Identification and characterization of novel components of the *Drosophila* piRNA pathway

A Dissertation Presented

by

Paloma Maria Guzzardo

Leslie Quick, Jr. Fellow

and

William Randolph Hearst Scholar

to

The Watson School of Biological Sciences

at

Cold Spring Harbor Laboratory

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Biological Sciences

Cold Spring Harbor Laboratory

January 2013

Contents

List of Figures

List of Tables

List of abbreviations

cDNA:	complementary DNA
LTR:	long terminal repeats
Gag:	group specific antigen
Pol:	polymerase
RNA:	ribonucleic acid
mRNA:	messenger RNA
DNA:	deoxyribonucleic acid
dsDNA:	double stranded DNA
env:	envelope
ORF:	open reading frame
UTR:	untranslated region
LINE-1:	long interspersed nucleotide elements-1
SINE:	short interspersed elements
TPRT:	target-site-primed reverse transcription
RISC:	RNA interfering silencing complex
PAZ:	Piwi Argonaute Zwille
PIWI:	P-element induced wimpy testes
sRNA:	small RNA
miRNA:	micro RNA
siRNA:	small interfering RNA
piRNA:	Piwi-interacting RNA
Ago:	Argonaute
dsRNA:	double stranded RNA
Aub:	Aubergine
Bam:	bag of marbles
fGS:	female germ line stem cells
OSS:	ovarian somatic sheet
nt:	nucleotide
rRNA:	ribosomal RNA
qPCR:	quantitative PCR
EGFP:	enhanced green fluorescent protein
Ub:	ubiquitin promoter
KD:	knock down
RT:	reverse transcriptase
Ct:	threshold cycle
DNase:	deoxyribonuclease
Kb:	kilobase
Zuc:	zucchini
Armi:	armitage
bp:	base pair
APOBEC3:	apolipoprotein B mRNA editing enzyme
Dnmt:	DNA methyltransferase

Acknowledgements

I would like to thank Greg for all his advice and support throughout these five years. From the moment I stepped into McClintock, I knew it was where I wanted to do my PhD. Words cannot describe how grateful I am for having the opportunity to be in your lab.

I would like to thank Felix, my partner in science and in many other things. Getting this project off and running would have not been possible without you. Thank you for all your help and encouragement during the writing of this dissertation. You are pretty amazing.

This dissertation would have not been possible without guidance and advice from my thesis committee, Zach Lippman, Emily Bernstein, Leemor Joshua-Tor and Adrian Krainer. I would also like to thank Victor Corces for agreeing to be my external committee member.

I would like to thank all the members of the Hannon Lab, both past and present, for creating a fun and supportive working environment, even if it was crowded and messy. I would also like to that all the people who contributed to work done in this thesis: Astrid Haase, Assaf Gordon, Ben Czech and Jon Preall. I also owe many thanks to Jo Leonardo and Sabrina Boettcher, for making the lab run smoothly and helping me throughout my time in the lab.

I am very grateful to all the Watson School administration, Dawn, Leemor, Kim, Alyson, Keisha and Alex. You made WSBS a home away from home.

I would like to thank my classmates, who were responsible for me keeping my sanity during the first year. I will always cherish the memories of those late nights finishing up problem sets, although I am glad it is over. Thank you for all the wine brakes, tea breaks and baking breaks. You are the best classmates I could have wished for.

I was fortunate enough to have an amazing group of friends during my time in CSHL. You all made my time in grad school the most fun I've had in my life. Colin, Fred, and Hannah: thanks for being there in the beginning; they were some good times. Katie, Amy and Saya: thank you for the fancy dinner and the dance parties. Thank you to Leah, Antoine, Emily, Kaja, Ian, Katie, Shane and Simon. All of you make it fun to come into lab every day.

To my family, for always believing in me and encouraging me when I thought I would never finish.

Abstract

Transposable elements are part of the landscape of many eukaryotic genomes. Although sometimes considered genomic parasites, transposons can also benefit their host, playing roles in enhancing genetic diversity and driving evolution. However, when unhindered transposon activity will lead to a reduction in the host's fitness and fertility. Therefore, it is imperative to have transposon control mechanisms in place. The piRNA pathway is an evolutionarily conserved pathway in which Piwi clade proteins and their bound piRNA silence these repetitive elements in the germline. In *Drosophila melanogaster*, the piRNA pathway is principally studied in the female germline, where two distinct versions of the pathway are active, one in germ cells involving all three Piwi clade proteins, Piwi, Aubergine and Argonaute-3, and one in somatic cells of the ovary where Piwi acts alone. In this study, we aim to further our understanding of the piRNA pathway by performing an unbiased genome-wide screen for components of the somatic piRNA pathway. This screen was carried out in a *Drosophila* somatic ovarian cell line that had not been previously used for high-throughput experiments. We demonstrate that it is a good model for studying the piRNA pathway and highly suited for large-scale studies. Furthermore, we completed a comprehensive genome-wide RNAi screen of over 40,000 dsRNAs. Using qPCR as a read-out we identified and validated 87 genes that upon knockdown, lead to the derepression of somatically expressed transposons. As part of this list we confirm all known somatic piRNA pathway components. Among these newly identified genes, we reveal novel biogenesis factors and well as identify previously uncharacterized effector proteins. This comprehensive study will provide an important foundation for understanding many aspects of piRNA biology.

1 Introduction

Citation for review article:

Guzzardo PM, Muerdter F, Hannon GJ. 2013. The piRNA pathway in flies: highlights and future directions. *Current Opinion in Genetics & Development*.doi:10.1016/j.gde.2012.12.003

1.1 Transposable elements and their interaction with eukaryotic genomes

In the 1950s Barbara McClintock proposed the revolutionary idea that genomes were not static, but dynamic entities, subject to change and rearrangements (McClintock 1950). While this idea was received with much skepticism at first, research in later years showed her findings to be correct. McClintock had discovered transposons, or 'jumping genes', fragments of DNA capable of mobilizing from one location to another within a genome (McClintock 1956). These mobile elements exist in the genome of virtually every eukaryotic species and come in two general varieties based on their transposition mechanism and their encoded sequences: retrotransposons and DNA transposons (Levin and Moran 2011).

Retrotransposons, or class I elements, mobilize in a 'copy-and-paste' manner, where the transposon is transcribed, then reverse transcribed, and this cDNA intermediate is inserted into a new site in the genome, thereby duplicating the element. These elements are very important features of eukaryotic genomes, but so far have not been identified in prokaryotes. Retrotransposons can be further subdivided into two categories, LTR and non-LTR transposons, based on the presence, or absence of long terminal repeats (LTR) on both ends of the element. LTR retrotransposons have been extensively studied because of their similarities to retroviruses, which have only been found in vertebrates. Like retroviruses, LTR retrotransposons have a *gag* (group specific antigen) gene, which encodes the viral particle coat and *pol* (polymerase) gene, which encodes the reverse transcriptase, ribonuclease and integrase, all which mediate the retrotransposition process (Figure 1.1a). For transposition to occur, the host RNA Polymerase II recognizes a promoter sequence in the 5′ LTR and produces an mRNA of the transposon. The Gag proteins then assemble into a virus like particle, where the mRNA is reverse transcribed into dsDNA. This particle is then transported to the nucleus where the integrase mediates integration into a new target site (Figure 1.1a). Previously, it was believed that the main difference between retroviruses and retrotransposons was that retrotransposons lacked a functional *env* (envelope) gene, which allows the virus to move from on cell to another. However, it has since been discovered that a sub-group of retrotransposons known

as errantaviruses possess a functional *env* gene and are capable of infecting neighboring cells (Havecker et al. 2004; Song et al. 1994).

The second class of RNA transposons, non-LTR retrotransposons, lack LTR, as their name implies and usually contain two open reading frames (ORF) and a 5′ and 3′ UTR. These elements can be autonomous and encode proteins necessary for transposition, like mammalian LINE-1 elements (long interspersed nucleotide elements-1) or non-autonomous elements that use proteins encoded by other elements for their mobilization, like SINEs (short interspersed elements). Autonomous elements generally contain two open reading frames, one that encodes a nucleic acid binding protein and another encoding an endonuclease and reverse transcriptase (Figure 1.1b) These elements mobilize by target-site-primed reverse transcription (TPRT), where the endonuclease creates a nick in the target site DNA that is used to prime reverse transcription of the RNA, that is then integrated.

DNA transposons, or class II elements, do not require a reverse transcription step, and mobilize using a 'cut-and-paste' mechanism. A transposase, encoded by the element itself, recognizes sequences in the transposon and excises the element from one position in the genome and integrates it in another (Figure 1.1c). As evident by this mechanism, there is no copy number increase during DNA transposon transposition, as there is with retrotransposons.

Although a very large fraction of many animal genomes are composed of transposable elements, the majority are not intact elements, but rather fragments, which contain mutations that inhibit transposition (Biémont 2010; Brennecke et al. 2007). However, there are often a number of intact active copies present in the genome. These are usually silenced by defense mechanisms that the host has evolved (Slotkin and Martienssen 2007). Active transposition can cause detrimental effects on the host's fitness due to obvious reasons. Insertion in the vicinity of, or within a gene locus, can alter its expression and therefore affect fertility or viability, depending on the nature of the gene and the quality of the mutation (Bradley et al. 1993). Furthermore, transposition can alter chromosomal structure, by causing double stranded breaks (McClintock 1950). However, in some instances the changes brought on by

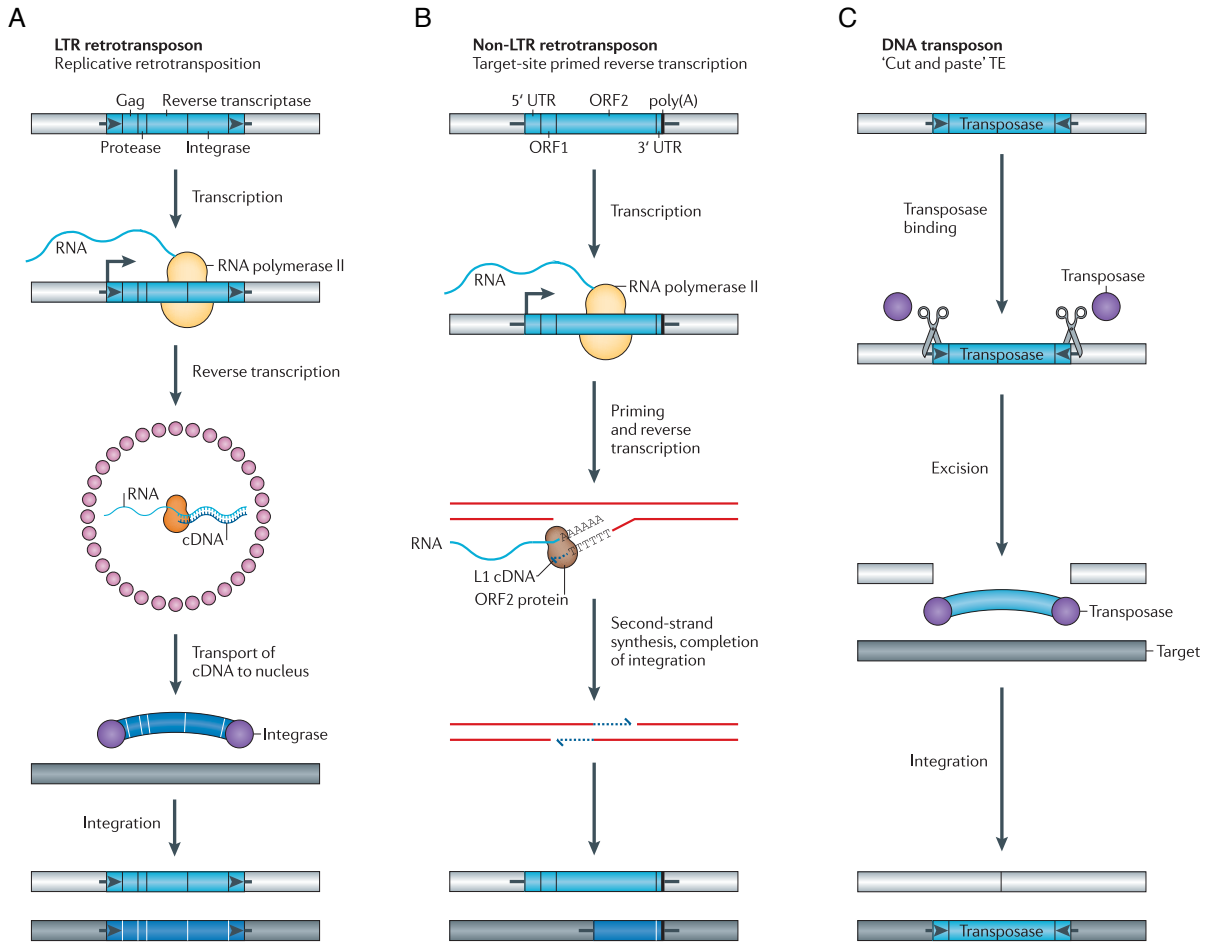


Figure 1.1: **The diverse mechanisms of transposon mobilization.** A) LTR retrotransposons contain two long terminal repeats (LTRs; black arrows) and encode Gag, protease, reverse transcriptase and integrase activities, all of which are crucial for retrotransposition. The 5′ LTR contains a promoter that is recognized by the host RNA polymerase II and produces the mRNA of the TE (the start-site of transcription is indicated by the right-angled arrow). In the first step of the reaction, Gag proteins (small pink circles) assemble into virus-like particles that contain TE mRNA (light blue), reverse transcriptase (orange shape) and integrase. The reverse transcriptase copies the TE mRNA into a full-length dsDNA. In the second step, integrase (purple circles) inserts the cDNA (shown by the wide, dark blue arc) into the new target site. B) Non-LTR retrotransposons lack LTRs and encode either one or two ORFs. As for LTR retrotransposons, the transcription of non-LTR retrotransposons generates a full-length mRNA (wavy, light blue line). However, these elements mobilize by target-site-primed reverse transcription (TPRT). In this mechanism, an element-encoded endonuclease generates a single-stranded 'nick' in the genomic DNA, liberating a 3′-OH that is used to prime reverse transcription of the RNA. The TPRT mechanism of a long interspersed element 1 (L1) is depicted in the figure; the new element (dark blue rectangle) is 5′ truncated and is retrotransposition-defective.

Figure 1.1: (Previous page.) C) Many DNA transposons are flanked by terminal inverted repeats (TIRs; black arrows), encode a transposase (purple circles), and mobilize by a 'cut and paste' mechanism (represented by the scissors). The transposase binds at or near the TIRs, excises the transposon from its existing genomic location (light grey bar) and pastes it into a new genomic location (dark grey bar). Adapted from Levin and Moran, *Nat Rev Genet*, 2011.

transposition can also be beneficial and serve to drive evolution (Kazazian 2004). There are also some situations in which transposable elements have been co-opted by the host for important cellular functions: For example, in *Drosophila* where three non-LTR retrotransposons, *HeT-A*, *TAHRE* and *TART*, are attached to chromosome ends and serve as telomeres (Pardue and DeBaryshe 2008).

In spite of their movement being occasionally advantageous, it is important to keep some degree of control over transposable elements. This is a formidable task due to the wide variety of transposable element families and diversity of transposition mechanisms. To address this challenge, the target of control mechanisms should be a feature common to all transposon types. One common aspect is that all these elements depend on factors encoded by the transposons themselves for movement. Therefore, targeting these transposon-encoded RNAs seems like a robust method to silence selfish elements, and indeed, one of the main mechanisms used to silence transposons is RNA interference (Malone and Hannon 2009).

1.2 RNA interference and small RNA pathways in flies

RNA interference is the process by which small RNA molecules bound to an RNA interfering silencing complex (RISC), target other RNAs using sequence complementarity and regulate their expression (Fire et al. 1998; Hannon 2002). To date, research has implicated RNAi in gene regulation by 'slicing' transcripts, binding transcripts to inhibit translation, and silencing targets by inducing changes in chromatin structure (Ghildiyal and Zamore 2009). RNAi has been implicated in an extensive number of cellular processes, with its targets varying from protein-coding genes, to transposable elements and viral RNAs (Hannon 2002). At the core of

this mechanism is the Argonaute (Ago) family of proteins, which bind small RNAs (Joshua-Tor and Hannon 2011). These proteins are characterized by the presence of two domains, PAZ and PIWI. These domains fold to form a channel where the sRNA can be held to use as a guide for targets. The PIWI domain also contains a ribonuclease H-like motif that is capable of cleaving RNA (Song et al. 2004). In *Drosophila* there are three main classes of small RNAs: micro RNAs (miRNAs), small interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs), each with a specific function and Argonaute binding partner. miRNAs, which arise from endogenous hairpin transcripts, bind Ago1 and target mRNAs for cleavage or translational repression (Lagos-Quintana et al. 2001; Okamura et al. 2004; Bushati and Cohen 2007). Ago2 binds siRNAs, which are generated from double stranded RNA (dsRNA) (Hammond et al. 2001; Okamura et al. 2004). This dsRNA can arise exogenously, from viruses or ectopically introduced long dsRNA, or endogenously, from transposon transcripts, long stem loops and sense-antisense transcript pairs (Czech et al. 2008). piRNAs bind to another clade of Ago family, the PIWI proteins (Brennecke et al. 2007). In *Drosophila*, this clade consists of Piwi, Aubergine (Aub) and Ago3 (Carmell et al. 2002). Most piRNAs arise from piRNA clusters, comprised of many small fragments of transposons, hence their main target is transposon transcripts, although there also are some arising from 3'UTRs of genes (Brennecke et al. 2007; Robine et al. 2009). These sRNAs are generated by at least two different biogenesis mechanisms. miRNAs and siRNAs are generated by a more elucidated mechanism that uses RNase III-type proteins, Dicer-1 or Dicer-2 (Bernstein et al. 2001; Lee et al. 2004). piRNA biogenesis is Dicer-independent and less understood, although several proteins have been implicated in the processes (Vagin et al. 2006; Ishizu et al. 2012) .

1.3 The piRNA pathway and the battle against transposons in the germline

Interestingly, while Ago1 and Ago2 are expressed ubiquitously throughout the organism, the PIWI clade is expressed specifically in gonadal tissue (Williams and Rubin 2002). Since the

piRNA pathway is the principal transposon control mechanism, this is not surprising. Germ cells have a particular need to protect their genome, since it will be transmitted to future generations. For this same reason, the germline genome is a desired target for selfish genetic elements like transposons. By integrating in germ cell DNA they will ensure propagation. When not controlled, transposable elements can compromise the genomic integrity of the germ cell, cause massive amounts of DNA damage, and trigger cellular checkpoints leading to a loss of fertility (Kidwell et al. 1977; Klattenhoff et al. 2007; Theurkauf et al. 2006). Silencing transposons represents a daunting task due to the incredible diversity of elements, transposition modes and patterns of expression. Specific transposons are expressed in a cell-specific fashion and at particular developmental times (Parkhurst and Corces 1987; Brookman et al. 1992; Ding and Lipshitz 1994). Therefore the piRNA pathway must adapt to these unique characteristics to maintain control over parasitic elements. For example, in *Drosophila*, the *gypsy* family of LTR retrotransposons is highly expressed in the ovarian somatic cells or follicle cells, which surround and provide support to the germ cells (Pelisson et al. 1994). This expression pattern may seem counter intuitive if the purpose of the retrotransposon is to increase its copy number in the germline genome. However, *gypsy* elements correspond to a class of errantivirus, which are related to retroviruses, and contain a gene encoding functional viral envelope, in addition to *gag* and *pol* genes. The *env* gene is expressed from a spliced version of the *gypsy* transcript, while the other two ORFs are translated from the unspliced transcript (Pelisson et al. 1994; Song et al. 1994). In this way, *gypsy* is able to create viral particles and infect its neighboring germ cells. In fact, in somatic cells that fail to silence *gypsy*, viral particles moving from follicle cells to germ cells have been detected (Song et al. 1994). To combat this threat, the piRNA pathway attacks *gypsy* at their source, in the somatic cells of the ovary (Brennecke et al. 2007; Malone et al. 2009). The piRNA pathway in the soma has unique characteristics that distinguish it from the germline piRNA pathway. First, a specific piRNA cluster, *flamenco*, is expressed exclusively in this cell type and is key in successful silencing of *gypsy* family transposons. Flies that harbor mutations in *flamenco* that inhibit this

cluster's transcription, have elevated levels of *gypsy* and are sterile, proving the importance of a functional piRNA pathway in both the germline and the somatic cells of the ovary (Bucheton 1995; Brennecke et al. 2007). Second, while all three Piwi clade proteins are expressed in germ cells, only Piwi is found in follicle cells (Brennecke et al. 2007). Therefore, a more simple, stripped down version of the piRNA pathway is active in somatic cells. More details on the piRNA pathway and its characteristics are explained in the following review article.

1.4 References

Bernstein E, Caudy AA, Hammond SM, Hannon GJ. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363-366.

Biémont C. 2010. A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics* **186**: 1085-1093.

Bradley D, Carpenter R, Sommer H, Hartley N, Coen E. 1993. Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the plena locus of Antirrhinum. *Cell* **72**: 85-95.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell* **128**: 1089-1103.

Brookman JJ, Toosy AT, Shashidhara LS, White RA. 1992. The 412 retrotransposon and the development of gonadal mesoderm in Drosophila. *Development* **116**: 1185-1192.

Bucheton A. 1995. The relationship between the flamenco gene and gypsy in Drosophila: how to tame a retrovirus. *Trends Genet* **11**: 349-353.

Bushati N, Cohen SM. 2007. microRNA functions. *Annu Rev Cell Dev Biol* **23**: 175-205.

Carmell MA, Xuan Z, Zhang MQ, Hannon GJ. 2002. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes & Development* **16**: 2733-2742.

Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, et al. 2008. An endogenous small interfering RNA pathway in Drosophila. *Nature* **453**: 798-802.

Ding D, Lipshitz HD. 1994. Spatially regulated expression of retrovirus-like transposons during Drosophila melanogaster embryogenesis. *Genet Res* **64**: 167-181.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**: 806-811.

Ghildiyal M, Zamore PD. 2009. Small silencing RNAs: an expanding universe. *Nat Rev Genet* **10**: 94-108.

Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ. 2001. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**: 1146-1150.

Hannon GJ. 2002. RNA interference. *Nature* **418**: 244-251.

Havecker ER, Gao X, Voytas DF. 2004. The diversity of LTR retrotransposons. *Genome Biol* **5**: 225.

Ishizu H, Siomi H, Siomi MC. 2012. Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. *Genes & Development* **26**: 2361-2373.

Joshua-Tor L, Hannon GJ. 2011. Ancestral roles of small RNAs: an Ago-centric perspective. *Cold Spring Harb Perspect Biol* **3**: a003772.

Kazazian HH. 2004. Mobile elements: drivers of genome evolution. *Science* **303**: 1626-1632.

Kidwell MG, Kidwell JF, Sved JA. 1977. Hybrid Dysgenesis in DROSOPHILA MELANOGASTER: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. *Genetics* **86**: 813-833.

Klattenhoff C, Bratu DP, McGinnis-Schultz N, Koppetsch BS, Cook HA, Theurkauf WE. 2007. Drosophila rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev Cell* **12**: 45-55.

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294**: 853-858.

Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, Carthew RW. 2004. Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**: 69-81.

Levin HL, Moran JV. 2011. Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet* **12**: 615-627.

Malone CD, Brennecke J, Dus M, Stark A, Mccombie WR, Sachidanandam R, Hannon GJ. 2009. Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. *Cell* **137**: 522-535.

Malone CD, Hannon GJ. 2009. Small RNAs as guardians of the genome. *Cell* **136**: 656-668.

McClintock B. 1956. Controlling Elements and the Gene. *Cold Spring Harb Symp Quant Biol* **21**: 197-216.

McClintock B. 1950. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci USA* **36**: 344-355.

Okamura K, Ishizuka A, Siomi H, Siomi MC. 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes & Development* **18**: 1655-1666.

Pardue M-L, DeBaryshe PG. 2008. Drosophila telomeres: A variation on the telomerase theme. *Fly (Austin)* **2**: 101-110.

Parkhurst SM, Corces VG. 1987. Developmental expression of Drosophila melanogaster retrovirus-like transposable elements. *EMBO J* **6**: 419-424.

Pelisson A, Song SU, Prud'homme N, Smith PA, Bucheton A, Corces VG. 1994. Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. *EMBO J* **13**: 4401-4411.

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

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**: 272-285.

Song J-J, Smith SK, Hannon GJ, Joshua-Tor L. 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**: 1434-1437.

Song SU, Gerasimova T, Kurkulos M, Boeke JD, Corces VG. 1994. An env-like protein encoded by a Drosophila retroelement: evidence that gypsy is an infectious retrovirus. *Genes & Development* **8**: 2046-2057.

Theurkauf WE, Klattenhoff C, Bratu DP, McGinnis-Schultz N, Koppetsch BS, Cook HA. 2006. rasiRNAs, DNA damage, and embryonic axis specification. *Cold Spring Harb Symp Quant Biol* **71**: 171-180.

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

Williams RW, Rubin GM. 2002. ARGONAUTE1 is required for efficient RNA interference in Drosophila embryos. *Proc Natl Acad Sci USA* **99**: 6889-6894.

1.5 The piRNA pathway in flies: highlights and future directions

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Genetics & Development

The piRNA pathway in flies: highlights and future directions

Paloma M Guzzardo, Felix Muerdter and Gregory J Hannon

Piwi proteins, together with their bound Piwi-interacting RNAs, constitute an evolutionarily conserved, germline-specific innate immune system. The piRNA pathway is one of the key mechanisms for silencing transposable elements in the germline, thereby preserving genome integrity between generations. Recent work from several groups has significantly advanced our understanding of how piRNAs arise from discrete genomic loci, termed piRNA clusters, and how these Piwi-piRNA complexes enforce transposon silencing. Here, we discuss these recent findings, as well as highlight some aspects of piRNA biology that continue to escape our understanding.

Address
Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, United States

Corresponding author: Hannon, Gregory J (hannon@cshl.edu)

Current Opinion in Genetics & Development 2012, **23**:xx–yy

This review comes from a themed issue on **Cancer genomics**

Edited by **Nahum Sonenberg** and **Nissim Hay**

0959-437X/$ – see front matter, Published by Elsevier Ltd.

http://dx.doi.org/10.1016/j.gde.2012.12.003

The piRNA pathway

Germ cells are the only cell type of an organism that contribute genetic material to future progeny. It is therefore essential that the integrity of this genome is preserved to protect reproductive success. One threat placed on germ cells is the movement of mobile genetic elements, or transposons, which correspond to a large fraction of the eukaryotic genome. Although transposons provide some benefits in driving evolution, their uncontrolled expression can lead to loss of genome integrity [1]. One of the major ways in which transposable elements (TEs) are kept under control is via Piwi-interacting RNAs (piRNAs) [2,3]. piRNAs are a class of small RNAs bound by the Piwi clade of Argonaute (Ago) proteins. As with all members of the Ago family, Piwi clade proteins rely on sequence complementarity to identify their targets, which for piRNAs are most commonly transposable elements. The importance of this pathway is evident; Piwi proteins are highly conserved throughout evolution, and their loss of function leads to gross defects in gametogenesis and to sterility.

With many aspects of this pathway being studied in a range of organisms, it is impossible to summarize all recent insights. Therefore, we will focus specifically on the piRNA pathway in the ovary of *Drosophila melanogaster*, which has been one of the main model organisms in the study of this pathway and which has helped establish the framework for how it functions.

An intriguing aspect of piRNA biology in *Drosophila* ovaries is that there are two distinct iterations of the pathway active in this tissue: one in the germ cells and one in the follicle cells, cells of somatic origin that surround and support the developing germ cells [4,5] (Figure 1a). Controlling TEs in both of these cell types is important, since active transposons found within follicle cells, such as those from the gypsy family of retrovirus-like transposons, can form viral particles and infect the oocyte [6]. The somatic and germline piRNA pathways are distinct mainly because of the different expression patterns of the three fly Piwi proteins. While Aubergine (Aub) and Argonaute (Ago3) are exclusively found in the nuage of germ cells, Piwi is found in the nuclei of both germ cells and follicle cells [7–10]. Therefore, the somatic pathway acts only through piRNAs generated by primary biogenesis, while in germ cells, in addition to primary biogenesis, a more complex piRNA amplification loop exists that depends on the slicer activity of Aub and Ago3 [9,10]. Understanding the less complex primary piRNA pathway acting in somatic cells has provided a basic mechanistic framework of piRNA biogenesis that is likely shared between both somatic and germline piRNAs.

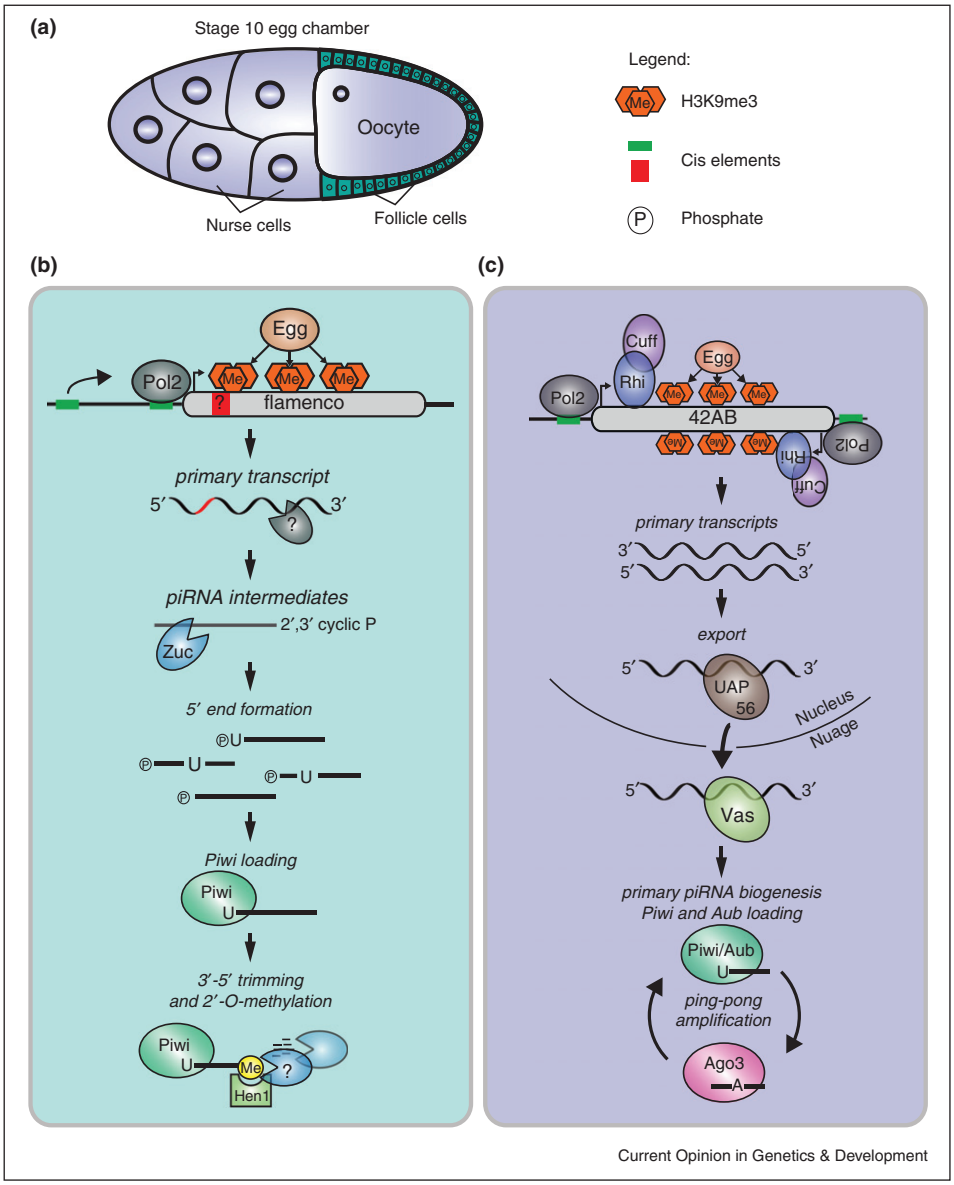
Taking advantage of the ease with which genetic manipulations can be done in *Drosophila*, studies of the small RNA populations in different piRNA mutants, together with other general molecular and cell biological analyses, such as localization studies and measurements of protein-protein interactions, have provided the main bulk of experimental data in the piRNA field [11]. The availability of cell lines derived from follicle cells (OSS/OSC) has also aided the study of the piRNA pathway [12–14]. To date, there are more than two-dozen proteins implicated in the piRNA pathway. However, many of the specific molecular steps that occur to generate a piRNA and that enable a piRNA to silence transposons remain unclear. In this review we will provide a brief summary of what is known about the piRNA pathway as well as discuss the open questions in the field.

How are piRNAs made?

The majority of piRNAs arise from specific genomic loci, known as piRNA clusters, which are found in pericentromeric heterochromatin [9]. Other sources of piRNAs

Please cite this article in press as: Guzzardo PM, et al.: The piRNA pathway in flies: highlights and future directions, Curr Opin Genet Dev (2013), http://dx.doi.org/10.1016/j.gde.2012.12.003

Figure 1



A model for piRNA biogenesis in the *Drosophila* ovary. **(a)** Two distinct piRNA pathways are active in a stage 10 egg chamber of the *Drosophila* ovary. The nurse cells that provide nutrients to the oocyte and the oocyte itself make up the germ cells of the ovary, shown in blue. The monolayer of somatic follicle cells surrounding the oocyte is shown in green. Nuclei are indicated as circles within each cell. **(b)** In follicle cells, primary piRNAs arise from *flamenco* and are processed through a cascade of enzymatic cuts. Transcription by RNA polymerase II (Pol2) depends on deposition of Histone 3 Lysine 9 trimethyl marks (H3K9me3) by Eggless (Egg). Regulatory *cis*-acting elements, indicated as green boxes, upstream of the transcriptional start site could affect Pol2 recruitment and transcription. Additionally, clusters could carry *cis* elements within themselves, shown in red, that affect downstream processing. After processing of the primary cluster transcript by unknown activities, piRNA intermediates are cleaved by the nuclease, Zucchini (Zuc). After 5′ end formation, transcripts with a U at the first position are preferentially loaded into Piwi. Trimming activity, which could be carried out by redundant nucleases, shortens the transcript to its mature length. This process is coupled to 2′-O-methylation by Hen1. **(c)** The transcription of clusters in germ cells can occur bidirectionally. In addition to Egg, the HP1 homolog Rhino (Rhi) and Cutoff (Cuff) are essential for transcription. Subsequently, the helicase UAP56 binds the primary transcript and escorts it to the nuclear periphery. There, it is handed over to another RNA helicase Vasa (Vas) and arrives at its site of biogenesis, the nuage. After primary processing by similar machinery as in (a), primary piRNAs are loaded into Piwi and Aub, and potentially Ago3. These primary piRNAs can be used to kick-start the ping-pong amplification cycle, which silences transposons post-transcriptionally.

Please cite this article in press as: Guzzardo PM, et al.: The piRNA pathway in flies: highlights and future directions, Curr Opin Genet Dev (2013), http://dx.doi.org/10.1016/j.gde.2012.12.003

do exist, such as the 3′UTRs of protein coding genes and dispersed euchromatic copies of TEs [9,14,15]. piRNA clusters contain remnants of transposons and serve as a catalog of sequences previously defined as targets for silencing. Exposure to a new transposon can lead to the expansion of this catalog and control of the TE, while omission from the catalog can mean that the element escapes repression [16•]. Brennecke *et al.* defined over 140 such clusters in *Drosophila* and saw that these clusters could be uni-directionally or bidirectionally transcribed [9]. Most of these clusters are active specifically in germ cells, while only a single major cluster (*flamenco*) drives transposon silencing in the soma. In general, germline clusters have two promoters, one on either side of the cluster (e.g. *cluster 42AB*), and are transcribed bidirectionally, while *flamenco* is uni-directionally transcribed.

Little is understood about what defines a piRNA cluster, how clusters are transcribed, and how this process is regulated. To date, we have no knowledge of transcription factors that regulate cluster expression. Clusters seem to be expressed in a cell-type specific manner, so there must be cell-type specific transcription factors enforcing this pattern. The promoters of clusters and their regulatory elements have not been defined, but in the case of *flamenco*, existing evidence suggests a single, discrete promoter, since a *P-element* insertion at the beginning of the cluster abolishes piRNA production, even ~200 kb away from the insertion point [9,17].

