentially expressed piRNA clusters maintained their distinct chromatin states in the hybrid progeny. Levels of H3K9me3 were higher on the A alleles compared to the I alleles in both the MD and the NMD ovaries (Fig. 4B, S5). However, MD progenies had higher levels of H3K9me3 mark on both the A and I alleles. This result indicates that maternally-inherited epigenetic signal in MD flies causes an increase in H3K9me3 methylation on the I alleles. Importantly, the increase in the H3K9me3 signal on the I alleles in the MD progeny correlated with the induction of piRNA generation from these alleles, whereas no piRNAs were generated from the I alleles in the NMD flies (Fig. 4C-D). Therefore, trans-generational inheritance of an epigenetic signal correlates with deposition of the H3K9me3 mark and induction of piRNA generation from previously inactive genomic loci (Fig. 4E).

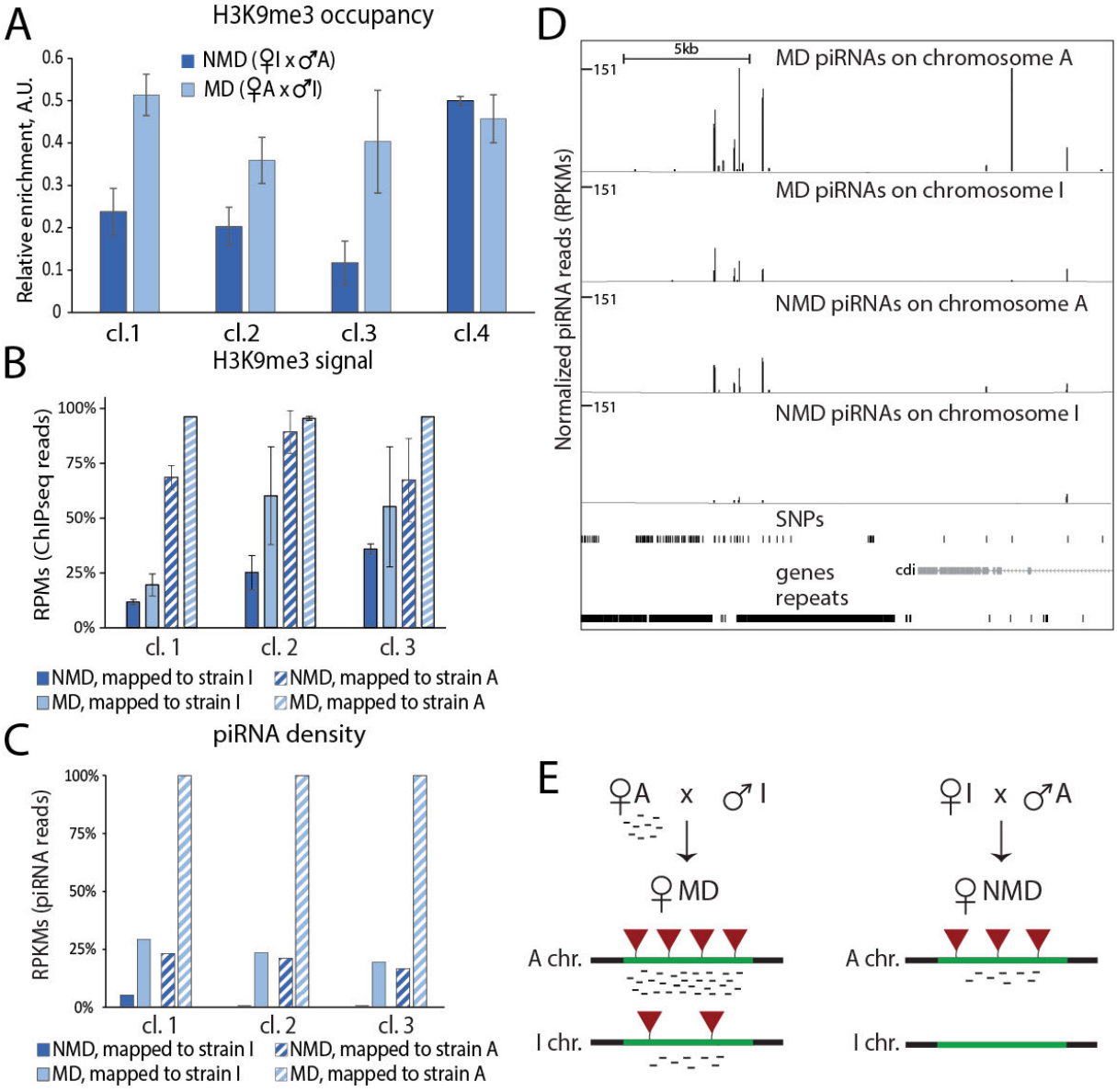


Figure 4. The effect of trans-generational epigenetic signal on the activity and the chromatin state of piRNA clusters.

(A) Trans-generational epigenetic signal leads to increase in H3K9me3 signal on piRNA clusters in the progeny. The input-normalized H3K9me3 signal as measured by ChIP-qPCR is higher in the MD progeny compared to the NMD progeny. ChIP-qPCR signals were normalized to respective inputs and compared to signal at a control genomic region located inside cluster #4 (scaffold_12823:176535-176654). The levels of H3K9me3 mark on the same clusters measured by ChIP-Seq performed on two independent biological samples are shown in Figure S4.

(B) Maternally inherited piRNAs lead to an increase in H3K9me3 levels on both the paternal and maternal genomes in the progeny. H3K9me3 ChIP-seq data was mapped to the individual alleles, considering only reads that could be unambiguously mapped to either the I or the A genome (see Methods for details). For each cluster the H3K9me3 levels are normalized to the signal on allele A in MD cross. Only reads mapping uniquely to one strain are shown. For both the I and the A chromosomes, H3K9me3 levels over the differentially expressed clusters increase when cognate piRNAs are inherited (MD vs. NMD progeny). At the same time each parental chromosome preserves its distinct chromatin state in the progeny: H3K9me3 levels are higher on the parental chromosome coming from strain A than on the one coming from strain I. The error bars represent the standard deviation of the mean between two independent biological replicates.

(C) Inheritance of piRNAs boosts piRNA production from both parental chromosomes in the progeny. piRNA density on each of the parental chromosomes was determined by mapping reads to the individual alleles, considering only reads that could be unambiguously mapped to either the I or the A genome. For each cluster the number of piRNAs are normalized to the piRNA derived from allele A in MD cross. For each differentially expressed cluster, more piRNAs are produced from each allele when piRNAs are maternally inherited (MD vs. NMD progeny). Importantly, inheritance of piRNAs activates piRNA generation from clusters that were inactive in the parental I strain (solid light blue vs. dark blue bars).

(D) The profiles of piRNA derived from A and I chromosomes in MD and NMD progenies on cluster #1. Only reads that mapped to SNPs that are distinct between strain A and I genomes are shown. Note that chromosome I generates piRNAs in MD, but not NMD flies. The distribution of SNPs is shown below piRNA tracks.

(E) Scheme of the cross between A and I strains along with the chromatin status (the level of H3K9me3 mark, red triangles) and the piRNA levels of each chromosome in the progeny.

Discussion

For the past 3 years, our results lead to three major advances in the piRNA field, that help us better understand the biogenesis of piRNAs:

1) **Piwi mediates transcriptional silencing through piRNA recognition of TE transcripts. The silencing is achieved by deposition of the repressive chromatin mark H3K9me3, and subsequently alteration of RNA Pol II transcription.**

2) **Germline piRNA clusters are fully heterochromatic and their promoters are involved in piRNA cluster transcription.**

3) **Trans-generationally inherited piRNAs are necessary to induce germline piRNA clusters formation at each new generation. They are responsible to recognize which genomic sequence is going to be generating new piRNA and process clusters transcripts into secondary piRNAs.**

Cluster chromatin and promoters:

Our results show that piRNA clusters are covered with repressive chromatin marks, and totally devoid of known active chromatin marks. It would suggest that piRNA clusters are being silenced by the cell machinery. However, bi-directional clusters are being effectively transcribed in nurse cell nuclei of the Drosophila ovary, do lead specifically to production of piRNAs, and have two distinct promoter regions (one if unidirectional, like X upstream), which are responsible for driving expression of both genomic strands: piRNAs are lost upon precise deletion of the promoter region. More importantly, these promoter regions present a stretch of DNA sequence that is highly similar in each one of them. Inside this homologous sequence we can identify a TATA box, an initiation motif, a CCAAT-like box motif, and especially the recurrence of one particular eight nucleotides motif, called DMv4. This motif is present three times in each promoter region and their deletion lead to loss of transcription. More work is needed to understand the exact role of these motif and especially what transcription factor might be binding this motif. In our hands, cluster promoters containing this motif were able to promote transcription in ovaries and S2 embryonic cells. Bi-directional clusters promoters present a

unique disposition of motifs: they have a special architecture where the three DMv4 motifs seem to hold a crucial role.

Trans-generational piRNAs in cluster activation:

De Vanssay and colleagues showed that a maternal factor supplied to the progeny by females expressing piRNAs from the T1 locus activates piRNA generation from the homologous inactive BX2 locus[40]. Furthermore, maintaining the activity of T1 in the subsequent generation also requires the maternal factor. We extended this observation to other systems and show that generation of piRNAs from a single-copy transgene, inserted into a telomeric piRNA cluster, also depends on a maternally-transmitted cytoplasmic factor. This maternal factor also activates piRNA generation from a single-copy euchromatic sequence, which simultaneously becomes the target of repression and the source of new piRNAs. In *D. virilis*, it is able to activate previously naive loci on I chromosomes. Finally, the activity of endogenous clusters in *D. melanogaster* also seems to require a maternally inherited factor: our analysis of previously published piRNA profiles in interspecies hybrids between *D. melanogaster* females and *D. simulans* males[58] showed that only *D. melanogaster* piRNA clusters generated piRNAs, while *D. simulans* piRNA clusters were inactive (Annexes – Chapter 3 Part 1 - Figure S12).

What is the nature of the maternally supplied epigenetic factor that triggers piRNA generation in the progeny? Multiple lines of evidence point to piRNAs themselves as the carriers of this epigenetic signal. As first shown by de Vanssay and colleagues, the epigenetic signal produced by the T1 locus does not require inheritance of the locus itself, indicating that the signal has non-chromosomal nature. This eliminates the possibility that the signal is any kind of chromatin mark linked to the active locus. Second, the process of BX2 activation genetically depends on piRNA pathway genes, but is independent of Dicer, which is required for siRNA biogenesis[40]. Third, both piRNAs and Piwi proteins are inherited from the maternal germline to the early embryos, while piRNAs are not transmitted through the sperm[28]. Finally, piRNAs can be sequence-specific guides to identify and activate homologous loci. Importantly, recent studies have shown that piRNAs and the nuclear Piwi protein trigger installation of the H3K9me3 mark on homologous targets providing a possible mechanism by which inherited piRNAs could lead to chromatin changes[49,59-61]. Together these results strongly support the role

of inherited piRNAs as a trans-generationally inherited epigenetic signal that activates piRNA generation from homologous loci in the progeny.

How can trans-generationally inherited piRNAs activate piRNA generation from homologous loci? Our results imply two mechanisms that cooperate and work at different steps of piRNA biogenesis. In the cytoplasm trans-generationally inherited piRNAs activate processing of complementary transcripts by the ping-pong amplification loop as evidenced by a dramatic increase in piRNAs generated by the ping-pong processing upon maternal deposition of cognate piRNAs.

In the ping-pong processing, initial piRNAs guide generation of secondary piRNAs from complementary sequences. Previously, it was proposed that the ping-pong cycle requires two types of piRNA precursors, cluster transcripts and transcripts from active transposons provided in *trans*[4,26]. Our results indicate that the ping-pong cycle can be activated by inherited piRNAs derived from the very same locus, provided that it is bi-directionally transcribed. Importantly, with the exception of one locus, all major piRNA clusters in the *D. melanogaster* germline are transcribed from both genomic strands providing an abundant source of complementary transcripts to be used by the ping-pong process.

The major players in ping-pong processing in Drosophila are two Piwi proteins, AUB and AGO3, while the third Piwi protein, PIWI itself, is not involved in this process (Fig. 2C, S2). AUB and AGO3 co-localize in cytoplasmic *nuage* granules, where the ping-pong processing is believed to take place[4]. Therefore, the effect of inherited piRNAs on ping-pong processing impacts a late step of piRNA biogenesis after piRNA precursor transcripts are exported to the cytoplasm.

Though enhancing the ping-pong processing is clearly an important mechanism by which trans-generationally inherited piRNAs boost piRNA biogenesis, it cannot explain all our observations, suggesting the existence of another mechanism. We found that maternal piRNAs are also required for the biogenesis of PIWI-associated piRNAs, although those are not generated by ping-pong processing. Using several genetic systems we showed that inheritance of piRNAs leads to an increase of the H3K9me3 mark on regions homologous to the piRNAs. Importantly, acquisition of the H3K9me3 mark by genomic regions that did not previously produce piRNAs, triggered piRNA generation in two transgenic systems. In contrast, absence of inherited piRNAs lead to a decrease H3K9me3 signal on homologous regions and

concomitant decrease of the corresponding piRNAs. These results suggest that modification of the chromatin structure of homologous genomic regions is the other mechanism by which trans-generationally inherited piRNAs turn on piRNA biogenesis in the progeny. Counter-intuitively, we found that enrichment of the H3K9me3 mark, which is generally assumed to be repressive, strongly correlates with enhanced piRNA biogenesis. In agreement with our results, a previous study showed that biogenesis of piRNAs from double-stranded clusters requires Eggless/SetDB1, one of the methyl-transferases responsible for installation of the H3K9me3 mark[33].

Our analysis of several transgenic piRNA clusters revealed differences in the impact of inherited piRNAs on the level of the H3K9me3 mark. The inherited piRNAs seem indispensable to maintain high H3K9me3 signal on the transgenic T1 and BX2* loci. However, the absence of maternal piRNAs leads to a relatively mild decrease in H3K9me3 on the telomeric piRNA cluster in the P1152 strain and *D. virilis* clusters. These results indicate that natural piRNA clusters are able to maintain a certain level of the H3K9me3 mark in a piRNA-independent fashion. This is not unexpected as natural piRNA clusters are located close to heterochromatin, which is known to have a high level of H3K9me3 signal. In contrast, the T1 and BX2 transgenes are inserted in a euchromatic site that is normally lacking this mark. Overall, our data strongly support an essential role of the H3K9me3 mark in piRNA generation. They further reveal that enrichment of this mark on regions that generate piRNAs at least partially depends on the inheritance of homologous piRNAs from the previous generation. Finally, acquisition of the H3K9me3 mark by a naïve locus as a result of exposure to homologous piRNAs strongly correlates with the initiation of de novo primary piRNA biogenesis from such a locus. The exact mechanism for piRNA-dependent deposition of the H3K9me3 mark on piRNA regions remains to be elucidated, however, recent studies suggest that it might occur through recognition of nascent transcripts by the nuclear PIWI/piRNA complex, which is known to be deposited by the mother into the developing egg[27,28] and which has been shown to install H3K9me3 on its genomic targets[49,59,60,62].

The proposal that inherited piRNAs trigger piRNA biogenesis by changing chromatin structure of homologues sequences raises the question of how the piRNAs distinguish a genuine transposon, a target that needs to be silenced, from a piRNA cluster that needs to be activated? Surprisingly, our results indicate that targeting by piRNAs leads to simultaneous repression of the target and activation of piRNA biogenesis from the same sequence. We found that targeting

of a unique sequence by piRNAs triggers accumulation of the H3K9me3 mark, a decrease in target expression and activation of piRNA biogenesis. Importantly, the target-derived piRNAs are not generated by the ping-pong mechanism (which would be a trivial explanation) as they are present in complex with PIWI, which does not participate in ping-pong processing. The similarity between transposon targets and piRNA-producing regions is supported by recent work that demonstrated that new transposon insertions in euchromatin start to generate piRNAs i.e. they are becoming de novo piRNA clusters[63]. By extension, we can propose that transposon remnant copies present inside piRNA clusters may not all be inactive. The few copies still active, most likely scattered along clusters, would participate in the establishment of H3K9me3 over the cluster: transcript recognition by Piwi-bound piRNAs mediating SetDB1 deposition of the repressive mark. Cluster promoters might avoid repression thanks to the presence of multiple Dmv4 motifs in its core, allowing for a strong binding of a yet-to-be-found TF increasing transcription. Careful consideration suggests that the requirement to 'silence' a genuine transposon target vs. 'activate' a piRNA cluster is a false dichotomy: if nascent transcripts generated from piRNA target loci are channeled into the piRNA processing machinery instead of the standard mRNA processing pathway, the transcript will be effectively silenced since no full-length mRNA will accumulate. The idea that target of piRNA repression becomes a source of new piRNAs makes the distinction between piRNA clusters (source of piRNAs) and targets obsolete. Furthermore, our results expose a case in which the same genomic region is 'silenced' and 'activated' at the same time, depending on the exact output the researcher is looking at (generation of full-length mRNA or piRNAs). Similar phenomena might be more widespread than previously suspected as studies in yeast suggest a very similar model, where centromeric repeats are 'silenced' and generate siRNAs at the same time[64].

How can the high level of the allegedly repressive H3K9me3 mark enhance piRNA biogenesis? Our results suggest that, counterintuitively, high levels of H3K9me3 correlate with elevated transcription in the context of piRNA clusters. Indeed we found that high levels of H3K9me3 on differentially expressed piRNA clusters in *D. virilis* correlated with increased RNA Pol II occupancy and with the generation of more precursor transcripts. We propose that changes in chromatin state associated with high levels of H3K9me3 lead to the recruitment of non-conventional readers of this mark, which subsequently affect transcription in these regions, providing more precursors for piRNA biogenesis. Our results show that the H3K9me3 mark provides a platform for the binding of Rhino, a chromodomain protein that shows specific enrichment over piRNA clusters. As high levels of H3K9me3 mark is also present in other

genomic regions, it is possible that recognition of H3K9me3 is not sufficient for Rhino stable binding and it interacts with other proteins to achieve its localization on chromatin of double-stranded clusters. Rhino forms a complex with Cutoff, a protein that is also required for piRNA biogenesis[51]. The H3K9me3 mark, Rhino and Cutoff co-localize at double stranded piRNA clusters and Cutoff is de-localized from nuclear foci in Rhino mutants[51] suggesting that it is recruited to piRNA clusters through its interaction with Rhino. Taken together our results suggest that Rhino and Cutoff, which were previously shown to be indispensable for piRNA generation from double-stranded piRNA clusters[32,51], are recruited to cluster chromatin through the H3K9me3 mark.

The exact molecular mechanism by which the Rhino/Cutoff complex activates piRNA biogenesis in the nucleus remains to be elucidated, however, two not necessarily mutually exclusive hypotheses can be proposed. First, Cutoff might bind and target nascent transcripts generated from piRNA clusters to the piRNA processing machinery instead of the normal pre-mRNA processing. In support of this idea we found that association of Cutoff with chromatin is RNA-dependent. We have previously shown that inserting intron-containing heterologous gene sequences into piRNA clusters results in abundant piRNAs from both the exonic and intronic sequences indicating that normal splicing is perturbed[45]. According to the second hypothesis, the Rhino/Cutoff complex might enhance transcription of piRNA clusters, hence providing more precursors for piRNA biogenesis. Indeed, run-on experiment showed that Rhino is required for efficient transcription of dual-stranded piRNA clusters. Furthermore, in agreement with an effect on transcription we found that in the Cutoff mutant both siRNAs and piRNAs as well as long RNAs are eliminated from double-stranded piRNA clusters arguing against a role of Cutoff exclusive to piRNA processing.

The counter-intuitive idea that the H3K9me3 mark might enhance rather than suppress transcription through binding of non-conventional epigenetic ‘readers’ has interesting parallels in yeast. In *S. pombe*, H3K9 methylation induces binding of Swi6/HP1, which then recruits the Jumonji protein Epe1 that promotes nucleosome turnover, resulting in increased transcription of heterochromatic repeats and generation of siRNAs[65,66]. One possible mechanism by which Cutoff might enhance cluster transcription is by suppressing RNA Pol II termination. Indeed, transgenic insertions that contain polyA cleavage/termination signals into piRNA clusters generate piRNAs downstream of the polyA signal indicating that not only splicing, but transcription termination is also suppressed in piRNA clusters[45]. Ignoring transcription

termination signals is likely an important feature of piRNA clusters, as otherwise, multiple signals within transposon sequences present in the clusters would terminate transcription and disrupt piRNA generation.

Overall, our data revealed that regions that produce piRNAs in Drosophila germ cells are defined by the epigenetic process of the trans-generational inheritance of cognate small RNAs. We found that inherited piRNAs trigger piRNA generation in the progeny by two mechanisms that seem to work simultaneously and cooperate to shape the final piRNA population (next Figure). In the nucleus inherited piRNAs mark genomic regions that will give rise to new piRNAs and enhance early steps of piRNA biogenesis. In the cytoplasm inherited piRNAs further trigger the post-transcriptional processing of cluster transcripts through the ping-pong amplification loop.

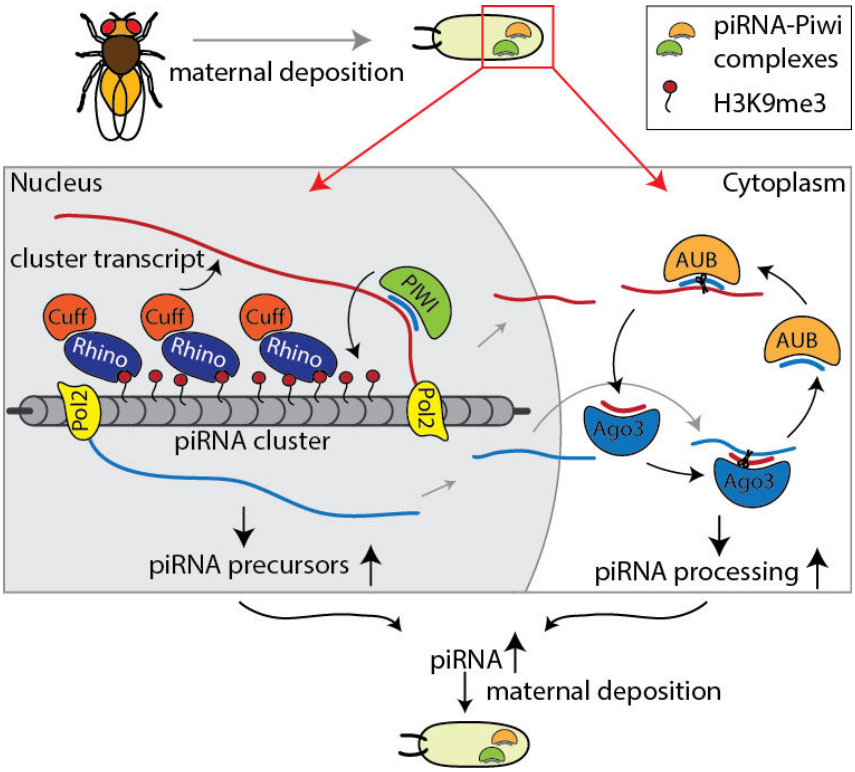


Figure: A model for the role of trans-generationally inherited piRNAs in determining piRNA biogenesis in the next generation.
The inherited piRNAs act in two ways: In the nucleus piRNAs mediate installment of the H3K9me3 mark over piRNA producing loci. Increased level of H3K9me3 is required for recruitment of the Rhino/Cutoff complex. Cutoff might target nascent transcripts to the piRNA processing machinery and/or enhance transcription. In the cytoplasm, piRNAs initiate the ping-pong amplification cycle thereby boosting processing of piRNA cluster transcripts into mature piRNAs.

Material and methods:

Drosophila stocks

D. virilis strain 160 (A) and strain 9 (I) were a generous gift from M. Evgeyev. Strains T1 and BX2 described in[40,42]. The strain P-1152, that carries insertion of P{lArB} construct in telomeric sequences of X chromosome (site 1A) is described in [67]. The strain BC69 that has insertion of P{A92} construct at a euchromatic location on chromosome 2L (site 35B10-35C1) is described in [68]. Both stocks were a generous gift from S. Ronsseray. The Rhino-GFP fly line (GFP-tagged Rhino driven by the Rhi promoter) was a generous gift from W. Theurkauff. The Cutoff-EGFP fly line (Nanos-GAL4/UASp-Cutoff-GFP), Cuff^{wm25} and Cuff^{qq37} were a generous gift from T. Schupbach. The Rhino-BioTAP flies were made by fusing the BioTAP tag[69] to the C-terminal region of the Rhino gene under the UASp promoter. shRhino flies were obtained from Bloomington stock center (stocks 34071 and 35171) and driven by Nos-Gal4 (stock 4937).

piRNA isolation, immunoprecipitation of PIWI proteins and small RNA cloning

Immunoprecipitations of PIWI, AUB and AGO3 proteins from *D. melanogaster* ovaries were carried out using rabbit polyclonal antisera directed against the N-terminal 14–16 AA according to previously described procedures[4]. Small RNAs from immunoprecipitates and total RNA extracts were cloned as previously described in[4,70]. Briefly, small RNAs within a 19-29nt window were isolated form 12% polyacrylamide gels. 3' and 5' linkers were ligated, and products were reverse transcribed using Superscript III (Invitrogen). Following PCR amplification, libraries were submitted for sequencing using the Illumina platform.

Analysis of the ping-pong processing

To determine the fraction of piRNAs that participate in ping-pong pairs we counted uniquely mapped piRNA reads that map to opposite strands of each other and have a 10 bp distance between their 5' ends. To account for variable sequencing depth, we carried out the analysis by sampling reads to ~25% of the read counts in the library with the fewest reads in a region and repeating this 1000-times.

ChIP and ChIP-qPCR

ChIPs were carried out using commercially available antibodies (anti-H3K9me3 (ab8898), anti-RNA Pol II (ab5408), and anti-GFP (DHSB-4C9[71]). ChIPs for Rhino and Cutoff were performed with anti-GFP antibodies on ovaries from transgenic animals expressing GFP-tagged proteins.

Prior to any ChIP experiment, 100 flies, per IP condition, were put on yeast for 2 to 3 days before ovary dissections. Ovaries were fixed using Paraformaldehyde (PFA) at a final concentration of 1% and incubated for 10min at RT. Samples were quenched by adding directly glycine (final concentration 25mM) for 5min at RT, and then washed 3 times in PBS. Ovaries were afterwards slightly dounced in Farnham Buffer (5mM HEPES pH8.0; 85mM KCl; 0.5% NP-40/Igepal; Protease Inhibitor Cocktail; NaF 10mM; Na3VO4 0.2mM) followed by strong douncing in RIPA Buffer (20mM Tris pH7.4; 150mM NaCl; 1% NP-40/Igepal; 0.5% Sodium Deoxycholate; 0.1% SDS; Protease Inhibitor Cocktail; NaF 10mM; Na3VO4 0.2mM) prior to sonication. Sonication was done using a Bioruptor from Diagenode on Medium power for 20 cycles (30sec ON, 30sec OFF). Samples were centrifuged, supernatant collected and pre-cleared for 2h at 4C using *Dynabeads Protein G* (Invitrogen) beads. If RNase treatment was required, half of the sample was incubated with 1µl of RNase A (10mg/ml) during pre-clearing and all samples incubated 2h at RT. Meanwhile, in parallel, antibodies were conjugated to *Dynabeads Protein G* for 2h at 4C too. 5% of pre-cleared samples were saved for Input fraction and the rest was then transferred to the antibody-conjugated beads and incubated for 2h at 4C. Beads were then washed 5 times at 4C using LiCl IP Buffer (10mM Tris ph7.5; 500mM LiCl; 1% NP-40/Igepal; 1% Sodium Deoxycholate), then rinsed in TE and finally resuspended in Proteinase K Buffer (200mM Tris ph7.4; 25mM EDTA; 300mM NaCL; 2% SDS) with 100µg of proteinase K. Samples were incubated 3h at 55C then overnight at 65C. DNA was then extracted following standard phenol-chlorophorm extraction and concentration was measured by Qbit.