Some studies suggest a role for chromatin context in regulating cluster transcription. Deposition of Histone 3 Lysine 9 trimethyl marks (H3K9me3) was proposed to be necessary for cluster transcription, since mutations in Eggless (Egg, dSETDB1), a histone methyltransferase, lead to decreases in H3K9me3 deposition, and in the levels of cluster transcripts within both germ cells and somatic cells [18] (Figure 1b and c). As expected, these decreases in cluster transcription led to a reduction of mature piRNAs and upregulation of TEs. Rhino, a Heterochromatin Protein 1 homolog, and Cutoff (Cuff), a yeast Rai1-like nuclease, physically interact, and together bind specifically to bidirectionally transcribed clusters in the germline to promote their transcription [19,20]. Both proteins are found in nuclear foci in germ cells and depend on each other for their nuclear localization. How these factors promote cluster transcription remains unclear. Although Rhino and Cutoff are predominantly nuclear, their depletion is sufficient to disrupt Aub and Ago3 localization in nuage [19,20].

In another study addressing the role of chromatin context in cluster identity, Muerdter and colleagues found that when a cluster was taken out of its normal heterochromatic genomic context and placed in a euchromatic locus, it is still able to produce piRNAs [21]. This implies that clusters themselves contain sufficient information, possibly through *cis*-elements or secondary structure, to trigger piRNA production. However, it is also possible that information in the modified cluster is capable of recreating the chromatin context necessary for its expression, since the authors did not verify the euchromatic status of the cluster after insertion. In summary, more research is needed to understand the determinants of cluster identity; whether it be the chromatin context of the cluster, sequences in or surrounding the cluster that are important for transcription, or if it is sequences recognized within the transcript after transcription that then mark it as a piRNA producing transcript.

Following cluster transcription, the current model states that the primary transcript is exported to the cytoplasm, where it is processed into primary piRNAs that are loaded into Piwi or Aub. A recent study by Zhang *et al.* shed some light on how cluster transcripts are escorted from the transcription site to the nuage where processing is thought to occur [22]. The study shows that UAP56, a putative helicase, co-localizes with Rhino in nuclear foci. Mutation of UAP56 leads to germline transposon upregulation, decrease of piRNAs mapping to germline clusters, and disruption of Aub, Ago3, and Vasa from nuage. Based on how the Rhino-UAP56 foci are positioned next to the nuclear pore, and the finding that UAP56 and Vasa bind germline cluster transcripts, the authors proposed a model in which UAP56 escorts the primary transcript through the nuclear pore to nuage, where the transcript is handed over to Vasa and funneled into the biogenesis machinery. Since UAP56 is believed to be germ cell specific, factors that mediate export in the follicle cells remain a mystery. Whether the cluster transcript is exported as one long RNA or if some processing occurs in the nucleus to generate smaller piRNA intermediates to be exported, remains unknown.

After the cluster transcript is exported, it must be processed into piRNAs. Since Piwi-bound piRNAs have a strong preference for a uridine at the 5′ end (1 U) [9], this suggests a model of primary piRNA biogenesis wherein the 5′ end of the piRNA is generated first, followed by preferential loading of piRNA intermediates with a 5′ U into Piwi, followed by 3′ trimming. The variable lengths of primary piRNAs (23–29nt) could result from a footprint specific to the Piwi protein into which the intermediate is loaded, since the size of the RNA binding pocket probably varies slightly between each protein, and Aub, Ago3 and Piwi associated piRNAs are of slightly different lengths.

The factors responsible for 5′ and 3′ end formation have yet to be uncovered. However, recent advancements were made in our understanding of one piRNA protein that may be involved with end formation. Nishimasu *et al.* and Ipsaro *et al.* both revealed the crystal structure of the piRNA pathway protein Zucchini (Zuc) [23••,24••].

Please cite this article in press as: Guzzardo PM, et al.: The piRNA pathway in flies: highlights and future directions, Curr Opin Genet Dev (2013), http://dx.doi.org/10.1016/j.gde.2012.12.003

Based on its structure, Zuc shows a preference for binding specifically single stranded RNA. *In vitro* studies demonstrated that both the mouse and *Drosophila* Zuc protein had endoribonuclease activity [23••,24••], contradictory to previous reports implicating Zuc as a phospholipase [25,26]. The cleaved RNA product bore a 5′-monophosphate group, a characteristic of mature piRNAs. These data make Zuc the principal candidate for 5′ end formation. Both studies failed to show association of Zuc with piRNA precursors, which would have made the argument for its role as the 5′ nuclease much stronger, given that it shows no sequence preferences. Unlike most other piRNA factors, Zuc localizes to the mitochondrial membrane, and loss of this nuclease in either the germline, or the soma, results in a dramatic reduction of piRNAs [4,25–28]. The role that mitochondria could play in the piRNA pathway remains enigmatic, though its ancient connections to antiviral responses, for example it serving as the location at which the RIG-I pathway operates, is provocative [29]. In flies and mice, Piwi proteins are localized to discrete cytoplasmic structures associated with mitochondria [3], but whether this is purely to allow compartmentalization of the pathway, or whether it implies a further role of mitochondrial activity in the piRNA pathway is unclear.

The precise biochemical mechanism of piRNA 3′ end formation remains a mystery. Recent work in a cell line derived from silkworm ovaries, BmN4, has brought the field closer to identifying the 3′ generating enzyme [30••]. Kawaoka and colleagues established an *in vitro* 3′ trimming assay using BmN4 cell extracts. The authors found that Siwi (silkworm Piwi) binds transcripts with a bias toward 1 U, and that extended precursor transcripts could be trimmed in extracts, in a Mg 2+ dependent manner, to mature piRNA length. It had been determined previously that piRNAs are 2′-O-methylated at their 3′ termini by Hen1, and the addition of this modification was observed to be coupled with the trimming activity [31,32]. The importance of the 3′ terminal modification remains uncertain, because mutants of Hen1 have no detectable phenotype [31,32]. These findings are in accordance with the model that piRNA precursors bind to Piwi in the cytoplasm, and then are trimmed and methylated at the 3′ terminus. Unfortunately, the molecular nature of the trimming activity remains enigmatic; 'trimmer' could not be purified due to its insoluble nature. Moreover, no exonuclease has yet emerged as a candidate trimmer from genetic screen, which could indicate that multiple redundant trimmers exist or that trimmer has essential functions that mask an ability to isolate it as a piRNA pathway mutant.

Our current model follows the idea that Piwi must be loaded with a mature piRNA in order to be imported into the nucleus. Successful loading of Piwi-family proteins with primary piRNAs requires several other players. Although there are some distinguishing factors between the loading process in somatic and germ cells, many proteins are shared between the two pathways. The common proteins involved in biogenesis are Armitage (Armi), an RNA helicase, Shutdown (Shu), a cochaperone, and Vreteno (Vret) a TUDOR domain containing protein [27,28,33–37]. Although we understand little of the precise role of any of these proteins, mutation of any one disrupts localization of Piwi, and levels of associated piRNAs decrease dramatically [4,28,34–36]. It is important to note that mutations in Shu and Vret lead to delocalization of all three Piwi proteins in the germline, while Zuc and Armi mutants delocalize Piwi, but not Aub and Ago3. This could mean that Shu and Vret play a more general role in primary biogenesis involving Piwi and Aub, while Armi only aids Piwi in the piRNA loading process.

In the soma, Yb, a TUDOR-domain protein that also contains an RNA helicase motif, is an important additional factor for primary biogenesis. This protein localizes to foci in the cytoplasm, together with all other known loading components [27,33,38]. Zuc, the putative 5′ nuclease, localizes to mitochondria, many of which are adjacent to Yb bodies, supporting the role of these structures in Piwi RISC assembly. In Zuc mutants, Vret, Armi, Shu, and Yb all accumulate in enlarged Yb bodies with Piwi, suggesting that when the 5′ end of the piRNA cannot be generated, the loading machinery accumulates in the foci in response to a stall in biogenesis [27,28,33,35]. In the germline, there are no Yb bodies, and Yb is not expressed. Current evidence suggests that two Yb-related proteins, Brother of Yb and Sister of Yb, might serve the role played by Yb in the cytoplasm [28].

In germ cells, the loading process seems to occur in the nuage, where Aub and Ago3 localize. The function of the nuage is unknown, but many piRNA factors are found there, suggesting an important role in the piRNA pathway. One important difference between germ cells and the soma is that in germ cells, Aub and Ago3 engage in an adaptive, slicer-dependent loop termed the ping-pong cycle, which specifically amplifies the piRNA response against active elements [9,10]. In this model, Aub, bound to cluster-derived piRNAs, recognizes an active transposon transcript and cleaves it, generating the 5′ end of a new sense piRNA, which associates with Ago3. Subsequently, sense strand piRNA-loaded Ago3 can recognize complementary sequences in cluster transcripts, and through its slicer activity can generate a new antisense Aub bound piRNA, completing the cycle. According to the ping-pong model of piRNA amplification, Aub and Ago3 must be catalytically active in order to cleave new piRNAs from expressed transposons or piRNA cluster transcripts. However, the phenotypes of catalytically inactive mutants have never been described. While Aub and Ago3 seem to be responsible for generating

Please cite this article in press as: Guzzardo PM, et al.: The piRNA pathway in flies: highlights and future directions, Curr Opin Genet Dev (2013), http://dx.doi.org/10.1016/j.gde.2012.12.003

the 5′ end of each piRNA amplified through ping-pong, how the 3′ end is generated remains unknown, though it may proceed through the action of the same trimmer that is used for primary biogenesis.

In order to initiate the ping-pong cycle, piRNAs loaded into Aub are required. These come from two sources. One is primary biogenesis. The second is maternally deposited Aub, as the protein is loaded into developing oocytes along with associated piRNAs [8,39,40]. The importance of maternally deposited piRNAs is evident from analyses of hybrid dysgenesis models. In these cases, maternal deposition of piRNAs, produced by ping-pong and corresponding to the *I-element* or *P-element*, correlates with initiation of ping-pong in progeny and with effective element silencing [40]. For the *I-element*, as mothers age, their progeny have a reduced probability of being sterile even in the absence of the ability of the mother to use active *I-elements* as ping-pong substrates [16•,41]. For *P-elements*, even the dysgenic progeny can regain some fertility as the animals age. This suggests that perhaps primary piRNAs corresponding to those elements accumulate with age in the mother or offspring to a level sufficient to confer resistance.

How do piRNAs silence transposons?

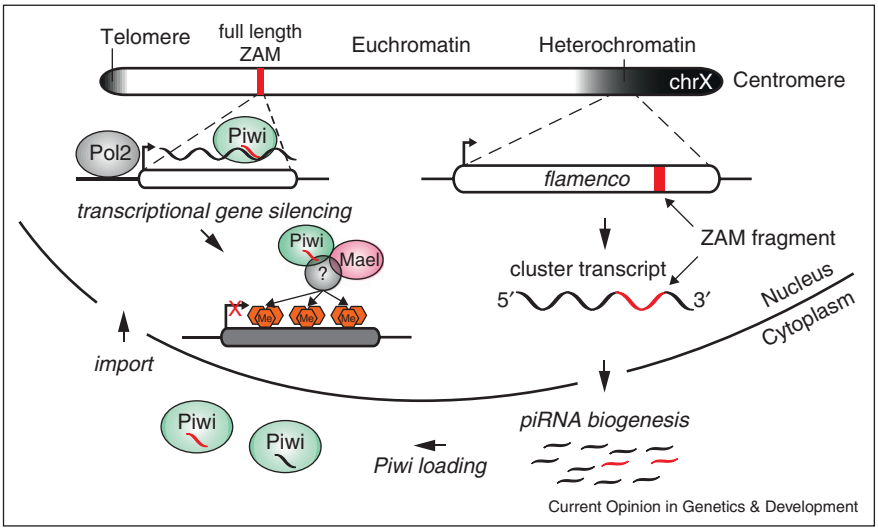
It seems evident that in germ cells Aub and Ago3 silence transposons through post-transcriptional gene silencing (PTGS). These two proteins posses slicer activity and cleave active TE transcripts during the ping-pong amplification cycle. By using the cleavage products to make more piRNAs, this cycle is able to amplify its response to actively transcribed elements [9,10].

The mechanism by which Piwi silences transposons proved much more difficult to dissect. It had long been suspected that Piwi mediates transcriptional gene silencing (TGS) of TEs through impacts on chromatin, due mainly to several provocative clues. First, Piwi is a nuclear protein, and this localization is essential to its silencing capability. A mutant Piwi lacking its nuclear localization signal is found in the cytoplasm and is incapable of silencing TEs but binds piRNAs to wildtype levels [14,27,42]. In addition, Piwi's slicer activity is not necessary for silencing, as a catalytically dead Piwi mutant rescues the null mutant phenotype [14,27].

Many studies have suggested that Piwi could silence transposons at a transcriptional level by inducing changes in histone marks, much like the mechanism by which small RNAs induce heterochromatin formation in yeast [43]. In fact, the murine piRNA pathway silences transposable elements by inducing chromatin changes, ultimately resulting in DNA methylation [44,45]. In *Drosophila*, several studies support a role for Piwi in acting through TGS in the ovary; multiple groups have reported changes in histone marks on a handful of transposons upon disruption of the piRNA pathway [42,46,47], and a study by Shpiz *et al.* detected an increase in several nascent TE transcripts upon Piwi knockdown (KD) [48]. However, it was the recent study by Sienski and colleagues that definitively demonstrated that Piwi silences transposons at the transcriptional level, triggering changes in chromatin state genome wide (Figure 2). The authors took advantage of the OSS/OSC cell line and did side-by-side comparisons of RNA Polymerase II (Pol2) occupancy, trimethylation of H3K9 (a common mark of heterochromatin), nascently transcribed RNA, and steady state RNAs at a global level in Piwi KD versus control cells [49••]. They observed that in the absence of Piwi, Pol2 occupancy on transposons increased, along with an increase of nascent TE transcripts and steady state RNA levels. Furthermore, levels of H3K9me3 marks on transposons dropped in the Piwi KD as compared to controls. Interestingly, the authors also observed that many TE sequences dispersed in euchromatin trigger the formation of an H3K9me3 island that is dependent on Piwi and on transcription of the locus. This strongly implicates an RNA-recognition mode for Piwi-dependent silencing. The study also identified Maelstrom (Mael), a protein previously implicated in the germline piRNA pathway, as playing a role in transcriptional silencing of transposons [50,51]. Upon Mael KD, there was no change in levels of mature piRNAs, but there were increases in Pol2 occupancy on TEs and nascent transcripts. Interestingly, levels of H3K9me3 did not decrease when Mael was depleted; rather, H3K9 methylation appeared to spread downstream of the TE insertion, in some cases for up to 30 kb. This places Mael downstream of Piwi in silencing of TEs. The precise mechanism by which Piwi influences chromatin state remains elusive. Other than Mael, no other effector protein has been identified. One likely candidate to play a role in this process is Heterochromatin Protein 1a (HP1a), which is believed to bind H3K9 methyl groups [52,53]. HP1a has been shown to interact with Piwi, and its depletion leads to TE derepression [47,54]. The current model of piRNA-mediated TGS proposes that Piwi RISC recognizes nascent transposon transcripts by sequence complementarity and then, with the help of Mael, recruits silencing machinery to trigger histone modifications at the site of transcription (Figure 2). The association of Piwi with chromatin seems to be unstable, as the authors were unable to map it to TE loci using chromatin immunoprecipitation. It is clear that other silencing effectors in addition to H3K9 are necessary because Mael mutants do not lose H3K9me3, but have upregulation of TE transcripts. Further experiments are needed to fully understand this process. Even though it seems likely that TGS is the main silencing mode for Piwi, there remains a possibility that it is also acting through PTGS at some level. This study did not address the role of Piwi in the germline nucleus but it

Please cite this article in press as: Guzzardo PM, et al.: The piRNA pathway in flies: highlights and future directions, Curr Opin Genet Dev (2013), http://dx.doi.org/10.1016/j.gde.2012.12.003

Figure 2



Transcriptional silencing of transposable elements by Piwi-piRNA complexes in the soma. The X chromosome of *Drosophila melanogaster* (chrX) is shown. A simplistic view of its chromatin state is indicated in shades of gray. The transcriptionally active euchromatin in white harbors a full-length copy of the retroelement, *ZAM* (indicated as a red box). An inactive remnant of the same element (in red) can be found within the *flamenco* piRNA cluster in pericentromeric heterochromatin. After transcription and processing of *flamenco*, this fragment gives rise to antisense piRNAs that are loaded into Piwi in the cytoplasm (indicated as red piRNA species). Upon reimport into the nucleus, these Piwi-piRNA complexes recognize active transcription of the full-length *ZAM* copy by RNA polymerase II (Pol2) based on sequence complementarity. This recognition leads to the recruitment of additional factors such as Maelstrom (Mael) and unknown chromatin remodelers. Ultimately, the deposition of H3K9me3 marks leads to loss of Pol2 occupancy and the transcriptional silencing of *ZAM*.

seems likely that it will also silence TEs by TGS in that setting.

Germ cells might prove to be more complex because of the presence of Aub and Ago3. Although these two proteins are engaged in the ping-pong cycle in the nuage, spatially separated from Piwi in the nucleus, there seems to be a more intimate connection between these proteins than has been generally appreciated. A strong indication of this connection is that in Aub and Ago3 mutants, levels of Piwi protein decrease [5]. Furthermore, in an Ago3 mutant, the levels of Piwi-bound piRNAs decrease and there is a shift in their sense versus antisense bias [5]. Considered together, these data indicate that there is significant crosstalk between Piwi and the ping-pong cycle. One point to remember is that, although ping-pong is thought to occur mainly between Aub and Ago3, there are a significant number of Piwi:Ago3 ping-pong pairs detected in ovaries [9]. Further studies will be critical in understanding the relationship between Piwi and ping-pong, and which mechanisms are employed to silence TEs in the germline.

What is the function of maternally deposited Piwi RISC complexes?

Piwi and Aubergine, together with their bound piRNAs, are maternally deposited in the embryo and accumulate in the pole plasm, which gives rise to the future germline [8,39,40]. These maternally contributed complexes are thought to be essential in priming the piRNA pathway to be able to successfully silence elements. Previous studies have revealed that hybrid dysgenesis is caused by the failure to maternally deposit piRNAs corresponding to a paternally contributed transposon [40]. These maternally contributed Piwi and Aub RISCs may serve to jump-start the silencing pathway to target elements even before zygotic transcription has begun. Therefore, maternally deposited complexes could be one of the triggers to initiate the ping-pong cycle, which will continue throughout the life of the organism.

A recent study offers another important role for these inherited complexes. de Vanssay *et al.* found that maternally deposited piRNAs could be involved in the specification of a piRNA cluster [55•]. In a previous study, the group characterized a phenomenon known as trans-silencing effect (TSE) in which *P-element* derived transgenes inserted in a heterochromatic region can silence a distinct *P-element* derived transgene inserted at a euchromatic locus. Using this system the authors found that a transgene cluster that induces strong silencing can convert a separate, homologous locus that is normally incapable of trans-silencing, into a strong silencer, in a heritable manner. This effect is dependent on maternally deposited piRNA complexes. This implies that the inherited piRNA complexes are needed to reestablish piRNA

Please cite this article in press as: Guzzardo PM, et al.: The piRNA pathway in flies: highlights and future directions, Curr Opin Genet Dev (2013), http://dx.doi.org/10.1016/j.gde.2012.12.003

cluster definitions in the progeny. Consequently, the piRNA pathway may completely reset and cluster identity be re-acquired between each generation. This concept is analogous to piRNA-driven transposon silencing in mammals; during primordial germ cell development, the germline is stripped of all DNA methylation, which is then reacquired on TEs through the action of piRNA-driven *de novo* methylation [45]. Since *Drosophila* lacks the ability to methylate DNA, maternally deposited piRNA complexes may serve a similar role in identifying TEs in the developing progeny. However, further work is necessary to evaluate this hypothesis. For instance, it would be interesting to specifically eliminate the maternally inherited pool of Piwi RISCs to observe if cluster definitions are lost.

In embryos, although maternally deposited Piwi and Aub are both localized to pole plasm in early embryogenesis, their localization patterns rapidly change during the cellularization of the embryo. While Aub continues to reside exclusively in pole cells, Piwi localizes to the nuclei of every cell of the embryo, and continues to do so until ~12 hours after egg laying [40,54]. What role might Piwi play in somatic nuclei during embryogenesis? One interesting possibility, especially considering the recent findings implicating Piwi in TGS, is that the protein is establishing silencing marks on transposons throughout the somatic compartment. In fact, many studies have implicated Piwi in positional effect variegation (PEV), a clearly somatic effect, and have observed Piwi binding on polytene chromosomes [54,56,57]. Perhaps the suppression of transgenes observed during PEV is mediated by Piwi-induced chromatin silencing in early embryogenesis, and is maintained throughout development. Extensive additional work will be necessary to fully understand the role of maternally deposited Piwi and Aub, but there is no doubt that there are many fascinating discoveries to be made in this area.

Conclusions

It has been almost a decade since the discovery of piRNAs, and many advances have been made toward understanding the general function of the pathway. However, surprisingly little is known about several key aspects of piRNA biology, such as the mechanistic details of piRNA biogenesis and how the downstream targets of the pathway are silenced. This is because many of the important players in the pathway still remain unknown. A genome-wide screen for piRNA pathway factors would aid in identifying all proteins involved, so that a full genetic framework could finally and conclusively be established. There is also an overwhelming need to develop biochemical assays that recapitulate several aspects of the piRNA pathway *in vitro*. These could bring much needed mechanistic insights into precisely how the pathway operates. Some progress has been made in this direction with the development of the silkworm trimming assay [30••]. Following the introduction of the Drosophila OSS/OSC cell lines by Niki *et al.*, both genome wide screens and *in vitro* assays have become feasible [12]. Given these tools and recent advances described here, it is easy to imagine that we will see many more exciting discoveries and insights into how small RNAs provide an immune defense against mobile elements.

Acknowledgements

We would like to thank Clare Rebbeck and Leah Sabin for critical comments on the manuscript and helpful discussion. We would also like to thank Julius Brennecke for sharing of data before publication. P.M.G. is a NIH trainee on a CSHL WSBS NIH Kirschstein-NRSA pre-doctoral award (T32 GM065094), a William Randolph Hearst Scholar and a Leslie Quick Junior Fellow. Work in the Hannon laboratory is supported by grants from the NIH and by a kind gift from Kathryn W. Davis. G.J.H. is an investigator of the HHMI.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Levin HL, Moran JV: **Dynamic interactions between transposable elements and their hosts**. *Nat Rev Genet* 2011, **12**:615-627.
2. Malone CD, Hannon GJ: **Small RNAs as guardians of the genome**. *Cell* 2009, **136**:656-668.
3. Siomi MC, Sato K, Pezic D, Aravin AA: **PIWI-interacting small RNAs: the vanguard of genome defence**. *Nat Rev Mol Cell Biol* 2011, **12**:246-258.
4. Malone CD, Brennecke J, Dus M, Stark A, Mccombie WR, Sachidanandam R, Hannon GJ: **Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary**. *Cell* 2009, **137**:522-535.
5. Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, Seitz H, Horwich MD, Syrzycka M, Honda BM *et al.*: **Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies**. *Cell* 2009, **137**:509-521.
6. Lécher P, Bucheton A, Pelisson A: **Expression of the Drosophila retrovirus gypsy as ultrastructurally detectable particles in the ovaries of flies carrying a permissive flamenco allele**. *J Gen Virol* 1997, **78(Pt 9)**:2379-2388.
7. Cox DN, Chao A, Lin H: **piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells**. *Development (Cambridge, England)* 2000, **127**:503-514.
8. Harris AN, Macdonald PM: **Aubergine encodes a Drosophila polar granule component required for pole cell formation and related to eIF2C**. *Development (Cambridge, England)* 2001, **128**:2823-2832.
9. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ: **Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila**. *Cell* 2007, **128**:1089-1103.
10. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC: **A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila**. *Science (New York, NY)* 2007, **315**:1587-1590.
11. Ishizu H, Siomi H, Siomi MC: **Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines**. *Genes Dev* 2012, **26**:2361-2373.
12. Niki Y, Yamaguchi T, Mahowald AP: **Establishment of stable cell lines of Drosophila germ-line stem cells**. *Proc Natl Acad Sci USA* 2006, **103**:16325-16330.

Please cite this article in press as: Guzzardo PM, et al.: The piRNA pathway in flies: highlights and future directions, Curr Opin Genet Dev (2013), http://dx.doi.org/10.1016/j.gde.2012.12.003

13. Lau NC, Robine N, Martin R, Chung W-J, Niki Y, Berezikov E, Lai EC: **Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line**. *Genome Res* 2009, **19**:1776-1785.

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

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

16. • Khurana JS, Wang J, Xu J, Koppetsch BS, Thomson TC, Nowosielska A, Li C, Zamore PD, Weng Z, Theurkauf WE: **Adaptation to P element transposon invasion in *Drosophila melanogaster***. *Cell* 2011, **147**:1551-1563.
This paper demonstrates how the piRNA system can evolve its silencing repertoire in response to challenge by a new transposon.

17. Robert V, Prud, apos, homme N, Kim A, Bucheton A, Pelisson A: **Characterization of the flamenco region of the *Drosophila melanogaster* genome**. *Genetics* 2001, **158**:701-713.

18. Rangan P, Malone CD, Navarro C, Newbold SP, Hayes PS, Sachidanandam R, Hannon GJ, Lehmann R: **piRNA production requires heterochromatin formation in Drosophila**. *Curr Biol* 2011, **21**:1373-1379.

19. Pane A, Jiang P, Zhao DY, Singh M, Schüpbach T: **The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline**. *EMBO J* 2011, **30**:4601-4615.

20. Klattenhoff C, Xi H, Li C, Lee S, Xu J, Khurana JS, Zhang F, Schultz N, Koppetsch BS, Nowosielska A *et al.*: **The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters**. *Cell* 2009, **138**:1137-1149.

21. Muerdter F, Olovnikov I, Molaro A, Rozhkov NV, Czech B, Gordon A, Hannon GJ, Aravin AA: **Production of artificial piRNAs in flies and mice**. *RNA (New York, NY)* 2012, **18**:42-52.

22. Zhang F, Wang J, Xu J, Zhang Z, Koppetsch BS, Schultz N, Vreven T, Meignin C, Davis I, Zamore PD *et al.*: **UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery**. *Cell* 2012, **151**:871-884.

23. •• Ipsaro JJ, Haase AD, Knott SR, Joshua-Tor L, Hannon GJ: **The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis**. *Nature* 2012.
This paper reports the biochemistry and three dimensional structure of the mouse Zucchini protein and implicates it as the nuclease that forms piRNA 5′ ends.

24. •• Nishimasu H, Ishizu H, Saito K, Fukuhara S, Kamatani MK, Bonnefond L, Matsumoto N, Nishizawa T, Nakanaga K, Aoki J *et al.*: **Structure and function of Zucchini endoribonuclease in piRNA biogenesis**. *Nature* 2012.
Along with the Ipsaro paper, this report of the structure and biochemistry of Drosophila Zucchini may have solved the mystery of piRNA 5′ processing.

25. Huang H, Gao Q, Peng X, Choi S-Y, Sarma K, Ren H, Morris AJ, Frohman MA: **piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling**. *Dev Cell* 2011, **20**:376-387.

26. Watanabe T, Chuma S, Yamamoto Y, Kuramochi-Miyagawa S, Totoki Y, Toyoda A, Hoki Y, Fujiyama A, Shibata T, Sado T *et al.*: **MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline**. *Dev Cell* 2011, **20**:364-375.

27. Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, Siomi H, Siomi MC: **Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila**. *Genes Dev* 2010, **24**:2493-2498.

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

29. Arnoult D, Soares F, Tattoli I, Girardin SE: **Mitochondria in innate immunity**. *EMBO Rep* 2011, **12**:901-910.

30. •• Kawaoka S, Izumi N, Katsuma S, Tomari Y: **3′ end formation of PIWI-interacting RNAs in vitro**. *Mol Cell* 2011, **43**:1015-1022.
This paper reports one of the few successful attempts at addressing questions of piRNA biology using *in vitro*, biochemical approaches and provides a mechanism, if not the enzyme, for piRNA 3′ end formation

31. Horwich MD, Li C, Matranga C, Vagin V, Farley G, Wang P, Zamore PD: **The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC**. *Curr Biol* 2007, **17**:1265-1272.

32. Saito K, Sakaguchi Y, Suzuki T, Suzuki T, Siomi H, Siomi MC: **Pimet, the Drosophila homolog of HEN1, mediates 2′-O-methylation of Piwi-interacting RNAs at their 3′ ends**. *Genes Dev* 2007, **21**:1603-1608.

33. Olivieri D, Sykora MM, 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* 2010, **29**:3301-3317.

34. Zamparini AL, Davis MY, Malone CD, Vieira E, Zavadil J, Sachidanandam R, Hannon GJ, Lehmann R: **Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in Drosophila**. *Development (Cambridge, England)* 2011, **138**:4039-4050.

35. Olivieri D, Senti K-A, Subramanian S, Sachidanandam R, Brennecke J: **The cochaperone shutdown defines a group of biogenesis factors essential for all piRNA populations in Drosophila**. *Mol Cell* 2012, **47**:954-969.

36. Preall JB, Czech B, Guzzardo PM, Muerdter F, Hannon GJ: **shutdown is a component of the Drosophila piRNA biogenesis machinery**. *RNA (New York, NY)* 2012, **18**:1446-1457.

37. Xiol J, Cora E, Koglgruber R, Chuma S, Subramanian S, Hosakawa M, Reuter M, Yang Z, Berninger P, Palencia A *et al.*: **A role for Fkbp6 and the chaperone machinery in piRNA amplification and transposon silencing**. *Mol Cell* 2012, **47**:970-979.

38. Szakmary A, Reedy M, Qi H, Lin H: **The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in *Drosophila melanogaster***. *J Cell Biol* 2009, **185**:613-627.

39. Megosh HB, Cox DN, Campbell C, Lin H: **The role of PIWI and the miRNA machinery in Drosophila germline determination**. *Curr Biol* 2006, **16**:1884-1894.

40. Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ: **An epigenetic role for maternally inherited piRNAs in transposon silencing**. *Science (New York, NY)* 2008, **322**:1387-1392.

41. Grentzinger T, Armenise C, Brun C, Mugat B, Serrano V, Pelisson A, Chambeyron S: **piRNA-mediated transgenerational inheritance of an acquired trait**. *Genome Res* 2012, **22**:1877-1888.

42. Klenov MS, Sokolova OA, Yakushev EY, Stolyarenko AD, Mikhaleva EA, Lavrov SA, Gvozdev VA: **Separation of stem cell maintenance and transposon silencing functions of Piwi protein**. *Proc Natl Acad Sci USA* 2011, **108**:18760-18765.

43. Huisinga K, Elgin S: **Small RNA-directed heterochromatin formation in the context of development: what flies might learn from fission yeast**. *Biochim Biophys Acta* 2008, **1789**:3-16.

44. Aravin AA, Sachidanandam R, Girard A, Fejes Toth K, Hannon GJ: **Developmentally regulated piRNA clusters implicate MILI in transposon control**. *Science (New York, NY)* 2007, **316**:744-747.

45. Aravin AA, Bourc'his D: **Small RNA guides for de novo DNA methylation in mammalian germ cells**. *Genes Dev* 2008, **22**:970-975.

46. Klenov MS, Lavrov SA, Stolyarenko AD, Ryazansky SS, Aravin AA, Tuschl T, Gvozdev VA: **Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the *Drosophila melanogaster* germline**. *Nucleic Acids Res* 2007, **35**:5430-5438.

Please cite this article in press as: Guzzardo PM, et al.: The piRNA pathway in flies: highlights and future directions, Curr Opin Genet Dev (2013), http://dx.doi.org/10.1016/j.gde.2012.12.003

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

48. Shpiz S, Olovnikov I, Sergeeva A, Lavrov S, Abramov Y, Savitsky M, Kalmykova A: **Mechanism of the piRNA-mediated silencing of Drosophila telomeric retrotransposons**. *Nucleic Acids Res* 2011, **39**:8703-8711.

49. •• Sienski G, Dönertas D, Brennecke J: **Transcriptional silencing of transposons by Piwi and Maelstrom and its impact on chromatin state and gene expression**. *Cell* 2012, **151**:964-980.
This paper provides a genome-wide view of changes in chromatin structure upon Piwi silencing and provides concrete evidence that Piwi operates in the soma by regulating transposons at the level of their transcription.

50. Findley SD, Tamanaha M, Clegg NJ, Ruohola-Baker H: **Maelstrom, a Drosophila spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage**. *Development (Cambridge, England)* 2003, **130**:859-871.

51. Lim AK, Kai T: **Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster***. *Proc Natl Acad Sci USA* 2007, **104**:6714-6719.

52. Lachner M, apos O, Carroll D, Rea S, Mechtler K, Jenuwein T: **Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins**. *Nature* 2001, **410**:116-120.

53. Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T: **Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain**. *Nature* 2001, **410**:120-124.

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

55. • de Vanssay A, Bougé A-L, Boivin A, Hermant C, Teysset L, Delmarre V, Antoniewski C, Ronsseray S: **Paramutation in Drosophila linked to emergence of a piRNA-producing locus**. *Nature* 2012, **490**:112-115.
The authors implicate maternally deposited piRNAs as being important for specifying the identity of piRNA clusters.

56. Haynes K, Caudy A, Collins L, Elgin S: **Element 1360 and RNAi components contribute to HP1-dependent silencing of a pericentric reporter**. *Curr Biol* 2006, **16**:2222-2227.

57. Pal-Bhadra M: **Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery**. *Science (New York, NY)* 2004, **303**:669-672.

Please cite this article in press as: Guzzardo PM, et al.: The piRNA pathway in flies: highlights and future directions, Curr Opin Genet Dev (2013), http://dx.doi.org/10.1016/j.gde.2012.12.003

2 Harnessing the Ovarian Somatic Sheath cell line for genome-wide studies

Chapter contributions:

Work presented in this chapter was a collaboration between Felix Muerdter, Astrid Haase and myself. Work done regarding the maintenance of the OSS cell line and experiments testing transfection reagents was a joint effort. Felix Muerdter made the small RNA library. I constructed RNA-Seq libraries and performed western blots. The knockdown experiments and qPCRs were done by Felix Muerdter and myself.

2.1 Introduction

The *Drosophila* ovary, the model used for the majority of the piRNA pathway studies, is a heterogeneous mixture of different cell types and developmental stages. A female *Drosophila* has 2 ovaries, each composed of approximately 18 ovarioles. Every ovariole contains several egg chambers at consecutive stages of development. Hence each ovariole can be seen as an egg production line, with the germarium, containing somatic and germline stem cells, at the anterior end, and a mature egg at the posterior end. One egg chamber contains 15 nurse cells and one oocyte, the germ cells, encircled by a monolayer of ~1,000 somatic cells known as follicle cells (Bastock and St Johnston 2008). The inherent complexity of this tissue makes it difficult to study specific aspects of the piRNA pathway. To complicate things further, several studies have shown that there are two distinct versions of the piRNA pathway active in germ cells and follicle cells of the ovary (Malone et al. 2009; Li et al. 2009). Examining the somatic and germline pathway using ovary tissue is problematic due to the close association of the cell types and that several proteins are common to both pathways. Working with tissue also proves difficult for biochemical studies and high-throughput experiments. For these reasons, a cell line derived from germ cells or follicle cells would aid research a great deal. Fortuitously, the Mahowald group recently developed two *Drosophila* ovarian cell lines. Using ovaries from *bag of marbles* (*bam*) mutant females, Niki et al derived a cell line containing a mixture of germ cells and somatic cells (fGS/OSS) and one purely of somatic cells (OSS) (Niki et al. 2006). Flies lacking functional Bam protein, which is necessary for differentiation, accumulate germline stem cell-like cells. When put in culture, these germline stem cells could not survive without somatic stem cells surrounding them; however, when somatic cells were alone in culture, they would form a monolayer of cells and could be successfully passaged. This purely somatic cell line provides the opportunity to study the primary piRNA pathway in a more direct manner, and allows for many previously infeasible experiments. With still many open questions regarding primary piRNA biogenesis and how Piwi silences transposons, a powerful way to identify all involved proteins would to carry out a genome-wide screen searching for

factors involved in the primary piRNA pathway. However, before beginning a screen, it was necessary to evaluate the cell line and see if it was suitable for genome-wide studies, as well as perform general characterization to ensure that these cells recapitulate the piRNA pathway from follicle cells of the ovary.