In case of the Cutoff ChIP-seq, the experiment was performed using an altered protocol: ovaries were treated with Collagenase (2mg/ml) for 3 minutes at room temperature (RT), washed once in PBS, then fixed using fresh EGS (Thermo Scientific, 21565) solution (1,5mg/ml final in PBS) for 30 min at RT. Paraformaldehyde (PFA) was then added to the solution to a final

concentration of 1% and incubated for 10min at RT. Consecutive steps were as previously described. H3K9me3 and Rhino quantitative ChIP-PCR (ChIP-qPCR) experiments were performed in at least two biological replicates with three technical replicates each. Cutoff ChIP-qPCR was performed in two technical replicates. Values were normalized to respective inputs and to a genomic region known to be enriched in H3K9me3, Rhino, and Cutoff (chr2R: 2336913-2337023). Cutoff ChIP-qPCR including RNase treatment was normalized to respective inputs and to the control region RP49. For primers see Table S2. Error bars represent the standard error of the mean.

ChIP-seq and RNA-seq library construction and high-throughput data analysis

ChIP-seq and RNA-seq library construction and sequencing were carried out using standard protocols following the general principles described by Johnson et al.,2007[72] and Mortazavi et al., 2008[73], respectively. Libraries were sequenced on the Illumina HiSeq 2000 (50-bp reads) platform. Datasets for piRNAs were extracted from available GEO data[57,58]: *D. melanogaster* SRX205641, *D. simulans* SRX205642, and F1 SRX205643, *D. virilis* strain A (160) SRX023769, strain I (9) SRX023754, strain NMD SRX023737, and strain MD SRX023735. The resulting sequencing reads were mapped against the genome using Bowtie[74] v0.12.7 with the following settings: "-v 0 --best --strata", retaining only uniquely mappable reads with zero mismatches. Read mapping statistics for ChIP-seq datasets processed this way are presented in Tables S1 (Chapter 2 – Part 1) and S2 (Chapter 2 – Part 2). Data analysis was carried out using a combination of publicly available software tools and custom-written python scripts.

D. virilis genome sequencing and allele-specific mapping

Genomic libraries were generated from each *D. virilis* strain and sequenced (as 2x100 reads) on the Illumina HiSeq 2000 platform. Sequencing coverage was ~82x for strain A and ~71x for strain I. Reads were mapped against the droVir2 version of the *D. virilis* genome using bwa mem (Li and Durbin, 2009) (version 0.7.5a). Variant calling was carried out using GATK (DePristo et al., 2011) in two steps. First, an initial set of variants was obtained and filtered for high-confidence variants (GQ > 90; DP >15). That set of variants was then used for base-quality recalibration and a second round of variant calling. Single-nucleotide polymorphisms with GQ >30 were retained. Next, a diploid version of the droVir2 assembly was constructed using the resulting set of SNPs. A Bowtie[74] (v0.12.7) index containing two versions of each droVir2

chromosome (one for each strain) was created and ChIP-seq, RNA-seq, and piRNA reads were mapped to it with the following settings: "-v 0 -k 3 -m 2 -t --best --strata" (i.e., allowing for up to 2 locations in the diploid genome that a read can map to and zero mismatches). The alignments were further filtered to exclude any reads that map to 2 different positions. Detailed statistics on the number of SNPs identified genome-wide and in the particular piRNA clusters studied here can be found in Table S1 (Chapter 2 – Part 2).

Rhino and Cutoff co-immunoprecipitation

Immunoprecipitation and western blots were performed as previously described[51,75] on ovarian extracts from transgenic flies expressing Rhino-BioTAP and Cutoff-EGFP using anti-BioTAP antibody (Sigma, P1291) for immunoprecipitating Rhino-BioTAP. Western blots were carried on with anti-BioTAP antibody at 1:1000 and anti-GFP antibody at 1:1000 dilution in 5% non-fat milk.

Cloning, expression, and purification of Rhino

The *Drosophila melanogaster* Rhino chromodomain (19-85) was inserted into the pETSumo expression vector (Invitrogen) which fuses a hexa-His tag plus a yeast sumo tag at the N-terminus to the target protein. The plasmid was transformed into the *E. coli* BL21(DE3) strain (Stratagene). The cells were cultured at 37 °C until OD600 reached 0.8, and then the protein expression was induced with 0.2 mM IPTG at 18 °C overnight. The hexa-His-Sumo tagged protein was purified using a HisTrap FF column (GE Healthcare). The tag was cleaved by Ulp1 protease and further removed by a second step HisTrap FF column (GE Healthcare) purification. The target protein was further purified by a Hiload Superdex G200 16/60 column (GE Healthcare).

Crystallization and diffraction data collection

The purified Rhino protein was concentrated to about 15 mg/ml in a storage buffer of 150 mM NaCl, 20 mM Tris, pH 7.5, and 5 mM DTT. Synthetic H3(1-15)K9me3 peptide was added at a threefold excess molar ratio to Rhino chromodomain and incubated at 4 °C for 30 min. Crystallization was conducted at 20 °C using the hanging drop vapor diffusion method by mixing 1 µl protein solution and 1µl reservoir solution (0.2 M ammonium nitrate, and 20% PEG3350) and equilibrated against 0.5 ml reservoir solution. The crystals were soaked into the reservoir solution supplemented with 15% glycerol before flash-cooled into liquid nitrogen for

diffraction data collection. The diffraction data were collected at beamline 24-ID-E at the Argonne National Laboratory, Chicago, and processed using the HKL2000 suite[76]. The statistics of the diffraction data are summarized in Table S4.

Structure determination and refinement

The structure of Rhino chromodomain was solved using the molecular replacement method as implemented in Phaser[77] using the structure of the HP1 chromodomain as the search model (PDB ID: 1KNE). The model building was carried out using Coot[78] and the structure refinement was conducted using Phenix[79]. Throughout the refinement, a free *R* factor was calculated using 5% randomly chosen reflections. The stereochemistry of the structure model was analyzed using Procheck[80]. The structure of Rhino-H3(1-15)K9me3 complex was solved using the same protocol as the free form structure. The statistics of the refinement and the structure model are summarized in Table S4. All the molecular graphics were generated using Pymol (DeLano Scientific LLC).

Isothermal titration calorimetry

All the peptides, including H3(1-15), H3(1-15)K9me1, H3(1-15)K9me2 and H3(1-15)K9me3, were purchased from Tufts University Peptide Facility. Protein concentration was determined by absorbance spectroscopy. The Rhino chromodomain (23-85) was dialyzed against 25 mM sodium chloride, 20 mM Tris, pH 8.0, 2 mM b mercaptoethanol. The peptide was dissolved in the same buffer. ITC was conducted at a Microcal calorimeter ITC 200 instrument at 20 °C. Binding curves were analyzed by non-linear least-squares fitting of the data using Origin 7.0 software.

Nuclear Run-On / GRO-Seq

Nuclear run-on / GRO-Seq experiments were performed on ovaries of shRhino flies (Bloomington stock 3407) driven by Nos-Gal4 (Bloomington stock 4937). Nos-Gal4 flies were used as a control. The nuclear run-on procedure was carried out as previously described[81] with slight modifications. BrUTP (5'-Bromouridine-5'-triphosphate, Sigma, B7166) labeled NRO-RNA was filtered through Illustra MicroSpin G25 columns (27-5325-01) twice to remove unincorporated BrUTP. The NRO-RNA was captured using the anti-BrdU antibody (Sigma, 032M 4753) for 1 hour followed by incubation with ProteinG beads (Dynabeads, Invitrogen, 1003D) for 1 hour. The immunoprecipitation procedure was sequentially performed three times

to yield highly enriched BrUTP RNA. As a negative control the same procedure was performed on non-labeled, total *Drosophila* ovary RNA. As a quality control RT-qPCR was performed on 10% of the purified RNA with primers for Vasa, Rp49 and selected piRNA clusters. Libraries were cloned with NEBNext Ultra Directional RNA Library Kit (E7420S) and sequenced on the Illumina HiSeq 2000 (50-bp reads) platform. Reads were mapped uniquely to the dm3 genome using Bowtie 0.12.7 allowing for no mismatches. The numbers of reads from the respective clusters are shown in Table S5.

BAC recombineering

Region targeted in the 38C upstream cluster for LacZ fragments insertion: BAC CH322-76F01, insertion in chr2L:20101698-20119654.

See published paper for detailed protocol information[82]:

A Framework for piRNA Cluster Manipulation
Olovnikov, I., Le Thomas, A., Aravin, A.A.
Methods in Molecular Biology, Volume 1093, *PIWI-Interacting RNAs*, 47-58.

References

1 Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811, doi:10.1038/35888 (1998).

2 Ghildiyal, M. & Zamore, P. Small silencing RNAs: an expanding universe. *Nature reviews. Genetics* **10**, 94-108, doi:10.1038/nrg2504 (2009).

3 Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. *Curr Biol* **11**, 1017-1027 (2001).

4 Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**, 1089-1103, doi:10.1016/j.cell.2007.01.043 (2007).

5 Vagin, V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science (New York, N.Y.)* **313**, 320-324, doi:10.1126/science.1129333 (2006).

6 Bastock, R. & St Johnston, D. Drosophila oogenesis. *Curr Biol* **18**, R1082-1087, doi:10.1016/j.cub.2008.09.011 (2008).

7 Fedoroff, N. V. Transposable elements, epigenetics, and genome evolution. *Science* **338**, 758-767 (2012).

8 Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8**, 973-982, doi:10.1038/nrg2165 (2007).

9 Bucheton, A. I transposable elements and I-R hybrid dysgenesis in Drosophila. *Trends Genet* **6**, 16-21 (1990).

10 Kazazian, H. H., Jr. Mobile elements: drivers of genome evolution. *Science* **303**, 1626-1632, doi:10.1126/science.1089670 (2004).

11 Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. PIWI-interacting small RNAs: the vanguard of genome defence. *Nature reviews. Molecular cell biology* **12**, 246-258, doi:10.1038/nrm3089 (2011).

12 Malone, C. D. & Hannon, G. J. Small RNAs as Guardians of the Genome. *Cell* **136**, 656-668, doi:http://dx.doi.org/10.1016/j.cell.2009.01.045 (2009).

13 Gunawardane, L. *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science (New York, N.Y.)* **315**, 1587-1590, doi:10.1126/science.1140494 (2007).

14 Robine, N. *et al.* A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Current biology : CB* **19**, 2066-2076, doi:10.1016/j.cub.2009.11.064 (2009).

15 Malone, C. D. *et al.* Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell* **137**, 522-535, doi:10.1016/j.cell.2009.03.040 (2009).

16 Saito, K. *et al.* A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. *Nature* **461**, 1296-1299, doi:10.1038/nature08501 (2009).

17 Haase, A. *et al.* Probing the initiation and effector phases of the somatic piRNA pathway in Drosophila. *Genes & development* **24**, 2499-2504, doi:10.1101/gad.1968110 (2010).

18 Ipsaro, J., Haase, A., Knott, S., Joshua-Tor, L. & Hannon, G. The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature* **491**, 279-283, doi:10.1038/nature11502 (2012).

19 Nishimasu, H. *et al.* Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature* **491**, 284-287, doi:10.1038/nature11509 (2012).

20 Kawaoka, S., Izumi, N., Katsuma, S. & Tomari, Y. 3' end formation of PIWI-interacting RNAs in vitro. *Mol Cell* **43**, 1015-1022, doi:10.1016/j.molcel.2011.07.029 (2011).

21 Aravin, A. *et al.* A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* **31**, 785-799, doi:10.1016/j.molcel.2008.09.003 (2008).

22 Saito, K. *et al.* Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. *Genes & development* **21**, 1603-1608, doi:10.1101/gad.1563607 (2007).

23 Frank, F., Hauver, J., Sonenberg, N. & Nagar, B. Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. *The EMBO journal* **31**, 3588-3595, doi:10.1038/emboj.2012.204 (2012).

24 Frank, F., Sonenberg, N. & Nagar, B. Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465**, 818-822, doi:10.1038/nature09039 (2010).

25 Olivieri, D., Sykora, M., Sachidanandam, R., Mechtler, K. & Brennecke, J. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. *EMBO J* **29**, 3301-3317, doi:10.1038/emboj.2010.212 (2010).

26 Aravin, A., Hannon, G. & Brennecke, J. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science (New York, N.Y.)* **318**, 761-764, doi:10.1126/science.1146484 (2007).

27 Megosh, H., Cox, D., Campbell, C. & Lin, H. The role of PIWI and the miRNA machinery in Drosophila germline determination. *Current biology : CB* **16**, 1884-1894, doi:10.1016/j.cub.2006.08.051 (2006).

28 Brennecke, J. *et al.* An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science (New York, N.Y.)* **322**, 1387-1392, doi:10.1126/science.1165171 (2008).

29 Lin, H. & Spradling, A. A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. *Development (Cambridge, England)* **124**, 2463-2476 (1997).

30 Darricarrère, N., Liu, N., Watanabe, T. & Lin, H. Function of Piwi, a nuclear Piwi/Argonaute protein, is independent of its slicer activity. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 1297-1302, doi:10.1073/pnas.1213283110 (2013).

31 Moshkovich, N. & Lei, E. HP1 recruitment in the absence of argonaute proteins in Drosophila. *PLoS genetics* **6**, doi:10.1371/journal.pgen.1000880 (2010).

32 Klattenhoff, C. *et al.* The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* **138**, 1137-1149, doi:10.1016/j.cell.2009.07.014 (2009).

33 Rangan, P. *et al.* piRNA production requires heterochromatin formation in Drosophila. *Current biology : CB* **21**, 1373-1379, doi:10.1016/j.cub.2011.06.057 (2011).

34 Kharchenko, P. V. *et al.* Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**, 480-485, doi:10.1038/nature09725 (2011).