2.2 Results

To begin, we decided to analyze patterns of gene expression in OSS cells by studying their transcriptome. Studies on expression patterns of Piwi clade proteins in flies have concluded that Piwi is the only member expressed in follicle cells, while all three, Piwi, Aub and Ago3, are expressed in germ cells (Brennecke et al. 2007). The transcriptome of OSS cells confirmed these findings, with very high levels of Piwi mRNA detected, and very low levels of both Aub and Ago3 (Figure 2.1). OSS cells also express known piRNA components, Armitage (Armi), Zucchini (Zuc) and fs(1)Yb, which have been described to be active in the somatic compartment, by research from our lab as well as other groups (Figure 2.1, Appendix 1) (Haase et al. 2010; Saito et al. 2010; Olivieri et al. 2010). Furthermore, we found that components of the canonical RNAi machinery, both for siRNAs and miRNAs, are also expressed (Figure 2.1). Knowing that Piwi is expressed, we next wanted to verify if the protein product could be detected and where it localizes. In ovaries, Piwi localizes to the nuclei of somatic and germ cells; when we performed immunofluorescence analysis we observed that in OSS cells, Piwi is also nuclear (Figure2.2).

After establishing that Piwi is expressed and has the expected nuclear localization, we set out to analyze the small RNA populations (19-28nt) present in this cell line. For this purpose, we cloned 19-28nt RNAs from OSS total RNA. The resulting library contained 5.7 million reads, which collapsed into 2 million unique sequences. Following mapping and sequence annotation, there were approximately 4 million reads, made up of 1.1 million sequences that mapped unambiguously to the genome. These sequences were comprised of several annotation categories, with the vast majority, 61%, being transposon derived (Figure 2.2). The next

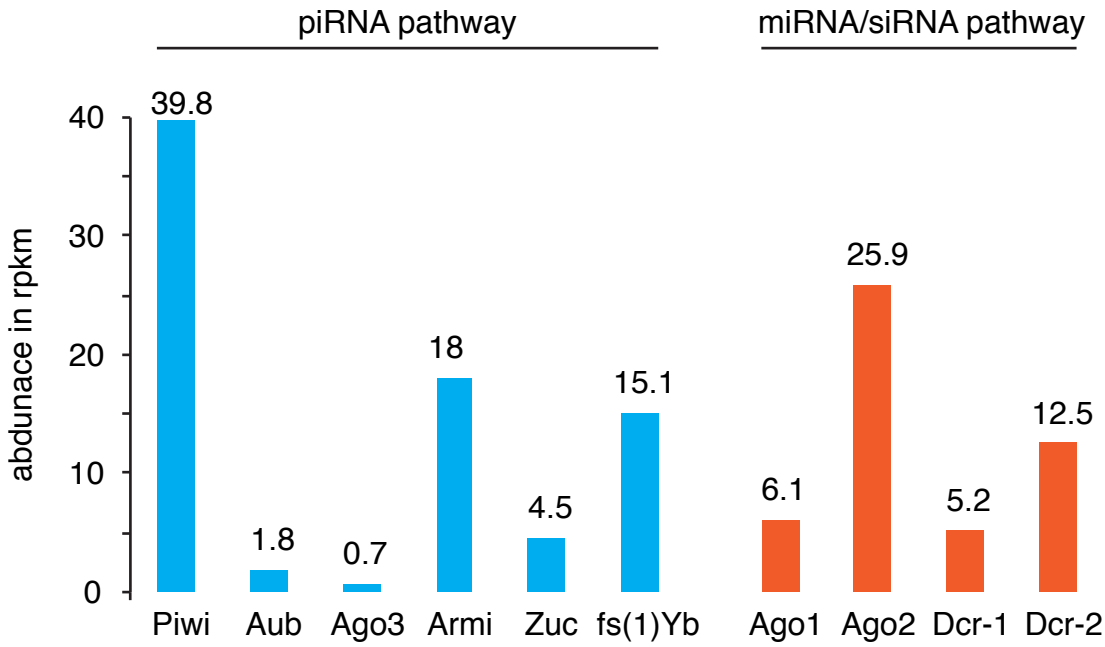


Figure 2.1: **Expression of piRNA pathway components and key microRNA and siRNA pathway members in OSS cells.** Reads from a poly-A selected transcriptome library were mapped to the coding sequence of each specified gene. The normalized number of mapped reads in rpkm is shown.

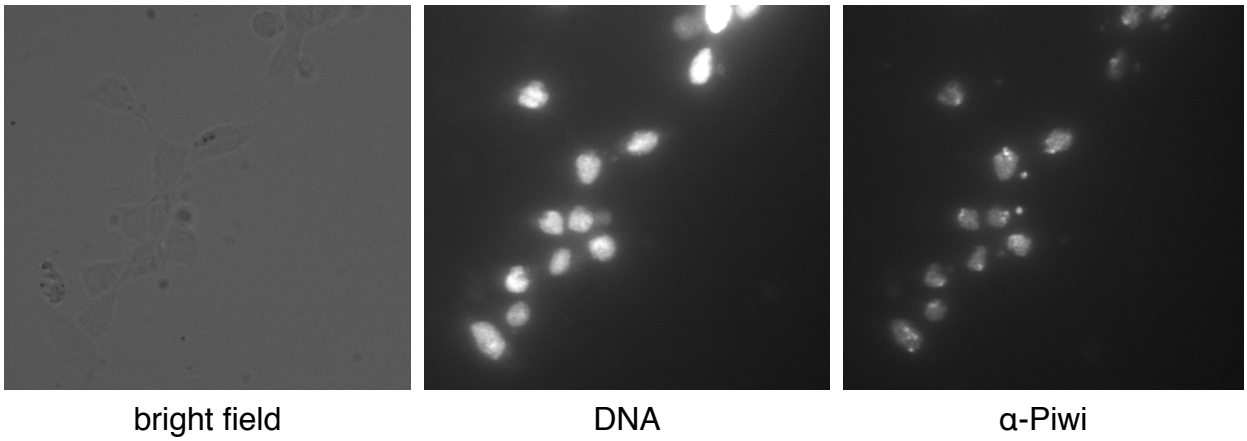


Figure 2.2: **Piwi protein localization in OSS cells.** The bright field image shows OSS cell morphology. OSS cells were stained with DAPI to show nuclei localization. Sub-cellular localization of endogenous Piwi protein was assessed by immunofluorescence using Piwi antibody.

most abundant categories, each representing 10% to 12% of all mapped reads, were miRNAs, structural RNAs, and sRNAs arising from coding sequences (gene). The structural RNA category was mainly composed of 2S rRNA, which migrates at about 28nt. Thus, it is not surprising to clone these sequences. However, the amount of contamination is small enough to not affect sequencing depth of the remainder.

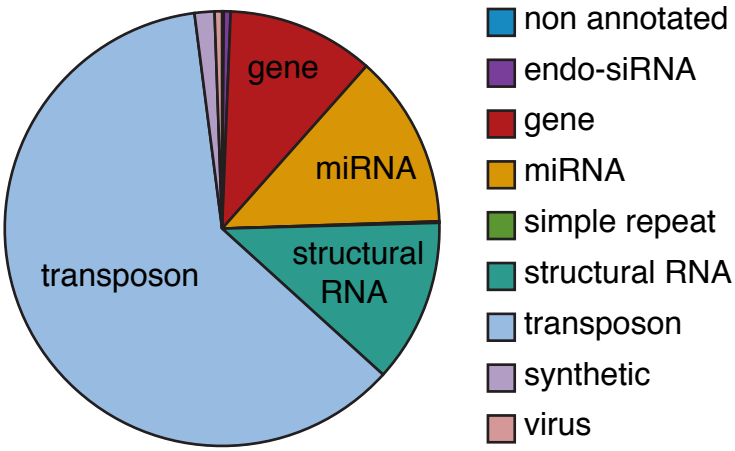


Figure 2.3: **Small RNA population in OSS cells.** A pie chart displaying the percentage of reads corresponding to each annotation category from an OSS total RNA library is shown.

Next, we verified the size profiles of several of these different classes of sRNAs. Every sRNA pathway is associated with a specific sRNA length distribution, due to the biogenesis pathway and Ago protein binding partner. Therefore, size profiles have become a way to determine what class the sRNA belongs too. miRNA show their characteristic length peak around 22-23 nt, while endo-siRNAs show a very tight peak around 21nt, as has been previously described (Figure 2.4) (Kim et al. 2009). This data, taken together with the transcriptome expression data showing the expression of RNAi machinery, suggests that OSS cells have functional miRNA and siRNA biogenesis pathways. With these pathways in place, theoretically, it is possible to shutdown gene using RNAi by introducing dsRNAs, siRNAs or short hairpin RNAs targeting specific genes. Undoubtedly, being able to manipulate gene expression in this way is essential for successful execution of a genome-wide RNAi screen.

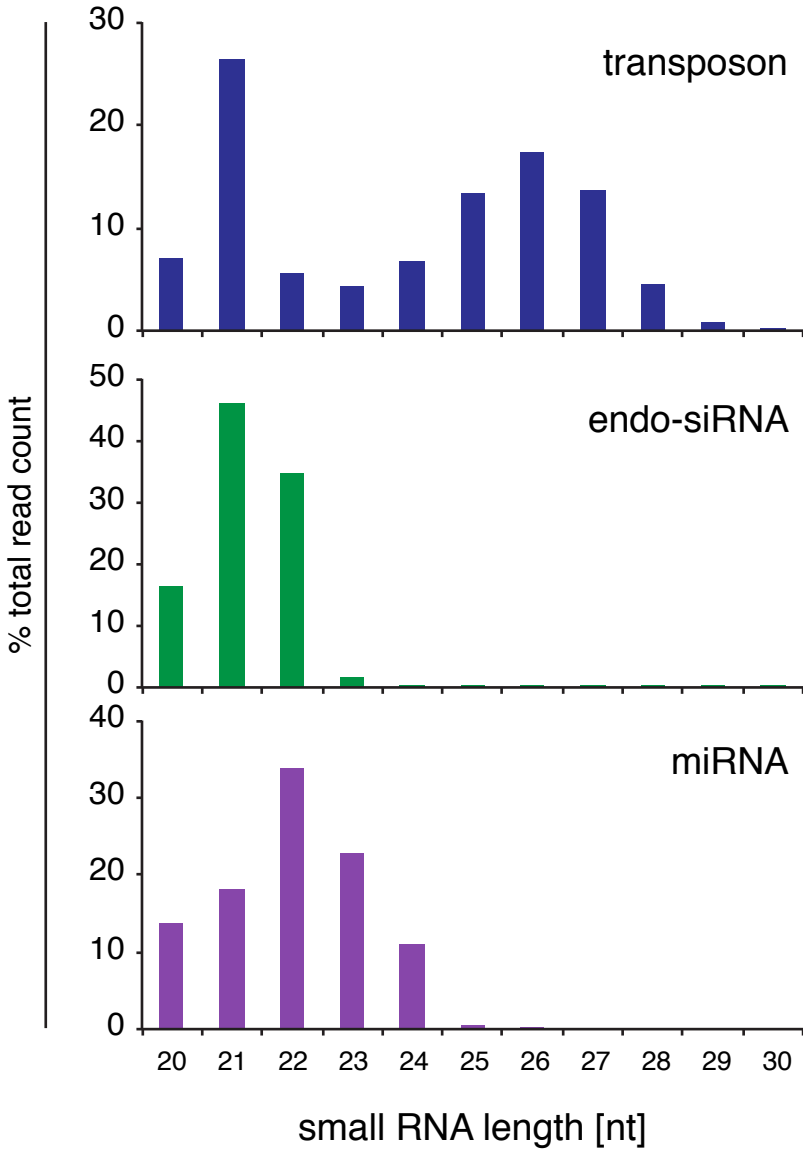


Figure 2.4: **Size distribution of several classes of sRNAs in OSS cells.** Size distributions of small RNA populations mapping to transposons, endo-siRNAs and miRNAs are shown.

We next observed the size distribution of sRNAs mapping to transposon consensus sequences. These sRNAs showed two peaks, one at 21nt, which corresponds to siRNAs and a broader peak around 26nt, characteristic of piRNAs (Figure 2.4). In ovaries, piRNA size profiles vary slightly depending on the Piwi protein they bind to, with Ago3 binding the smallest piRNAs (average of 24.1nt), followed by Aub (24.7nt) and Piwi (25.7nt) (Brennecke et al. 2007). In OSS cells we observe an average piRNA size of ~26nt, which agrees with Piwi-bound piRNAs being the only piRNAs present. Another trademark of Piwi bound piRNAs in ovaries is their bias towards being antisense to transposons. When we studied strand bias of the sRNA population mapping to transposons, we observed that while siRNAs (21nt peak) showed similar amounts of sense and antisense transcripts, piRNAs were predominantly antisense (Figure 2.5), similarly to ovaries. In conclusion, OSS cells contain populations of miRNAs, siRNAs and predominantly piRNAs, as would be expected.

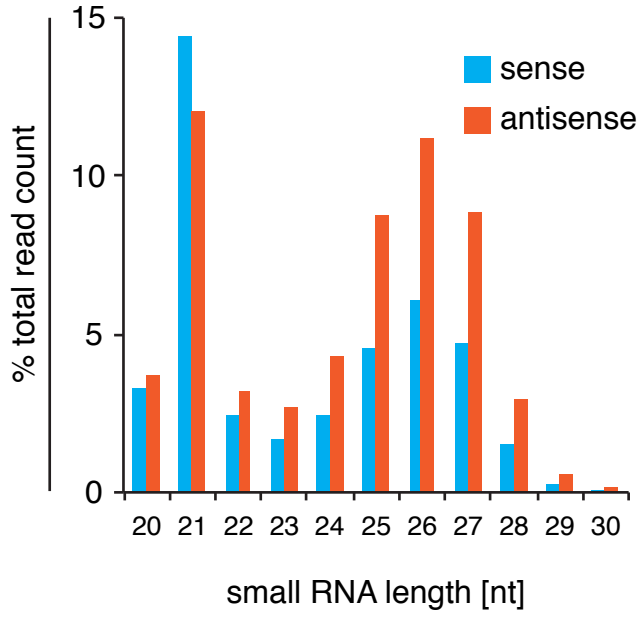


Figure 2.5: **Amount of sense and antisense sRNAs mapping to transposons.** Size profiles of small RNA populations mapping in sense and antisense direction to transposons are shown.

One important characteristic of piRNAs, particularly Piwi-bound piRNAs, is they have enrichment for uracil at the 5' end of the mature piRNA (1U) (Brennecke et al. 2007; Gunawardane

et al. 2007). When we studied nucleotide composition of piRNAs in OSS cells, we observed that more than 70% of all piRNAs have this 1U trademark (Figure 2.6). In germ cells, an active ping-pong cycle results in piRNA pairs with 10nt overlap, with the 1U causing a bias for adenine in the 10^{th} position (10A) of the secondary piRNA (Brennecke et al. 2007; Gunawardane et al. 2007). Since Aub and Ago3 are the proteins that mainly engage in ping-pong, and these two proteins are not expressed in OSS cells, we did not expect any ping-pong pairs. However, there is still the possibility that Piwi could ping-pong with itself. To study this possibility, we searched for complementary ping-pong pairs in the library, but found none. In fact, when we looked for presence of sense piRNAs mapping to the transposons, minimal amounts were detected. Based on this, we can conclude that piRNAs generated in OSS cells arise thorough primary piRNA biogenesis.

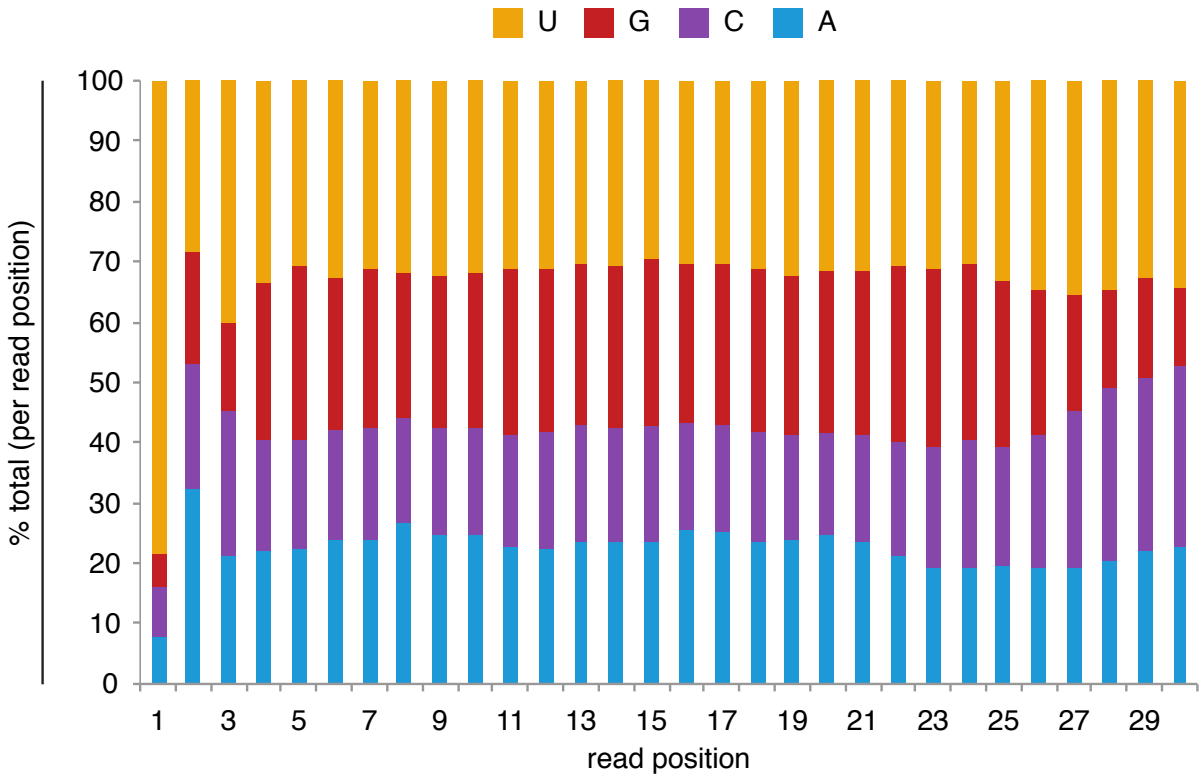


Figure 2.6: **Nucleotide distribution over piRNA sequences in OSS cells.** Percentage of nucleotide composition of piRNA reads over each position is shown.

Several studies have been published characterizing the somatic piRNA pathway in ovaries. Li et al., analyzed the phenotype of an *ago3* loss-of-function mutant where there was a collapse of germline piRNAs, and only somatic piRNAs were unaffected (Li et al. 2009). Malone et al. used a different approach, comparing piRNAs bound to Piwi proteins in ovaries versus 0-2 hour embryos (Malone et al. 2009). The assumption was that in 0-2 hour embryos, since zygotic transcription has not commenced, the populations of piRNAs are only those that nurse cells deposited into the oocyte. Therefore, piRNA populations that are not deposited would represent those present in the follicle cells of the ovary. Both these studies found that there was enrichment in the somatic compartment for piRNAs derived from the *flamenco* piRNA cluster, which specifically targets *gyspy* family retrotransposons. With this in mind, we set out to see if OSS piRNAs were enriched for this transposons class (Figure 2.7). Appreciably, when we list the top 20 highest expressed piRNAs, we see that all but two are members of the *gypsy* family.

Another important implication of the Li et al. and Malone et al. studies is that piRNA clusters are expressed in a cell-type specific fashion. Just as *flamenco* seems to be enriched in the soma, other clusters, such as *42AB*, are predominantly expressed in the germline. To study piRNA cluster expression in OSS cells, we measure the levels of piRNAs mapping to all 170 defined clusters (Brennecke et al. 2007). We found that, in fact the major piRNA producing cluster is *flamenco*, making up for 76% of all reads uniquely mapping to clusters (Figure 2.8). *42AB* mapping piRNAs, on the other hand, are barely detectable, with only 0.5% of all cluster reads. Other than *flamenco*, several piRNA clusters are expressed in OSS cells, albeit to a much lower level: *X-upstream*, *cluster 16* and *cluster 18* represent 4% to 7% of all cluster piRNAs (Figure 2.8). Given this data showing the similarity in protein and piRNA expression patterns between OSS cells and follicle cells, we conclude that OSS cells contain an active Piwi-associated primary piRNA pathway.

The arrival of a *Drosophila* cell line that recapitulates the piRNA pathway in follicle cells makes a number of experiments feasible that were technically impossible to accomplish in the past. We decided that a powerful way to use this cell line would be to perform an unbiased

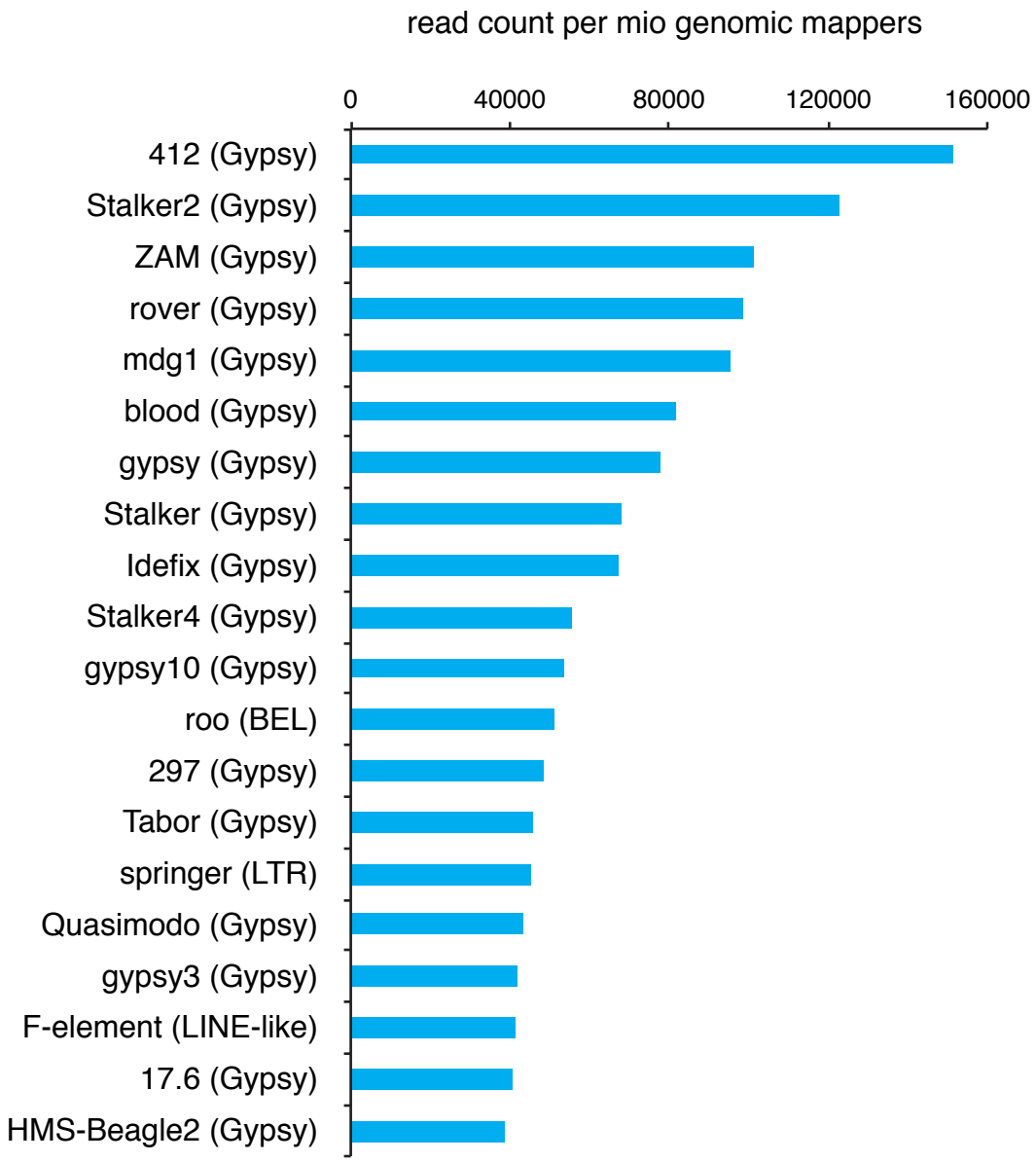


Figure 2.7: **Most abundant piRNAs in OSS cells correspond to *gypsy* family transposons.** Shown are the 20 most targeted transposons in OSS cells based on unique mapping counts to consensus sequences.

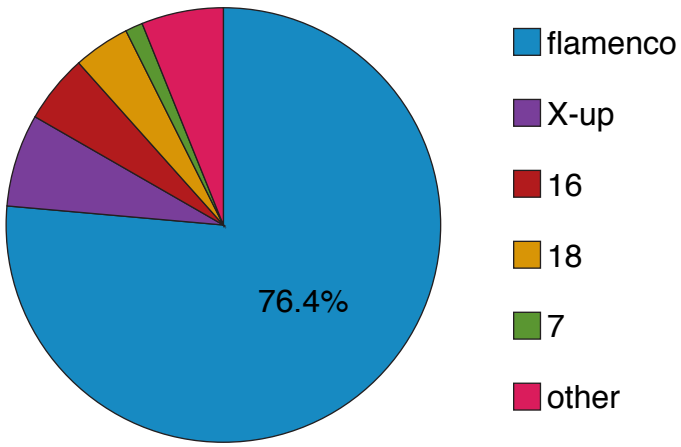


Figure 2.8: **piRNA clusters expressed in OSS cells.** The abundance of piRNA clusters is displayed as a percentage of all reads mapping uniquely to all piRNA clusters. The five most abundant clusters are shown, all other clusters are grouped together under 'other'.

genome-wide screen for novel piRNA components. There are still many poorly understood parts of the pathway, and by completing an unbiased screen we could discover a comprehensive list of proteins implicated in the pathway. The general screening scheme designed was to use a qPCR based readout to measure transposon levels upon knockdown of a gene. If the targeted gene were important in the piRNA pathway it would lead to elevated levels of transposons upon knockdown. Although this screening plan seemed relatively straightforward, there were several essential questions that had to be answered before beginning the screen. First of all, we had to establish if it was possible to deliver nucleic acids into the cells to be able to target genes for knockdown. We also had to determine if we would detect derepression of transposons upon disruption of a piRNA pathway component. Lastly, we had to develop a robust high-throughput screening assay. With this plan in place, we began to test transfection methods to deliver nucleic acids to these cells. Using a plasmid expressing EGFP driven by a ubiquitin promoter (Ub-GFP), we tested a wide variety of transfection reagents. Only two resulted in GFP expressing cells, the calcium phosphate transfection method and the Xfect transfection reagent. To assess the efficiency of transfection, cells were transfected with the GFP expressing plasmid using either transfection method, and analyzed using flow cytometry. Using calcium phosphate we obtain 67% GFP positive cells, while with Xfect transfection efficiency was higher

with almost 85% positive cells (Figure 2.9). Furthermore, Xfect had a lower cytotoxicity when compared to calcium phosphate and was more consistent regarding transfection efficiency. Therefore, we continued using Xfect for all further experiments.

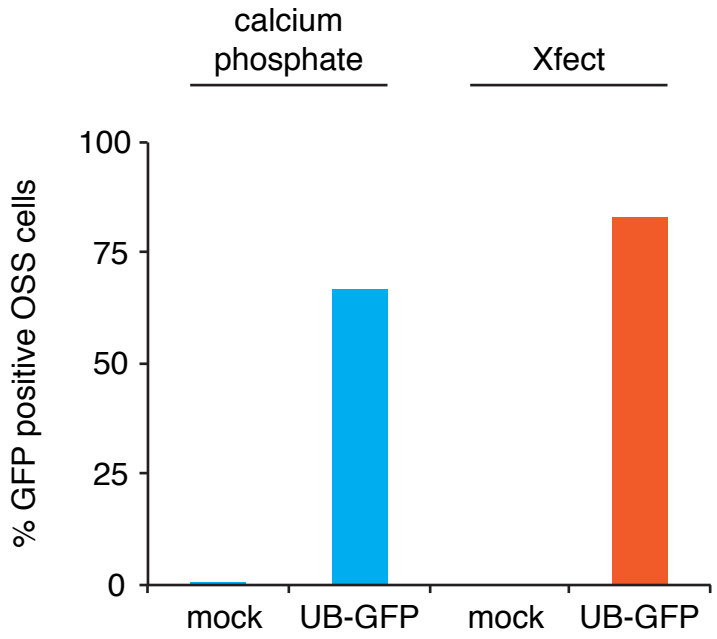


Figure 2.9: **Transfection efficiency in OSS cells.** Cells were transfected with a plasmid containing GFP under a ubiquitin promoter (UB-GFP) using either calcium phosphate or Xfect reagent. Cells were analyzed by flow cytometry 48 hours after transfection. Percentage of GFP positive cells in shown for each transfection method and corresponding mock transfection.

Next, we wanted to test if it was possible to efficiently deliver dsRNA and siRNAs into OSS cells to knock-down exogenous sequences. We co-transfected the Ub-GFP plasmid with dsRNA and siRNAs targeting GFP. Upon knockdown with dsRNA, the percentage of GFP positive cells decreased dramatically when compared to the control (Figure 2.10). Using siRNAs, there was a decrease in GFP positive cells, but it was not as dramatic as with dsRNA (Figure 2.10). This result was verified with two additional siRNAs, and similar results were obtained (data not shown). One possibility for the apparent inability of siRNAs to efficiently silence EGFP could be due to problems with siRNA delivery into cells. However, when fluorescently labeled siRNAs (siGLO) were transfected into cells, they could be visualized within the cell (data not shown). Since dsRNA worked so efficiently, we decided to continue experiments using these

instead of siRNAs. It is worth noting that after we conducted these experiments, another group successfully achieved knockdown with siRNAs using Amaxa Nucleofector technology, where siRNAs are delivered to the nucleus through electroporation (Saito et al. 2009). This method is very costly and has large cell viability problems; therefore, for our purposes Xfect continued to be the most suited method.

After establishing that it is possible to knockdown plasmid DNA using dsRNA, we set out to attempt knockdown of endogenous genes. We transfected cells with several different dsRNAs targeting Piwi and took timepoints after transfection to see if a decrease in protein and mRNA levels could be detected. When Piwi protein levels were measured upon knockdown, we saw that by the fourth day timepoint lower levels of Piwi compared to the control were detected (Figure 2.11). Timepoints from day five and six show similarly low levels of Piwi protein. Next, we measured Piwi mRNA levels in these samples by qPCR. On day four post-transfection, there was 10 to 12 fold less Piwi mRNA than in the control. By day six, transcript levels were already beginning to increase (Figure 2.12). To test if depletion of Piwi from these cells resulted in derepression of transposons, we measured levels of two retrotransposons: *mdg1* and *gypsy*, known targets of the follicle cell specific *flamenco* piRNA cluster. Four days after transfection there is no discernable upregulation for either transposon, however by day six post transfection *gypsy* transcripts are 4 to 10 fold upregulated and *mdg1* shows a more modest effect with 2 to 4 fold upregulation (Figure 2.12). In order to get a clearer picture of the global changes that occurred upon Piwi KD, we made RNA-seq libraries of OSS cells transfected with control dsRNA and a dsRNA targeting Piwi. These libraries showed that many transposable elements were upregulated in response to Piwi KD, and the most strongly derepressed were *gypsy* family retrotransposons (Figure 2.13).

The fact that we see these effects when transiently knocking down Piwi in this cell line is quite remarkable. It implies that constant surveillance by Piwi is necessary to maintain transposons silenced, and that it is not a mark that was established in the fly and that can be maintained throughout development without further targeting by Piwi.

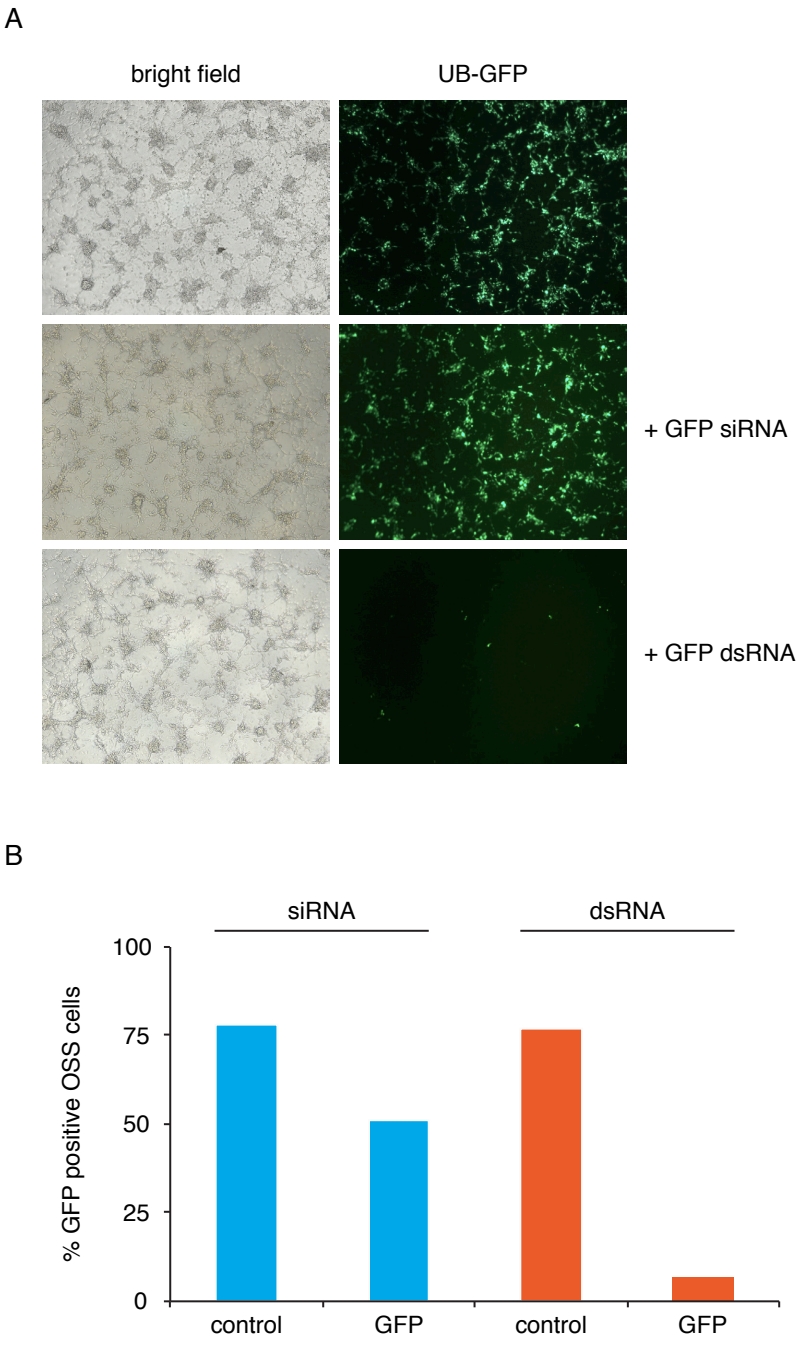


Figure 2.10: **Knockdown of GFP expression using dsRNA and siRNA in OSS cells.** A) Images of cells transfected with UB-GFP alone, and in the presence of either dsRNA or siRNA targeting GFP were taken 48 hours after transfection. B) Transfected cells were analyzed by flow cytometry and the percentage of GFP positive cells for each sample was assessed.

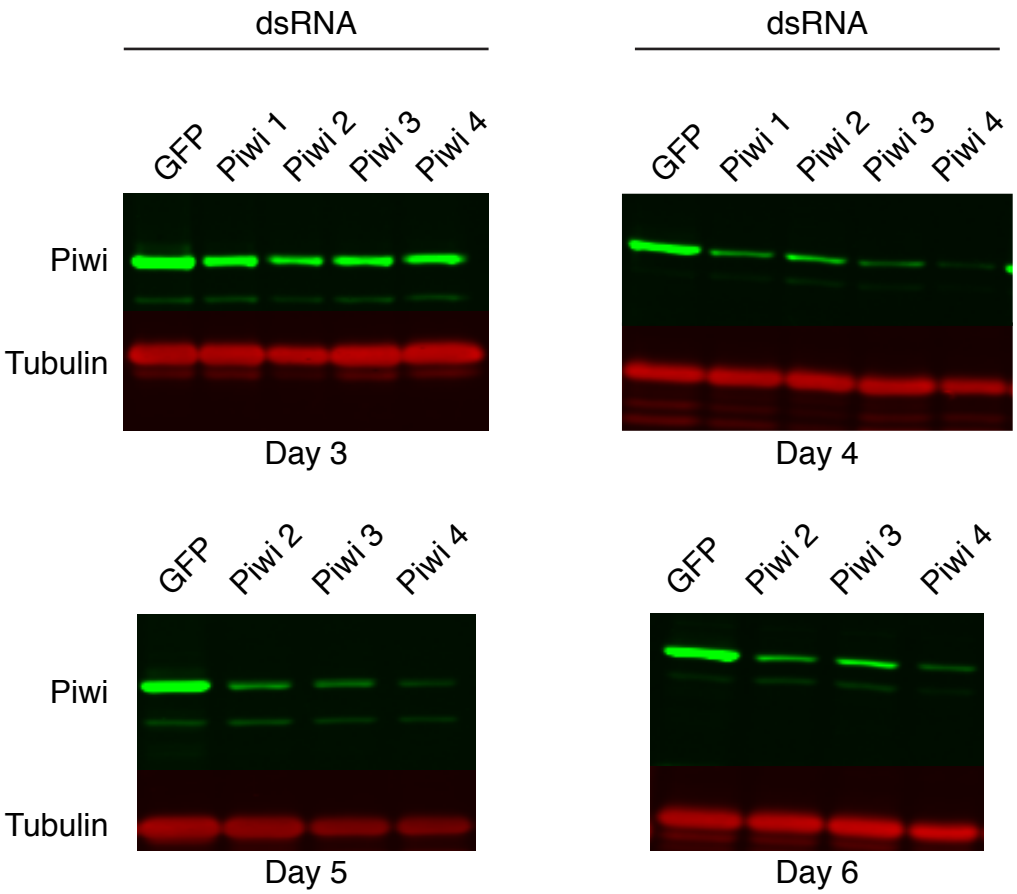


Figure 2.11: **Piwi protein levels upon Piwi knockdown in OSS cells.** Piwi protein levels were measured by western blot, using Piwi antibody. Westerns were performed on whole protein lysates from samples from day 3, 4, 5, and 6 days post-transfection with several Piwi dsRNAs (dsRNA 1-4). Tubulin was used as a loading control.

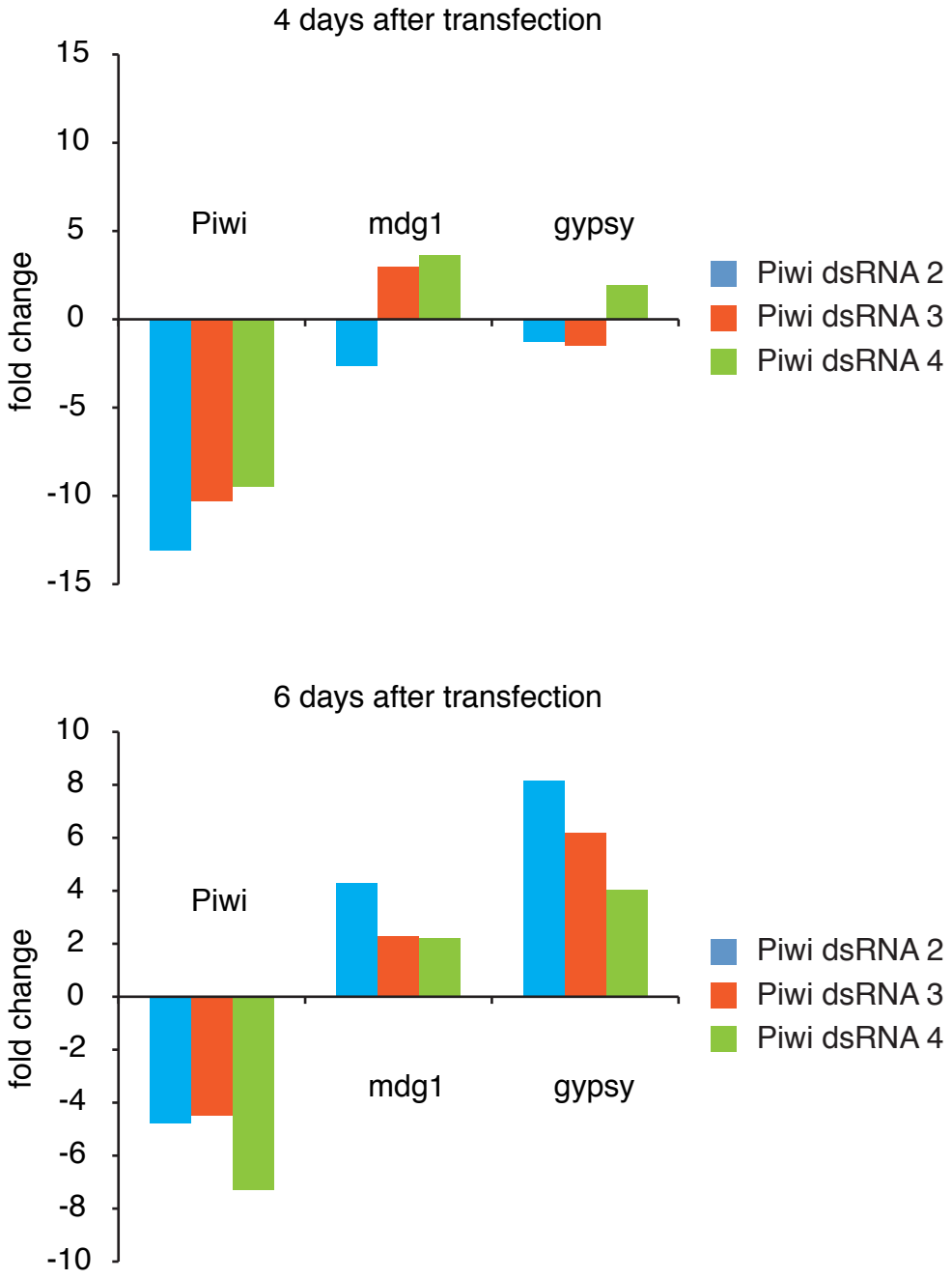


Figure 2.12: **Levels of Piwi mRNA and transposon transcripts upon Piwi knockdown.** Quantitative PCR (qPCR) was used to assess levels of Piwi, *mdg1* and *gypsy* transcripts at 4 and 6 day timepoints after transfection with several Piwi dsRNAs. Fold change was calculated using the delta delta Ct method, where each sample was normalized to house keeping gene, rp49, and compared to samples transfected with GFP dsRNA.

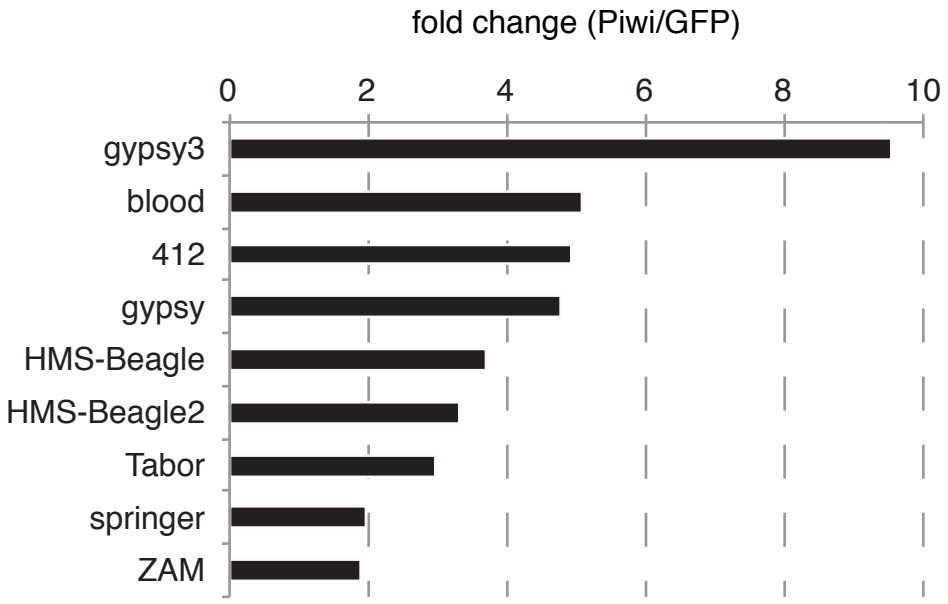


Figure 2.13: **Expression levels of transposons upon Piwi KD.** RNA-seq libraries were made from OSS cells transfected with either GFP or Piwi dsRNA, six days post-transfection. Libraries were normalized to number of total genomic mappers and reads were mapped to all transposon consensus sequences. Fold change was calculated by comparing number of reads mapping to transposons in Piwi versus GFP sample. Nine representative transposons are shown.

Although it was very encouraging to see these results, more optimization would be necessary for this assay to be robust enough to use in a genome-wide screen. First, it would be ideal to have a higher dynamic range for transposon derepression. It seems likely that the level of transposon upregulation observed upon Piwi KD is the upper limit of what can be detected since Piwi is the core protein of the piRNA pathway. With all the variation that can occur within a screening setting, a more sensitive assay would allow us to detect higher fold changes of derepression. Second, in order to perform qPCRs to measure transposon levels, we needed an efficient way to generate the input for the reverse transcription (RT) reaction in a high-throughput fashion. Using Trizol RNA extraction techniques, as well as other RNA extraction alternatives were very time-consuming and difficult to adapt to a high-throughput method. Therefore, we began to test using crude lysate as input into the reverse transcription reactions instead of pure RNA. The Cells-to-Ct reagent (Ambion) uses this method; cells are lysed directly in the tissue culture vessel, mixed by pipetting, and following addition of a

reagent to stop the lysis reaction, ready to be used as input in the RT reaction. This method was very simple and adaptable to a high-throughout assays. However, using a crude lysate resulted in a high genomic DNA contamination in the samples, which interfered greatly with accurately determining *gypsy* levels. When we lysed increasing numbers of cells and preformed qPCR for *gypsy* and a control gene, we observed that samples where no RT enzyme was added into the cDNA synthesis reaction, resulted in similar threshold cycle (Ct) values as samples were the enzyme was added (Figure 2.14). We attempted to make the DNase treatment more efficient by increasing DNase concentration and extending treatment time. However, this lead to inconsistent qPCR results, possibly due to inefficient inactivation of DNase contaminating the RT.

We began thinking of alternative approaches to not have a problem with genomic DNA contamination. Several different qPCR primers for *gypsy* had been tested and had given similar results; however, they all targeted *gypsy* ORF1. When searching the published literature on *gypsy*, we found that in addition to the full length *gypsy* transcript, there is also spliced subgenomic transcript generated from the *gypsy* loci (Figure 2.15A). *Gypsy* contains two LTRs flanking three ORFs in the center. This locus is transcribed into a 7kb transcript from which ORF1 and ORF2 are expressed. ORF3 is expressed from another smaller, 2kb transcript in which a 5kb intron containing most of ORF1 and ORF2 are spliced out (Pelisson et al. 1994). Several studies on the subgenomic transcript have shown that it is a sensitive measure of *gypsy* derepression (Pelisson et al. 1994). Presumably, by designing a qPCR assay specifically for the spliced transcript, with primers that flank the splice site, the amplification of genomic DNA contaminating the sample would be avoided. We designed primer pairs that spanned the splice site and verified *gypsy* subgenomic levels in mutant flies for *zuc* (Pane et al. 2007). A massive derepression of the subgenomic transcript in *zuc* mutants was observed when compared to sequencing strain, and a modest derepression in heterozygous *zuc* mutant flies (Figure 2.15B). These results were verified by running the amplification products of the qPCR by gel electrophoresis, and it was clear that there was no *gypsy* subgenomic amplification in

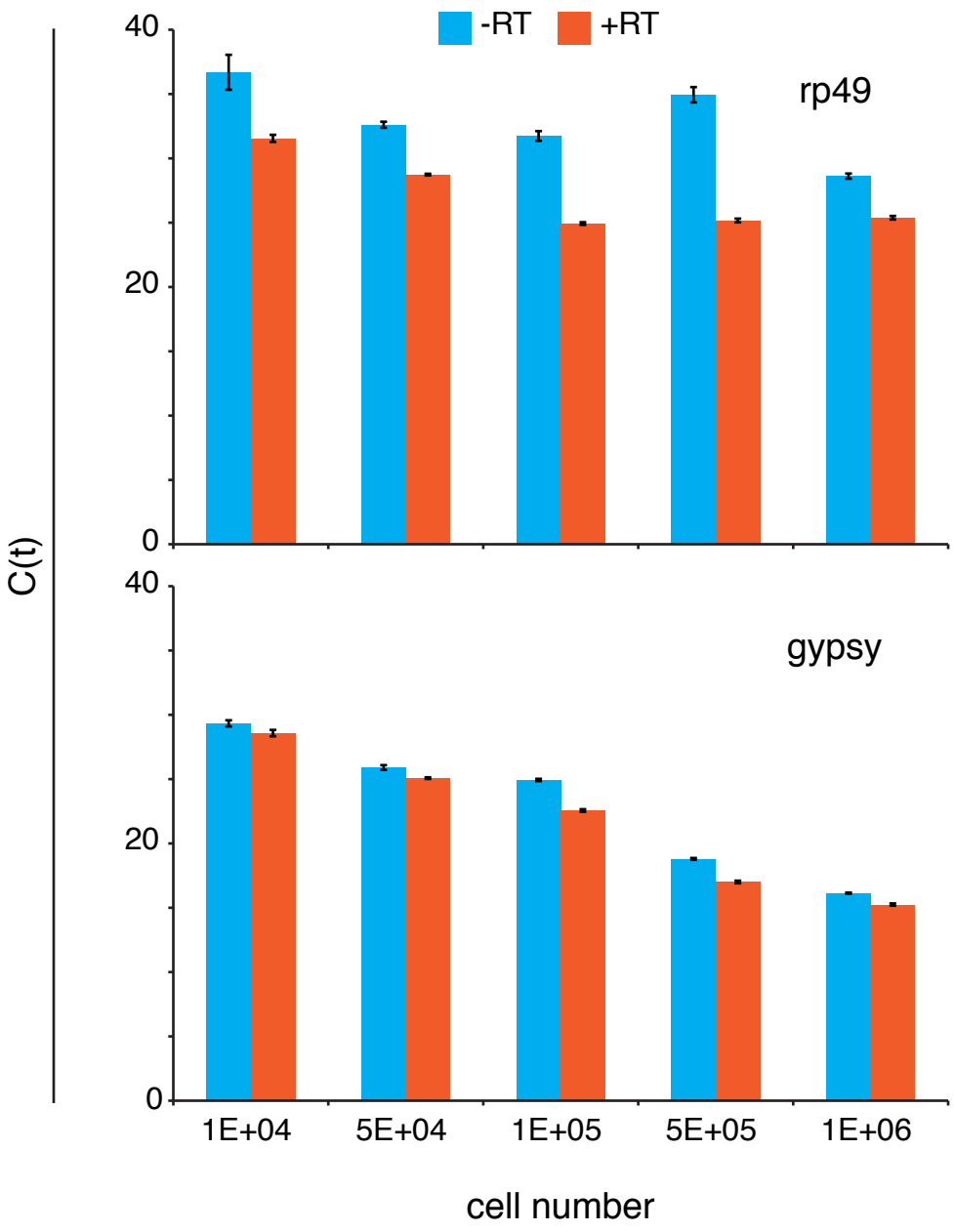


Figure 2.14: **Genomic DNA contamination interferes with qPCR measurement.** Increasing number of cells, specified on X-axis, were processed using Cells-to-Ct reagent and the levels of a reference gene, rp49, and the *gypsy* transposon were measured. The threshold cycle (Ct) for reactions with reverse transcriptase (+ RT) and without enzyme (-RT) are shown for each condition.

the sequencing strain, while the *zuc* mutant shows a strong band at the expected size (Figure 2.15C).

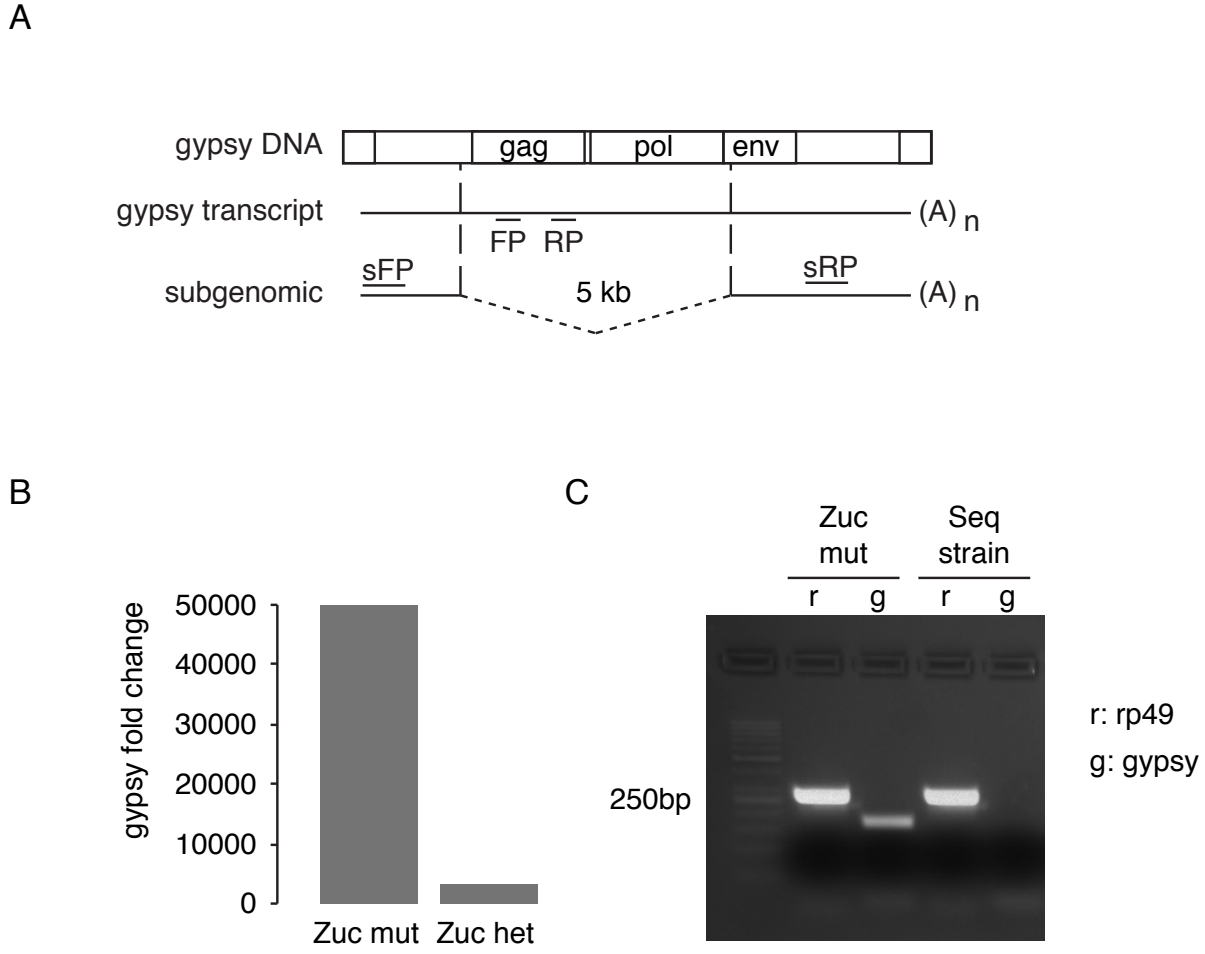


Figure 2.15: ***Gypsy* subgenomic transcript is a sensitive measure of piRNA pathway disruption in vivo.** A) Model of the structure and transcripts of the *gypsy* retrotransposon. Location of forward and reverse primers used in this study to measure levels of the large *gypsy* transcript (FP, RP) and the subgenomic transcript (sFP, sRP) is shown. B) Levels of *gypsy* subgenomic transcript measured by qPCR in zuc mutant flies and zuc heterozygous siblings. Fold change compared to sequencing strain was calculated using the delta delta Ct method. C) RT-PCR products showing levels of *gypsy* subgenomic transcript and housekeeping gene (*rp49*) in zuc mutant flies.

Knowing that the subgenomic transcript was derepressed upon disruption of the piRNA pathway *in vivo*, we wanted to see if a similar derepression could be observed in OSS cells. We measured levels of the subgenomic transcript four, five, and six days after transfection with Piwi dsRNA. In all these timepoints a marked increase in *gypsy* subgenomic transcript

was observed, with the highest derepression occurring five days post transfection, with *gypsy* upregulated by almost 70 fold (Figure 2.16A). We also wanted to verify that knockdown of other piRNA pathway components would lead to similar results. Upon knockdown of Armi or Zuc, there was a very strong derepression of this *gypsy* transcript, even stronger than had been measured previously for *gypsy* ORF1 (Figure 2.16B). Based on these findings, the *gypsy* subgenomic transcript seemed to be a good measure for disruption of the piRNA pathway. However, all these experiments were performed on extracted RNA, where the interference with genomic DNA is negligible. For the subgenomic transcript to be useful for a screening assay, it had to be insensitive to genomic DNA contamination. When Piwi KD experiments were performed in OSS cells using Cell-to-Ct as an extraction method, we observed that the minus RT sample had no signal, while the sample with RT detected a product (Figure 2.17). This indicates that *gypsy* subgenomic transcript is a sensitive measure of piRNA pathway disruption, but it is not affected by presence of genomic DNA contamination, making it an ideal read-out for our screening assay.

It was also necessary to measure levels of a control gene in order to normalize each individual sample for RNA input amount, as well as to have a measurement for cell viability. In all previous experiments we had used *rp49*, a ribosomal protein. However this gene proved not to be a suitable control, since its levels were skewed by presence of genomic DNA, due to the presence of only a small 25bp intron. Due to our high genomic DNA background, we needed a gene that had a large intron to avoid background signal. One of the highest expressed genes in the OSS cell line is *ciboulot*, an actin binding protein, contained a large 3kb intron. This gene was highly expressed, and its levels did not change upon Piwi knockdown (Figure 2.17).

We decided to repeat this experiment in a smaller well format, which would be more practical for a screening assay. However, when knockdown experiments were performed in 96-well plates, and *gypsy* subgenomic levels were measure using Cells-to Ct lysate, the level of derepression was very low and there was a great deal of variation (4-10 fold) (Figure 2.18). It seemed as though the level of derepression was inversely proportional to the growth area of the

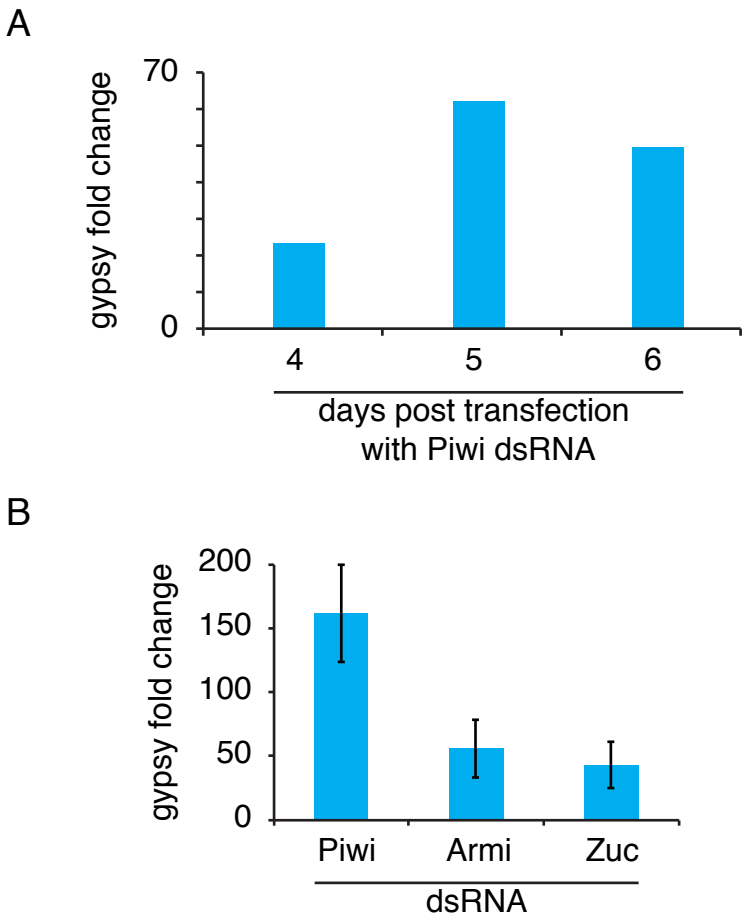


Figure 2.16: **Levels of *gypsy* subgenomic transcript upon impairment of the piRNA pathway in OSS.** A) Levels of *gypsy* subgenomic transcript 4, 5 and 6 days post transfection with Piwi dsRNA are shown. B) *Gypsy* subgenomic levels upon knockdown of Piwi, Armi and Zuc. Cells were collected five days post-transfection. Error bars represent one standard deviation across three biological replicates. For A) and B) fold changes were calculated using the delta delta Ct method and compared to cells transfected with GFP dsRNA.

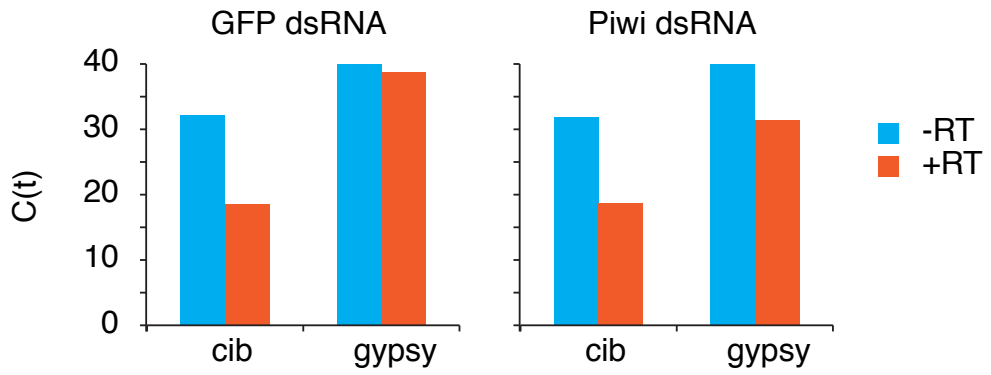


Figure 2.17: **Measurement of levels of *gypsy* subgenomic and ciboulot transcripts is not affected by presence of genomic DNA contamination.** Cells transfected with Piwi or GFP dsRNA were lysed using Cells-to-Ct reagent five days after transfection and levels of cib and *gypsy* were measured by qPCR. Ct values for +RT and –RT samples are shown.

cell culture vessel. To test this specifically we performed the same experiments using different size tissue culture plates (6-well to 96-well) and found that, indeed, levels of derepression decrease as the well size decreases (Figure 2.18). We concluded that using a 48-well format, where we saw 50 fold upregulation of *gypsy* subgenomic transcript was robust enough for a screening assay, yet small enough to be managed in a high-throughput fashion.

There were also several options of how to perform the qPCR and which reagents to use. We compared several methods and observed that most reproducible, consistent, and fast procedure was performing both the RT and qPCR in 96-well plates, followed by multiplexed Taq-man qPCR using *cib* and *gyspy* hydrolysis probes. In conclusion, we successfully developed a sensitive and robust screening assay. In this assay, cells plated on 48-well plates were transfected with dsRNAs using Xfect reagent and after five days lysed using Cells-to-Ct reagent followed by qPCR for the *gypsy* subgenomic transcript.

In high-through put methods, such as a genome-wide screen, the use of automation methods helps improve consistency throughout the process. For this reason, we decided to use the Epmotion robot and the Biomek robot. The Epmotion would aid us in the transfection step, where a 96-well plate of dsRNAs was transfected into two 48-well plates and to lyse the cells 5 days after transfection and transfer lysates back into a 96-well plate. The Biomek robot

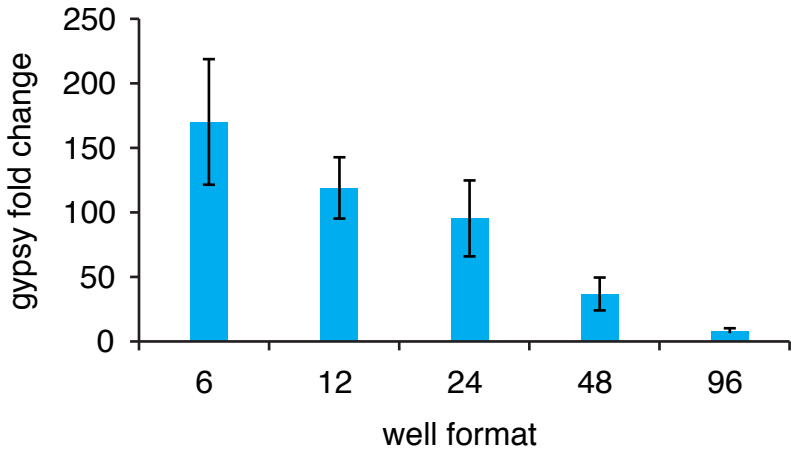


Figure 2.18: **Tissue culture well size affects level of transposon derepression upon Piwi KD.** Cells were plated on several tissue culture well formats and transfected with Piwi dsRNA. Cells were lysed five days after transfection using Cells-to-Ct reagent. Fold change of *gypsy* subgenomic transcript was calculated using the delta delta Ct method. Error bars represent one standard deviation between two biological replicates.

was only used to transfer the input cDNA from the 96-well RT plate to the 96-well qPCR plate. Using these robots, we developed a schedule that would allow us to screen 5,000 genes in a 10 day period.

2.3 Experimental Procedures

OSS cell culture and maintenance

OSS cells were grown at 25ºC in Shields and Sang M3 media (Sigma) supplemented with 10% FBS, 5% fly extract, 0.6 mg/ml glutathione and 10 mg/ml insulin. For fly extract, 1 to 3 day old OregonR flies were collected and flash frozen in liquid nitrogen. Frozen flies were homogenized using a Retsch Mixer Mill 400. Grinding jars and all other tools used were cooled in liquid nitrogen prior to use. After grinding, fly powder was weighed and resuspended in 6.8ml of M3 medium per gram of flies. A blender was used to aid resuspension. Mixture was aliquoted in 50ml conical tubes and centrifuged at 4ºC at 1,500Xg for 15 minutes. Supernatant was filtered through Miracloth, transferred to a fresh conical tube and incubated at 60ºC for 30 minutes. After heat inactivation, tubes are centrifuged at 4ºC at 4,000Xg for 90 minutes.

After the spin, supernatant is filtered using 0.8uM syringe filters, aliquoted and stored at -80ºC until needed. To passage cells, media was aspirated from flask and cells were washed with PBS. TrypLE 1X was pipetted gently over cells and aspirated off after immediately. After five minutes, media was added to the flask to detach cells. Cells were then split to the desired confluence in fresh media, but not below 1:5.

In-vitro transcription of dsRNA for knockdown

To generate dsRNA, a ~500bp region of the gene of interest was PCR amplified. The primers used for amplification also contained a T7 RNA polymerase at their 5' end, for bi-directional transcription of the PCR product. PCR reactions were carried out using standard Taq DNA polymerase. PCR products were examined on an agarose gel to verify size of the fragment and purified using QUIquick PCR purification kit (Quigen). In vitro transcription was carried out with the MEGAscript T7 kit (Ambion) according to the manufacturers protocol and 200ng of PCR product was used as template. All primer sequences used for *in-vitro* transcribing dsRNAs are shown in Table 2.1.

Table 2.1: Table of oligos

Name	Primer sequence
dsRNA Piwi 1 fwd	TAATACGACTCACTATAGGGCGCTGCAGACGAACTTTTTCC
dsRNA Piwi 1 rev	TAATACGACTCACTATAGGGCACCCGTACTTCGTCCTGATG
dsRNA Piwi 2 fwd	TAATACGACTCACTATAGGGCGCTGCAGACGAACTTTTTCC
dsRNA Piwi 2 rev	TAATACGACTCACTATAGGGCACCCGTACTTCGTCCTGATG
dsRNA Piwi 3 fwd	TAATACGACTCACTATAGGGCCATCAGGACGAAGTACGGGT
dsRNA Piwi 3 rev	TAATACGACTCACTATAGGGCGTCTCTGAAGTGCCTTTGCC
dsRNA Piwi 4 fwd	TAATACGACTCACTATAGGGCGGCAAAGGCACTTCAGAGAC
dsRNA Piwi 4 rev	TAATACGACTCACTATAGGGCATTGAGGCAATCAATGCTCC
dsRNA Armi fwd	TAATACGACTCACTATAGGGCGCTATCTCAGCGAAACCGAC
dsRNA Armi rev	TAATACGACTCACTATAGGGCGTCGGCTTTTCGTTCTTCAG

Table 2.1: Table of oligos (continued)

Name	Primer sequence
dsRNA Zuc fwd	TAATACGACTCACTATAGGGAGGTGATTTGGAAGCTGGTG
dsRNA Zuc rev	TAATACGACTCACTATAGGGGTTGTGCATCAAGTTCGTGG
dsRNA GFP fwd	TAATACGACTCACTATAGGGCGCAAGCTGACCCTGAAGTT
dsRNA GFP rev	TAATACGACTCACTATAGGGCGGACTGGGTGCTCAGGTAG
Piwi fwd	TCGTACCCAATGATAACGCCGAAAG
Piwi rev	AGTCCGGACAAGGGTAGTTCGATCA
Armi fwd	CGGTTCTATGCCAACCAAGT
Armi rev	TGAATAGTGCTGCTCGATGG
Zuc fwd	GGCACTCCTTTGTGTGTGAA
Zuc rev	GGCACTCCTTTGTGTGTGAA
rp49 fwd	ATGACCATCCGCCCAGCATAC
rp49 rev	CTGCATGAGCAGGACCTCCAG
mdg1 fwd	AACAGAAACGCCAGCAACAGC
mdg1 rev	CGTTCCCATGTCCGTTGTGAT
gypsy ORF1 fwd	AGGCAAGGATTGGAAATGGTTAGGC
gypsy ORF1 rev	CCTTTTTGAGCCCCGAAATAAAAGC
gypsy subgenomic fwd	AGTACCCGCCACAACCTTTAAG
gypsy subgenomic rev	AGTACCCGCCACAACCTTTAAG
gypsy subgenomic probe	CAAACAGGGTAGTTAAGTTAG
ciboulot fwd	GCCAGCATCCCAGCTTAGTAGT
ciboulot rev	GCTGGGGCGGCCATCTT
ciboulot probe	CGCTTCGCCAATCCA

siRNA preparation

For siRNAs, single stranded, complementary RNA oligos were designed using the Design of small interfering RNAs (DSIR) online tool and ordered from IDT. Oligos were resusupended in Duplex Buffer (100 mM Potassium Acetate; 30 mM HEPES, pH 7.5; available from IDT) to a final concentration of 100uM. Both strands were mixed together in equal molar amounts and annealed by heating the oligo mixture to 94ºC, removing it from heating source and allowed to cool to room temperature on the bench-top. Annealed siRNAs were stored at -20ºC.

DNA plasmids and purification

For transfections, Ubiquitin-EGFP plasmid (Astrid Haase) was extracted from *E.Coli* cells using the EndoFree Plasmid Midi Kit (Quiagen), which yields a high amount of pure, endotoxin free plasmid DNA.

Transfection methods

The day prior to transfection, OSS cells were plated at 25% confluence. Numbers of cells that were plated according to tissue culture vessel size, as well as the amount of plasmid DNA or dsRNA that was transfected are shown in Table 2.2. For the Calcium Phosphate transfection method the Calcium Phosphate Transfection Kit (Invitrogen) was used following manufacturer's guidelines. In brief, one tube containing the nucleic acid, and $CaCl_2$ is added dropwise to another tube containing HEPES Buffered Saline, while bubbling air through solution B with another pipette. This mixture was incubated at room temperature for 30 minutes and then added dropwise to the media of the cells that were being transfected. Media was changed 12 hours after transfection. These transfections were carried out in 6-well dishes and 5μg of plasmid DNA was transfected. Xfect transfections were conducted as recommended by the manufacturer. Briefly, Xfect polymer diluted in dilution buffer was added to a tube containing the nucleic acid. For every 1 μg of nucleic acid, 0.3μl of Xfect polymer was used. This mixture

was vortexed for 10 seconds, followed by a 10 minute incubation at room temperature and the added dropwise to the media of cells. Media was changed 12 hours after transfection.