35 Dolfini, D., Gatta, R. & Mantovani, R. NF-Y and the transcriptional activation of CCAAT promoters. *Crit Rev Biochem Mol Biol* **47**, 29-49, doi:10.3109/10409238.2011.628970 (2012).

36 Fleming, J. D. *et al.* NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors. *Genome Res* **23**, 1195-1209, doi:10.1101/gr.148080.112 (2013).

37 Down, T. A., Bergman, C. M., Su, J. & Hubbard, T. J. Large-scale discovery of promoter motifs in Drosophila melanogaster. *PLoS Comput Biol* **3**, e7, doi:10.1371/journal.pcbi.0030007 (2007).

38 FitzGerald, P. C., Sturgill, D., Shyakhtenko, A., Oliver, B. & Vinson, C. Comparative genomics of Drosophila and human core promoters. *Genome Biol* **7**, R53, doi:10.1186/gb-2006-7-7-r53 (2006).

39 Kondo, S. & Ueda, R. Highly improved gene targeting by germline-specific Cas9 expression in Drosophila. *Genetics* **195**, 715-721, doi:10.1534/genetics.113.156737 (2013).

40 de Vanssay, A. *et al.* Paramutation in Drosophila linked to emergence of a piRNA-producing locus. *Nature* **490**, 112-115, doi:10.1038/nature11416 (2012).

41 Harris, A. N. & Macdonald, P. M. Aubergine encodes a Drosophila polar granule component required for pole cell formation and related to eIF2C. *Development* **128**, 2823-2832 (2001).

42 Dorer, D. R. & Henikoff, S. Transgene repeat arrays interact with distant heterochromatin and cause silencing in cis and trans. *Genetics* **147**, 1181-1190 (1997).

43 Ronsseray, S., Boivin, A. & Anxolabehere, D. P-Element repression in Drosophila melanogaster by variegating clusters of P-lacZ-white transgenes. *Genetics* **159**, 1631-1642 (2001).

44 Josse, T. *et al.* Telomeric trans-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation. *PLoS genetics* **3**, 1633-1643, doi:10.1371/journal.pgen.0030158 (2007).

45 Muerdter, F. *et al.* Production of artificial piRNAs in flies and mice. *RNA (New York, N.Y.)* **18**, 42-52, doi:10.1261/rna.029769.111 (2012).

46 Ronsseray, S., Josse, T., Boivin, A. & Anxolabehere, D. Telomeric transgenes and trans-silencing in Drosophila. *Genetica* **117**, 327-335 (2003).

47 Olovnikov, I. A. & Kalmykova, A. I. piRNA clusters as a main source of small RNAs in the animal germline. *Biochemistry. Biokhimiia* **78**, 572-584, doi:10.1134/S0006297913060035 (2013).

48 Li, C. *et al.* Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137**, 509-521, doi:10.1016/j.cell.2009.04.027 (2009).

49 Le Thomas, A. *et al.* Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev* **27**, 390-399, doi:10.1101/gad.209841.112 (2013).

50 Jacobs, S. A. & Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080-2083, doi:10.1126/science.1069473 (2002).

51 Pane, A., Jiang, P., Zhao, D. Y., Singh, M. & Schupbach, T. The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline. *EMBO J* **30**, 4601-4615, doi:10.1038/emboj.2011.334 (2011).

52 Jiao, X., Chang, J. H., Kilic, T., Tong, L. & Kiledjian, M. A mammalian pre-mRNA 5' end capping quality control mechanism and an unexpected link of capping to pre-mRNA processing. *Mol Cell* **50**, 104-115, doi:10.1016/j.molcel.2013.02.017 (2013).

53 Jiao, X. *et al.* Identification of a quality-control mechanism for mRNA 5'-end capping. *Nature* **467**, 608-611, doi:10.1038/nature09338 (2010).

54 Kim, M. *et al.* The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**, 517-522, doi:10.1038/nature03041 (2004).

55 Xue, Y. *et al.* Saccharomyces cerevisiae RAI1 (YGL246c) is homologous to human DOM3Z and encodes a protein that binds the nuclear exoribonuclease Rat1p. *Molecular and cellular biology* **20**, 4006-4015 (2000).

56 Czech, B. *et al.* An endogenous small interfering RNA pathway in Drosophila. *Nature* **453**, 798-802, doi:10.1038/nature07007 (2008).

57 Rozhkov, N. *et al.* Small RNA-based silencing strategies for transposons in the process of invading Drosophila species. *RNA (New York, N.Y.)* **16**, 1634-1645, doi:10.1261/rna.2217810 (2010).

58 Kelleher, E., Edelman, N. & Barbash, D. Drosophila interspecific hybrids phenocopy piRNA-pathway mutants. *PLoS biology* **10**, doi:10.1371/journal.pbio.1001428 (2012).

59 Rozhkov, N., Hammell, M. & Hannon, G. Multiple roles for Piwi in silencing Drosophila transposons. *Genes & development* **27**, 400-412, doi:10.1101/gad.209767.112 (2013).

60 Sienski, G., Dönertas, D. & Brennecke, J. Transcriptional silencing of transposons by piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151**, 964-980, doi:10.1016/j.cell.2012.10.040 (2012).

61 Wang, S. & Elgin, S. Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 21164-21169, doi:10.1073/pnas.1107892109 (2011).

62 Kalmykova, A. I., Klenov, M. S. & Gvozdev, V. A. Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. *Nucleic Acids Res* **33**, 2052-2059, doi:10.1093/nar/gki323 (2005).

63 Shpiz, S., Ryazansky, S., Olovnikov, I., Abramov, Y. & Kalmykova, A. Euchromatic transposon insertions trigger production of novel Pi- and endo-siRNAs at the target sites in the drosophila germline. *PLoS Genet* **10**, e1004138, doi:10.1371/journal.pgen.1004138 (2014).

64 Grewal, S. & Jia, S. Heterochromatin revisited. *Nature reviews. Genetics* **8**, 35-46, doi:10.1038/nrg2008 (2007).

65 Grewal, S. I. RNAi-dependent formation of heterochromatin and its diverse functions. *Curr Opin Genet Dev* **20**, 134-141, doi:10.1016/j.gde.2010.02.003 (2010).

66 Zofall, M. & Grewal, S. I. Swi6/HP1 recruits a JmjC domain protein to facilitate transcription of heterochromatic repeats. *Mol Cell* **22**, 681-692, doi:10.1016/j.molcel.2006.05.010 (2006).

67 Roche, S. E. & Rio, D. C. Trans-silencing by P elements inserted in subtelomeric heterochromatin involves the Drosophila Polycomb group gene, Enhancer of zeste. *Genetics* **149**, 1839-1855 (1998).

68 Lemaitre, B., Ronsseray, S. & Coen, D. Maternal repression of the P element promoter in the germline of Drosophila melanogaster: a model for the P cytotype. *Genetics* **135**, 149-160 (1993).

69 Alekseyenko, A. A., Gorchakov, A. A., Kharchenko, P. V. & Kuroda, M. I. Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by BioTAP-XL cross-linking and affinity purification. *Proc Natl Acad Sci U S A* **111**, 2488-2493, doi:10.1073/pnas.1400648111 (2014).

70 Aravin, A. A. *et al.* A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* **31**, 785-799, doi:S1097-2765(08)00619-9 [pii]
10.1016/j.molcel.2008.09.003 (2008).

71 Sanchez, P., Daniels, K. J., Park, Y. N. & Soll, D. R. Generating a battery of monoclonal antibodies against native green fluorescent protein for immunostaining, FACS, IP, and ChIP using a unique adjuvant. *Monoclonal antibodies in immunodiagnosis and immunotherapy* **33**, 80-88, doi:10.1089/mab.2013.0089 (2014).

72 Johnson, D., Mortazavi, A., Myers, R. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science (New York, N.Y.)* **316**, 1497-1502, doi:10.1126/science.1141319 (2007).

73 Mortazavi, A., Williams, B., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-628, doi:10.1038/nmeth.1226 (2008).

74 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, doi:10.1186/gb-2009-10-3-r25 (2009).

75 Chen, Y., Pane, A. & Schupbach, T. Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in Drosophila. *Curr Biol* **17**, 637-642, doi:10.1016/j.cub.2007.02.027 (2007).

76 Otwinowski, Z. & Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology, Macromolecular Crystallography.* **Volume 276**, p.307-326 (1997).

77 McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta crystallographica. Section D, Biological crystallography* **63**, 32-41, doi:10.1107/S0907444906045975 (2007).

78 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).

79 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).

80 Laskowski, R. A., Moss, D. S. & Thornton, J. M. Main-chain bond lengths and bond angles in protein structures. *Journal of molecular biology* **231**, 1049-1067, doi:10.1006/jmbi.1993.1351 (1993).

81 Shpiz, S. *et al.* Mechanism of the piRNA-mediated silencing of Drosophila telomeric retrotransposons. *Nucleic Acids Res* **39**, 8703-8711, doi:10.1093/nar/gkr552 (2011).

82 Olovnikov, I., Le Thomas, A. & Aravin, A. A. A Framework for piRNA Cluster Manipulation. *Methods in molecular biology* **1093**, 47-58, doi:10.1007/978-1-62703-694-8_5 (2014).

Annexes – Supplementary Tables

Chapter 1

Supplemental materials available online at the address:

http://genesdev.cshlp.org/content/suppl/2013/01/31/gad.209841.112.DC1.html

Chapter 3 - Part 1

Table S1. Read mapping statistics

Library	Read length	Unique mappers
RNA-seq (rRNA minus) *Cutoff*$^{+}$*/Cutoff*$^{qq37/wm25}$, *D. mel.*	50	6,870,786
RNA-seq (rRNA minus) *Cutoff*^{qq37}*/Cutoff*^{wm25}, *D. mel.*	50	583,407
RNA-seq (rRNA minus) shWhite, *D. mel.*	50	5,688,673
RNA-seq (rRNA minus) shRhino, *D. mel.*	50	5,075,564

Library	Cloning range	Unique mappers
SmallRNA-seq *D. mel* mapped on *D. sim*. genome	19-29	617,944
SmallRNA-seq *D. sim* mapped on *D. sim*. genome	19-29	715,023
SmallRNA-seq F1 mapped on *D. sim*. genome	19-29	860,898
SmallRNA-seq *D. mel* mapped on *D. mel*. genome	19-29	1,224,035
SmallRNA-seq *D. sim* mapped on *D. mel*. genome	19-29	715,023
SmallRNA-seq F1 mapped on *D. mel*. genome	19-29	575,365
SmallRNA-seq *Cutoff*$^{+}$*/Cutoff*$^{qq37/wm25}$, *D. mel.*	19-29	796659
SmallRNA-seq *Cutoff*^{qq37}*/Cutoff*^{wm25}, *D. mel.*	19-29	1012788
SmallRNA-seq MD:♀ P1152 x ♂ BC69 progeny	19-29	2,184,242
SmallRNA-seq NMD: ♀ BC69 x ♂ P1152 progeny	19-29	1,782,627
SmallRNA-seq Total-BQ16, *D. mel.*	19-29	1,246,326
SmallRNA-seq Total-P1152, *D. mel.*	19-29	1,232,610
SmallRNA-seq Total-NMD, *D. mel.*	19-29	1,340,181
SmallRNA-seq Total-MD, *D. mel.*	19-29	1,155,386
SmallRNA-seq Piwi-P1152, *D. mel.*	19-29	1,651,641
SmallRNA-seq Piwi-MD, *D. mel.*	19-29	2,184,242
SmallRNA-seq Piwi-NMD, *D. mel.*	19-29	1,782,627
SmallRNA-seq Aub-P1152, *D. mel.*	19-29	574,772
SmallRNA-seq Aub-MD, *D. mel.*	19-29	644,996
SmallRNA-seq Aub-NMD, *D. mel.*	19-29	370,322
SmallRNA-seq Ago3-P1152, *D. mel.*	19-29	390,085
SmallRNA-seq Ago3-MD, *D. mel.*	19-29	426,663
SmallRNA-seq Ago3-NMD, *D. mel.*	19-29	124,222

Table S2. qPCR primers

LacZ-a-f	GTTACGGCCAGGACAGTCGTTTGC
LacZ-a-r	TCCAACGCAGCACCATCACC
LacZ-b-f	TATTGAAACCCACGGCATGGTGC
LacZ-b-r	GCTGCACCATTCGCGTTACG
LacZ-c-f	CCTTACTGCCGCCTGTTTTGACC
LacZ-c-r	CGTTTTCGCTCGGGAAGACG
LacZ-control-f	GTGGAGTTTGGTGCAGAAGC
LacZ-control-r	AGCCGTGCTTTATGCTTTAC
S1-f	GCCTTTGACCCGTGTGC
S1-r	CAGGAAATTCAGGGAGACGGG
S2-f	GTAATACTGTGACTGGGAGCCAAC
S2-r	CTGTTACGAAAGCACGGGATCTG
T1-f	ATCCGTCGACCTGCAGCCAAG
T1-r	CGTGGCATGCAATGCATTCTG
T2-f	TGCACATTTTGCAGGAGTACGGC
T2-r	GATTTCGGCGCGACTGCTACC
RP49-f	CCGCTTCAAGGGACAGTATCT
RP49-r	ATCTCGCCGCAGTAAACG
42AB-f	CTTCAGCTTGGAGGAAACCAG
42AB-r	CAGCGACTTAAAGAGTGCAGTG
ST-control-f	GCTTTGGAGGAAGGGAGACAC
ST-control-r	GCATCCAGCCAATTTACCCTC

Table S3. piRNA cluster coordinates

Cluster ID	**Chromosome**	**Start**	**Stop**
42AB:	chr2R	2144349	2382500
20A:	chrX	21391192	21431907
38C:	chr2L	20148652	20223262
80EF:	chr3L	23281510	23310943
Flamenco:	chrX	21502935	21844576
	chr3LHet	1402377	1521600
38C-up:	chr2L	20104577	20116331

Table S4. Summary of diffraction data and structure refinement statistic.