Table 2.2: Table of transfection conditions.

vessel	area (cm2)	volume (ml)	# of cells to plate	plasmid DNA (ug)	dsRNA (ug)
T75	75	15	3.8E+00	80	40
10cm	56.75	10	2.5E+00	60	30
6 well	9.50	2.50	6.3E-01	10	5
12 well	3.80	1.00	2.5E-01	4	2
24 well	1.90	0.50	1.3E-01	2	1
48 well	0.95	0.25	6.3E-02	1	0.5
96 well	0.32	0.10	2.5E-02	0.35	0.15

FACS sorting

48 hours after transfection with UB-GFP plasmid, cells were washed twice with PBS. Then, cells were mechanically detached from the tissue culture vessel by pipetting and filtered through a mesh filter to break up clumps of cells. Cells were then analyzed using the LSR II Flow Cytometer (BD Biosciences) and after 10,000 events the percentage of GFP positive cells was determined.

RNA extraction and Reverse Transcription

For RNA extractions, cells were washed once with PBS and lysed in 1 mL TRIzol reagent. For ovary tissue, ovaries were dissected in PBS and homogenized in 200µl of TRIzol using a pestle, followed by the addition of 800µl of TRIzol. Lysate was transferred to a microtube and incubated for 5 minutes at RT. Following incubation, 200µl of chloroform was added to the TRIzol and sample was vortexed for 15 seconds and incubated for 3 minutes at RT before centrifugation at 4ºC at 12,000Xg for 15 minutes. The upper aqueous phase (~450µl) was transferred to a fresh tube. RNA was precipitated by adding an equal volume of isopropanol. Samples were incubated for 10 minutes at RT and the centrifuged for 10 minutes at 4ºC at a speed of 12,000Xg. RNA pellet was washed once with 70% ethanol and then air-dried for 5

minutes. RNA pellet was then resuspended in nuclease free water. 1ug of total RNA was DNase treated for 30 minutes at 37ºC using the Turbo DNA-free kit (Ambion) prior to reverse transcription to eliminate all traces of genomic DNA. cDNA was then synthesized using the 1ug of DNase treated RNA as input using Superscript III Reverse Transcriptase (Life Technologies) and a mixture of oligo dT primers (dT20) and random hexamers. cDNA synthesis reaction was incubated at 50ºC for 50 minutes followed by 15 minutes at 70ºC.

Quantitative PCR and analysis

For SYBR Green chemistry qPCRs, synthesized 2µl of a 1:5 dilution of the cDNA was mixed with 2X Power SYBR Green Mastermix (Life Technologies) and primers to a final concentration of 500nM for the desired target. For Taq-Man chemistry 2µl of a 1:5 dilution of the cDNA was used as input in a multiplexed reaction measuring levels of *ciboulot* and *gypsy* subgenomic transcript using Taq-Man Fast Advanced Mastermix (Life technologies). All primers and probes used for qPCRs are found listed in Table 3. Changes in gene expression were calculated using the delta delta Ct method (Livak and Schmittgen 2001), where a control gene (rp49 or cib) was used to normalize the values of the other samples (delta Ct). Normalized samples were then compared to the control group, cells treated with GFP dsRNA, to generate the delta delta Ct measurement.

Western Blotting

To confirm decrease in Piwi protein levels upon knockdown, cells were lysed using RIPA Buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 14mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% DOC, protease inhibitors) and lysate was cleared by centrifugation at 4ºC. Protein concentration was determined using the Bradford Protein Assay (Bio-Rad). SDS-PAGE gel was loaded with equal amounts of protein for each sample. The gel was then transferred to a nitrocellulose membrane using a Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad). The membrane was blocked in Odyssey Blocking Buffer for 1 hour at RT, followed by an

overnight incubation with primary antibodies at 4ºC. Piwi antibody (Brennecke, 2007) was used at a 1:1,000 dilution and alpha-Tubulin antibody (Sigma) was used at a 1:10,000 dilution. Odyssey Imaging System secondary antibodies (LI-COR) were used: Goat anti-rabbit IRDye 680 for Piwi and goat anti-mouse IRDye800CW for Tubulin. All antibody dilutions were made in Odyssey blocking solution. Signal intensities were detected using the Odyssey Imaging System (LI-COR).

Immunofluorescence

To detect Piwi protein localization, OSS cells were plated on chamber slides (Lab-Tek). After cells had attached, they were washed with PBS and fixed in 2% paraformaldehyde for 15 minutes without rocking. Cells were then gently washed with PBS twice and permeabilized for 10 minutes in PBS + 0.2% TritonX-100. After two washes with PBS + 0.1% Tween-20 (PBS-T) cells were blocked in PBS + 2%BSA for one hour, and incubated with anti-Piwi antibody (1:500 in PBS + 2%BSA) overnight at 4°C. After three washes with PBS-T, cells were incubated with secondary goat anti rabbit Alexa Fluor 488 goat anti-rabbit IgG (Invitrogen) (1:400) for one hour, followed by three washes with PBST. In the last wash, DAPI (1:1000) was added to stain nuclei. Cells were visualized under a fluorescence microscope.

Fly strains

All fly stocks were kept at 25ºC. OregonR flies (Bloomingtom Stock Center), a laboratory wild-type strain was used for fly extract preparation. For analysis of the *gypsy* subgenomic transcript upon impairment of the piRNA mutants, for zucchini mutants transheterozygotes zuc^{HM27}/ Df(2I)PRL, where zuc^{HM27} contains a stop codon at residue 5 were used (Pane et al. 2007). Heterozygous flies zuc^{HM27}/+ were used as a control. Both strains were compared to the *Drosophila* sequencing strain, the isogenized y^1; cn^1 bw^1 sp^1 strain (Adams et al. 2000).

RNA-seq library preparation

For transcriptome libraries, 10 ug of total RNA from OSS cells transfected with GFP control dsRNA or Piwi dsRNA was used as input for the Illumina mRNA-Seq sample prep kit (catalog no. RS-930-1001). Libraries were made following the instructions provided by the manufacturer. In brief, poly-A containing mRNA was purified using magnetic beads with an attached poly-T oligo. After purification, mRNA is fragmented and these RNA fragments are copied into first strand cDNA using reverse transcriptase and random hexamers. The second cDNA strand is the synthesized using DNA Polymerase I and RNaseH. The cDNA fragments are then end-repaired, adenylated at the 3' end and adaptors are ligated to the 5' and 3' end of the fragments. Ligation products are purified by gel-electrophoresis and enriched by PCR resulting in the final cDNA library. Libraries were sequenced on the Illumina Genome Analyzer II platform for 76 cycles.

Bioinformatic analysis of RNA-seq

After collapsing all reads into a non-redundant list, they were mapped to *Drosophila* viral, tRNA and miscRNA (rRNA, snoRNA etc) sequences using the short read aligner bowtie (Langmead et al. 2009). Only sequences that did not map to either of these contaminants were then mapped to the D. melanogaster genome (D. melanogaster Apr. 2006 [BDGP R5/dm3]) with up to two mismatches. Additionally, only uniquely mapping sequences were considered for further analysis. The same reads were mapped to a custom index of transposon consensus sequences with up to 2 mismatches (http://www.fruitfly.org/p_disrupt/TE.html). Reads mapping to up to 2 locations were considered for further analysis. For expression analysis of transposons we aggregated read counts mapping to these consensus sequence in sense orientation. For expression analysis of genes, we used htseq-counts (Part of the 'HTSeq' framework, version 0.5.3p3) to assess rpkm values.

sRNA library preparation

Small RNA libraries were generated as described in Malone et al, 2012 with some changes. (Malone et al. 2012) In brief, total RNA from OSS cells extracted as described above, was size selected on a denaturing polyacrylamide gel. Size selection was aided by the presence of radioactively labeled 19nt and 28nt RNA oligonucleotide cloning markers. RNA was eluted from gel slice by electro-elution using D-tube Dialyzers (Millipore). After elution, the adenylated Modban Linker-1 adaptor (IDT) was ligated to the 3′ end of the sRNA. For this ligation step, T4 RNA Ligase 2 truncated (NEB) and a buffer not containing ATP was used. Ligation reaction was examined on a polyacrylamide gel and only fragments of the expected ligated size were cut out and electro-eluted. The Solexa 5′ adapter oligo was ligated to the eluted RNA using T4 RNA Ligase. After another gel extraction of ligated products, RNA was eluted from the gel, and RNA fragments with both adaptors ligated were reverse transcribed. Resulting cDNA was the amplified by PCR followed by Pme1 digestion to eliminate all RNA oligonucleotide cloning markers. Library was single end sequenced on an Illumina Genome Analyzer II for 36 cycles.

Bioinformatic analysis of small RNAs

After FASTQ to FASTA conversion, the Illumina dapter (CTGTAGGCACCATCAATTC) was clipped from the 3′ end of the read and sequences shorter than 15 nt were discarded from further analysis. The remaining sequences were collapsed into a nonredundant list and mapped to the D. melanogaster genome (D. melanogaster Apr. 2006 [BDGP R5/dm3]) allowing only exact matches. Annotations were extracted from UCSC Genome Browser (http://genome.ucsc.edu, Meyer et al. 2013), Flybase (http://flybase.org, McQuilton et al. 2012) and miRBase (http://www.mirbase.org, Griffiths-Jones et al. 2006). A collection of commonly used cloning oligonucleotides served as an annotation database for synthetic sequences. For transposon abundance analysis, reads were mapped to transposon consensus sequences (http://www.fruitfly.org/p_disrupt/TE.html) using the short read aligner bowtie (Langmead et al. 2009). Up to two mis-

matches and two mapping locations were allowed. Read counts were normalized to reads per million genomic mappers.

2.4 References

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, et al. 2000. The genome sequence of Drosophila melanogaster. *Science* **287**: 2185-2195.

Bastock R, St Johnston D. 2008. Drosophila oogenesis. *Curr Biol* **18**: R1082-7.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell* **128**: 1089-1103.

Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**: D140-4.

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

Haase AD, Fenoglio S, Muerdter F, Guzzardo PM, Czech B, Pappin DJ, Chen C, Gordon A, Hannon GJ. 2010. Probing the initiation and effector phases of the somatic piRNA pathway in Drosophila. *Genes & Development* **24**: 2499-2504.

Kim VN, Han J, Siomi MC. 2009. Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol* **10**: 126-139.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, Seitz H, Horwich MD, Syrzycka M, Honda BM, et al. 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137**: 509-521.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402-408.

Malone C, Brennecke J, Czech B, Aravin A, Hannon GJ. 2012. Preparation of Small RNA Libraries for High-Throughput Sequencing. *Cold Spring Harbor Protocols* **2012**: pdb.prot071431-pdb.prot071431.

Malone CD, Brennecke J, Dus M, Stark A, Mccombie WR, Sachidanandam R, Hannon GJ. 2009. Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. *Cell* **137**: 522-535.

McQuilton P, St Pierre SE, Thurmond J, FlyBase Consortium. 2012. FlyBase 101–the basics of navigating FlyBase. *Nucleic Acids Res* **40**: D706-14.

Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, Sloan CA, Rosenbloom KR, Roe G, Rhead B, et al. 2013. The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res* **41**: D64-9.

Niki Y, Yamaguchi T, Mahowald AP. 2006. Establishment of stable cell lines of Drosophila germ-line stem cells. *Proc Natl Acad Sci USA* **103**: 16325-16330.

Olivieri D, Sykora MM, 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.

Pane A, Wehr K, Schüpbach T. 2007. zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. *Dev Cell* **12**: 851-862.

Pelisson A, Song SU, Prud'homme N, Smith PA, Bucheton A, Corces VG. 1994. Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. *EMBO J* **13**: 4401-4411.

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

Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, Siomi H, Siomi MC. 2010. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. *Genes & Development* **24**: 2493-2498.

3 A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in *Drosophila*

Chapter contributions:

The genome-wide screen, both primary and validation, was done equally by Felix Muerdter and myself. Felix Muerdter wrote the scripts for analysis of the primary screen dataset. Jesse Gillis conducted statistical analysis for GO term enrichment and gene network analysis of the dataset. For follow-up experiments, I was responsible for construction of all RNA-seq libraries. Felix Muerdter and I both constructed small RNA libraries. I was responsible for the ChIP-seq libraries of the *asterix* mutant. Felix Muerdter was responsible for all bioinformatical analysis on RNA-seq, sRNA and ChIP seq libraries. Yang Yu and Yicheng Luo performed OSS immunofluorescence experiment. Felix Muerdter and I contributed equally to the writing of the manuscript, on which we are co-first authors. The manuscript is being prepared for submission to *Cell*.

Manuscript citation:

Muerdter, F. *, Guzzardo, P. M. *, Gillis, J., Luo, Y., Yang, Y. & Hannon, G. J. A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in Drosophila. *In preparation*. *authors contributed equally

3.1 Summary

A large fraction of our inheritable genome consists of mobile genetic elements. Governing these transposable elements is of utmost importance and failure to do so can compromise genome integrity and ultimately lead to sterility. Similar to an adaptive immune system, Piwi proteins together with their bound piRNAs are the key to proper transposon constraint, yet their precise molecular means are poorly understood. In an effort to identify general requirements for transposon control and novel components of the piRNA pathway, we undertook a genome-wide RNAi screen in *Drosophila* ovarian somatic sheet cells. We identify and validate 87 genes required for transposon silencing. Among these genes we uncover several novel piRNA biogenesis factors. We also identify CG3893 (Asterix) as being essential for transposon silencing, most likely at the transcriptional level. Loss of its function leads to drops in H3K9me3 silencing marks on certain transposons, but has no effect on piRNA levels.

3.2 Introduction

Transposable elements populate virtually every eukaryotic genome. Although their presence imparts many beneficial effects, when out of control, they can compromise the genomic integrity of their host and its offspring (Levin and Moran 2011). Mobilization of transposons can lead to double stranded DNA breaks and deleterious mutations (McClintock 1950). Hence, in all higher animals there are control mechanisms in place, which prevent transposons from mobilizing (Malone and Hannon 2009).

In *Drosophila*, the main pathway that protects the inheritable genome is comprised of a catalogue of small RNAs that interact with a subclade of Argonaute family proteins, the Piwi proteins (Ishizu et al. 2012). This pool of Piwi-interacting RNAs (piRNAs), which contains millions of distinct sequences bearing homology to transposable elements, functions as a molecular memory of self versus non-self (Brennecke et al. 2007). Using their bound piRNA

as a guide, Piwi proteins can recognize their targets and silence them. Failure to do so leads to defects in germline development and sterility (Khurana et al. 2011).

Genetic studies have uncovered some of the loci that are essential for proper function of the piRNA pathway. Besides the core proteins of the Piwi clade, *flamenco* has long been implicated in transposon control (Pelisson et al. 1994). This discrete locus on the X-chromosome of *Drosophila* was found to be a major determinant for silencing of the retroelement *gypsy* almost two decades ago (Bucheton 1995). Sequencing small RNAs bound to Piwi proteins and mapping these sequenced reads back to the genome revealed the true nature of the flamenco locus: rather than being a protein coding gene, the *flamenco* transcript produces the majority of piRNAs expressed in follicle cells of the ovary, which are used to recognize *gypsy* transposition (Brennecke et al. 2007). However, what distinguishes such a piRNA cluster from other transcribed loci and how the primary transcript gets funneled into piRNA production remains unclear. Several studies have shed some light on this topic by identifying some of the protein factors that play a role in piRNA cluster transcription and transport. Rhino and Cutoff, as well as histone methylation marks deposited by Eggless (Egg), are necessary for cluster transcription (Rangan et al. 2011; Pane et al. 2011; Klattenhoff et al. 2009). In addition, UAP56, a helicase implicated in splicing and RNA export, recently was found to bind a germline piRNA cluster and may escort the transcript from the nucleus to the nuage for processing (Zhang et al. 2012). Intriguingly, Rhino, Cutoff and UAP56 all were reported to be specific to germline piRNA clusters, leaving factors involved in somatic cluster determination a mystery.

Mutagenesis screens for sterility phenotypes in *Drosophila* also discovered factors that later were found to be members of the piRNA pathway (Schüpbach and Wieschaus 1989; 1991). Some of these factors can be placed at a specific step of the pathway, such as being required for biogenesis or downstream silencing of transposons. However, very little is known about specific biochemical functions or enzymatic activities within each step. Some of the more detailed knowledge of piRNA biogenesis came from bioinformatical studies of small RNA populations bound to the Piwi-clade proteins active in the nurse cells and the oocyte:

Aubergine (Aub) and Argonaute3 (Ago3). When comparing possible partners between each population, an unusually high number of small RNA pairs with a 10nt overlap was revealed (Brennecke et al. 2007; Gunawardane et al. 2007). It was proposed that cleavage of active transposable elements through primary piRNAs derived from cluster transcripts could lead to the biogenesis of secondary piRNAs. This target-dependent amplification loop was termed the ping-pong cycle and highlights the necessity of the slicer activity of Aub and Ago3.

Regarding primary piRNA biogenesis, less is understood. One major insight came from the discovery of a trimming activity in insect cell lysates, which shortens the 3′ end of putative piRNA precursors to their mature length (Kawaoka et al. 2011). However, the genetic identity of this activity remains unknown. Biochemical data and structural knowledge of the nuclease Zucchini (Zuc), which was previously implicated in the piRNA pathway, suggests it as a promising candidate for the enzyme that creates the 5′ end of mature piRNAs (Ipsaro et al. 2012; Nishimasu et al. 2012). The requirement for other endo- or exonucleolytic cuts needed to create a piRNA is unknown but remains a possibility.

Another aspect of the pathway that remains enigmatic is how Piwi-piRNA complexes silence their targets. In the case of somatic cells of the ovary it has become clear that control of transposons happens at the transcriptional level through modification of epigenetic marks. Recently, a conclusive study showed that upon Piwi depletion euchromatic copies of transposons engage in active transcription and show a depletion of H3K9 trimethylation (H3K9me3) (Sienski et al. 2012). The authors also place Maelstrom (Mael) at the silencing step of the piRNA pathway. Interestingly Mael separates transcriptional silencing from H3K9me3 deposition, indicating that this modification is not the final silencing mark. What the final silencing mark may be, and what proteins are responsible for establishing these marks are yet to be identified.

With the advent of an invaluable resource, the ovarian somatic sheet (OSS) cell line derived from follicle cells of the ovary, it has become feasible to answer some of these open questions (Niki et al. 2006). This cell line only expresses Piwi but not Aub or Ago3 (Lau et al. 2009; Saito

et al. 2009). Consequently, these cells do not show signatures of secondary piRNA biogenesis or the ping-pong cycle. However, with an active primary piRNA pathway in place, genetic requirements for primary piRNA biogenesis and transposon silencing can be tested.

Here we describe a genome-wide screen that addresses some of the important open questions in the piRNA field. By individually assaying more than 41,000 dsRNAs targeting every gene in the *Drosophila* genome and examining their effect on transposons levels we draw a comprehensive genetic framework for transposon control. We reveal novel piRNA biogenesis factors and place previously unknown proteins at the effector step of the pathway. In addition, by validating a large subset of the candidate hits *in vivo*, we demonstrate the strength and relevance of the primary data set. This study not only provides an important resource for the scientific community but also a solid foundation on which future research can be built.

3.3 Results

Primary Screen

In order to assay derepression of transposons upon knockdown of any given target gene, we established a sensitive assay for *gypsy* mRNA levels. Based on quantitative PCR (qPCR) with hydrolysis probes, this assay specifically detects the spliced subgenomic transcript of the retrotransposon (Figure 3.1A). The expression of this transcript is known to be highly sensitive to disruption of the piRNA pathway, even more so than its full-length counterpart (Pelisson et al. 1994).

Knockdown of target genes was accomplished by transfecting long double-stranded RNAs (dsRNAs) from two genome-wide libraries with a total of 41,342 dsRNAs. The average count of dsRNAs per gene was 2.28, targeting 13,914 genes with valid IDs in Flybase (McQuilton et al. 2012). Additionally, the two libraries contained 1,045 negative controls, 2,097 dsRNAs without annotated target and 2,301 dsRNAs targeting the Heidelberg collection of predicted genes.

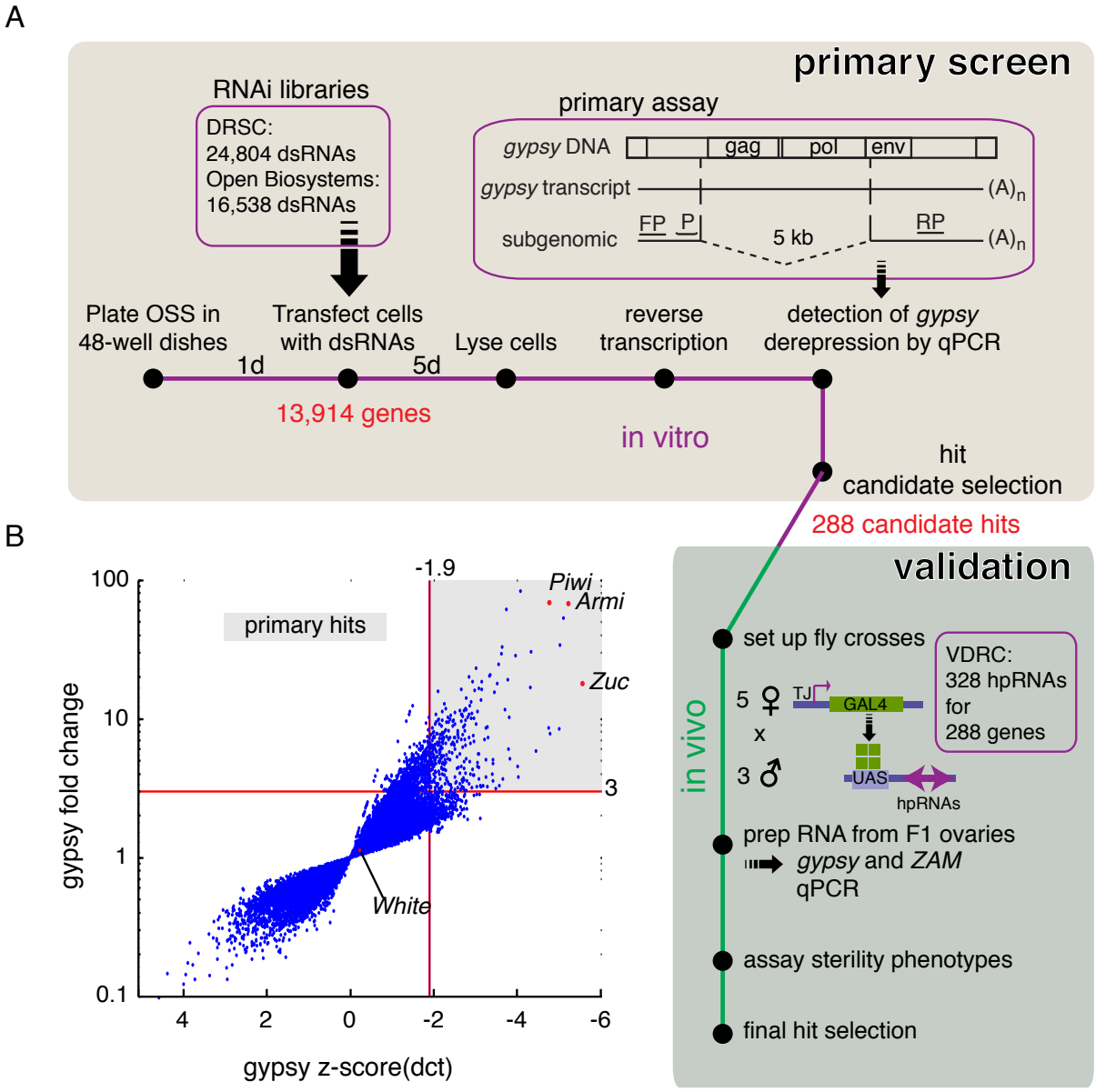


Figure 3.1: **A genome-wide RNAi screen in the somatic compartment of *Drosophila* ovaries tests for derepression of the *gypsy* retrotransposon** A) A workflow of the primary RNAi screen in ovarian somatic sheet cells (OSS) and validation of primary hit candidates in vivo is shown. Each gene in the *Drosophila* genome was knocked down with long double-stranded RNAs (dsRNAs). The origins of the dsRNA libraries are indicated (DRSC: Drosophila RNAi screening center). 5 days after transfection, cells were tested for increased levels of the *gypsy* retrotransposon. The primers and the hydrolysis probe used for the qPCR-based readout are shown in relation to the subgenomic *gypsy* transcript and its splice sites (FP: forward primer, P: hydrolysis probe, RP: reverse primer). The dashed line indicates the 5kb segment of the full-length transcript not present in the subgenomic transcript. After primary hit candidate selection, 288 genes were further tested in vivo.

Figure 3.1: (Continued) The validation panel shows a schematic representation of the Vienna Drosophila RNAi Center (VDRC) Gal4/UAS system used to drive hairpin RNAs (hpRNAs) specifically within the *traffic jam* (TJ) expression domain (Dietzl, 2007). As in the primary screen, the readout was qPCR based, except that an additional transposable element (*ZAM*) and effects on fertility were assayed. B) All transfected wells were assayed for levels of *gypsy* and one housekeeping gene for normalization. The levels of *gypsy* expression are displayed as z-scores (distance of standard deviations from the median) and fold change (absolute distance from the median). The cutoffs for both z-score (<-1.9) and fold change (>3) are indicated as red lines. The shaded area shows the selection of primary hit candidates. Three positive controls (Piwi, Armi, Zuc) and one negative control (White) are marked as red dots. Only wells that passed the filter for primary datapoint selection are shown (-2<=z-score-cib<=2, ct-gypsy<38). For all primary datapoints see Table 3.3.

Because of the sheer number of dsRNAs to be assayed, providing clean RNA input without genomic DNA contamination for the reverse transcription was impractical. By assaying the subgenomic transcript, crude lysates from transfected OSS cells could be used as input, which made it feasible to assay on a genome-wide scale.

We transfected OSS cells in 48-well plates, lysed the cells five days later and used the lysate for reverse transcription (Figure 3.1A). Since the smallest experimental unit of this assay is a 48 well plate, the qPCR results ($ct_{transposon}$-$ct_{reference}$) were normalized to their respective plate using z-scores (Ramadan et al. 2007). This normalization method is a simple measure of distance in standard deviations (SD) from the plate median, which we used in order to account for extreme outliers due to experimental error. As a secondary metric we calculated the absolute distance of each data point within a plate as a fold change in relation to the median. The knockdown of Piwi in this experimental setting led to a *gypsy* mRNA signal that was detectable much earlier by qPCR than the average of the plate. In four biological replicates, the average normalized signal for *gypsy* was almost five standard deviations away from the median of its plate (Figure3.2A). Hence, our assay for transposon derepression is both sensitive and robust, given that the RNAi trigger is of good quality. When comparing several independent dsRNAs, we saw that there is considerable variance: dsRNAs against known components of the pathway such as Armitage (Armi) led to consistent derepression of *gypsy*. However, fold changes of the *gypsy* transcript levels varied from 25 to 70 fold (Figure 3.2B).

Nevertheless, by assaying several dsRNAs against each gene of the genome, we are confident to have overcome this hurdle.

Out of 41,342 tested dsRNAs, 33,780 met our criteria for data point selection: We ignored extreme outliers for the reference gene *ciboulot* and wells in which *gypsy* could not be detected after 38 cycles of qPCR. Out of these 33,780 datapoints, 320 dsRNAs met the criteria for primary hit selection (Figure 3.1B). Included in this list were 18 dsRNAs without annotated primary target, which were disregarded. We identified all known components of the pathway, which were previously implicated in *gypsy* control as outliers of the primary screen (Figure 3.2B). Knockdown of pathway members such as Capsuleen, Hen1, Egg or Squash was not expected to cause strong *gypsy* derepression based on existing literature (Horwich et al. 2007; Saito et al. 2010; Haase et al. 2010; Rangan et al. 2011; Nishida et al. 2009; Kirino et al. 2009). After choosing z-score and fold change cutoffs for hit selection based on 217 GFP negative controls, only 3 out of 645 (0.5%) additional negative controls (empty wells, Rho1 and Thread) were weak hits (Figure 3.2C).

To be called a primary hit candidate, only one of the assayed dsRNAs against each gene had to meet our selection criteria. This approach can be misleading if the underlying library already has a higher than average number of dsRNAs for a subset of genes, since these will be more likely to score. Indeed, while the average of dsRNAs per gene was 2.28, the primary hit candidates each had a representation of 2.64 dsRNAs, which is significantly higher. This representation bias has potential implications for the analysis of hit lists, since the over represented dsRNAs tend to be part of the same functional annotation groups. Even weak correlations between representation and the number of functional annotations ($r \sim 0.08$ in our libraries) can create strong biases in downstream analyses, sometimes called annotation or multifunctionality bias. In fact, performing over-representation analysis on the 50 most represented genes in the library already shows functional enrichment of 29 terms ($p<0.05$ after multiple test correction). Using an ROC based threshold free approach yields enormous enrichment (955 terms enriched at $p<0.05$ after multiple test correction). One straightforward way to control for this

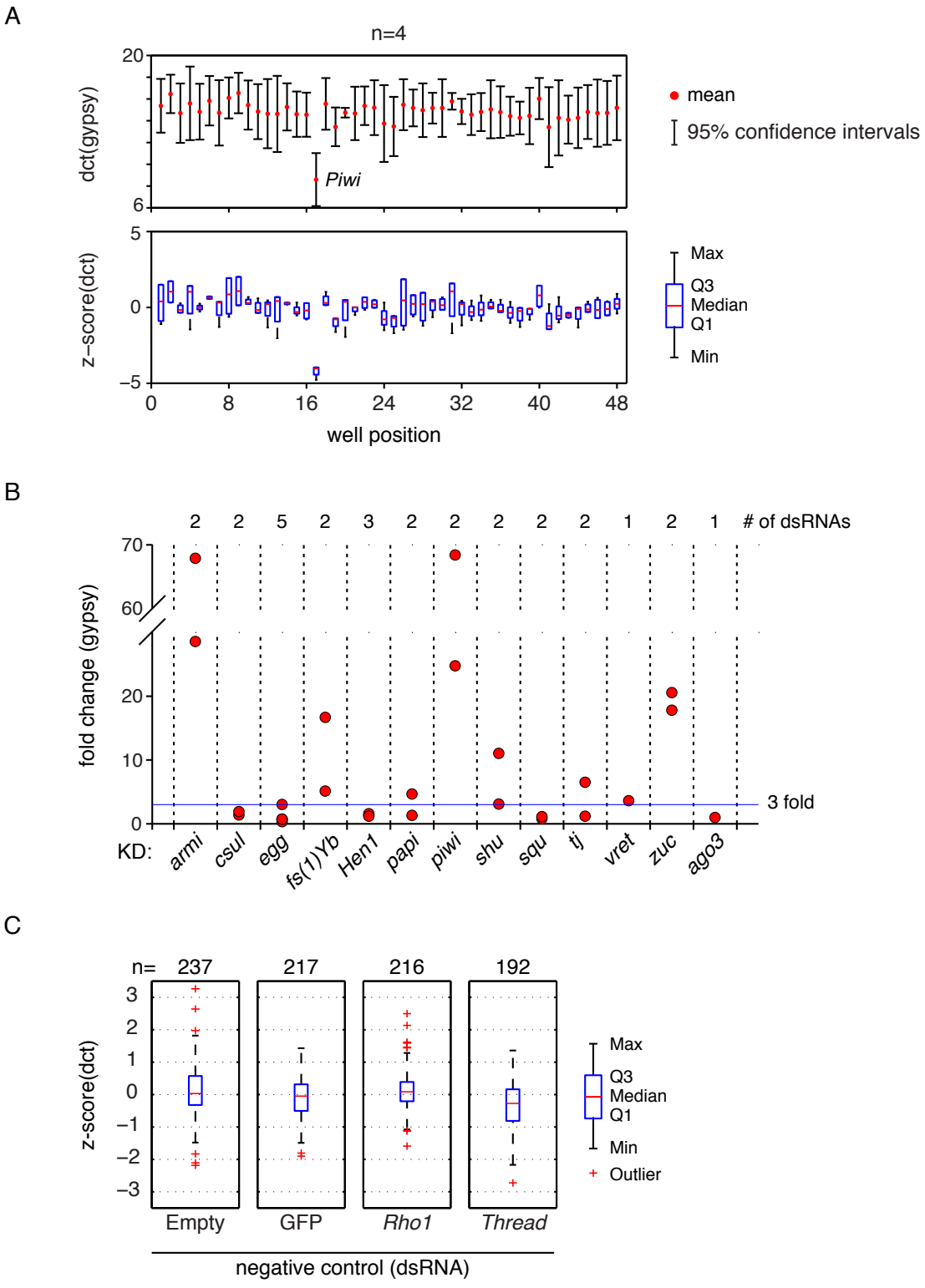


Figure 3.2: **Performance and controls of the primary screen, related to Figure 3.1**

Figure 3.2: (Continued) A) For 48 genes, including the positive control Piwi, dsRNAs were transfected in four independent biological replicates. The upper graph shows the means and 95% confident intervals of *gypsy* levels relative to a reference gene. The lower graph shows the individual z-scores as box plots for all 48 wells after normalization to the median of the plate. Piwi is a clear outlier in all four independent experiments. B) The primary screen results in fold changes for all known somatic piRNA pathway components and Ago3 as a negative control are shown. The number of independent dsRNAs against each gene is indicated on top of the graph. The threshold for primary hit selection (3 fold upregulation of *gypsy*) is marked in blue. C) The primary screen results in z-scores for 862 negative controls are shown as boxplots. Outliers are indicated as red crosses. The number of independent transfections of each dsRNA is indicated above.

was to aggregate data for each gene. This would be methodologically conservative since it decreases the variance of the most commonly annotated genes. We take this approach in the subsequent analysis for enrichment of biological functions. Although we use the aggregated data for enrichment analysis, we decided to test all 320 original primary hits for candidate validation, since potential false positives can later be weeded out given a robust *in vivo* assay.

To see whether genes that affect transposon control show preference towards specific annotation groups, we performed functional enrichment analysis on our primary dataset. After multiple test correction, 215 functions were enriched (corrected $p<0.05$), many with strong potential relevance and very significant enrichment (the top 20 functions have a corrected $p<1E-6$, Table 3.1). Among the most significant, we find expected cellular components like the 'Yb body', but also more surprising functions like 'regulation of growth of symbiont in host'. While several of these enrichments are driven by genes implicated in the piRNA pathway, all scoring highly in the primary screen, novel genes intersect with these in some of the enriched functions. For example under 'dorsal appendage formation' Smt3 (SUMO) joins Armi and Zuc (Nie et al. 2009). Further research will be necessary to determine to what extent these unexpected intersections relate to biologically relevant connections.

Next, we compared protein interaction data to the full ranked fold change list. For every gene in the genome, the degree to which that gene's interaction partners scored highly on the fold change list was measured (as ROC). Of the top 20 genes with interaction partners significantly elevated in the fold change list, 18 are annotated as belonging to the proteasomal

Table 3.1: Top enriched GO terms in the primary screen

GO term	p-value	GO ID
Yb body	0	GO:0070725
negative regulation of growth of symbiont in host	0	GO:0044130
stem cell development	0	GO:0048864
negative regulation of multi-organism process	0	GO:0043901
positive regulation of Ras protein signal transduction	0	GO:0046579
dorsal appendage formation	0	GO:0046843
germ-line stem cell maintenance	8.00E-10	GO:0030718
regulation of mRNA 3-end processing	6.00E-09	GO:0031440
male germ-line stem cell division	6.00E-09	GO:0048133
gene silencing by RNA	2.00E-08	GO:0031047
negative regulation of transposition	2.00E-07	GO:0010529
imaginal disc-derived wing expansion	3.00E-07	GO:0048526

complex (which has 65 genes in total). This observation was remarkably significant, which may be partially due to the correlated interaction profiles of the proteasomal complex. The two remaining genes not belonging to the proteasomal complex were Bx42, a homolog of mammalian Skip (SKI-interacting protein), a protein implicated in splicing and Calypso, a Histone 2A deubiquitinase (Scheuermann et al. 2010; Makarov et al. 2002). Both genes are highly expressed in ovaries according to the modENCODE tissue expression data. Whether their interaction with particularly high scoring genes is indicative of any regulatory function remains to be seen.

Validation Screen

We obtained 328 fly lines from the Vienna Drosophila Resource Center (VDRC) containing inducible hairpin RNAs (hpRNAs) against 288 out of our top 302 primary hit candidates (Dietzl et al. 2007). When crossed to males expressing a follicle cell specific Gal4 driver (*traffic jam*), these hpRNAs can effectively knockdown any given target gene within the same expression domain (Tanentzapf et al. 2007). Using hpRNAs against Aub as a negative control, we observed highly significant changes in *gypsy* expression by qPCR when knocking down Armi (Figure 3.4A). Furthermore, not only *gypsy* levels were detectable to a much higher

degree: We confirmed the observed derepression with two independent transposons, *ZAM* and *gypsy3*. All pathway components that were identified as primary hits were validated using this approach (Figure3.4B). Harnessing this *in vivo* system, we validated 87 out of the 288 primary hit candidates (Figure 3.3A). In order to validate, knockdown of the target gene had to cause upregulation of *gypsy* or *ZAM* by at least two fold. For crosses with male flies originating from the GD library from VDRC we used *ZAM* as a metric, for the KK library we measured *gypsy* levels. This decision was based on the finding that negative controls from the GD library already showed higher levels of *gypsy* subgenomic transcript when compared to the KK library negative controls.