Summary of diffraction data		
Crystal	**Chromodomain**	**Chromodomain + H3(1-15)K9me3**
Beamline	NSLS-X29A	APS-24ID-E
Wavelength (Å)	1.029	0.9792
Space group	$P3_221$	$P2_1$
Cell parameters		
a, b, c (Å)	36.5, 36.5, 76.1	42.3, 79.9, 81.2
β (°)	90	96.4
Resolution (Å)	50.0-1.5 (1.55-1.50)[a]	50.0-2.5 (2.59-2.5)
R_{merge} (%)	5.8 (65.3)	13.1 (81.8)
Observed reflections	95,734	75,287
Unique reflections	9,293	18,675
Redundancy	10.3 (9.5)	4.0 (4.0)
Average I/σ(I)	61.9 (3.1)	13.3 (2.0)
Completeness (%)	94.0 (85.0)	100.0 (99.9)
Refinement and structure model		
R factor / Free *R* factor (%)	21.1 / 25.3	19.5 / 24.5
Number of non-H atoms	506	3,506
Protein / peptide	461 / -	2,946 / 357
Water	45	203
Average B factors (Å^2)	40.6	30.3
Protein / peptide	36.8 / -	28.7 / 44.2
Water	44.5	28.9
RMS deviations		
Bond lengths (Å)	0.010	0.008
Bond angles (°)	1.252	1.233

[a] Values in parentheses are for highest resolution shell

Table S5. Read-mapping statistics for GRO-Seq experiment

Cluster	Numbers or reads	
	shRhi_GRO-Seq	control_GRO-Seq
42AB	1,220	2,642
80EF	1,413	1,405
38C	293	452
Xup	3,846	2,824
TJ	440	273
total mapped reads	4,443,868	3,280,199
uniquely mapped reads	3,151,761	2,438,699

Chapter 3 - Part 2

Table S1. SNPs identified in genomes of strains A and I

Region	SNPs					
	Strain A	Strain I	Common	Unique to Strain A	Unique to Strain I	Strain A vs. I
Whole genome	326,026	1,086,963	169,192	156,834	917,771	1,074,605
Cl. 1. Scaffold 12822: 5314-28206	7	27	6	1	21	22
Cl. 2. Scaffold 13052: 1782637-1820594	2	6	1	1	5	6
Cl. 3. Scaffold 10322: 427899-451171	13	9	9	4	0	4
Cl. 4. Scaffold 12823: 172786-186120	4	4	3	1	1	2

Table S2. Read mapping statistics

Library	Read length	Unique mappers
ChIP-Seq H3K9me3 strain A rep1	50	7,648,612
ChIP-Seq Pol II strain A rep1	50	4,991,905
ChIP-Seq input strain A rep1	50	6,548,402
ChIP-Seq H3K9me3 strain I rep1	50	3,572,652
ChIP-Seq Pol II strain I rep1	50	3,740,529
ChIP-Seq input strain I rep1	50	5,673,571
ChIP-Seq H3K9me3 MD progeny rep1	50	11,756,746
ChIP-Seq input MD progeny rep1	50	23,944,512
ChIP-Seq H3K9me3 NMD progeny rep1	50	11,016,563
ChIP-Seq input NMD progeny rep1	50	27,633,846
ChIP-Seq H3K9me3 strain A rep2	50	8,416,288
ChIP-Seq Pol II strain A rep2	50	13,697,961
ChIP-Seq input strain A rep2	50	10,310,302
ChIP-Seq H3K9me3 strain I rep2	50	5,126,455
ChIP-Seq Pol II strain I rep2	50	7,236,748
ChIP-Seq input strain I rep2	50	8,749,926
ChIP-Seq H3K9me3 MD progeny rep2	50	43,753,436
ChIP-Seq input MD progeny rep2	50	11,394,244
ChIP-Seq H3K9me3 NMD progeny rep2	50	37,452,873
ChIP-Seq input NMD progeny rep2	50	6,996,146
RNA-seq (rRNA minus) strain A	50	12,164,323
RNA-seq (rRNA minus) strain I	50	6,707,230

Library	Cloning range	Unique mappers
SmallRNA-Seq strain A	19-29	566,504
SmallRNA-Seq strain I	19-29	737,247
SmallRNA-Seq MD progeny	19-29	499,334
SmallRNA-Seq NMD progeny	19-29	770,258

Table S3. qPCR primers

cl.1-a-f	CGGCCACTTATGTTTCAAGC
cl.1-a-r	TAGTTCTTGCTATGTGTTGG
cl.1-b-f	ACAGCAGACTACACAACAGC
cl.1-b-r	CCTAACAGCAAGTTTTAGCTCC
cl.1-c-f	CGTCCGTTCACAATTAGACC
cl.1-c-r	AAGATTAAGCTAGCAATCGC
cl.2-a-f	AAGTATTGGCACTTGGGTCC
cl.2-a-r	CTATTTCGGATACTAGCAGG
cl.2-b-f	GAAGCTGTTCCGTTCATTGG
cl.2-b-r	CCACCACTTTTTGGGTTGCG
cl.2-c-f	CTTGGACACCCAATAACTGC
cl.2-c-r	TTTTGTAGGCACGTGATTGG
cl.3-a-f	ACCACAAAGGTACTCAGTCC
cl.3-a-r	GGCAACGTATATCTGAAATGC
cl.3-b-f	TCACAACGCCCAGTAATACC
cl.3-b-r	GATTTGGCATCTCACTGTGG
cl.3-c-f	CGCATTTCGATAGCTCATGG
cl.3-c-r	AGCACCATATTATTTGCCGA
cl.4-f	AGGTGAAATCGCGCAAATGG
cl.4-r	TGTTGGGTTTGTTGACTTGG
vir-control-f	ATGAGCAAGAGGCTGAAAGG
vir-control-r	CTTTAAGTTTGCGACGAAGC

Table S4. piRNA cluster coordinates

Cluster ID	Scaffold	Start	Stop
cl.1	12822	5314	9259
cl.2	13052	1764724	1822123
cl.3	10322	427899	451171
cl.4	12823	172751	189360

Annexes – Supplementary Figures

Chapter 1

Supplemental materials available online at the address:

http://genesdev.cshlp.org/content/suppl/2013/01/31/gad.209841.112.DC1.html

Chapter 3 - Part 1

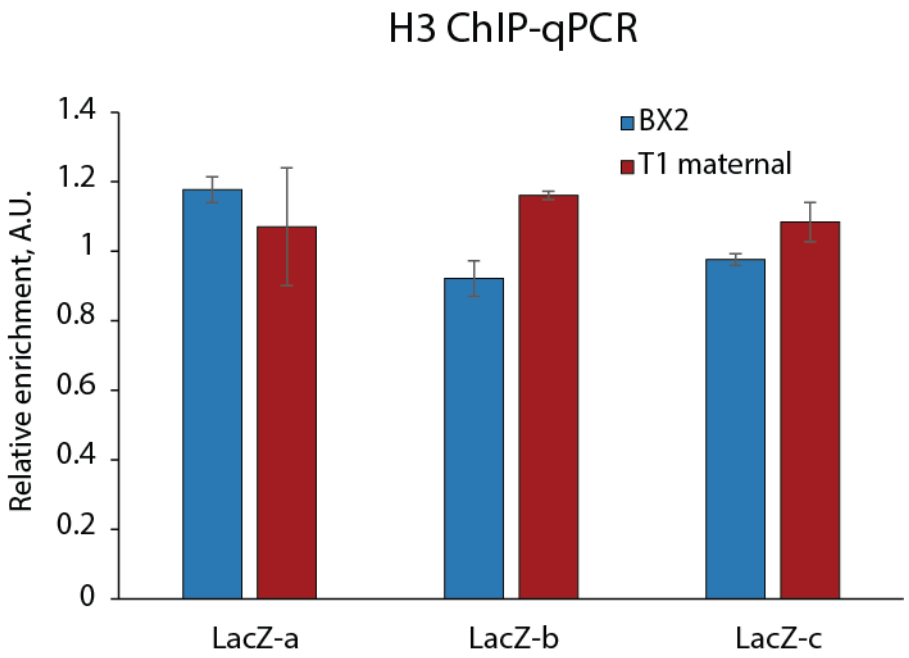


Figure S1. H3 ChIP-qPCR over T1 and BX2 transgenes.
H3 signal on T1 and BX2 as measured by ChIP-qPCR on three different regions of the *lacZ* sequence shown on Fig. 1A. Signal was normalized to input and compared to signal at the same control genomic region (chr2R: 2336913-2337023) used in Fig. 1B-D. Error bars represent the standard error of the mean between two technical replicates.

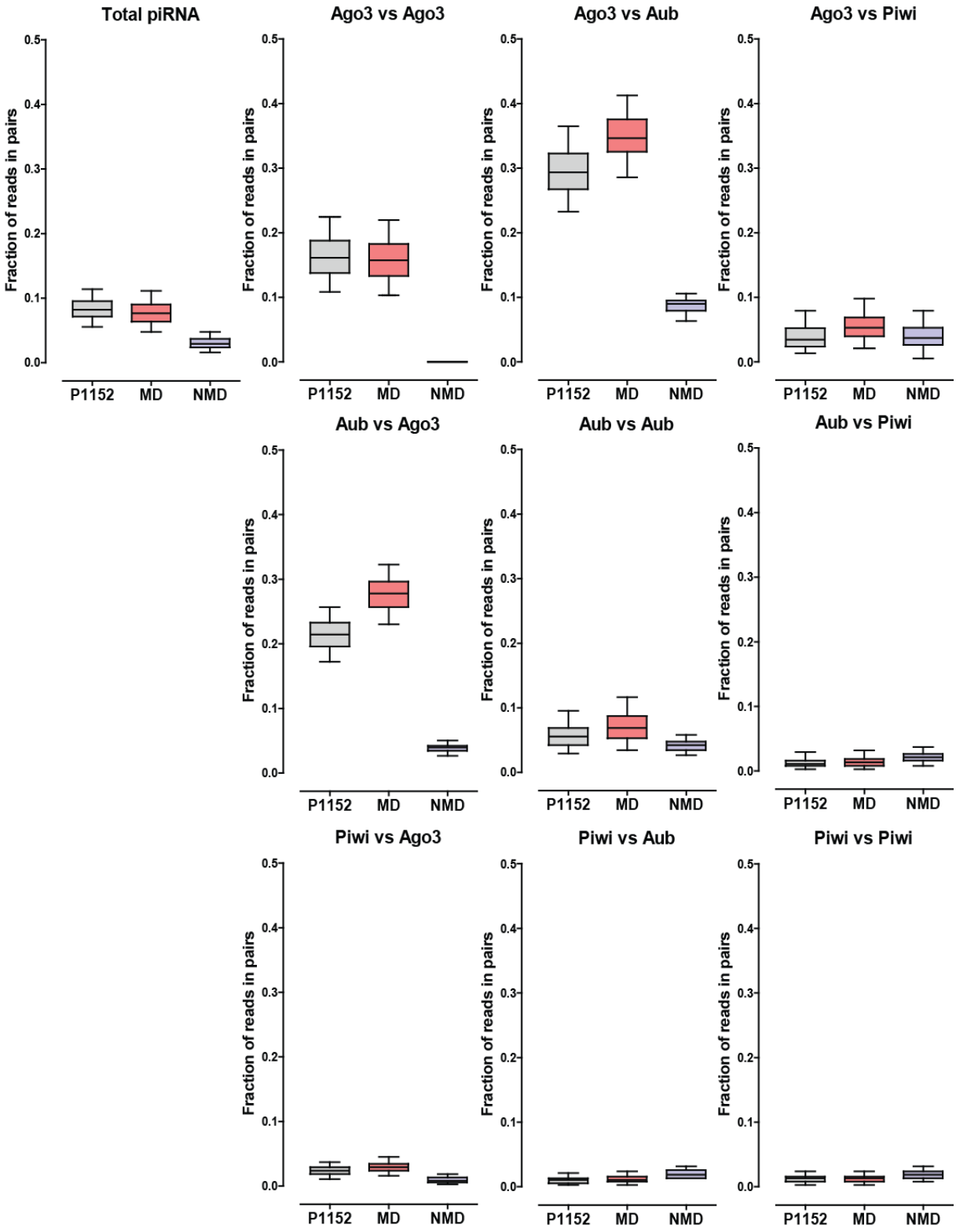


Figure S2. Analysis of the ping-pong processing.
An equal number of piRNA reads mapping to the P{lArB} transgene in each library was sampled 1000-times. Shown are the distributions of the fractions of reads participating in ping-pong piRNA pairs.

		piRNAs from total ovaries RNA	piRNAs cloned after immunoprecipitation					
sequence	origin	Total-MD	Piwi-MD	Aub-MD	Ago3-MD	Piwi-NMD	Aub-NMD	Ago3-NMD
TGCAGCCAAGCTTCTTTTGATTGAGCT	plus strand	0	1	0	0	0	0	0
AATCAAAAGAAGCTTGGCTGCAGGTC	minus strand	0	1	0	0	0	0	0
GCAGCCAAGCTTCTTTTGATTGAGC	plus strand	0	1	0	0	0	0	0
AGCTCAATCAAAAGAAGCTTGGCT	minus strand	0	1	0	0	0	0	0
AATCAAAAGAAGCTTGGCTGCAGG	minus strand	1	0	0	0	0	0	0
GAAGCTTGGCTGCAGGTCGACGGAT	minus strand	0	1	0	0	0	0	0
CTGCAGCCAAGCTTCTTTTGATT	plus strand	0	1	0	0	0	0	0
GCTCAATCAAAAGAAGCTTGGC	minus strand	0	1	0	0	0	0	0
TCTTTTGATTGAGCTTATATTTTCTTG	plus strand	1	0	0	0	0	0	0
total number of sequences		2	7	0	0	0	0	0



Figure S3. piRNA sequences that mapped to BC69 target locus.
The junction between *hsp70* and *rosy* segments is present in BC69, but not in P1152.