Out of the 288 candidate hits, knockdown of 52 genes led to such severe developmental defects, that dissection of ovaries and confirmation of the initial screen result was not possible. However, several arguments can be made for this category to harbor a substantial amount of 'true' hits. First, the genes in this category had an average *gypsy* fold change of 5.8 in the primary dataset, as compared to 3.3 for the non-validated genes. This average fold change was even higher than the validated subset (5.1 fold). Since primary fold changes as well as z-scores are a function of precision (the likelihood of a primary hit to be validated), this is indicative of the biological relevance of these hits (Figure 3.4C).

Second, the previously mentioned overrepresentation bias for dsRNAs against primary hits was significantly lower for the developmental defects subset than for the non-validated: While the non-validated genes had a representation count of 2.84, the developmental defect set had a count of 2.42, which is significantly different for the two sets ($p \sim 0.0035$, ranksum test). In other words, the genes of the developmental defect category were disadvantaged to be a primary hit in comparison to the non-validated, yet had a much higher average fold change.

Both validated and developmental defect sets were significantly enriched for genes with higher expression levels in ovaries when compared to whole fly (Figure 3.3B). While knockdown of genes within the non-validated category only led to sterility of 9% of the crosses, the fraction was 16% for the validated and 94% for the developmental defect set. The fact that we

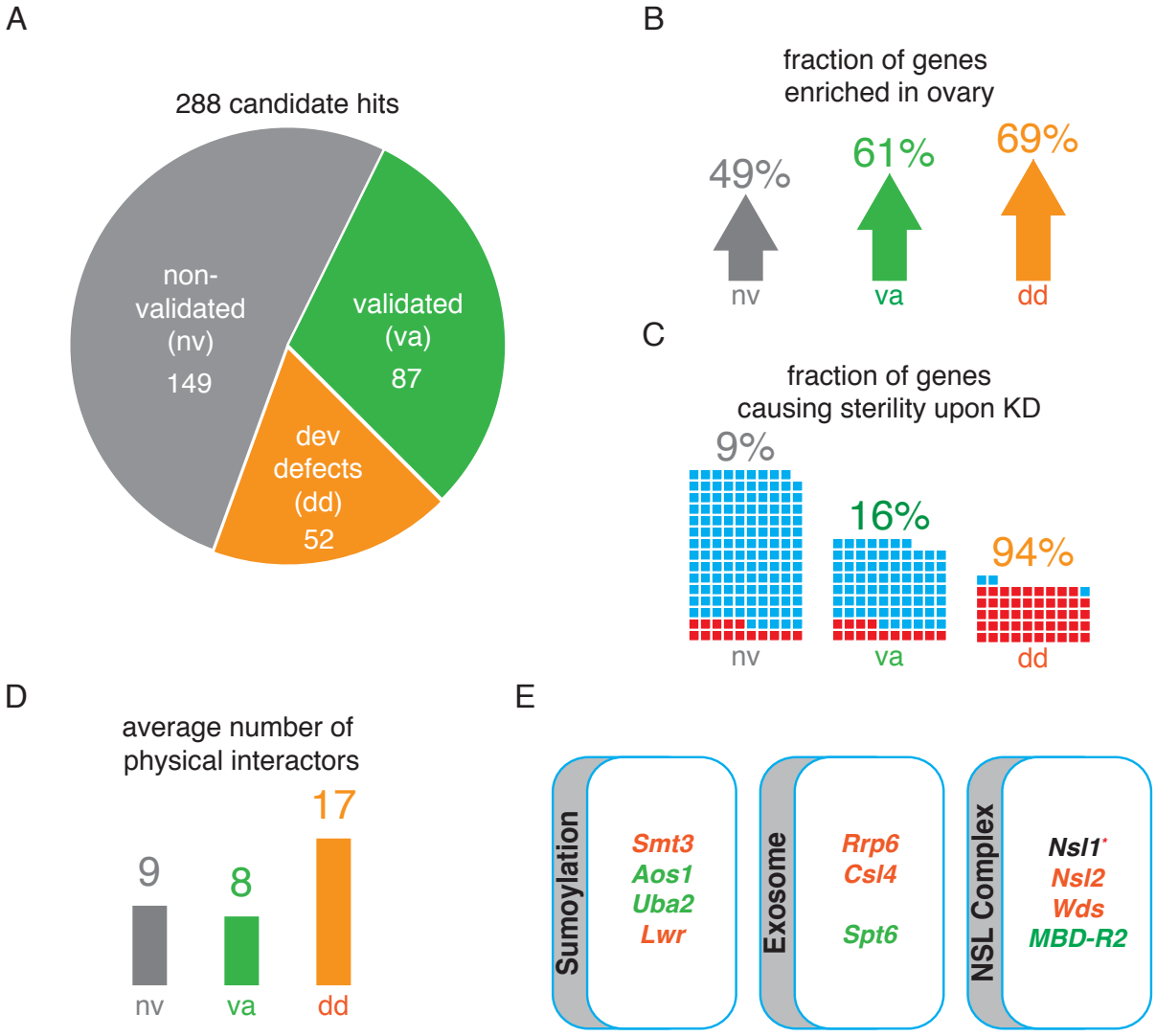


Figure 3.3: **Primary hit candidates are validated *in vivo*** A) The number of hit candidates that validated (va) or did not validate (nv) in vivo is shown. Genes that upon knockdown caused severe developmental defects and therefore could not be assayed are also indicated (dd). A full list of validated fly lines and corresponding transposon derepression information is available in Table 3.4. B) Validated hits are preferentially expressed in ovaries. The percentage of genes that are enriched in ovaries compared to whole fly is shown. This data is based on mRNA signals on Affymetrix expression arrays and is available from FlyAtlas (Chintapalli, 2007). The percentages are shown for all three categories (nv, va, and dd). C) The fraction of genes causing sterility upon knockdown is shown for all three categories. Each small box represents one gene, with blue and red indicating if flies were fertile or sterile, respectively, upon knockdown. Fertility was defined as presence of any larvae in the vial, 8 days after egg lay. D) Node degree of genes in each of the classifications measured by number of physical interactors is shown as a bar graph. Interaction data from BIOGRID was used for this analysis.

Figure 3.3: (Continued) E) A selection of several protein complexes identified in the validation screen is shown. Members of the *Drosophila* sumoylation pathway, the exosome and the nonspecific lethal (NSL) complex are primary hits that validate in vivo. NSL1 could not be validated in vivo because no RNAi fly was available at the time of publication (red asterisk). The text coloring of each gene indicates the result of the validation screen and is consistent with the categories in panel A.

do find fertile flies within the developmental defect set can most likely be attributed to human error during ovary dissections. The extreme phenotypes we observe in the developmental set imply more generic functions for these genes. Indeed, around 90% have an average of 17 physical interactors, which is significantly higher than the other categories (Figure 3.3D). Given these observations, one might expect the average fold changes within the primary dataset to be correlated overall with node degree within a protein interaction network. However, using protein-protein interaction data from BIOGRID, totaling 34,523 unique interactions over 7,895 genes, revealed that this was not true ($|r|<0.01$, $p>0.5$). To further test whether the developmental set should be treated as non-validated, we excluded these genes from the fold change data (aggregated across dsRNA to genes) and performed threshold-free ROC based enrichment analysis. This yielded 74 enriched terms ($p<0.05$, after multiple test correction) of which 33 overlapped with the enriched terms of the developmental set alone; thus, the functional categories predominately characterizing the developmental defect list are broadly present in the data overall.

When ranked by their fold changes in the validation round, the validated genes display some remarkable properties. All known components of the piRNA pathway were among the top 20 hits (Table 3.2). Nxt1, a nuclear export factor ranks first with *gypsy* levels almost 2,500 fold higher than the negative control (Herold et al. 2001). In addition, depletion of Nxt1 also led to sterility. Interestingly, Nxt1 is part of a heterodimer together with Nxf1. While Nxf1 was not a hit in the primary screen, knockdown of UAP56, which acts in the same export pathway, showed consistent derepression of *gypsy* (Herold et al. 2003). First implicated in splicing, this RNA helicase was recently shown to be involved in transport of the primary piRNA transcript of dual-stranded clusters to the nuclear pore (Zhang et al. 2012). Unfortunately,

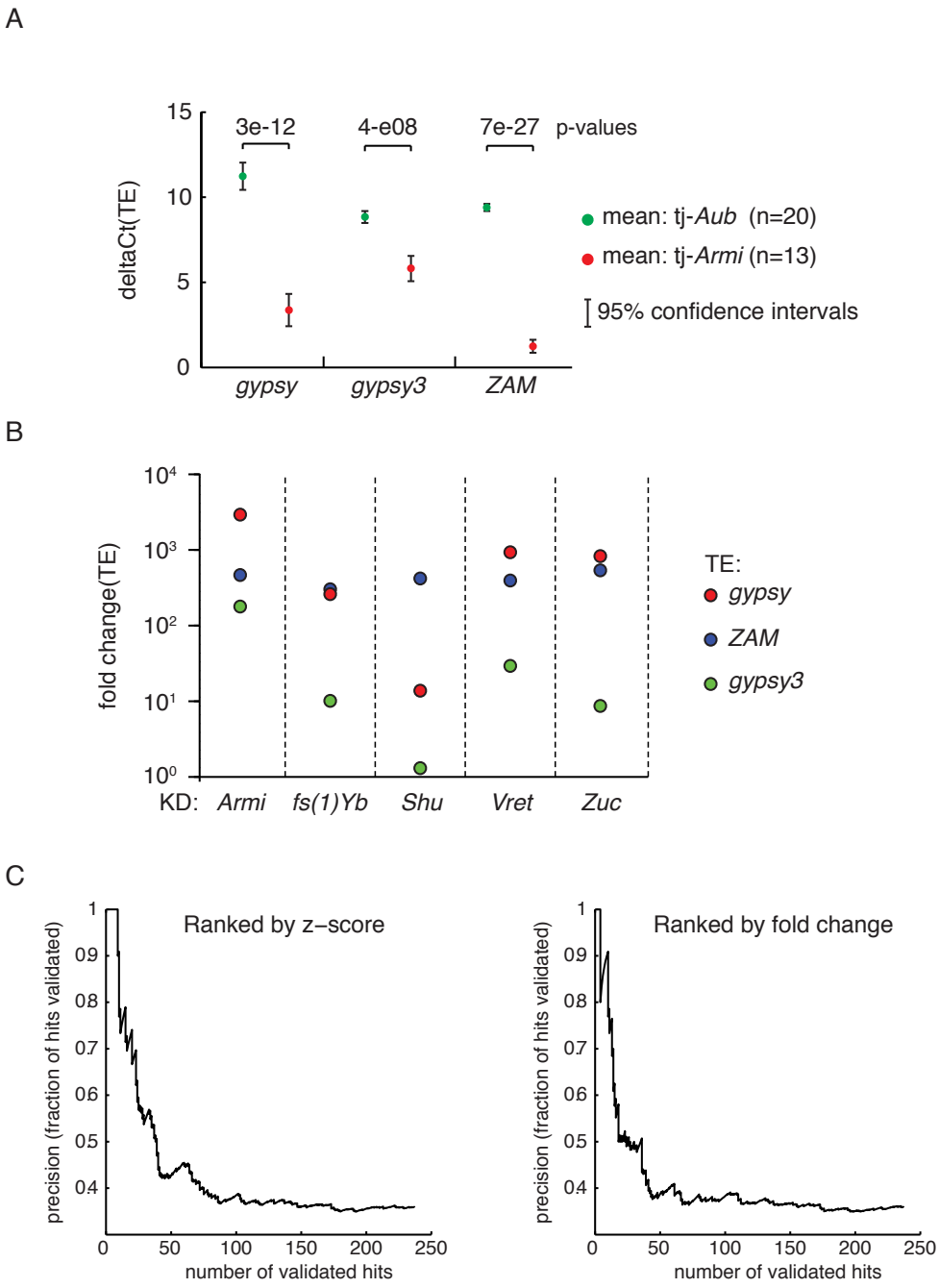


Figure 3.4: **Performance and controls of the validation screen, related to Figure 3.3** A) Knockdown of Armi leads to highly significant differences in transposon expression when compared to a negative control (Aub). Shown are mean delta-Ct values and 95% confidence intervals for three transposons assayed by qPCR. The number of biological replicates is indicates in brackets. The results of a t-test for significance are indicated as p-values for each transposon.

Figure 3.4: (Continued) B) Knockdown of all components of the somatic piRNA pathway, which scored in the primary screen, has strong effects on the expression levels of three transposons in vivo. The fold change of each transposon upon knockdown is displayed on a log scale. C) Z-scores and fold changes are a function of precision. Precision is the fraction of validated hits out of the total number of hits (validated and non-validated). The number of validated hits is shown on the x-axis. All dsRNAs for validated genes were used to cover the depicted range. Thus, if genes had dsRNAs producing z-scores or fold changes outside the range needed for primary hit selection, the genes' final annotation as validated or non-validated was assigned to those dsRNAs.

the knockdown of UAP56 in follicle cells affects germline development to such an extent that verification of the primary screen results was not possible.

Nxt1 was previously reported to affect interactions with the nuclear pore complex as well (Lévesque et al. 2001). Hence, the presence of several nuclear pore components within the top 20 is not surprising: Both Nup154 and Nup43 show similar derepression of *gypsy*. Additionally, Nup154 deficient flies are sterile in our assay.

Another two genes ranking among the top 10 are as of yet uncharacterized: CG3893 and CG2183. The latter is predicted to be a homolog of GASZ, a gene previously implicated in the piRNA pathway in mice (Ma et al. 2009). CG3893 shows homology to mammalian Gtsf1. Even though no direct link to the piRNA pathway has been shown, this germline specific factor also seems to be indispensible for transposon control in mice (Yoshimura et al. 2009).

Table 3.2: Top 20 validate hits

Symbol	Primary screen fold change	Validation screen fold change		
		Gypsy	Zam	Gypsy3
Nxt1	2	2452	3566	41
fs(1)Yb	11	96	700	335
armi	48	197	846	112
zuc	19	809	549	9
vret	4	74	315	22

Table 3.2: Top 20 validate hits (continued)

Symbol	Primary screen fold change	Validation screen fold change		
		Gypsy	Zam	Gypsy3
CG3893 (Asterix)	42	80	207	10
mael	3	159	452	16
CG2183	4	173	158	11
lin-52	5	153	153	8
MBD-R2	16	85	48	4
Uba2	3	84	12	8
CG9754	2	26	61	14
wde	3	40	120	12
Nup154	6	30	186	3
Su(var)2-10	2	9	20	4
dlg1	2	16	2	1
shu	7	14	416	1
CG4686	4	13	1	1
Nup43	3	12	3	1
Cchl	0	12	1	2

Smt3 (SUMO), which was one of the highest scoring genes in the primary screen, could not be validated *in vivo* because of developmental defects upon knockdown (Talamillo et al. 2008). However, the depletion of the E1 activating enzymes Aos1 and Uba2 caused consistent transposon derepression in the validation screen. Knockdown of the E2 conjugating enzyme Lesswright also caused developmental defects and could not be validated. Another notable validated hit is Windei (Wde), which was previously reported as a cofactor of Egg in histone 3 lysine 9 trimethylation (H3K9me3) (Koch et al. 2009). While Egg depletion had no effect

on *gypsy* expression, knockdown of Wde caused derepression of *gypsy*, although to a lower extend than *ZAM*.

Lin-52, which is part of the dREAM transcriptional regulator complex, also scored highly with *gypsy* and *ZAM*. This complex is a highly conserved multi-subunit complex that functions in both transcriptional activation and repression (Lewis et al. 2004). Interestingly, it was previously described as a transcriptional activator of Piwi (Georlette et al. 2007).

Follow-up and novel piRNA pathway components

In order to place some of the validated hits within the piRNA pathway, we constructed RNA-seq libraries from biological replicates of the same crosses used in the validation round. Mapping RNA-seq reads to transposon consensus sequences revealed the same levels of *gypsy* and *ZAM* derepression observed by qPCR (Figure 3.5A). When analyzed on a global scale, only transposons dominant within the somatic compartment of the ovary reacted significantly to the respective knockdown of the target gene (FDR<0.05) (Malone et al. 2009). Transposons like *HeT-A*, *roo* or *Rt1b*, which were previously shown to be germline dominant, did not change expression levels (Malone et al. 2009). The patterns of derepression that we observe in knockdowns of known components of the pathway (Armi and Mael) remarkably resemble those observed in CG3893 and Wde knockdown. Uba2 can be considered an outlier in this analysis, since it is the only gene not affecting *ZAM* expression. This is consistent with a much lower fold change of *ZAM* in comparison to *gypsy* when assayed by qPCR (Table 3.2).

We observe a similar clustering affect of CG3893 and known piRNA pathway components in the numbers of significantly differentially expressed genes. While targets like Nxt1 or Uba2, with potentially more general functions, impact the expression of hundreds to thousands of genes, this is not the case for Armi, Mael or CG3893 (Figure 3.5B). However, the fact that so many genes are differentially expressed in the other knockdowns might simply be a consequence of more severely impacted gonadal development. Considering that follicle cells provide the structural niche for the germ cells, the observed changes could therefore reflect

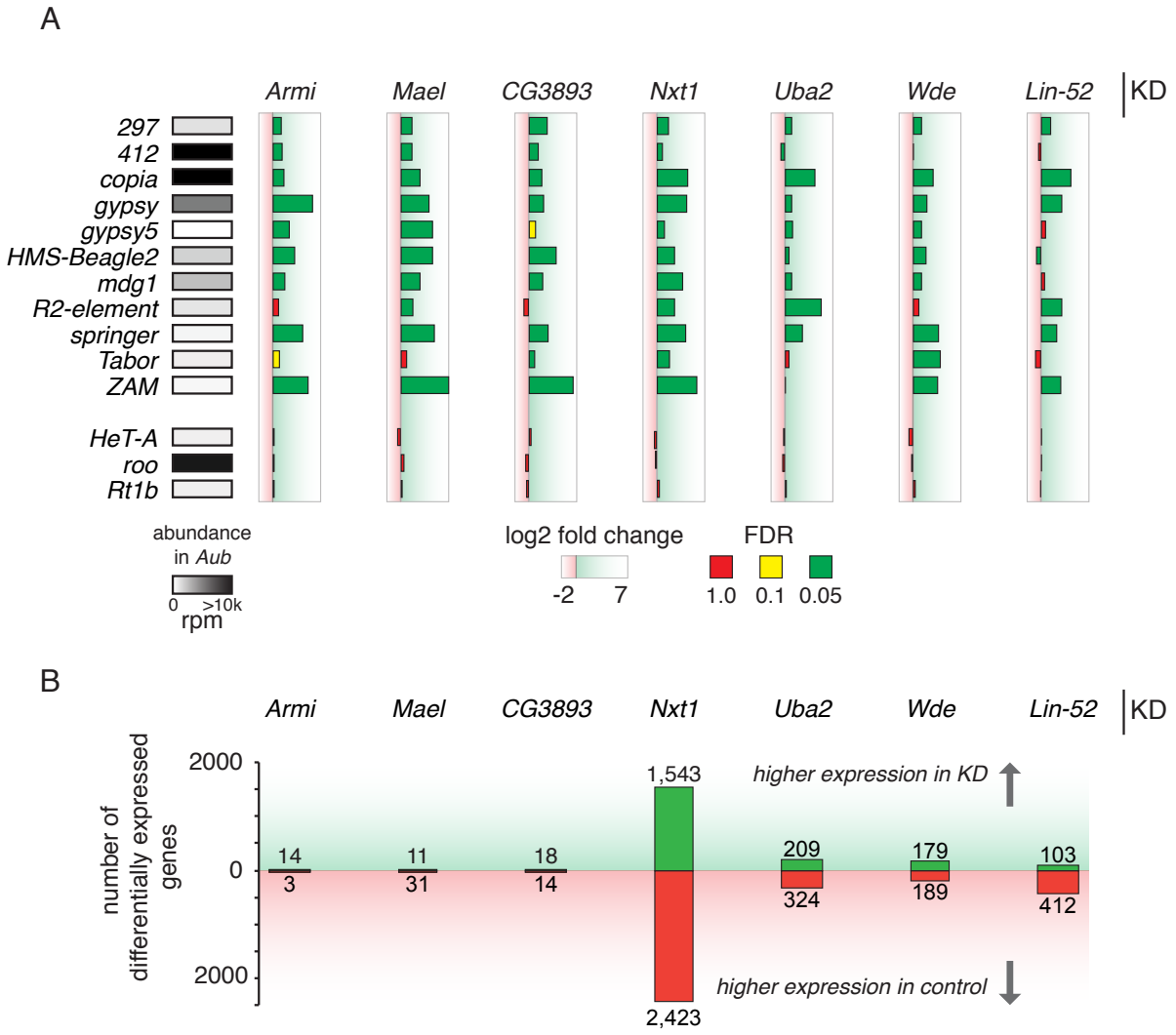


Figure 3.5: **RNA-seq shows massive changes in gene and transposon expression upon knockdown of top candidates *in vivo*** A) A subset of somatically expressed transposons is derepressed when gene expression of a subset of top hits is disrupted. The absolute abundance of reads in Aub control mapping to each transposon is shown in shades of grey. The log2 fold change of each target gene versus a negative control (Aub) is shown. Color of the bars represent the significance of these fold changes and are indicated as an adjusted p-value (FDR). Green indicates highly significant differences (p<=0.05), yellow moderately significant changes (0.05<p<=0.1) and red non-significant changes (0.1<p<=1), based on two biological replicates. Each knockdown is normalized to Aub knockdown controls from their corresponding library (GD or KK). For differences in transposon abundance levels between both Aub controls see Fig S3. B) Number of genes differentially expressed (p-adj<0.05) in each knockdown with respect to the control is observed. Green bars indicate number of genes that have higher expression levels in the knockdown fly line, while red bars designate number of genes with higher levels in the Aub negative control. The precise numbers of differentially expressed genes are indicated above and below each bar.

the transcriptional levels of the nurse cells and oocyte, and not the expression domain of the knockdown. Indeed, *nanos* expression is significantly altered in Nxt1 knockdowns (data not shown). One interesting observation from analyzing the RNA-seq datasets is a striking difference in *gypsy* levels between Aub negative controls originating from the two available VDRC libraries (GD and KK). While technical and biological replicates of both library types show highly correlated levels of transposons overall, we saw significant differences in *Tirant* and *gypsy* (Figure 3.6 A-C), between the two libraries. The Aub fly line from the GD library has almost 10-fold higher levels of *gypsy* compared to the KK library. In hindsight, this justifies our decision during the validation screen of using *ZAM* derepression as a metric for the GD library, rather than *gypsy*.

Given a possible involvement of all these genes in the piRNA pathway based on their transposon derepression phenotype, we then searched for changes in mature piRNA populations. By constructing small RNA libraries from knockdown flies, we were able to ask several questions: Do levels of mature piRNAs change? Are processing patterns of piRNA clusters altered? Does this correlate with changes of length profiles and nucleotide biases of the remaining piRNAs? Interestingly, we saw a severe drop in numbers of piRNAs uniquely mapping to the soma dominant flamenco piRNA cluster in the Nxt1, Uba2 and Wde knockdowns (Figure 3.7A). To avoid skewing this result based on normalization to a piRNA producing locus, which theoretically should not change in soma specific knockdowns (i.e. cluster 1/42AB), we looked at the internal rankings each cluster is assigned based on overall abundance. Given that we only knock down each gene in the somatic compartment, only clusters within the same expression domain (i.e. *flamenco*) should change their ranking. And indeed, in the Armi knockdown *flamenco* is only the 10^{th} most abundant cluster while it is the third most abundant in total RNA libraries from negative control ovaries. Conversely, 42AB and X-up stay on top of the list in all tested knockdowns (Figure 3.7B). The two genes that mirror Armi are Nxt1 and Uba2. Wde and Lin-52 show different rankings for *flamenco*, however, not to the same extent. In none of the knockdowns do the length profiles of the remaining piRNAs from *flamenco* change

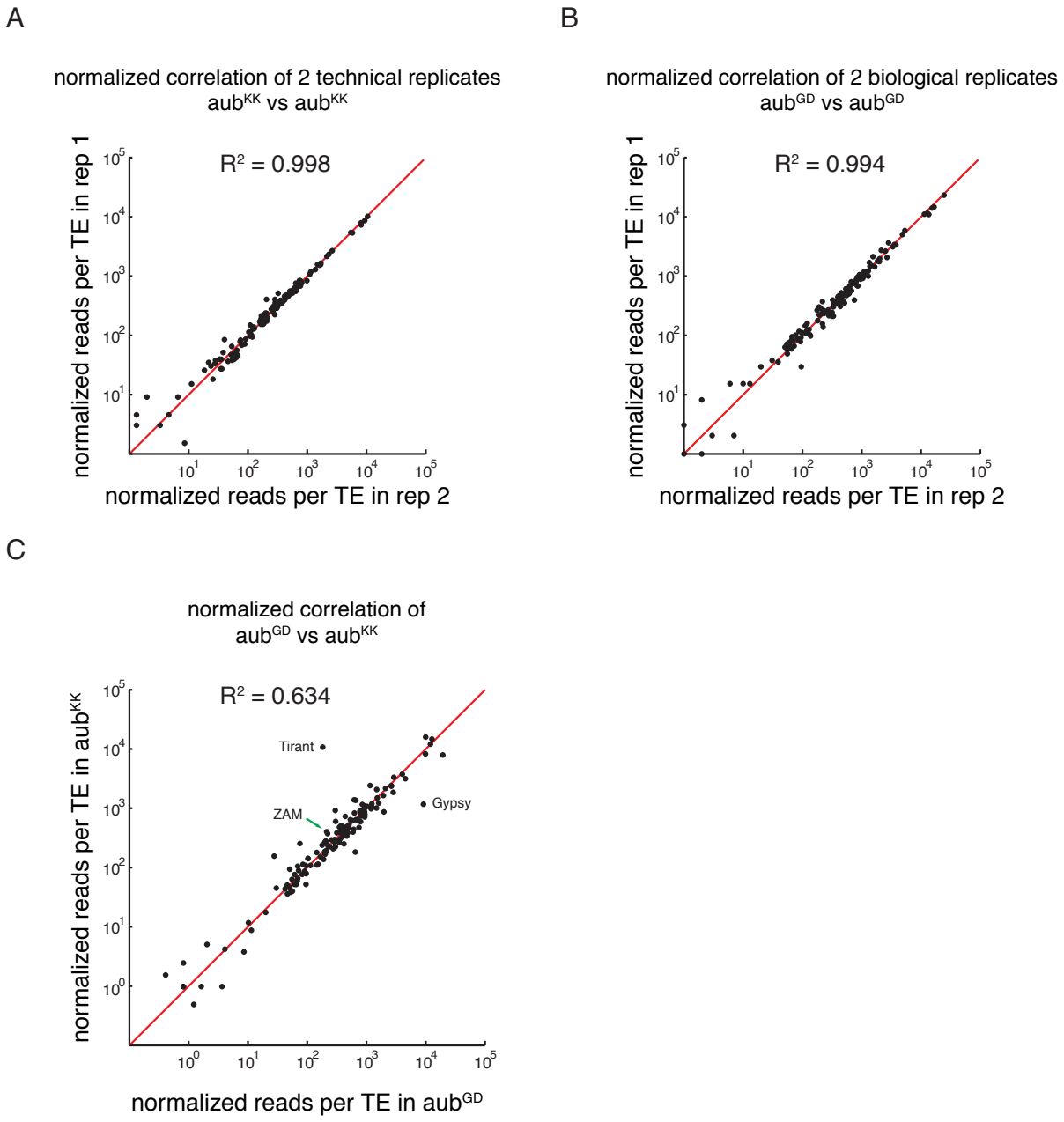


Figure 3.6: **Two negative control lines from two available VDRC fly libraries show different transposon expression levels for *gypsy* and *Tirant*, related to Figure 3.5** A) Scatter plots of normalized reads mapping to TE consensus from RNA-seq data is shown. The squared correlation coefficient for two technical replicates of the KK line is indicated. The red line indicates where data points would show equal numbers in both samples. B) The results for two biological replicates of Aub flies from the GD library are shown. C) Two Aub hpRNA lines from the KK and the GD library are compared. Data points for three transposons are highlighted: *gypsy* and *Tirant* as significantly differentially expressed and *ZAM* as the transposon used for hit calling in the validation screen.

(Figure 3.7C). When compared to their negative control, piRNA levels do not seem to change in the Mael and CG3893 knockdown. The same conclusions can be drawn when mapping to transposons consensus sequences: Antisense populations of piRNAs with homology to soma dominant transposons show severe reductions in the Nxt1, Uba2 and Wde knockdowns, which resembles patterns seen for the biogenesis factor Armi (Figure 3.7D). Depletion of Mael, Lin-52 and CG3893 does not show the same effect. Intriguingly, in the case of Lin-52 this does not coincide with the effects seen for *flamenco* mappers. None of the assayed knockdowns show any changes in mature miRNA levels (Figure 3.8).

CG3893 is indispensible for transposon silencing in the germline

Piwi and Mael have recently been shown to silence transposons in the somatic compartment of the ovary through transcriptional gene silencing (TGS) (Sienski et al. 2012). CG3893, with patterns in global transposon derepression similar to Mael and unaffected mature piRNA populations, seemed to be a good candidate for a novel pathway component at the effector step. CG3893 is a 20kDa large member of a protein family with unknown function (UPF0224), characterized by the presence of highly conserved Zinc-finger domains (Figure 3.9A). All five proteins of this family are weakly expressed in OSS cells, and show germline specific moderate to high expression in the ovary (Figure 3.9B, modENCODE Tissue Expression Data (Graveley et al. 2011). CG34283, the one family member with only weak conservation of its Zinc-finger domain, is the only member showing expression specific to testes. Out of all five members of the family, only CG3893 shows strong effects on transposon control when knocked down in OSS cells (Figure 3.9C). In order to have a more reliable model for a loss of function of CG3893, we searched for available transposon insertion lines. We investigated two available lines (204406, DGRC Kyoto; 22464, Bloomington Drosophila Stock Center Indiana University). Line 204406 has a P-element insertion into the first exon of CG3893, disrupting its N-terminal CHHC Zinc-finger domain (Figure 3.10A). Consistent with the insertion site, we clone truncated mRNAs by RNA-seq library from homozygous animals. The levels of

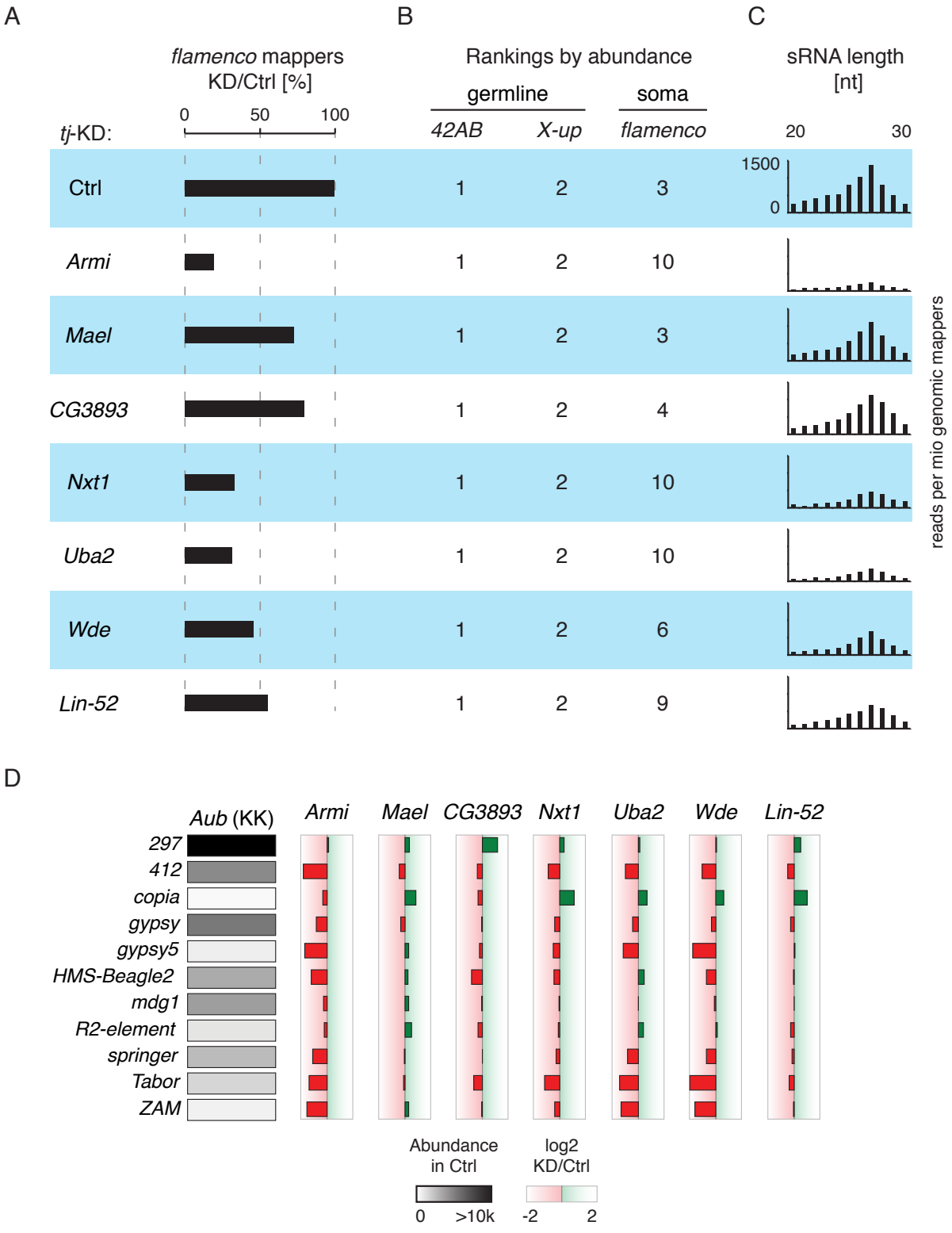


Figure 3.7: **Biogenesis of small RNAs from somatic clusters and transposons is affected in knockdowns of a subset of top candidate genes**

Figure 3.7: (Continued) A) Percentages of total unique mappers (sense species, >23nt) to *flamenco* in each knockdown (as indicated) in relation to the control knockdown are shown. B) The internal rankings for three representative piRNA clusters based on their total abundance are displayed. Expression bias towards either domain (soma or germline) is indicated. Cluster definitions are in concordance with Brennecke et al., 2007. C) The size profiles of piRNAs in each knockdown (as in A) are plotted as total read count per million genomic mappers. As a control, we show that levels of microRNAs do not change in knockdowns versus Aub negative control (Figure 3.8).

mRNA expression are also reduced in animals homozygous for this mutation, when compared to heterozygous siblings. The results obtained by RNA-seq were confirmed by qPCR (Figure 3.10B). The second line (22462) has a P-element insertion upstream of the first exon in either the promoter or the 5' UTR of CG3893. Our RNA-seq data indicates a slightly extended transcript when compared to the gene model status according to Flybase. Even though this insertion is not in any coding sequence, the observed phenotypes are severe: homozygous females are completely sterile and are characterized by a complete absence of ovarian structures. This corresponds to undetectable levels of CG3893 transcript when assayed by qPCR indicating a complete loss-of-function (Figure 3.10B). The phenotypes observed in females homozygous for the first insertion (204406) are milder with ovaries developing to a rudimentary stage (Figure3.9D). Nevertheless, this potentially hypomorphic mutation causes females to be sterile, demonstrating the negative impact of the insertion on CG3893 function. According to our current model of transposon silencing as a nuclear phenomenon, effectors at this step are expected to be nuclear as well. The mouse homolog of CG3893, Gtsf1, is reported to be mainly cytoplasmic in adult testes (Yoshimura et al. 2007). However, when overexpressed in OSS cells, GFP fusion proteins of CG3893 co-localize with Piwi in the nucleus (Figure 3.9E). This localization pattern is independent of Piwi, since expression of Piwi bearing an N-terminal deletion, which localizes to the cytoplasm, does not alter CG3893-GFP localization (Saito et al. 2009).