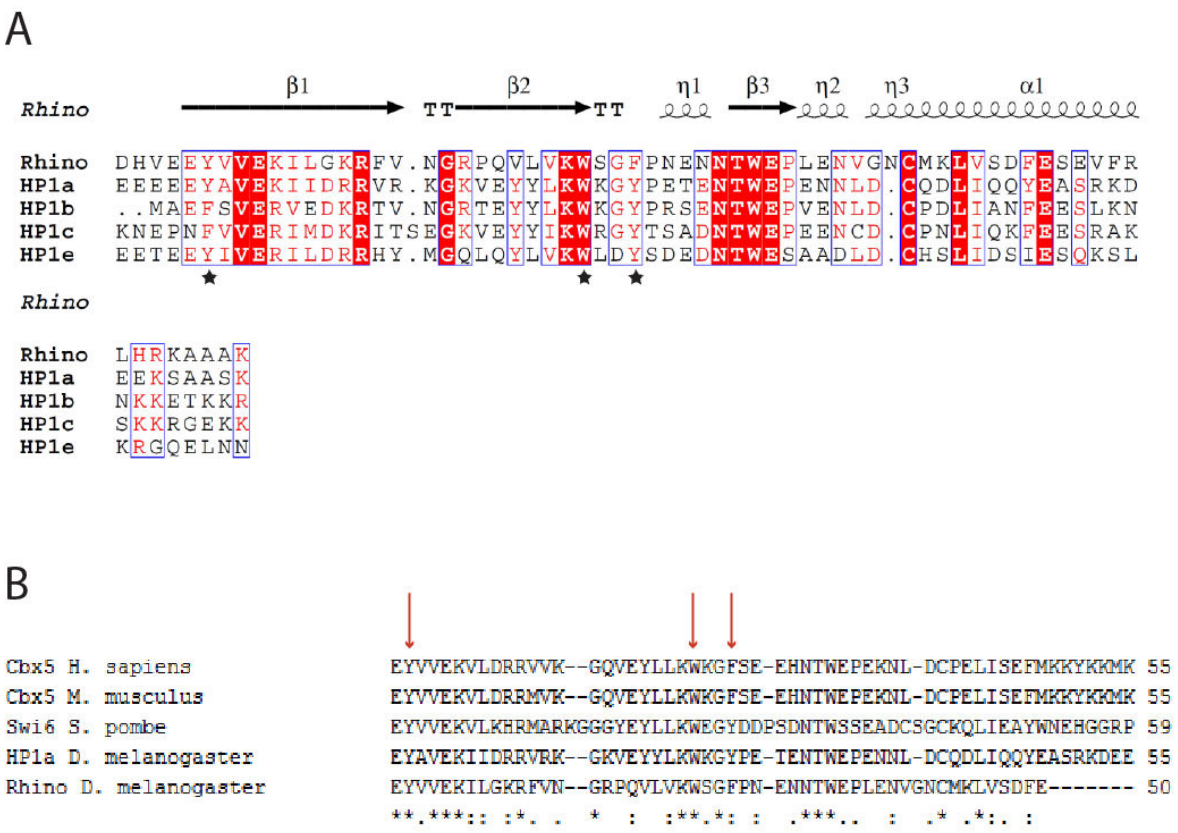


Figure S4. Sequence alignment of the Rhino chromodomain.
(A) Structure based sequence alignment of HP1 chromodomain from *D. melanogaster*. The secondary structure of Rhino is shown on the top of alignment. The three conserved aromatic residues involved in methyl-lysine recognition are highlighted with asterisk at the bottom of the alignment.
(B) Alignment of the Rhino chromodomain with HP1 family chromodomains from human, mouse, *Drosophila* and *S. pombe.* The three conserved aromatic residues involved in methyl-lysine recognition are marked with arrows.

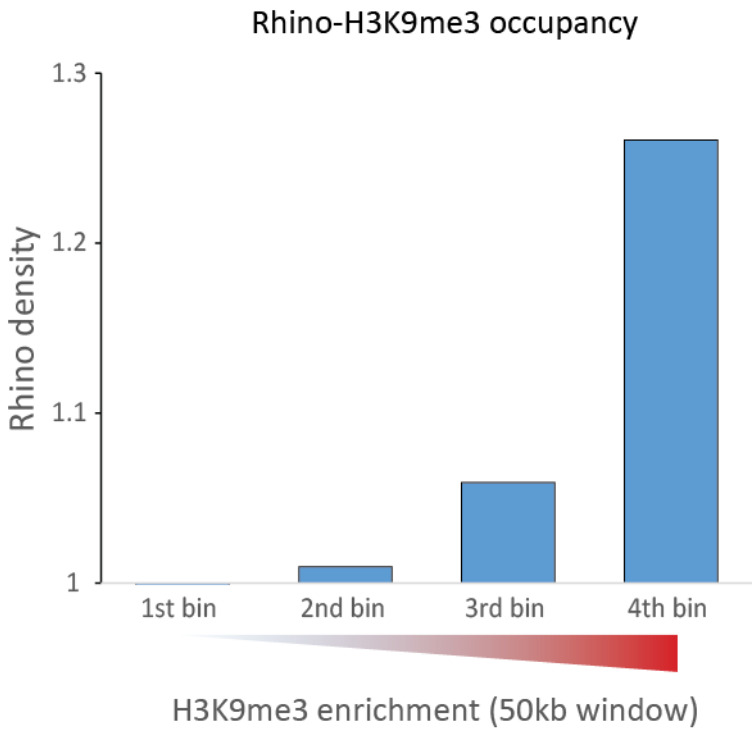


Figure S5. Genome-wide analysis of the correlation between Rhino and H3K9me3 occupancy. Shown are Rhino ChIP-Seq enrichments (RPMs, ChIP/Input) for genomic regions that have different level of H3K9me3 mark. The analysis was performed on 50 kB genomic windows that were binned into four groups according to the level of H3K9me3 ChIP-Seq enrichment.

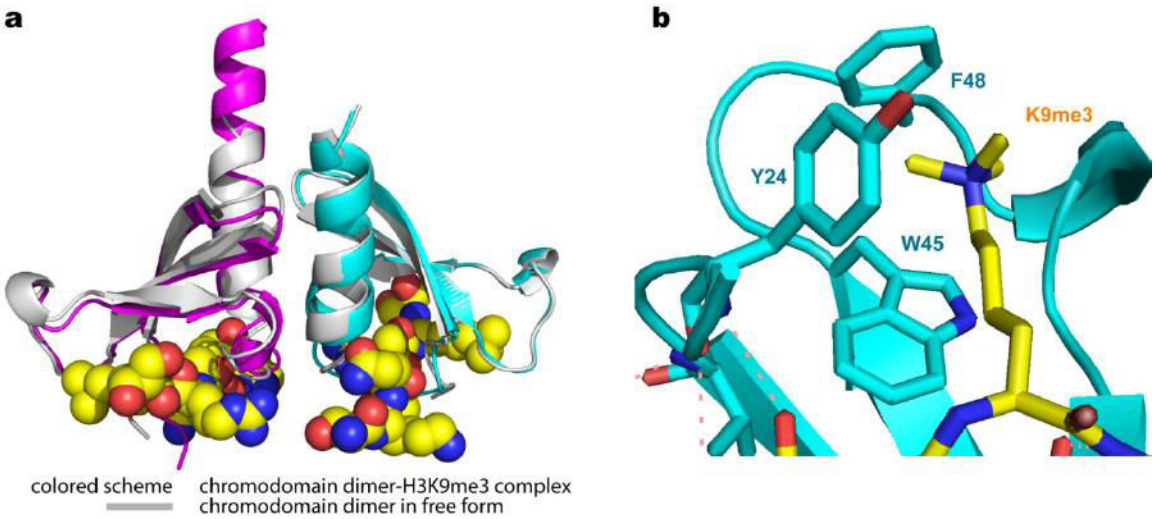


Figure S6. Recognition of the H3K9me3 peptide by the Rhino chromodomain.
(A) Superposition of the H3K9me3 peptide-bound Rhino chromodomain (in color) with the free form chromodomain (in silver) reveals no conformational change.
(B) The K9me3 group of the H3 peptide is recognized by three conserved aromatic residues.

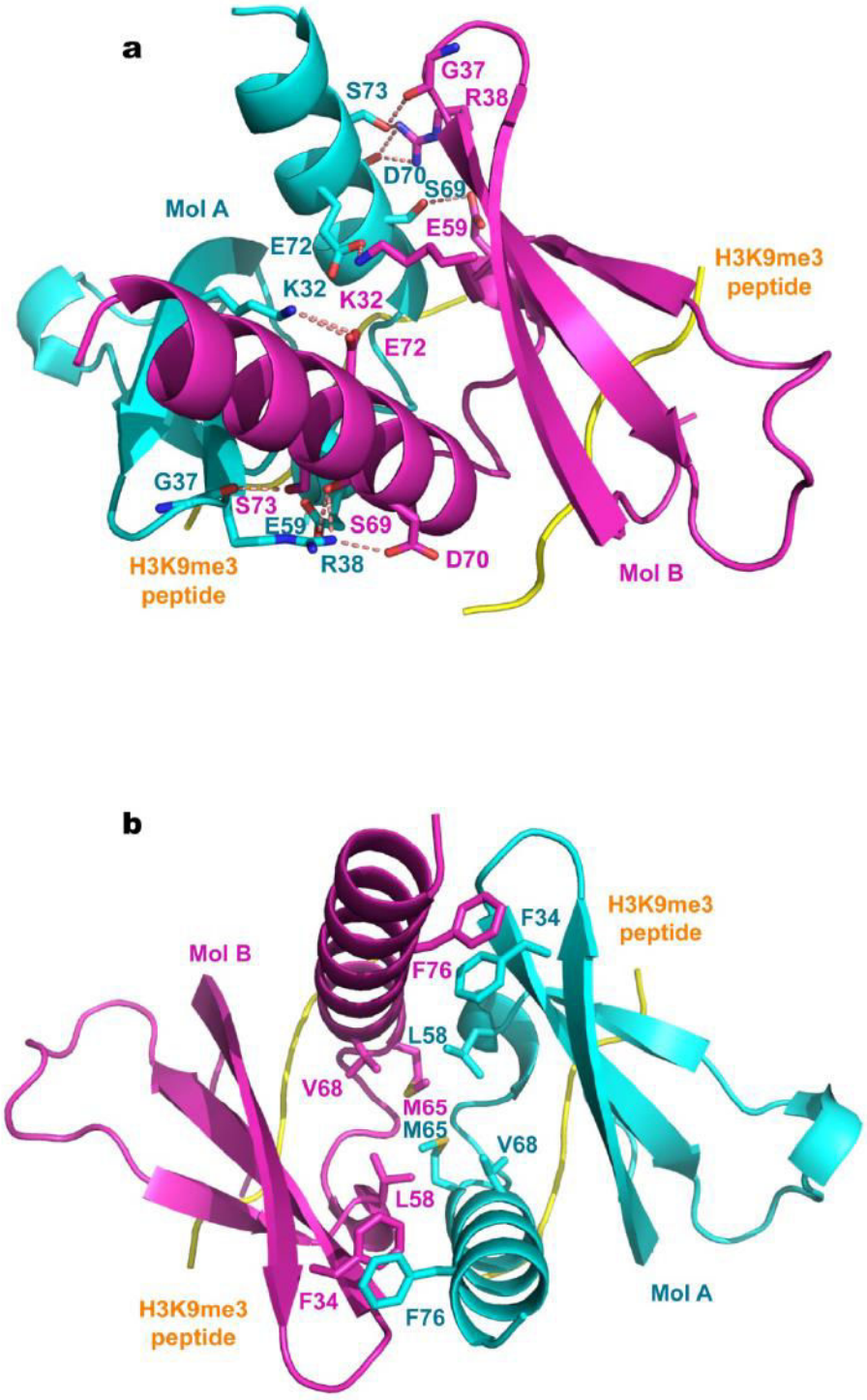


Figure S7. Intermolecular recognition contacts at the dimer interface of the Rhino-H3(1-15)K9me3 complex.
(A) Hydrophobic interactions at the dimer interface.
(B) Hydrogen bonding interactions at the dimer interface.

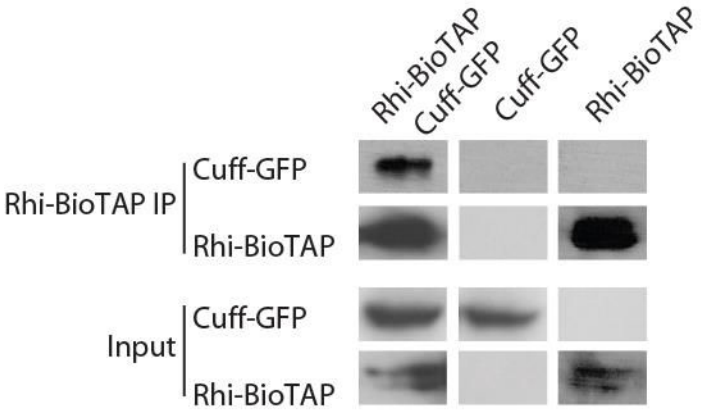


Figure S8. Rhino and Cutoff form a complex in vivo.
Co-immunoprecipitation from ovary extract shows specific interaction between tagged Rhino-BioTap and Cutoff-eGFP proteins.

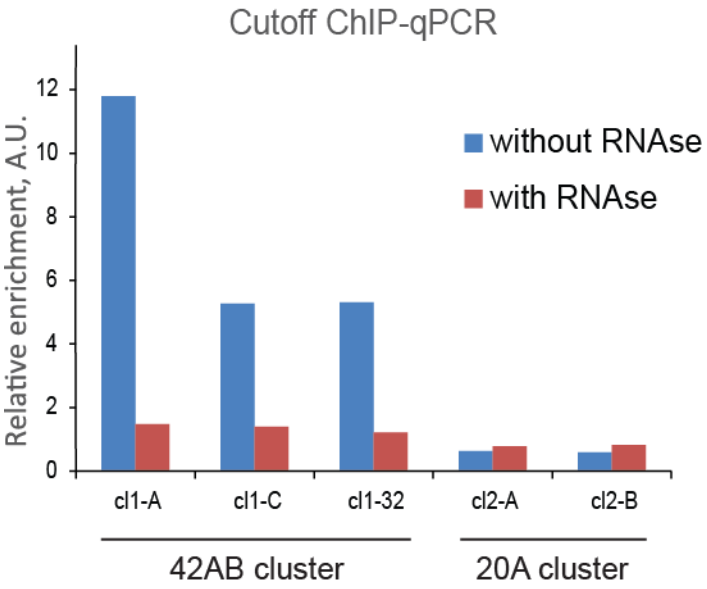


Figure S9. Association of Cutoff with piRNA cluster chromatin depends on RNA. Cutoff signal as measured by ChIP-qPCR on several regions of cluster 42AB and control region cluster 20A (n=2). Signal was normalized to input and compared to signal at a control genomic region (rp49 gene). Where indicated, RNase treatment was performed on the lysate prior to immunoprecipitation.

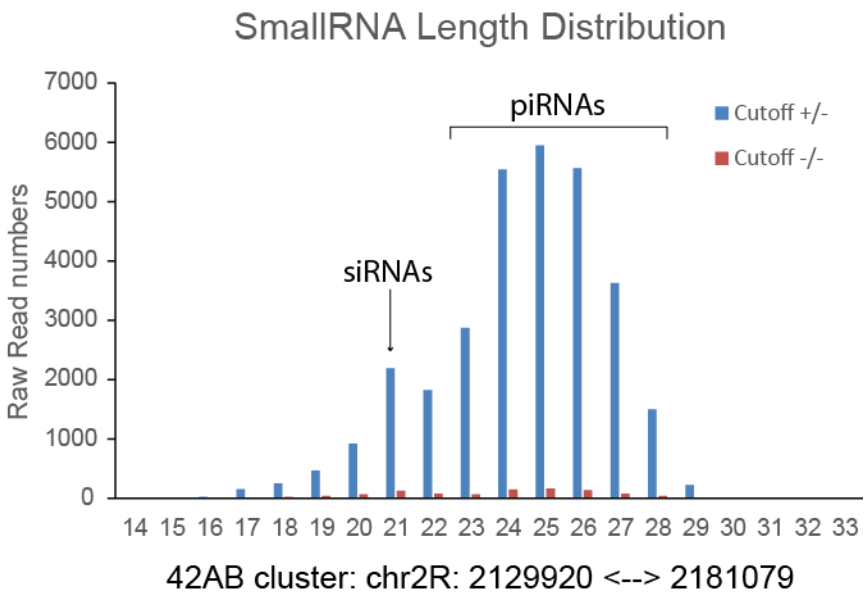


Figure S10. Size profile of small RNAs generated from the 42AB piRNA cluster.
Total small RNAs of the size range between 18 and 29 nt were cloned from *Drosophila* ovaries. The peak at 21 nt corresponds to siRNAs. The broader peak at 23-28nt corresponds to piRNAs.

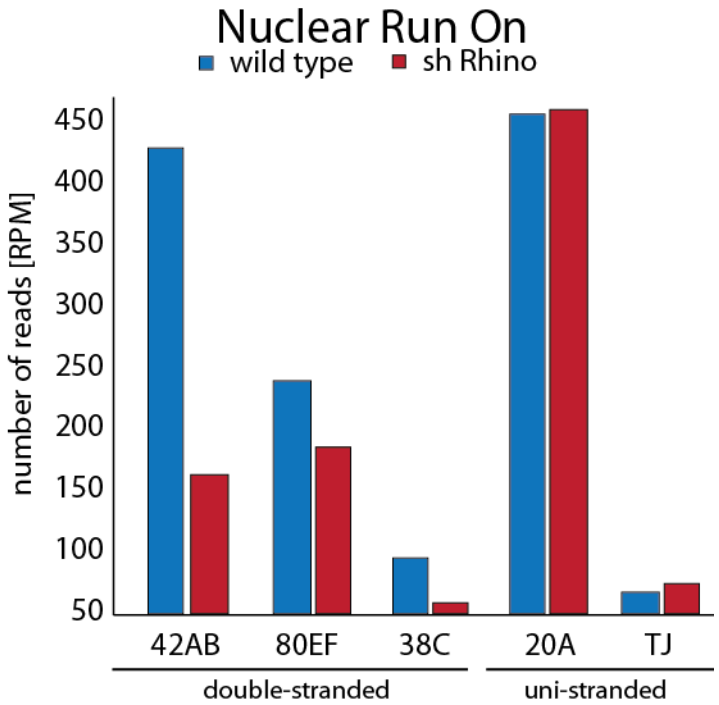


Figure S11. GRO-Seq analysis of transcriptional output from double-stranded piRNA clusters upon Rhino depletion.
Shown are RPM (reads mapping to the piRNA cluster normalized to total number of mapped reads in the respective library) of GRO-Seq on ovaries from wildtype and Rhino knockdown flies.

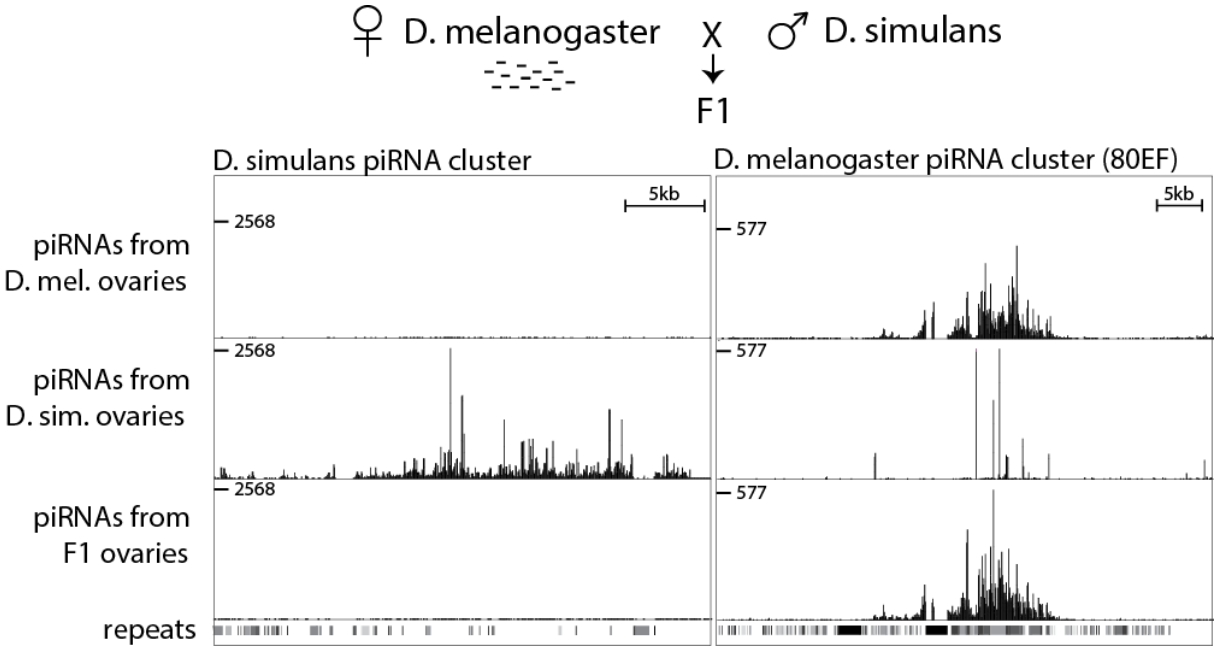


Figure S12. Specification of piRNA clusters by trans-generationally inherited piRNAs in interspecies hybrids between *D. melanogaster* and *D. simulans*. Shown are piRNAs (RPMs) expressed in ovaries of *D. melanogaster*, *D. simulans*, and F1 hybrids between *D. melanogaster* females and *D. simulans* males that inherited the piRNAs from their mothers. The left panel shows piRNAs mapping to a *D. simulans* piRNA cluster (chr2L:15156-36342), and the right panel shows mapping of piRNAs to a *D. melanogaster* piRNA cluster (chr3L:23257666-23337605). The *D. melanogaster* piRNA cluster is expressed in the F1 progeny (F1), but the *D. simulans* piRNA cluster is not.

Chapter 3 - Part 2

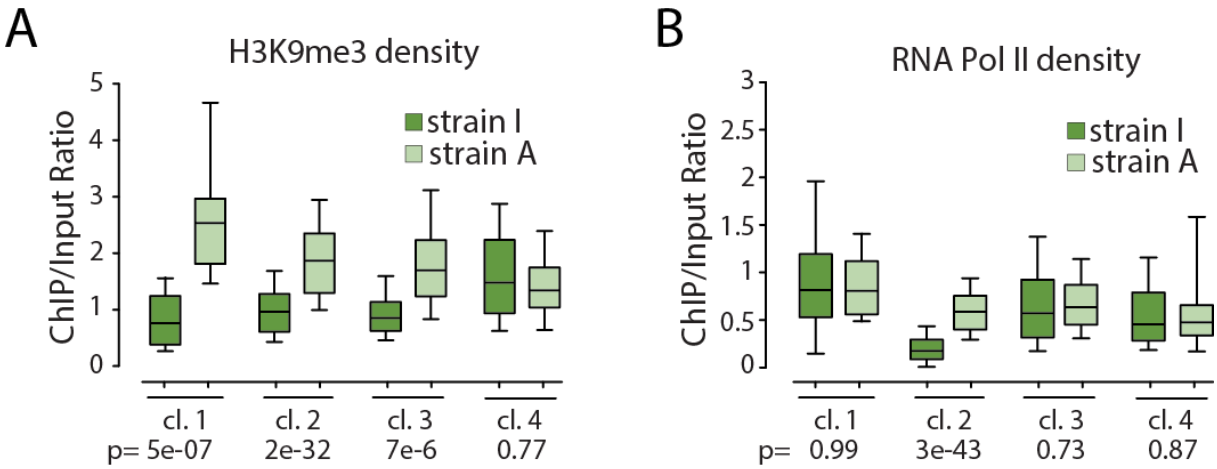


Figure S1. H3K9me3 and RNA pol II occupancy over *piRNA* clusters in parental strains and MD and NMD progenies.

(A) H3K9me3 over piRNA clusters in strains A and I as measured by ChIP-seq (an independent biological replicate for the experiment shown in Figure 2B). Cluster genomic regions were divided into 200 bp windows and a sequencing-depth normalized RPM (Reads Per Million) score was calculated for each window. The Mann-Whitney U test was used to calculate the p-values and evaluate the significance of the observed differences.

(B) RNA polymerase II density over piRNA clusters in the two strains as measured by ChIP-seq (an independent biological replicate for the experiment shown in Figure 2C). Cluster genomic regions were divided into 200 bp windows and a sequencing-depth normalized RPM (Reads Per Million) score was calculated for each window. The Mann-Whitney U test was used to calculate the p-values and evaluate the significance of the observed differences.

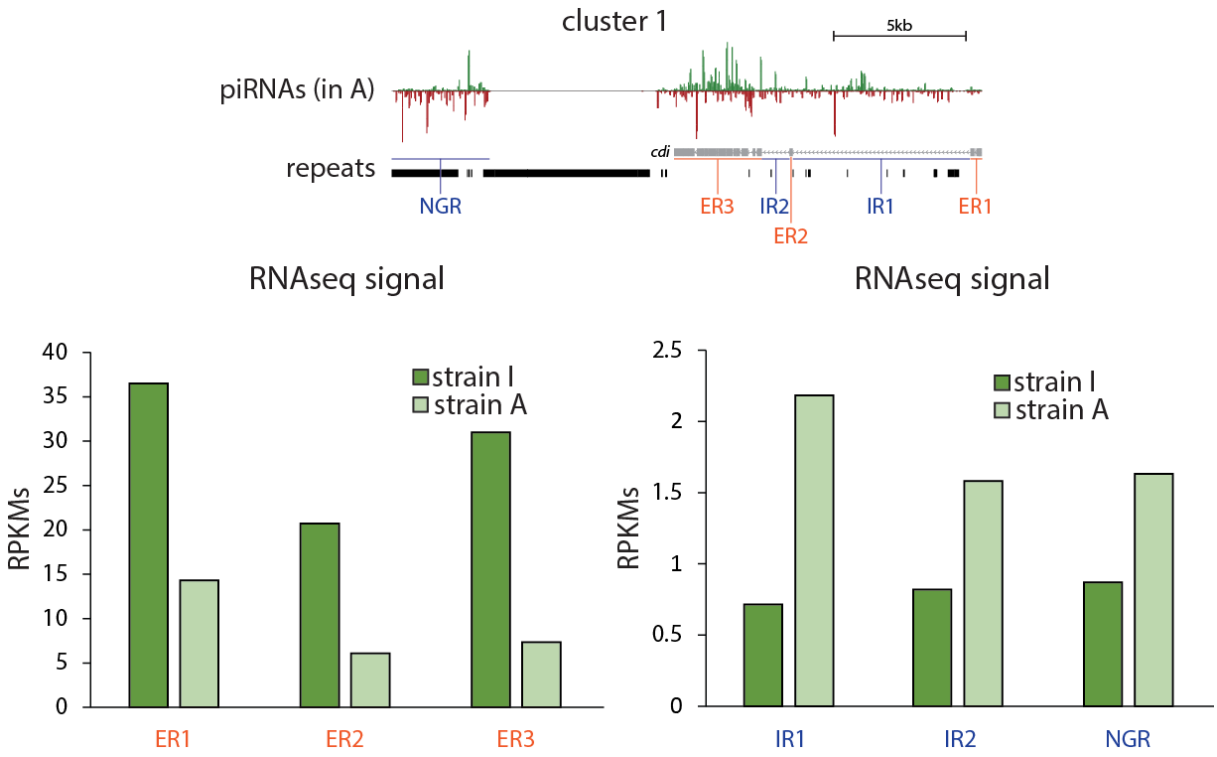


Figure S2. RNA-Seq profiles on the *cdi* gene present in cluster 1.
Production of piRNAs in strain A correlates with a higher level of RNA-Seq signals in intron regions (IR) as well as downstream sequences (NGR) of the *cdi* gene. In contrast, RNA-Seq levels at exons (ER1-3) are higher in strain I that does not generate piRNAs from this region.