So far we demonstrated CG3893's involvement in the somatic compartment of the ovary. In order to investigate its role in all tissues of the female germline, we cloned RNA-seq and small

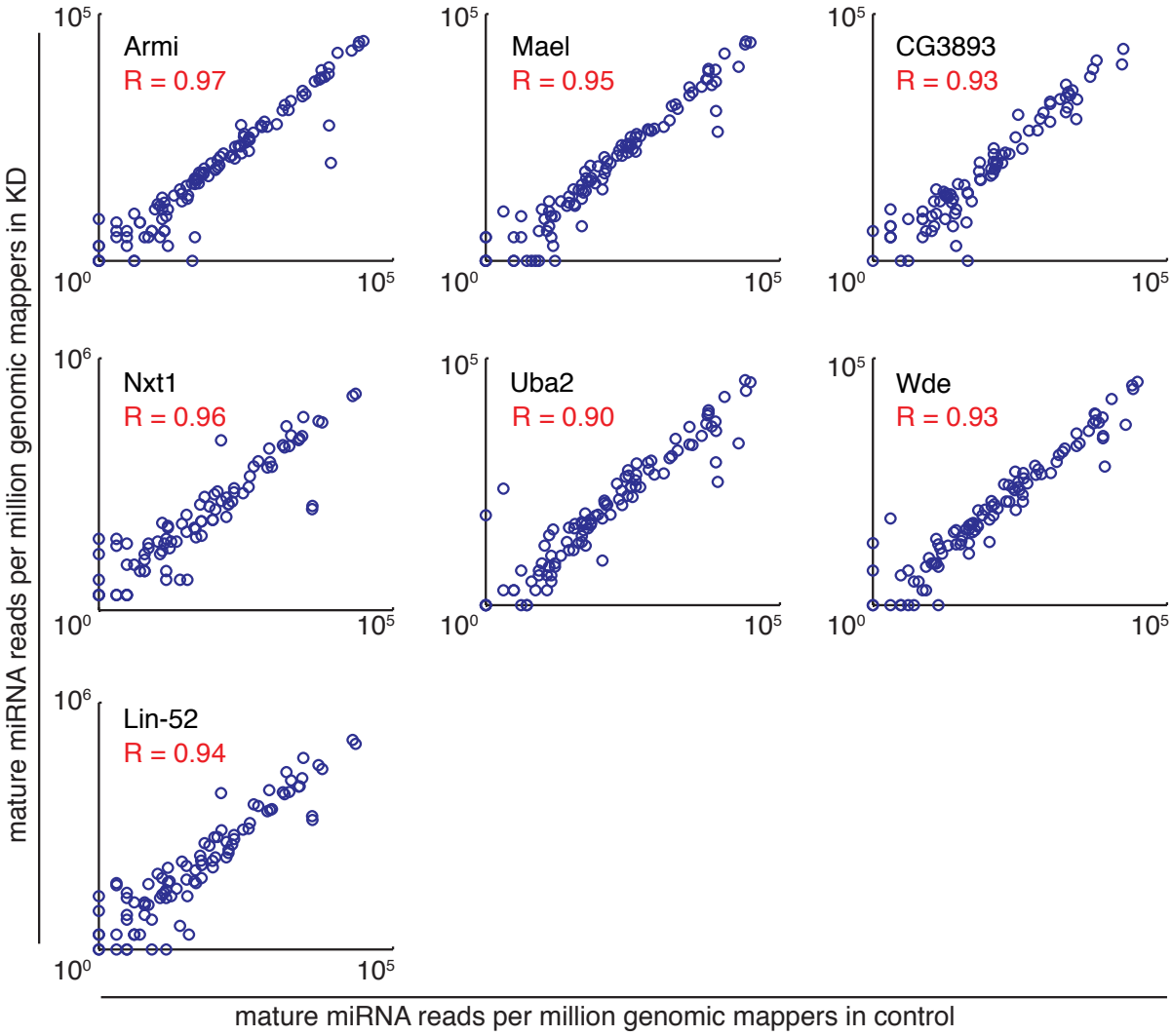


Figure 3.8: **miRNA populations in knockdowns, related to Figure 3.7** There are no significant changes in microRNA level upon knock down of a number of validated hits. Scatter plots for mature microRNA in reads per million genomic mappers are shown for all follow-up genes compared to the Aub negative control. Pearson correlation coefficients are shown for each population.

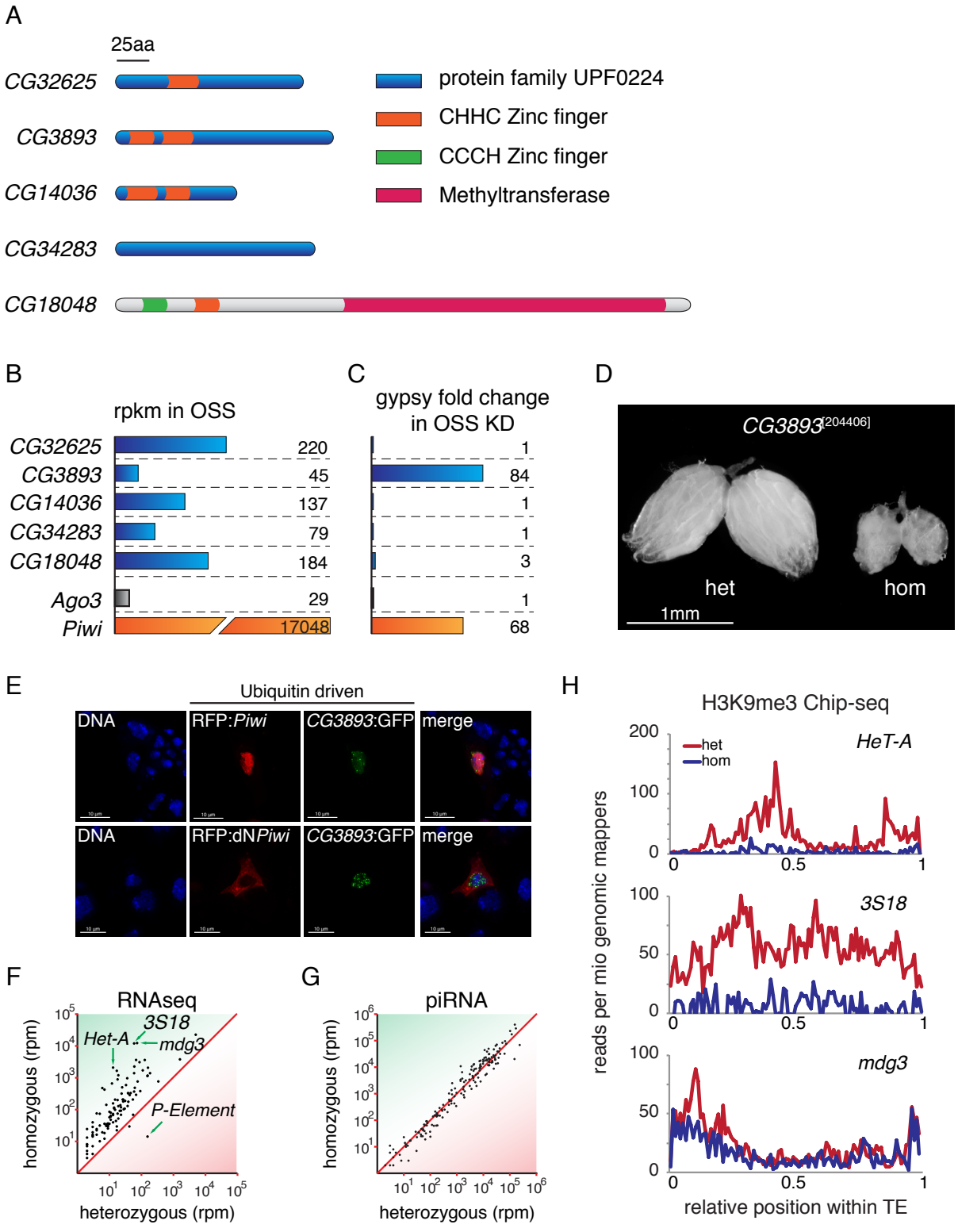
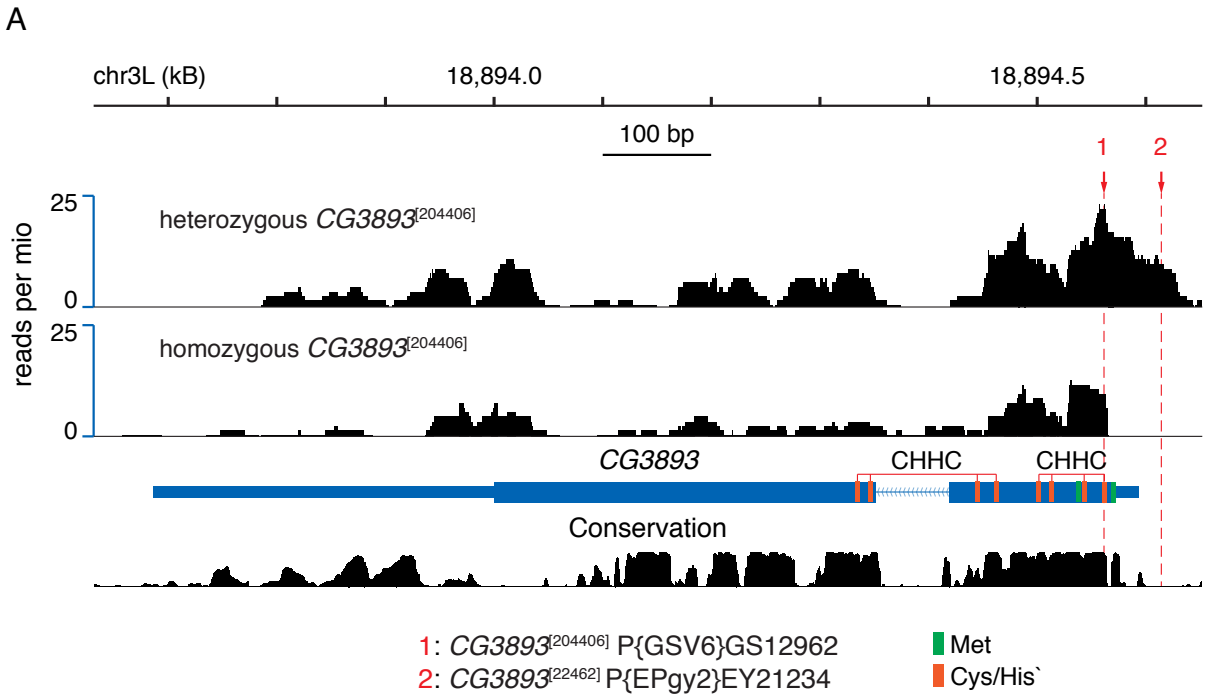


Figure 3.9: **Disruption of CG3893 function has a severe impact on transposon silencing**

Figure 3.9: (Continued) A) The five members of the *Drosophila* uncharacterized protein family UPF0224 and their domain structures. The conserved domains are highlighted as colored boxes. B) All five family members are weakly expressed in OSS cells. Piwi and Ago3 expression levels are shown for comparison. Expression levels are based on the modENCODE cell line expression data and are displayed as reads per kilobase per million mapped reads (rpkm). C) CG3893, but no other members of its protein family, has a strong impact on transposon silencing upon knockdown in OSS cells. Effects of knockdown of Ago3 and Piwi are shown for comparison. Numbers represent fold changes of *gypsy* levels in respect to the median fold change of the corresponding plate in the primary screen. D) Ovarian morphology of flies heterozygous or homozygous for a P-element insertion in CG3893 (204406, Kyoto Drosophila Genetic Resource Center). For a more detailed view of the insertion and expression levels see Figure 3.10. E) Tagged CG3893 co-localizes with Piwi in the nucleus of OSS cells when overexpressed in transient transfections. Nuclear Hoechst staining is blue, GFP tagged CG3893 is green and RFP tagged Piwi or Delta-NT-Piwi is shown in red (Saito, 2009). F) Transposons are highly upregulated upon disruption of CG3893 in the P-element insertion line. A scatter plot of reads per million (rpm) values for all transposons mapped to consensus is shown for RNA-seq of heterozygous versus homozygous flies. Each dot represents aggregated data for one consensus sequence. Only sequences mapping in sense direction are taken into account. G) piRNA levels are not affected by CG3893 disruption. Scatter plot showing levels of piRNA reads mapped to the same transposon consensus sequences as in F) are expressed in reads per million. H) Levels of H3K9me3 on a subset of transposons decrease dramatically upon depletion of CG3893. Density plots for normalized H3K9me3 ChIP-seq reads over three transposons, *HeT-A*, *3S18* and *mdg3* are shown. Red lines correspond to levels in heterozygous flies and blue lines to the homozygous state.

RNA libraries from females heterozygous and homozygous for the exonic P-element insertion. RNA-seq revealed a remarkable change in global transposon expression. Almost all classes of annotated transposons populating the *Drosophila* genome except for the *P-element* itself show upregulation in homozygous females when compared to their heterozygous sisters (Figure 3.9F). This derepression effect is equally strong for germline and soma dominant transposon classes. Yet, when mapping antisense piRNA reads to transposons consensus, we see no change in the homozygous animals (Figure 3.9G). In their recent publication, Brennecke and colleagues not only show that piRNA mediated transposon silencing is a nuclear phenomenon occurring through TGS, but also that it acts through deposition of silencing H3K9 trimethyl marks on active copies of transposons (Sienski et al. 2012). Given its potential involvement in this step, we sought to investigate the effects of CG3893 disruption on this histone mark by performing ChIP-seq analysis for H3K9me3 on ovaries from heterozygous and homozygous



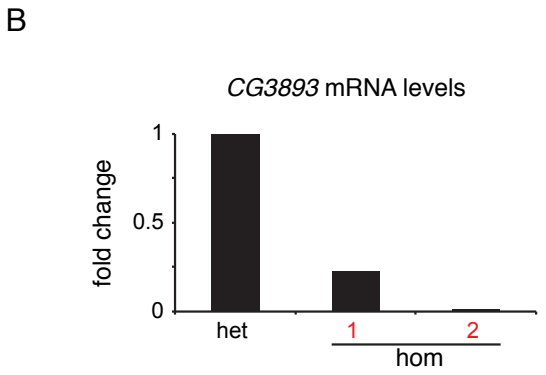


Figure 3.10: **Two P-element insertion disrupt CG3893 function, related to Figure 3.9** A) Density plots of reads mapping to CG3893 from RNA-seq libraries corresponding to the heterozygous and homozygous CG3893[204406] insertion line are shown. 1 and 2 shown in read designate the insertion point of P-element in CG3893[204406] and CG3893[22462], respectively. Beneath, the FlyBase gene model for CG3893 is shown with green boxes designating Met translation start sites and in red boxes are positions of Cys and His amino acids that make up the CHHC zinc fingers. Under the gene model, conservation is shown. B) Both CG3893 P-element insertion lines disrupt expression of the transcript but to different extents. qPCR for levels of CG3893 transcript in heterozygous and homozygous flies are shown. Each homozygous fly is normalized to its corresponding heterozygous sibling. 1 corresponds to CG3893[204406] and 2 to CG3893[22462].

females. Strikingly, homozygous females showed marked reduction of H3K9me3 levels over a subset of transposons (Figure 3.9H). Remarkably, the subset with the most profound drop in H3K9me3 levels corresponds to the most highly derepressed elements in RNA-seq data. Even though a lower level of this mark on transposons was a general trend, not all elements showed changes in levels as pronounced as for *HeT-A*, *3S18* and *mdg3*. This is not surprising, since only young, active and potentially euchromatic copies of transposons are expected to show this effect. In fact, by mapping to consensus sequences we most likely mask some of the changes caused by the absence of CG3893. With this in mind, it is remarkable to detect changes on the order of magnitude so far only reported for disruption of Piwi itself. Because of its small size yet powerful role in transposon silencing we name CG3893 Asterix (Arx).

3.4 Discussion

Transposon control in the germline of animals is of critical importance in order to ensure the integrity of the inheritable genome and consequently the wellbeing of progeny (Malone and Hannon 2009). In our current understanding, the piRNA pathway is essential in establishing this control (Siomi et al. 2011). Even though discovered almost two decades ago, very little is understood about the precise molecular steps necessary for piRNA biogenesis and successful silencing of transposons (Guzzardo et al. 2013). In an effort to shed light on all possible steps from piRNA biogenesis to general transposon control, we preformed an unbiased, extensive genome-wide RNAi screen in ovarian somatic cells. The primary *in vitro* screen proved to be a robust and specific assay for transposon derepression, with all expected piRNA components scoring highly. We identify many enriched functions needed for *gypsy* control, such as genes participating in RNA splicing and sumoylation. Although their involvement was unexpected, these functions can easily be tested for their potential role in the piRNA pathway because both tools and experience regarding these complexes is plentiful within the scientific community. To test the validity of the primary dataset and to gain a more detailed picture, we further

examined our top candidate hits *in vivo*. Within our list of 87 validated genes, we have promising candidates to fill almost every gap in our current understanding of the pathway. For example Nxt1, a nuclear RNA export factor, could be responsible for the export of primary cluster transcripts to the cytoplasm. And indeed, in sRNA-seq from knockdowns of Nxt1, we see a drastic reduction of mature piRNAs, which would be the expected phenotype when compromising such a factor. It could also be hypothesized that export of Piwi mRNA or of other key players of the pathway is comprised upon Nxt1 knockdown. However, the severity of the observed phenotypes renders such an indirect involvement rather unlikely. UAP56, an RNA helicase implicated in the transport of the primary transcript to the nuclear pore, was previously shown to act in the same export pathway as Nxt1 (Herold et al. 2003). Given that we also identify two nuclear pore components, an obvious hypothesis is that these genes are acting together to export piRNA precursors to the cytoplasm.

We also identify genes that are likely to affect transcription of these precursors. Primary cluster transcripts are believed to be transcribed by RNA polymerase II. Even though the potentially indirect effects of knocking down this gene are self-evident, we do see transposon derepression upon depletion of the 140kD subunit RpII140. Two other genes possibly affecting transcription of piRNA clusters whose roles are easier to dissect are Lin-52 and Wde. The latter was shown to be a co-factor of Eggless, a gene required for transcription of clusters (Koch et al. 2009; Rangan et al. 2011). However, the severe effects that we see in Wde knockdown even for *gypsy*, which was not expected for Egg depletion based on existing literature, hints towards a role for Wde independent of Egg.

Once transcribed, it is hard to imagine that the primary transcript of clusters, sometimes as long as 200kB, shuttle through the nuclear pore without prior processing into smaller fragments. We observe an overall enrichment for splicing factors in the primary dataset, which may indicate co-transcriptional splicing of the primary transcript into smaller precursors.

After export, those smaller precursors could further be processed by an endonuclease to create the 5′ end of the mature piRNA. Zuc, which was recently shown to be a single-

stranded RNA specific endonuclease, is the most likely candidate for this function (Ipsaro et al. 2012; Nishimasu et al. 2012). The fact that we do not identify any other endonuclease with comparable derepression phenotypes in our screen further cements the role for Zuc in this step. RNase P and RNase Z, both endonucleases implicated in tRNA processing, did score in the primary screen, but could not be validated *in vivo* because of their severe developmental defects (Dubrovsky et al. 2004; Frank and Pace 1998). After 5′ end formation and loading into Piwi, each piRNA has to be further trimmed to its mature length. Work by Tomari and colleagues demonstrated that this is done by Mg^{2+} dependent exonuclease activity (Kawaoka et al. 2011). The only genes exhibiting exonuclease activity and scoring highly in our screen were Csl4 and Rrp6, both members of the exosome (Andrulis et al. 2002). However, neither of the two genes could be validated *in vivo* due to arrested gonadal development in knockdowns. While the exosome remains an intriguing possibility for being involved in piRNA 3′ end formation, recent work from Grewal and colleagues showed unexpected phenotypes in Rrp6 mutant larvae: Loss of Rrp6 caused accumulation of small RNAs and derepression of transposons (Yamanaka et al. 2012). These small RNAs had similar size profiles to piRNAs and a strong 5′ uracil bias. One would expect that disrupting the 3′ trimming activity leads to accumulation of longer species and depletion of mature piRNA sizes. However, since these experiments were done in tissues with very low Piwi expression, further investigation in germline tissues should yield exciting new insight into the exosome's function.

At the onset of the screen, we did expect to identify a single exonuclease directly necessary for 3′ end formation, similar to 5′ end generation and Zuc. Not identifying such a candidate has several possible explanations: First, there could be genetic redundancy within this function, as no existing data rules out this possibility (Kawaoka et al. 2011). Second, even though electronic annotations of molecular function based on sequence homology is a powerful tool, the 3′ exonuclease could be hidden within our hit lists, not annotated as such and thus escaping our attention. Third, though unlikely, it is still conceivable that loss of the trimming activity

does not lead to derepression of transposons. Piwi could still be loaded with longer, premature piRNA species, shuttle back into the nucleus and exert its function.

Transcriptional silencing of transposons through Piwi is a nuclear process and previous data demonstrated that unloaded Piwi remains in the cytoplasm (Saito et al. 2010). One protein possibly involved in this reimport of loaded Piwi is Karybeta3, a homolog of importin 5 (Mosammaparast and Pemberton 2004). Further examination of the role these genes may play in Piwi reimport into the nucleus or export of piRNA cluster transcripts to the cytoplasm could be one of the interesting projects emerging from this screen.

Upon reentry into the nucleus Piwi is able to recognize transcription of active transposons through its bound piRNA and consequently silence them (Sienski et al. 2012). So far, only Piwi itself and Mael have been implicated in this step. With Asterix, we present a promising new candidate indispensible for transposon silencing. The degree to which transposons are derepressed on a global scale in mutant animals is extraordinary. Together with its putative nuclear localization and since both knockdown and mutant flies do not show changes in mature piRNA levels, it was easy to hypothesize a possible involvement in deposition of H3K9me3 marks. However, even though we see lower levels of this mark in mutant animals compared to heterozygous siblings, Asterix most likely is not responsible for depositing these marks. The only conserved domains within the protein are predicted to be RNA binding. This still leaves the door open for chromatin remodelers and methyltransferases in the big picture of transposon silencing through TGS. The obvious choice for the latter, Eggless, had no strong effect on *gypsy* expression in mutant animals or our assays (Rangan et al. 2011). All other hits annotated as involved in heterochromatin formation (HP1 and Mi-2), could not be validated due to developmental defects (Vermaak and Malik 2009; Brehm et al. 2000). This does not necessarily rule them out for occupying this role in the pathway, but further investigation is needed in order to conclusively tackle this issue.

In summary, our unbiased, genome-wide approach was successful in identifying likely candidates to fill in many of the open questions regarding molecular mechanisms of transposon

control in *Drosophila*. We strengthen our current hypothesis that the piRNA pathway is the major force exerting this function, given that both the primary and validation screen were dominated by known components of the pathway. Our meta-analysis on the primary dataset as a whole, as well as the list of validated genes, provides both starting point and directions for future research. It is our hope that this study will serve as a resource to the community and lead to a plethora of new and insightful research in this exciting field of transposon biology.

3.5 Experimental Procedures

Cell culture

OSS cell were cultured as previously described and transfected using Xfect transfection reagent according to manufacturer's guidelines (Niki et al. 2006) (Clontech cat 631317).

DNA plasmids

Expression vectors of CG3893:GFP, RFP:Piwi and RFP:ΔNTPiwi were made using the Drosophila Gateway Collection.

Imaging of fluorescent fusion proteins in OSS

OSS cells were co-transfected with plasmids expressing the indicated fusion proteins using Cell Line Nucleofector kit V (Amaxa Biosystems; program T-029). Fixed cells were stained with Hoechst 33342 (Invitrogen, R37601).

RNAi libraries

Two *Drosophila* dsRNA libraries were used in this study, the Open Biosystems Drosophila RNAi Collection and the Drosophila RNAi Screening Center Genome-wide RNAi library (DRSC 2.0).

RNAi screening

A detailed description of the primary screen can be found in extended experimental procedures. A basic workflow is shown in Figure 3.1.

Drosophila stocks and husbandry

Fly stocks are listed in Table 3.4. A description of husbandry and validation screen procedures can be found in extended experimental procedures.

RNA isolation and qPCR assays

Total RNA from 10 ovaries was extracted with Trizol and purified by phenol chloroform extraction followed by isopropanol precipitation. After DNase treatment, cDNA was synthesized from 800ng RNA using oligo dT primers and Superscript III Reverse Transcriptase (Life Technologies). qPCR was performed to assay levels of *gypsy*, *ZAM*, *gypsy3* and rp49. Fold changes for transposons were calculated using the delta Ct method (Livak and Schmittgen 2001).

RNA-seq and analysis

For RNA-seq libraries, 2.5-5ug of total RNA was depleted of ribosomal RNA using the Epicenter Ribo-Zero rRNA Removal Kits (Human/Mouse/Rat), following the manufacturer's directions. Libraries were prepared using the Illumina Script Seq v2 RNA-Seq library preparation kit and were sequenced on an Illumina HiSeq platform. Details on Analysis can be found in extended experimental procedures.

Small RNA cloning and analysis

For small RNA libraries, total RNA was depleted of 2S rRNA and libraries were constructed using the Illumina Tru Seq small RNA sample Prep kit following the manufacturer's protocol. Details on Analysis can be found in extended experimental procedures.

ChIP-seq

ChIP from 50 ovaries was done as described in Ram et al and Garber et al, with some modifications (Ram et al. 2011; Garber et al. 2012). Details on the methodology and analysis can be found in extended experimental procedures.

Statistical procedures

Details on enrichment analysis and statistical procedures can be found in extended experimental procedures.

3.6 Acknowledgments

We are greatly indebted to Caifu Chen, Richard Fekete (Life Tech.) and Norbert Perrimon (DGRC) for giving us reagents. We are very grateful to Sabrina Boettcher who managed all logistics and supplies. We would like to thank Alon Goren for invaluable help with ChIP-seq cloning. Advice given by Molly Hammell has been a great help in analyzing RNA-seq and ChIP-seq data. Leah Sabin always provided helpful insight and advice. We wish to thank Stephanie Muller, Assaf Gordon and Astrid Haase for help with robotics and library normalization. We would like to thank Ben Czech, Jon Preall and Sho Goh for sharing data prior to publication. Assistance with sequencing provided by Emily Lee was greatly appreciated. We also would like to thank Jo Leonardo and all members of the Hannon lab for vital support. P.M.G. is a NIH trainee on CSHL WSBS NIH Kirschstein-NRSA pre-doctoral T32 GM065094, and is a William Randolph Hearst Scholar and Leslie Quick Junior Fellow. A grant from T. and V. Stanley supported J.G.'s work. Work in the Hannon laboratory is supported by grants from the National Institutes of Health (5R01GM062534) and by a kind gift from Kathryn W. Davis. G.J.H. is an investigator of the HHMI.

Table 3.3: Primary dataset. This table is available upon request.

Table 3.4: Validated dataset. This table is available upon request.

3.7 Supplemental Information

Extended Experimental Procedures

Cell culture

OSS cell were cultured in Shields and Sang M3 Insect media (Sigma) supplemented with 10% FBS, 5% fly extract, 0.6mg/ml glutathione and 10mg/ml insulin as previously described (Niki et al. 2006). Cells were transfected using Xfect transfection reagent according to manufacturer's guidelines (Clontech, 631317).

DNA plasmids

Expression vectors of CG3893:GFP, RFP:Piwi and RFP:ΔNTPiwi driven by an ubiquitin promoter were made using the Drosophila Gateway Collection (Terence Murphy, Carnegie Institute of Washington, Baltimore, MD). To construct expression clones, coding sequences of CG3893, Piwi and ΔNTPiwi (excluding the first 72 aa) were PCR-amplified from ovarian cDNA and cloned into pENTR/ D-TOPO, and then recombined with either destination vector pURW (DGRC1282), for Piwi and ΔNTPiwi or pUWG (DGRC 1284), for CG3893.

Imaging of fluorescent fusion proteins in OSS

OSS cells were co-transfected with plasmids expressing the indicated fusion proteins using Cell Line Nucleofector kit V (Amaxa Biosystems; program T-029). 48 hours after transfection, cells were plated on glass coverslips. 24 hours later, cells were stained with Hoechst 33342 (NucBlue live cell stain; Invitrogen, R37601) and immediately fixed in 2% formaldehyde/PBS at room temperature for 5min. After three 10min PBS washes, coverslips were mounted in proLong antifade (Invitrogen, P7481) and examined under a fluorescent microscope (Nikon Eclipse Ti).

Z-stack images were taken with 40X magnification and the final images were de-convoluted under the default manufacturer settings.

RNAi libraries

Two *Drosophila* dsRNA libraries were used in this study, the Open Biosystems (now Thermo Scientific) Drosophila RNAi Collection version 1.0/2.0 and the Drosophila RNAi Screening Center Genome-wide RNAi library (DRSC 2.0).

RNAi screening

OSS cells were plated in 48-well dishes (79,000 cells/well). The following day cells were transfected with 500ng of dsRNA, 0.3µl Xfect reagent and 9.7µl Xfect Buffer. To do this procedure in a robust way the Epmotion robot (Eppendorf) was used to prepare the transfection mixture in a 96-well plate and to pipette the mixture onto the cells. Approximately 12 hours post transfection, cells were washed with PBS and media was replaced. An additional media change was done on day 3 post-transfection to avoid drying of wells. On day 5 post-transfection cells were lysed with 150ul of Lysis Buffer (10mM KCl,10mM Tris pH8, 1.5mM MgCl2, 0.5% NP-40, 60 units RNasin) per well and shaken for 5min at 300 rpm. For the DRSC library, instead of the Lysis Buffer, Ambion Cells-to-Ct Lysis Reagent (Life Technologies cat 4391848M) was used to lyse cells. Following the 5 minutes of shaking, 15µl of Stop Solution (Life Technologies cat 4402960) was added to stop the lysis reaction, mixed by pipetting and left for 2 minutes at room temperature. Lysates were transferred to a 96-well PCR plate. 22.5µl of the lysate was used as input for a 50µl reverse transcription (RT) reaction and then incubated at 37ºC for 1 hour and 95ºC for 5 min. The RT master mix and enzyme used, were those provided in the TaqMan Gene expression Cells-to-CT kit (4399002). Both the transfer of the lysate to 96-well plates, as well as the RT reaction set-up was done using the Epmotion. After cDNA synthesis, 2µl of the cDNA was used as input in a qPCR reaction to assay levels of *gypsy* and cib in a multiplexed reaction, using TaqMan Fast Advanced Master Mix (Life Technologies cat 4444965) on an Eppendorf MasterCycler EP realplex machine. Levels of *gypsy* subgenomic

transcript were assayed using hydrolysis probes spanning the splice junction. Primers and probes are listed in Table 3.5.

Table 3.5: Table of oligos

Name	Primer sequence
Target	Sequence (5′-3′)
Primary Screen	
gypsy fwd	CCAACAATCTGAACCCACCAATCTA
gypsy rev	AGTACCCGCCACAACCTTTAAG
gypsy probe	CAAACAGGGTAGTTAAGTTAG
cib fwd	GCCAGCATCCCAGCTTAGTAGT
cib rev	GCTGGGGCGGCCATCTT
cib probe	CGCTTCGCCAATCCA
Validation Screen	
gypsy fwd	CAGGCGACAAACAGGGTAG
gypsy rev	GTTCAAACACCAGCACATCC
gypsy probe	ACACAGGAATGTAGTTGGCATGCGA
gypsy3 fwd.	GACATACTGAAGGGCGAGAAC
gypsy3 rev	TCAGGGTATCTAAGGGTGACG
gypsy3 probe	CAAGGTAGAATTTTCCGAAGCGCAGC
ZAM fwd	GGTATGGAAGATGTGGGTGTC
ZAM rev	TCCTCTTCACCGTATCCCTAG
ZAM probe	TCGCCGTAATACTCACCTGGACACT
rp49 fwd	GTCGGATCGATATGCTAAGCTG
rp49 rev	CAGATACTGTCCCTTGAAGCG
rp49 probe	TTGTCGATACCCTTGGGCTTGCG
General qPCR primers and probes	
CG3893 fwd	TCGTCATCCCAGTTCTCCT

Table 3.5: Table of oligos (continued)

Name	Primer sequence
CG3893 rev	CATTTGATACCAGAGCCCCAG
CG3893 probe	CGAAGACACCAGACACGCGAAGAT

***Drosophila* stocks and husbandry**

For crosses in the validation round we used tj-GAL4 (DGRC stock 10455); GS12962 (DGRC stock 204406) and EY21234 (Bloomington stock 22462) are P-element insertions into the CG3893 locus. The 328 fly stocks corresponding to the candidate hits were ordered from Vienna Drosophila Resource Center (VDRC) and the Drosophila RNAi Resource center. The trans-IDs used by VDRC are listed in Supplementary Table 3.4. Lines from the DGRC are indicated with the prefix TRIP. For all crosses performed during the validation screen, five tj-GAL4 females and three VDRC hpRNA males were crossed and left in vials for five days, when parental flies were removed from the vial. Eight days after, ten female and three male F1 flies were put into new vials with yeast. After two days, ovaries from female flies were dissected. Eight days later, we checked the vials for the presence of larvae to test for fertility.

RNA isolation and qPCR assays

Ovaries from 10 F1 flies were dissected for each cross. Ovaries were washed once with cold PBS and homogenized in 1 ml of Trizol reagent. Total RNA was purified by phenol chloroform extraction followed by isopropanol precipitation according to the Trizol protocol. RNA was then subjected to DNase treatment using Ambion Turbo DNA-free kit at 37ºC for 30 minutes according to the manufacturer's protocol (Life Technologies). cDNA was synthesized with 800ng RNA as input using oligo dT primers (dT20) and Superscript III Reverse Transcriptase (Life Technologies) at 50ºC for 50 minutes, followed by 15 minutes at 70ºC. Next, qPCR was performed to assay levels of *gypsy*, *ZAM*, *gypsy3* and *rp49*. Using hydrolysis probes with

FAM and HEX fluorescent reporters, we multiplexed the qPCR for the transposon and rp49. Primers and probes are listed in Table 3.5. Fold changes for transposons were calculated using the delta Ct method (Livak and Schmittgen 2001). In the case of the GD library we compared each knockdown to an average of 5 biological replicates of White negative controls, for the KK library we used 5 biological replicates of Aub negative controls. All primers were tested for efficiency in single and multiplexed reactions. Only primers for which efficiency was not impaired in the multiplexed reactions were used.

RNA-seq and analysis

For RNA-seq libraries, 2.5-5ug of total RNA was depleted of ribosomal RNA using the Epicenter Ribo-Zero rRNA Removal Kits (Human/Mouse/Rat), following the manufacturer's directions. Libraries were prepared using the Illumina Script Seq v2 RNA-Seq library preparation kit and were sequenced on an Illumina HiSeq platform for 36 cycles in a single end run. After collapsing all reads into a non-redundant list (cloning counts were preserved), they were mapped to *Drosophila* viral, tRNA and miscRNA (rRNA, snoRNA etc) sequences using the short read aligner bowtie (Langmead et al. 2009). Only sequences in each library that did not map to either of these contaminants were then mapped to the *Drosophila* genome with up to two mismatches. Additionally, only uniquely mapping sequences were considered for further analysis. The same reads were mapped to a custom index of transposon consensus sequences with up to 3 mismatches (Kaminker et al. 2002). Reads mapping to up to 2 locations were considered for further analysis. For differential expression analysis of transposons we aggregated read counts mapping to these consensus sequences in sense orientation. For differential expression analysis of genes, we used htseq-counts (Part of the 'HTSeq' framework, version 0.5.3p3) to assess read counts per gene. In both cases we used the R package DESeq to call differential expression at a FDR cutoff of 0.05 based on two biological replicates.