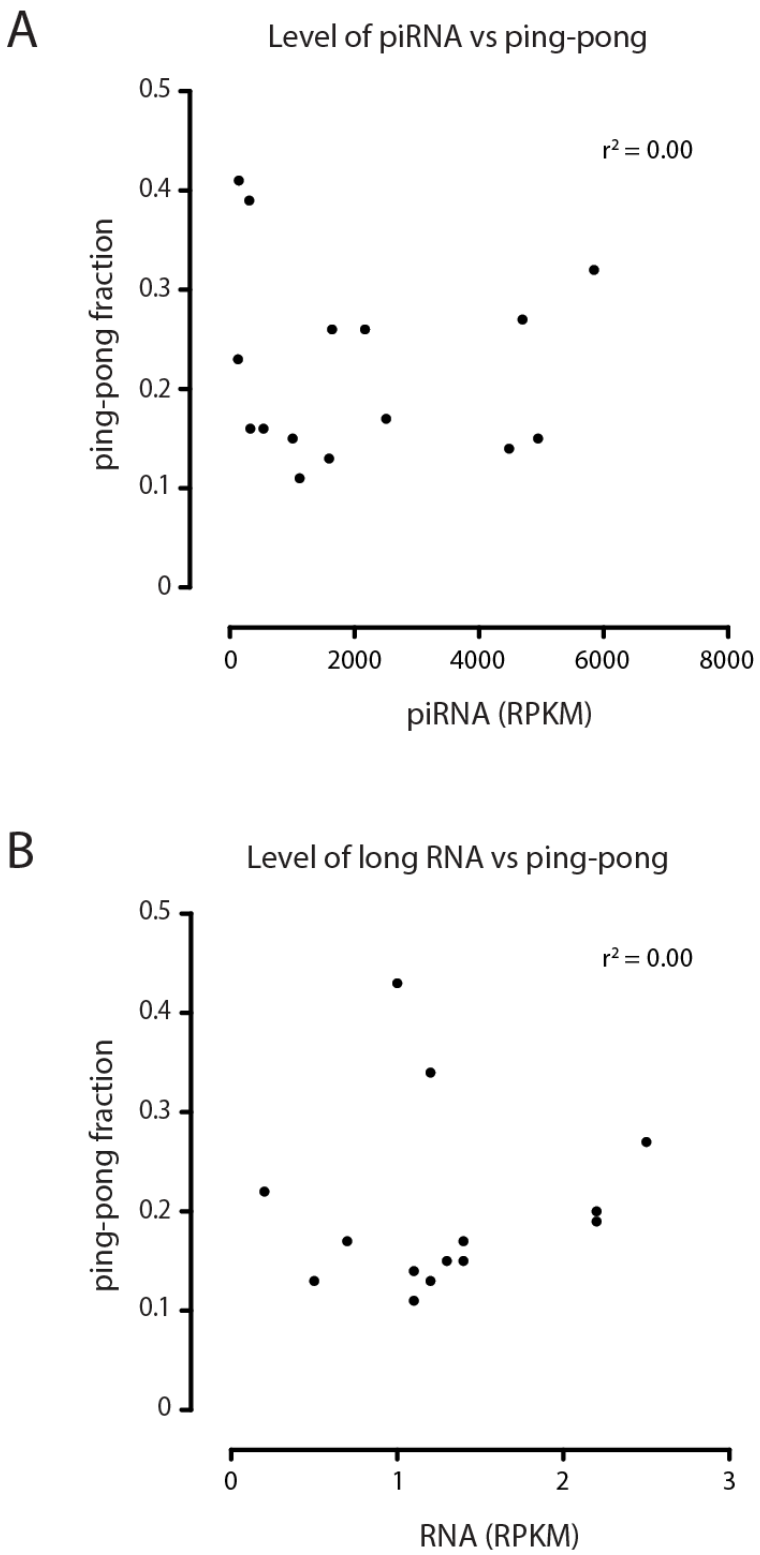


Figure S3. No correlation between the level of piRNA cluster expression and the ping-pong piRNA biogenesis.

Each dot represents individual double-stranded piRNA cluster in *D.virilis* genome. Expression of each piRNA cluster was measured by the level of piRNAs **(A)** or long RNAs **(B)**. The same number of piRNA from each cluster was sampled 1000 times to determine normalized fraction of piRNAs generated through ping-pong biogenesis (Y-axis).

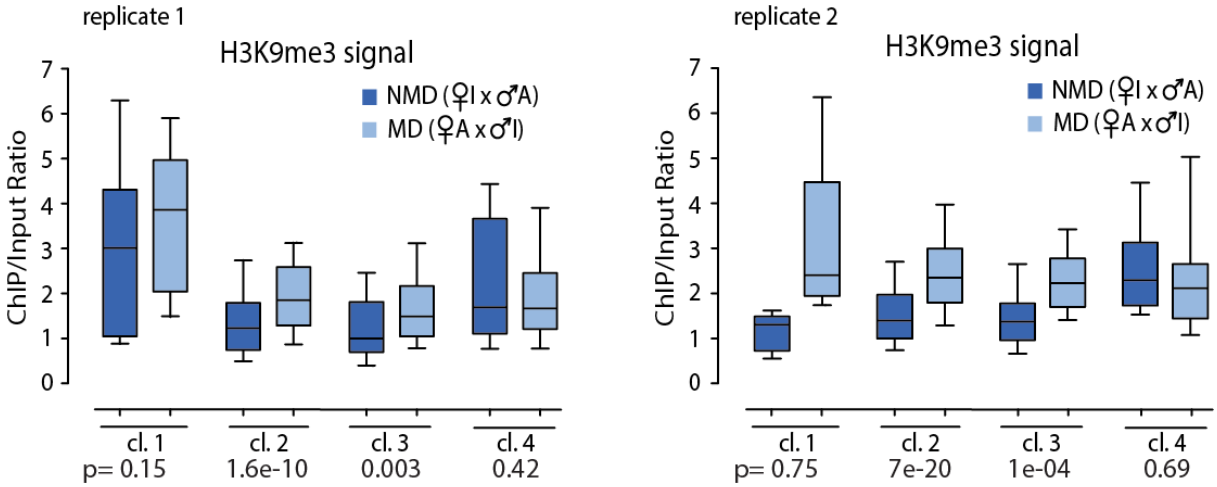


Figure S4. H3K9me3 over piRNA clusters in NM and NMD progenies.
The levels of H3K9me3 over piRNA clusters were measured by ChIP-seq on two independent biological samples. The input-normalized H3K9me3 signal is higher in the MD progeny compared to the NMD progeny. Signals on cluster #4, which is equally active in the two strains, are unchanged in the progeny. Cluster genomic regions were divided into 200 bp windows and a sequencing-depth normalized RPM (Reads Per Million) score was calculated for each window. The Mann-Whitney U test was used to calculate the p-values and evaluate the significance of the observed differences.

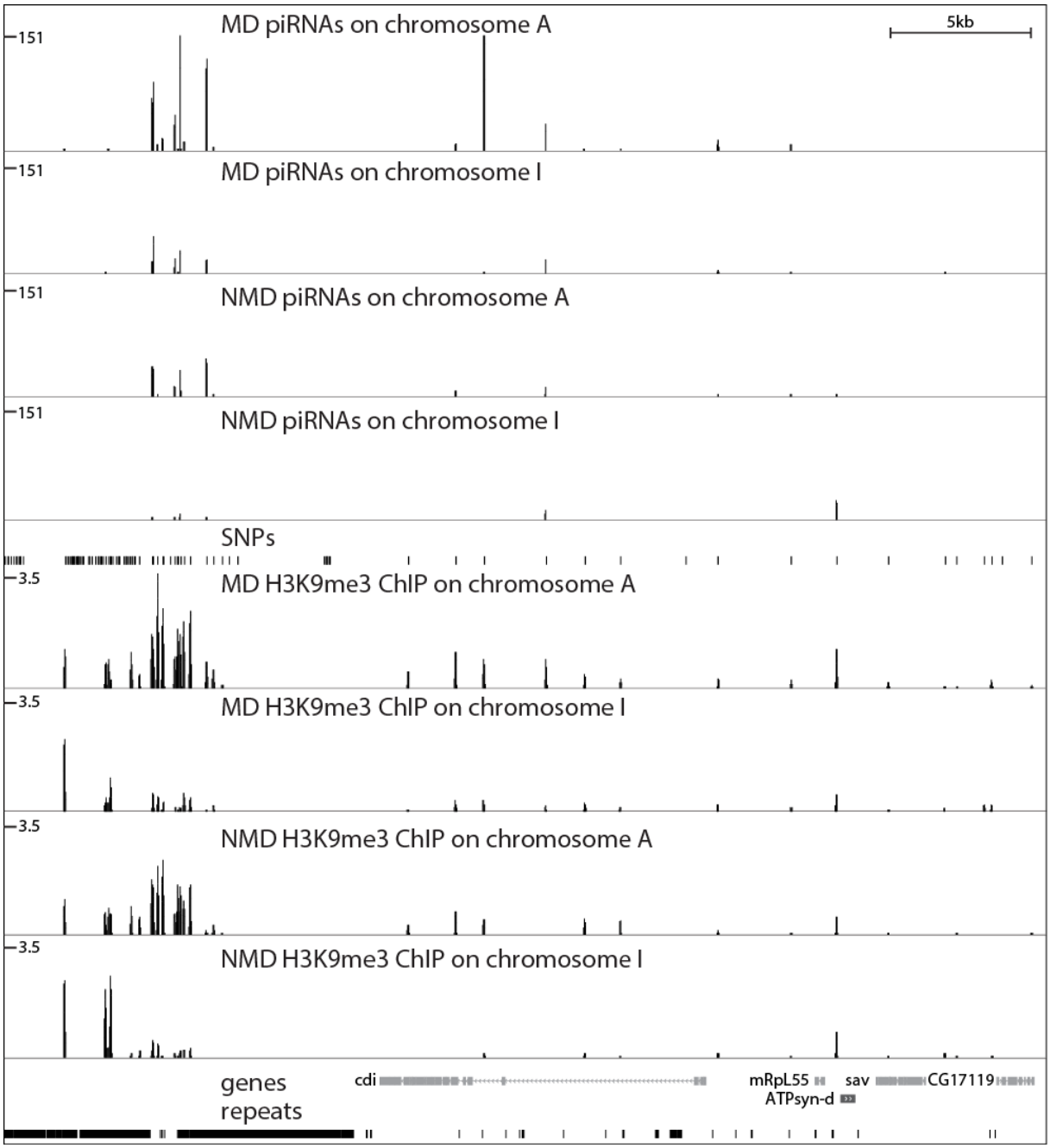


Figure S5. piRNA and H3K9me3 density over SNPs in cluster #1 in MD and NMD progenies. A higher level of piRNAs and H3K9me3 density is observed over the cluster sequence in strain MD at each chromosome versus strain NMD. Only reads that mapped to SNPs that distinct between A and I chromosomes are shown.

Le Thomas Adrien – Thèse de doctorat – Année 2014

Abstract:

piRNAs are an extremely diverse population of small RNA found in the animal germline to silence mobile genetic elements: loaded into Piwi proteins, they guide homology-dependent cleavage of active transposon mRNAs. In Drosophila, three Piwi proteins are expressed, from which two, AUB and AGO3, are known to destroy transposon transcripts in the cytoplasm. The third one, PIWI itself, is nuclear, and the molecular mechanism remains unknown. The main sources of piRNAs are discrete genomic loci called piRNA clusters, however it is not known what differentiate them from non-piRNA producing loci. During my PhD, I focused my work on two central questions in the field:

1) What is the role of Piwi in the nucleus?

We showed that Piwi is responsible for transcriptional silencing by mediating installment of repressive marks, especially H3K9me3, over active transposons copies in a piRNA dependent manner.

2) How are piRNA clusters defined, and what regulates their expression?

. We showed that maternally inherited piRNAs are responsible to define germline bi-directional clusters at the next generation through two mechanisms: in the nucleus, by deposition of H3K9me3 onto complementary genomic sequences, and, in the cytoplasm, by initiating the ping-pong cycle.

. We found that cluster promoters are essential to mediate full cluster transcription, which is allowed thanks to a very specific chromatin signature necessary to ensure piRNA production.

Résumé:

Les piRNAs sont une population de petit ARNs très diverse, que l'on retrouve dans la lignée germinale des animaux pour réprimer les éléments génétiques mobiles : agissant de pair avec les protéines Piwi, ils guident le clivage des transposons actifs. Chez la Drosophile, 3 protéines Piwi sont présentes, dont deux d'entre elles, AUB et AGO3, sont cytoplasmiques et la dernière, PIWI, est nucléaire, mais son mécanisme d'action reste inconnu. La source principale de piRNAs sont des régions du génome bien particulière, appelé cluster de piRNAs. Cependant, il n'est pas encore connu à ce jour ce qui différentie ces régions du reste du génome. Durant mon doctorant mon travail s'est focalisé sur ces deux questions centrales :

1) Quel est le rôle de PIWI dans le noyau?

Nous avons montré que PIWI était responsable de répression transcriptionelle des transposons par l'intermédiaire de la déposition de marques chromatiniennes répressives, H3K9me3, grâce à la spécificité des piRNAs.

2) Comment sont définit les régions générant des piRNAs et comment sont régulés leur expression ?

Nous avons trouvé que les piRNAs, transmis par la mère aux progénitures, sont responsables de l'identification des régions génomiques donnant naissance à de nouveaux piRNAs, grâce à la déposition de H3K9me3 dans le noyau et par l'initiation du cycle ping-pong dans le cytoplasme.

Nous avons aussi mis en évidence les régions promoteurs des clusters de piRNAs, et trouvé qu'elles sont nécessaires pour la production de piRNAs.

Keywords: piRNA, chromatin, transposons, transcription, epigenetic, Drosophila