Small RNA cloning and analysis

For small RNA libraries, 2.5 μg of total RNA was depleted of the 2S rRNA by annealing an

antisense primer (Table 3.5, 95ºC to 25ºC in ~1h) followed by RNase H digestion at 37ºC for 30 minutes in 5X FS buffer (from Superscript III Reverse Transcriptase Kit, Life Technologies; RNase H was from NEB, M0297S). The remaining RNA was used as input. Libraries were constructed using the Illumina TruSeq small RNA sample Prep kit following the manufacturer's protocol. For analysis of sRNA populations of CG3893 heterozygous and mutant animals, we used 50 ng of size selected RNA (19-28nt) as input. After sequencing on a Illumina HiSeq single-end 36 run, the TruSeq adapter (TGGAATTCTCGGGTGCCAAGGAACTCCAGTCAC) was clipped from the 3' end of the read and sequences shorter than 16 nt were discarded from further analysis. The remaining sequences were collapsed into a non-redundant list and mapped to tRNA and miscRNA sequences using the short read aligner bowtie (Langmead et al. 2009). Only non-mapping reads were consequently mapped to the D. melanogaster genome (D. melanogaster Apr. 2006 [BDGP R5/dm3]). Up to two mismatches were allowed. Read counts of uniquely mapping reads were normalized to reads per million genomic mappers and compared to a negative control: in the case of knockdowns using long hpRNAs from the VDRC KK libraries we used Aub (106999^{KK}), in the case of GD libraries we used White (30033^{GD}). The rankings displayed in Figure 3.7B are calculated based on aggregated read counts of unique mappers to piRNA clusters defined in Brennecke et al., 2007. For size profiles, we used the same negative control libraries for comparison, which were normalized to the same scale in order to accurately compare across knockdowns. For analysis of transposons we aggregated read counts mapping to consensus sequences in sense orientation.

ChIP-seq

ChIP was done as described in Ram et al and Garber et al, with some modifications (Ram et al. 2011; Garber et al. 2012). Approximately fifty ovaries were dissected from heterozygous or homozygous flies into cold PBS and washed once with PBS. Ovaries were then fixed in 1.8% formaldehyde for 10 minutes, then quenched by adding glycine to 0.125M and immediately placed on ice. Tissue was then homogenized by douncing five times with pestle A (Kontes).

Washed once with PBS supplemented with protease inhibitors (Roche) and pellet was flash frozen in liquid nitrogen. Pellets were then thawed on ice and resuspended in 1mL Lysis Buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl, pH 8.1) and lysed for 10 minutes in ice. Chromatin was sheared to 200-800bp using a Branson sonifier (model S-450D). After clearing lysate by centrifugation, 9mLs of Dilution Buffer (0.01% SDS, 1.1%Triton X-100, 1.2mM EDTA, 16.7mM Tris-HCl, pH 8.1, 167mM NaCl) were added to the lysate and 5mLs of the lysate were incubated with a 50ul of an equal mixture of conjuagted protein A and G Dynabeads (Invitrogen). To conjugate beads, they first had been washed once in Blocking Buffer (1X PBS, 0.5% TWEEN 20, 0.5% BSA), then coupled for 1 hour at 4ºC with 5ug of H3K9me3 antibody (Abcam 8898) and finally washed twice with Blocking Buffer to remove excess antibody. Lysate and conjugated magnetic beads were rotated at 4ºC overnight. Beads were then resuspended in 200ul cold RIPA buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 14 mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% DOC) and transfer to a 96-well plate. All further separation steps were preformed in the 96-well plate magnet. Beads were washed five times with 200ul cold RIPA, two times with RIPA buffer supplemented with 500 mM NaCl, two times with LiCl buffer (10 mM TE, 250mM LiCl, 0.5% NP-40, 0.5% DOC), and once with TE (10Mm Tris-HCl pH 8.0, 1mM EDTA). Samples were eluted in 50 µl of 0.5% SDS, 300 mM NaCl, 5 mM EDTA, 10 mM Tris HCl pH 8.0. The eluate was reverse cross-linked at 65ºC for 4 hours and then treated with 2ul of RNaseA (Roche, 11119915001) for 30 min followed by 2.5 µl of Proteinase K (NEB, P8102S) for two hours. Library preparation was done as indicated in Garber et al, but without automation. In brief, to purify DNA 120ul of Ampure XP beads (Agencourt) were added to the reverse cross-linked samples, mixed by pipetting and incubated for 2 minutes. Samples were then placed on the magnetic stand for 4 minutes to separate beads, followed by 2 washes with 70%ethanol and air dried for 4 minutes and eluted in 10mM Tris-HCl pH8.0. Library was constructed by performing DNA end-repair, A-base addition, adaptor ligation and enrichment PCR. After each step DNA was purified by adding 20% PEG to the reaction, to allow DNA to bind to Ampure XP beads already in the tube. Samples were not moved form their original

well position, until after PCR enrichment. The libraries were sequenced on the Illumina Miseq platform for 36 cycles in a single-end run. Mapping procedures were done as described for the RNA-seq libraries. For the analysis on differential H3K9me3 deposition over transposons, each consensus sequence for HeT-A, 3S18 and mdg3 was divided into 100 bins. Aggregating read counts over each bin allowed us to subtract input signal from IP for each bin. The difference was then plotted as a density curve over all bins. Bins with negative density (due to input subtraction) were set to a density of 0.

Statistical procedures

Enrichment analysis was conducted using 2714 gene sets from the gene ontology (Barrell et al. 2009; Ashburner et al. 2000). This constituted the complete complement of gene sets in GO with between five and 100 *Drosophila* genes annotated to them in either the cellular component or biological process branch of GO. Molecular function substantially overlapped with biological process in many top functions and was excluded to diminish redundancy. Significance was calculated using an adaptation of the ROC-based approach described in (Gillis et al. 2010) and elsewhere. To obtain a ranking for the genes, dsRNA z-scores and fold changes were independently averaged for each gene. These scores were then converted into ranks and averaged (effectively weighting them equally). Based on the ROC_{50} approach first described in (Gribskov and Robinson 1996), all scores outside of the top 50 were regarded as tied. Statistical enrichment of the GO functions was then calculated (Mann-Whitney U test) with multiple test correction (Benjamini and Hochberg 1995). Despite the significant penalty imposed by the large number of gene sets, 215 functions were enriched (corrected $p<0.05$), many with strong potential relevance and very significant enrichment (top 20 functions have corrected $p<1E-6$). In contrast, thresholding at the equivalent level (50) yielded only one significantly enriched function, Yb body. Using the entire ranking gave significance over 38 functions including exosome complex and proteasome complex, but otherwise tended to be dominated by the

aggregate weak effects of relatively low ranking genes (causing enrichment for, e.g., ribosomal functions).

3.8 References

Andrulis ED, Werner J, Nazarian A, Erdjument-Bromage H, Tempst P, Lis JT. 2002. The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. *Nature* **420**: 837-841.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25-29.

Barrell D, Dimmer E, Huntley RP, Binns D, O'Donovan C, Apweiler R. 2009. The GOA database in 2009–an integrated Gene Ontology Annotation resource. *Nucleic Acids Res* **37**: D396-403.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B ...* **57**: 289-300.

Brehm A, Längst G, Kehle J, Clapier CR, Imhof A, Eberharter A, Müller J, Becker PB. 2000. dMi-2 and ISWI chromatin remodelling factors have distinct nucleosome binding and mobilization properties. *EMBO J* **19**: 4332-4341.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell* **128**: 1089-1103.

Bucheton A. 1995. The relationship between the flamenco gene and gypsy in Drosophila: how to tame a retrovirus. *Trends Genet* **11**: 349-353.

Dietzl G, Chen D, Schnorrer F, Su K-C, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S, Scheiblauer S, et al. 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. *Nature* **448**: 151-156.

Dubrovsky EB, Dubrovskaya VA, Levinger L, Schiffer S, Marchfelder A. 2004. Drosophila RNase Z processes mitochondrial and nuclear pre-tRNA 3′ ends in vivo. *Nucleic Acids Res* **32**: 255-262.

Frank DN, Pace NR. 1998. Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu Rev Biochem* **67**: 153-180.

Garber M, Yosef N, Goren A, Raychowdhury R, Thielke A, Guttman M, Robinson J, Minie B, Chevrier N, Itzhaki Z, et al. 2012. A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol Cell* **47**: 810-822.

Georlette D, Ahn S, Macalpine DM, Cheung E, Lewis PW, Beall EL, Bell SP, Speed T, Manak JR, Botchan MR. 2007. Genomic profiling and expression studies reveal both positive and negative activities for the Drosophila Myb MuvB/dREAM complex in proliferating cells. *Genes & Development* **21**: 2880-2896.

Gillis J, Mistry M, Pavlidis P. 2010. Gene function analysis in complex data sets using ErmineJ. *Nat Protoc* **5**: 1148-1159.

Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, et al. 2011. The developmental transcriptome of Drosophila melanogaster. *Nature* **471**: 473-479.

Gribskov M, Robinson NL. 1996. Use of receiver operating characteristic (ROC) analysis to evaluate sequence matching. *Comput Chem* **20**: 25-33.

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

Guzzardo PM, Muerdter F, Hannon GJ. 2013. The piRNA pathway in flies: highlights and future directions. *Current Opinion in Genetics & Development*.

Haase AD, Fenoglio S, Muerdter F, Guzzardo PM, Czech B, Pappin DJ, Chen C, Gordon A, Hannon GJ. 2010. Probing the initiation and effector phases of the somatic piRNA pathway in Drosophila. *Genes & Development* **24**: 2499-2504.

Herold A, Klymenko T, Izaurralde E. 2001. NXF1/p15 heterodimers are essential for mRNA nuclear export in Drosophila. *RNA* **7**: 1768-1780.

Herold A, Teixeira L, Izaurralde E. 2003. Genome-wide analysis of nuclear mRNA export pathways in Drosophila. *EMBO J* **22**: 2472-2483.

Horwich MD, Li C, Matranga C, Vagin V, Farley G, Wang P, Zamore PD. 2007. The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr Biol* **17**: 1265-1272.

Ipsaro JJ, Haase AD, Knott SR, Joshua-Tor L, Hannon GJ. 2012. The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature*.

Ishizu H, Siomi H, Siomi MC. 2012. Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. *Genes & Development* **26**: 2361-2373.

Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, Frise E, Wheeler DA, Lewis SE, Rubin GM, et al. 2002. The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. *Genome Biol* **3**: RESEARCH0084.

Kawaoka S, Izumi N, Katsuma S, Tomari Y. 2011. 3′ end formation of PIWI-interacting RNAs in vitro. *Mol Cell* **43**: 1015-1022.

Khurana JS, Wang J, Xu J, Koppetsch BS, Thomson TC, Nowosielska A, Li C, Zamore PD, Weng Z, Theurkauf WE. 2011. Adaptation to P element transposon invasion in Drosophila melanogaster. *Cell* **147**: 1551-1563.

Kirino Y, Kim N, de Planell-Saguer M, Khandros E, Chiorean S, Klein PS, Rigoutsos I, Jongens TA, Mourelatos Z. 2009. Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. *Nat Cell Biol* **11**: 652-658.

Klattenhoff C, Xi H, Li C, Lee S, Xu J, Khurana JS, Zhang F, Schultz N, Koppetsch BS, Nowosielska A, et al. 2009. The Drosophila HP1 Homolog Rhino Is Required for Transposon Silencing and piRNA Production by Dual-Strand Clusters. *Cell* **138**: 1137-1149.

Koch CM, Honemann-Capito M, Egger-Adam D, Wodarz A. 2009. Windei, the Drosophila Homolog of mAM/MCAF1, Is an Essential Cofactor of the H3K9 Methyl Transferase dSETDB1/Eggless in Germ Line Development. *PLoS Genet* **5**: e1000644.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Lau NC, Robine N, Martin R, Chung W-J, Niki Y, Berezikov E, Lai EC. 2009. Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line. *Genome Res* **19**: 1776-1785.

Levin HL, Moran JV. 2011. Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet* **12**: 615-627.

Lewis PW, Beall EL, Fleischer TC, Georlette D, Link AJ, Botchan MR. 2004. Identification of a Drosophila Myb-E2F2/RBF transcriptional repressor complex. *Genes & Development* **18**: 2929-2940.

Lévesque L, Guzik B, Guan T, Coyle J, Black BE, Rekosh D, Hammarskjöld ML, Paschal BM. 2001. RNA export mediated by tap involves NXT1-dependent interactions with the nuclear pore complex. *J Biol Chem* **276**: 44953-44962.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402-408.

Ma L, Buchold GM, Greenbaum MP, Roy A, Burns KH, Zhu H, Han DY, Harris RA, Coarfa C, Gunaratne PH, et al. 2009. GASZ Is Essential for Male Meiosis and Suppression of Retrotransposon Expression in the Male Germline. *PLoS Genet* **5**: e1000635.

Makarov EM, Makarova OV, Urlaub H, Gentzel M, Will CL, Wilm M, Lührmann R. 2002. Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* **298**: 2205-2208.

Malone CD, Brennecke J, Dus M, Stark A, Mccombie WR, Sachidanandam R, Hannon GJ. 2009. Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. *Cell* **137**: 522-535.

Malone CD, Hannon GJ. 2009. Small RNAs as guardians of the genome. *Cell* **136**: 656-668.

McClintock B. 1950. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci USA* **36**: 344-355.

McQuilton P, St Pierre SE, Thurmond J, FlyBase Consortium. 2012. FlyBase 101–the basics of navigating FlyBase. *Nucleic Acids Res* **40**: D706-14.

Mosammaparast N, Pemberton LF. 2004. Karyopherins: from nuclear-transport mediators to nuclear-function regulators. *Trends Cell Biol* **14**: 547-556.

Nie M, Xie Y, Loo JA, Courey AJ. 2009. Genetic and proteomic evidence for roles of Drosophila SUMO in cell cycle control, Ras signaling, and early pattern formation. *PLoS ONE* **4**: e5905.

Niki Y, Yamaguchi T, Mahowald AP. 2006. Establishment of stable cell lines of Drosophila germ-line stem cells. *Proc Natl Acad Sci USA* **103**: 16325-16330.

Nishida KM, Okada TN, Kawamura T, Mituyama T, Kawamura Y, Inagaki S, Huang H, Chen D, Kodama T, Siomi H, et al. 2009. Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in Drosophila germlines. *EMBO J* **28**: 3820-3831.

Nishimasu H, Ishizu H, Saito K, Fukuhara S, Kamatani MK, Bonnefond L, Matsumoto N, Nishizawa T, Nakanaga K, Aoki J, et al. 2012. Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature*.

Pane A, Jiang P, Zhao DY, Singh M, Schüpbach T. 2011. The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline. *EMBO J* **30**: 4601-4615.

Pelisson A, Song SU, Prud'homme N, Smith PA, Bucheton A, Corces VG. 1994. Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. *EMBO J* **13**: 4401-4411.

Ram O, Goren A, Amit I, Shoresh N, Yosef N, Ernst J, Kellis M, Gymrek M, Issner R, Coyne M, et al. 2011. Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell* **147**: 1628-1639.

Ramadan N, Flockhart I, Booker M, Perrimon N, Mathey-Prevot B. 2007. Design and implementation of high-throughput RNAi screens in cultured Drosophila cells. *Nat Protoc* **2**: 2245-2264.

Rangan P, Malone CD, Navarro C, Newbold SP, Hayes PS, Sachidanandam R, Hannon GJ, Lehmann R. 2011. piRNA production requires heterochromatin formation in Drosophila. *Curr Biol* **21**: 1373-1379.

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

Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, Siomi H, Siomi MC. 2010. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. *Genes & Development* **24**: 2493-2498.

Scheuermann JC, de Ayala Alonso AG, Oktaba K, Ly-Hartig N, McGinty RK, Fraterman S, Wilm M, Muir TW, Müller J. 2010. Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* **465**: 243-247.

Schüpbach T, Wieschaus E. 1989. Female sterile mutations on the second chromosome of Drosophila melanogaster. I. Maternal effect mutations. *Genetics* **121**: 101-117.

Schüpbach T, Wieschaus E. 1991. Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* **129**: 1119-1136.

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 MC, Sato K, Pezic D, Aravin AA. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* **12**: 246-258.

Talamillo A, Sánchez J, Barrio R. 2008. Functional analysis of the SUMOylation pathway in Drosophila. *Biochem Soc Trans* **36**: 868-873.

Tanentzapf G, Devenport D, Godt D, Brown NH. 2007. Integrin-dependent anchoring of a stem-cell niche. *Nat Cell Biol* **9**: 1413-1418.

Vermaak D, Malik HS. 2009. Multiple roles for heterochromatin protein 1 genes in Drosophila. *Annu Rev Genet* **43**: 467-492.

Yamanaka S, Mehta S, Reyes-Turcu FE, Zhuang F, Fuchs RT, Rong Y, Robb GB, Grewal SIS. 2012. RNAi triggered by specialized machinery silences developmental genes and retrotransposons. *Nature*.

Yoshimura T, Miyazaki T, Toyoda S, Miyazaki S, Tashiro F, Yamato E, Miyazaki J-I. 2007. Gene expression pattern of Cue110: A member of the uncharacterized UPF0224 gene family preferentially expressed in germ cells. *Gene Expression Patterns* **8**: 27-35.

Yoshimura T, Toyoda S, Kuramochi-Miyagawa S, Miyazaki T, Miyazaki S, Tashiro F, Yamato E, Nakano T, Miyazaki J-I. 2009. Gtsf1/Cue110, a gene encoding a protein with two copies of a CHHC Zn-finger motif, is involved in spermatogenesis and retrotransposon suppression in murine testes. *Dev Biol* **335**: 216-227.

Zhang F, Wang J, Xu J, Zhang Z, Koppetsch BS, Schultz N, Vreven T, Meignin C, Davis I, Zamore PD, et al. 2012. UAP56 Couples piRNA Clusters to the Perinuclear Transposon Silencing Machinery. *Cell* **151**: 871-884.

4 Conclusions and Perspectives

Transposable elements constitute a major part of virtually every eukaryotic genome (Biémont 2010). Although McClintock had the foresight to call these fragments of DNA capable of moving form one genomic location to another, 'controlling elements', in subsequent years, transposable elements were considered to be parasites that only brought harm to the host organism. This over-simplified view of transposons has changed dramatically in recent years. It is now known that these elements are also beneficial to their host by creating diversity and driving evolution. Given both the good and the bad that can come from the movement of transposons in the genome, one can understand why control mechanisms have evolved that do not eliminate transposons completely, but rather attenuate their effects. The piRNA pathway, which controls transposable elements in the germline, achieves precisely that. The host is not alone in striving to find a balance in handling this double-edged sword. Transposons also have to find a balance between successfully propagating themselves, while not causing large amounts of damage to their host, since host's demise would mean demise for the element as well. Therefore the piRNA pathway and transposons each have characteristics that allow them to exist in an equilibrium, where piRNAs strive to control transposons without eliminating them completely and transposons struggle to spread to future generations, without affecting the host's viability.

One strategy some elements have evolved to keep this balance is specialized insertion patters to integrate at specific genomic loci, thus minimizing the damage they can inflict to the host. For example, in *Drosophila*, *P-element* DNA transposons avoid disrupting coding sequences by preferentially inserting upstream of the transcription start sites (Bellen et al. 2004). Other elements specifically target heterochromatic instead of gene-rich loci, like the *Ty5* retrotransposon in *Saccharomyces cerevisiae*. This element is specifically targeted to heterochromatin via a targeting domain in its encoded integrase, which binds Sir4, a component of heterochromatin (Xie et al. 2001; Zou et al. 1996). Interestingly, this targeting can be regulated. For the

integrase to bind Sir4, it needs to be phosphorylated at a certain residue (Dai et al. 2007). If not, *Ty5* will begin to integrate in expressed regions of the genome. Although the regulation of the site's phosphorylation status is not completely understood, under stress situations it seems as though phosphorylation is downregulated. This suggests that when the organism is under stress *Ty5* is able to transpose into euchromatic regions. This mechanism offers an interesting example of how the host can benefit from transposition under a stressful situation, where generation of diversity in gene expression could be valuable in adapting to stress. In fact, the idea of hosts utilizing transposon mobilization to reorganize the genome in result to environmental stress was proposed by Barbara McClintock many years ago (McClintock 1984). Thus, transposons represent a very complicated problem to tackle for the host; while they can sometimes provide benefits, they cannot be allowed to run amok. How do host organisms deal with this conflict and what mechanisms are in place to control the transposable elements?

Since its discovery, the piRNA pathway has proven to be an important mechanism for transposon control. Since the germline genome is the genomic material that will make up future generations, it is not surprising that there is a specialized transposon silencing mechanism in place specifically in germ cells. However, if other pathways existed to silence transposons in the germline in addition to the piRNA pathway is unknown. Results presented in this thesis demonstrate that piRNAs represent the core control of transposon activity in the germline.

Controlling transposons represents an incredible feat for the piRNA pathway. One way in which this pathway is able to silence such a variety of elements, is by targeting a molecule which is common to all transposons, the transposon mRNA. Retrotransposons have an RNA intermediate during the retrotransposition process and DNA transposons depend on transcription and translation of the transposase to jump to a new genomic location. In addition, targeting elements in this way without eliminating them from the genome, assures that if active transposition is needed the elements will still be capable of doing so. The benefit of maintaining transposable elements as a part of the genome is also suggested by the fact that

transposons comprise a large fraction of eukaryotic genomes, occupying as much as 50% of the human genome.

The piRNA pathway also has several other unique characteristics that make it an ideal system for dealing with selfish genetic elements. First, transposons contain great sequence variability, which constantly changes, especially with retrotransposons where reverse transcription can incorporate many mutations due to the RT enzyme's lack of proofreading capability. The piRNA pathway copes with this diversity by creating millions of unique piRNAs in gonadal tissue, as well as being malleable and adaptable to changes in elements the organism is exposed to. The basic scheme of how piRNAs are generated is very different from other endogenous sRNA pathways. microRNAs, for example, have a specific sequence and can even be conserved across species. piRNAs are generated from a designated cluster transcript, but they are not processed from this precursor with an emphasis on distinct sequences as occurs with miRNAs. However, piRNAs are not generated randomly, as isogenic flies show similar processing patterns, but these patterns can change dramatically from one strain to another. Each strain may face activation of different elements, so the piRNA pathway as evolved to be able to deal with this variability.

The precise degree of sequence complementarity necessary for a piRNA to efficiently target transposons is not known, but the pathway seems to make up for potential sequence variability in targets, with sheer number of available piRNAs that among themselves contain variability. The piRNA pathway is not a static entity, but always changing and evolving, learning to adapt to new transposon threats. How these new elements become part of the catalogue of piRNAs of an organism is not completely understood, but in some cases it seems as though the element transposes into a piRNA cluster, essentially causing piRNAs complementary to it to be generated. In addition, the pathway increases available piRNAs targeting active elements through ping-pong amplification. The piRNA pathway provides an ideal balance of controlling transposable elements, without eliminating them completely from the genome allowing for these elements to potentially aid in generating genetic diversity in a situation of stress.

Although the biological importance and function of piRNAs is clear, the specific molecular steps in the pathway are not nearly as well understood as in other sRNA pathways. Work in this thesis is aimed to identify components of the piRNA pathway, and in this way deepen our understanding of how the pathway functions.

First, we show that the OSS cell line is a good model in which to study the piRNA pathway because it recapitulates the primary piRNA pathway active in somatic cells of the ovary. Furthermore, this cell line is relatively easy to handle and can be passaged for several generations. We also show that one can obtain reproducible results when disrupting the piRNA pathway. Based on these findings, we demonstrate that the OSS cell line is highly suited for high throughput and genome-wide studies.

This cell line also offers the opportunity to perform more focused studies to better understand specific steps in the pathway as well as their biochemical aspects. In addition, one can study how the primary piRNA pathway specifically functions, without confounding results with the presence of the germline piRNA pathway. In focused experiments conducted in OSS, we were able to study the role of previously identified piRNA pathway components, Armi, Zuc and Squash, in the somatic piRNA pathway and place them as biogenesis or effector factors (see Appendix 1). Recently, several other groups have also conducted experiments utilizing this cell line that have greatly furthered our understanding of the pathway, demonstrating the strength and advantages it brings towards the study of the piRNA pathway.

Taking advantage of the possibilities this cell line offers, we performed an unbiased genome-wide screen and exhaustively searched for genes involved in transposon control in the *Drosophila* ovarian soma. Although the purpose of the screen was to uncover novel piRNA pathway factors, by using transposon derepression as a read-out, we would not only detected piRNA factors but also other genes involved in transposon control. Elevated transposons levels may also result from knockdown of genes involved in preserving genome integrity, which could explain the enrichment observed for the G2/M transition of mitotic cell cycle functional group within our candidate hits. There are several mechanisms, other than the

piRNA pathway, which are known to silence transposable elements in other organisms. For example, members of the apolipoprotein B mRNA editing enzyme 3 (APOBEC3) family of cytidine deaminases impede transposition of retrotransposons and retroviruses (Chiu and Greene 2008). These enzymes act by deaminating cytidines during cDNA synthesis during the transposition cycle of retroviruses and retrotransposons, which will lead to degradation or large amount of mutations, rendering the element inactive. However, there are no known homologs for these enzymes in *Drosophila*. Another mechanism to silence repetitive elements is through DNA methylation (Slotkin and Martienssen 2007). This process was been mainly studied in plants and mammals. In mice for example, disruption of both the maintenance DNA methyltranferase Dnmt1 and *de novo* methyltransferase Dnmt3 activity leads to transposable element upregulation (Bourc'his and Bestor 2004; Walsh et al. 1998). The current model of the piRNA pathway in mice proposes that it acts upstream of de novo DNA methylation, targeting transposons for silencing by this process (Aravin et al. 2008). In Arabidopsis this process is much more understood, where siRNAs derived mainly from transposable elements and tandem repeats bind to Ago4 and targets transposons loci for DNA methylation (Qi et al. 2006). The only DNA methyltransferase family member expressed in *Drosophila* is Dnmt2 (Lyko 2001). This protein seems to be specific for methylation of cytosines in tRNAs (Goll et al. 2006). Some studies have been published implicating it in DNA methylation, and specifically of transposable elements (Phalke et al. 2009). Several groups, including ours, have been unsuccessful in reproducing these results and our studies have shown no trace of DNA methylation in flies (See Appendix 3). Furthermore, Dnmt2 was not identified as a hit in our screen.

It seems as though cytidine deaminases and DNA methylation are not important for transposable element silencing in *Drosophila*. This leaves the piRNA pathway, as the principal mechanism for controlling transposons in the germline. This is further supported with the finding that many of the top hits of the screen are known piRNA pathway components and the majority of known pathway components were identified in the screen. This does not rule out

the possibility of the existence of additional uncharacterized transposon silencing pathways, however, the importance of the piRNA pathway remains clear.

With this in mind, it seems likely that many genes identified in this screen are novel components of the piRNA pathway. Since there are so many steps of the pathway that are still poorly understood, this genome-wide screen provides a list of genes that will aid research trying to uncover the molecular details of the pathway. Much investigation is still needed to decisively place any one of these genes at a precise step, but general analysis of some of the candidates shows that they have an effect on the piRNA pathway. Several novel biogenesis factors were identified. One such example is Shutdown, a member of the FK506-binding protein (FKBP) family of immunophilins, which has an important role in both the germline and the somatic piRNA pathway (See Appendix 2). Additionally, many interesting pathways were identified that had not been previously implicated in piRNA biogenesis. For example, several proteins involved in sumoylation were identified as hits in the screen. Further characterization of Uba2, the sumoylation E1 activating enzyme, was shown to have striking effects in piRNA biogenesis. Novel piRNA pathway effector genes were also identified, like Asterix, which has no effects on piRNA levels but seems be involved in downstream transposons silencing steps and has a dramatic effect on H3K9me3 marks over transposon loci. Intriguingly, Asterix contains two zinc-finger motifs that are predicted to bind RNA. Although we have not confirmed if this protein actively binds RNA, one possible model is that Asterix's RNA binding activity could anchor this protein to transcription sites where it could act as an anchor for other factors, such as Piwi, to recognize and silence the transposon. Much research is still needed to understand the specific roles that these proteins have in the pathway, but having a concise list of genes to focus studies upon will likely advance research a great deal. Hopefully, in subsequent years we will see the many studies surface as a product of research initiated based on our findings.

4.1 References

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.

Bellen HJ, Levis RW, Liao G, He Y, Carlson JW, Tsang G, Evans-Holm M, Hiesinger PR, Schulze KL, Rubin GM, et al. 2004. The BDGP gene disruption project: single transposon insertions associated with 40% of Drosophila genes. *Genetics* **167**: 761-781.

Biémont C. 2010. A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics* **186**: 1085-1093.

Bourc'his D, Bestor TH. 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**: 96-99.

Chiu Y-L, Greene WC. 2008. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu Rev Immunol* **26**: 317-353.

Dai J, Xie W, Brady TL, Gao J, Voytas DF. 2007. Phosphorylation regulates integration of the yeast Ty5 retrotransposon into heterochromatin. *Mol Cell* **27**: 289-299.

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh C-L, Zhang X, Golic KG, Jacobsen SE, Bestor TH. 2006. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science* **311**: 395-398.

Lyko F. 2001. DNA methylation learns to fly. *Trends in genetics : TIG*, April.

McClintock B. 1984. The significance of responses of the genome to challenge. *Science* **226**: 792-801.

Phalke S, Nickel O, Walluscheck D, Hortig F, Onorati MC, Reuter G. 2009. Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. *Nat Genet* **41**: 696.

Qi Y, He X, Wang X-J, Kohany O, Jurka J, Hannon GJ. 2006. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**: 1008-1012.

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**: 272-285.

Walsh CP, Chaillet JR, Bestor TH. 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* **20**: 116-117.

Xie W, Gai X, Zhu Y, Zappulla DC, Sternglanz R, Voytas DF. 2001. Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. *Molecular and Cellular Biology* **21**: 6606-6614.

Zou S, Ke N, Kim JM, Voytas DF. 1996. The Saccharomyces retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. *Genes & Development* **10**: 634-645.

Appendix 1: Probing the initiation and effector phases of the somatic piRNA pathway in *Drosophila*

Contributions to publication:

I contributed to experiments that helped determine how to grow and transfect the OSS cell line. I was also responsible for the western blots.

Citation:

Haase AD, Fenoglio S, Muerdter F, Guzzardo PM, Czech B, Pappin DJ, Chen C, Gordon A, Hannon GJ. 2010. Probing the initiation and effector phases of the somatic piRNA pathway in Drosophila. Genes & Development 24: 2499-2504.

RESEARCH COMMUNICATION

Probing the initiation and effector phases of the somatic piRNA pathway in *Drosophila*

Astrid D. Haase,[1] Silvia Fenoglio,[1] Felix Muerdter,[1] Paloma M. Guzzardo,[1] Benjamin Czech,[1] Darryl J. Pappin,[1] Caifu Chen,[2] Assaf Gordon,[1] and Gregory J. Hannon[1,3]

[1]Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA; [2]Genomic Assays R&D Life Technologies, Foster City, California 94404, USA

Combining RNAi in cultured cells and analysis of mutant animals, we probed the roles of known Piwi-interacting RNA (piRNA) pathway components in the initiation and effector phases of transposon silencing. Squash associated physically with Piwi, and reductions in its expression led to modest transposon derepression without effects on piRNAs, consistent with an effector role. Alterations in Zucchini or Armitage reduced both Piwi protein and piRNAs, indicating functions in the formation of a stable Piwi RISC (RNA-induced silencing complex). Notably, loss of Zucchini or mutations within its catalytic domain led to accumulation of unprocessed precursor transcripts from *flamenco*, consistent with a role for this putative nuclease in piRNA biogenesis.

Supplemental material is available at http://www.genesdev.org.

Received July 9, 2010; revised version accepted September 17, 2010.

Eukaryotic small RNAs regulate gene expression through various mechanisms, intervening at both transcriptional and post-transcriptional levels (for review, see Ghildiyal and Zamore 2009). Small RNAs are divided into classes according to their mechanism of biogenesis and their particular Argonaute protein partner. Piwi-interacting RNAs (piRNAs) bind Piwi-clade Argonaute proteins and act mainly in gonadal tissues to guard genome integrity by silencing mobile genetic elements (for review, see Malone and Hannon 2009).

Conceptually, the piRNA pathway can be divided into several different phases. During the initiation phase, small RNAs, called primary piRNAs, are produced from their generative loci, so-called piRNA clusters (Brennecke et al. 2007). These give rise to long, presumably single-stranded precursor transcripts, which are processed via an unknown biogenesis mechanism into small RNAs that are larger than canonical microRNAs (~24–30 nucleotides [nt]) (Aravin et al. 2006; Girard et al. 2006; Grivna et al. 2006; Lau et al. 2006; Vagin et al. 2006). Primary piRNAs become stably associated with Piwi proteins to form Piwi RISCs (RNA-induced silencing complexes), which also contain additional proteins that facilitate target recognition and silencing. During the effector phase, Piwi RISCs identify targets via complementary base-pairing. In some cases, for example, with Aubergine as a piRNA partner, there is strong evidence for target cleavage in vivo (Brennecke et al. 2007; Gunawardane et al. 2007). This nucleolytic destruction of transposon mRNAs is probably the main Aubergine effector mechanism, although this has not been rigorously demonstrated. Piwi also conserves the Argonaute catalytic triad; however, in this case, both its nuclear localization and its association with certain chromatin proteins suggest the possibility of transcriptional and post-transcriptional effector pathways (Brower-Toland et al. 2007; Klattenhoff et al. 2009; Saito et al. 2009). An additional phase, adaptation, is restricted to germ cells and constitutes the ping-pong cycle. During this phase, transposon mRNA cleavage directed by primary piRNAs triggers the production of secondary piRNAs, whose 5′ ends correspond to cleavage sites (Brennecke et al. 2007; Gunawardane et al. 2007). These generally join Ago3 and enable it to recognize and cleave RNAs with antisense transposon content, perhaps piRNA cluster transcripts. Cleavage by Ago3 RISC again triggers piRNA production from the target, closing a loop that enables the overall small RNA population to adjust to challenge by a particular transposon (for review, see Aravin et al. 2007). Finally, piRNA populations present in germ cells can be transmitted to the next generation to prime piRNA responses in progeny (Brennecke et al. 2008).

In *Drosophila* follicle cells, only the initiation and effector phases appear relevant (Brennecke et al. 2008). Here, the piRNA pathway relies on the coupling between a single Piwi protein (Piwi itself) and a principal piRNA cluster (*flamenco*) to silence mainly *gypsy* family retrotransposons (Sarot et al. 2004; Saito et al. 2006; Brennecke et al. 2007; Li et al. 2009; Malone et al. 2009). *Drosophila* ovarian somatic sheet cells (OSS) display many of the properties of follicle cells, and represent a convenient system to study the initiation and effector phases of the piRNA pathway without the complications inherent in the study of complex tissues in vivo (Niki et al. 2006; Lau et al. 2009; Saito et al. 2009). We therefore sought to leverage information derived from the use of RNAi in OSS cells with the analysis of ovaries derived from mutant animals to probe the roles of known piRNA pathway components in the initiation and effector phases of transposon silencing.

Results and Discussion

The piRNA pathway is continuously required for transposon silencing

Several prior studies have proposed models in which Piwi proteins silence targets by interfering with their transcription (Pal-Bhadra et al. 2004). Since piRNAs are largely absent from somatic tissues (Cox et al. 2000), impacts underlying these changes are presumed to have occurred during development and to have been epigenetically maintained in the adult. *Drosophila* Piwi protein is mainly localized to the nucleus and has been shown to

[*Keywords*: Piwi; Zucchini; Armitage; Squash; primary piRNA; *Drosophila*]
[3]Corresponding author.
E-MAIL hannon@cshl.edu; FAX (516) 367-8874.
Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.1968110.

interact with HP1, a core component of heterochromatin (Pal-Bhadra et al. 2004; Brower-Toland et al. 2007). Considered together, this body of evidence pointed strongly to an effector mechanism in which Piwi-associated small RNAs direct heterochromatin formation and silencing of targets.

Loss of *piwi* has dramatic effects on transposon expression in somatic follicle cells (Sarot et al. 2004; Brennecke et al. 2007; Klenov et al. 2007; Malone et al. 2009). Genetic mutants result in an absence of Piwi protein throughout development. This could lead to a failure to create heterochromatic marks that could have otherwise maintained epigenetic silencing of transposons in the absence of continuous Piwi expression. Alternatively, there could be an ongoing requirement for Piwi to maintain silencing, irrespective of whether it acted via transcriptional or post-transcriptional mechanisms.

To discriminate between these possibilities, we transfected OSS cells with dsRNAs corresponding to *piwi*, and followed impacts on Piwi mRNA and protein levels (Supplemental Fig. S1A; data not shown). Maximal suppression was reached by 3 d, and silencing persisted through day 6. At 6 d post-transfection, we probed impacts on two elements known to be derepressed in the follicle cells of *piwi* mutant ovaries: *gypsy* and *idefix*. Both showed derepression (up to 10-fold) (Fig. 1A) upon *piwi* silencing. Additional elements were also tested (Supplemental Fig. S1B), with *blood* being impacted strongly. Previous studies have also implicated *zucchini* (*zuc*) in the function of the somatic piRNA pathway (Pane et al. 2007; Malone et al. 2009). RNAi against this gene also increased *gypsy*, *blood*, and *idefix* expression (Fig. 1A; Supplemental Fig. S1B). Considered together, these results demonstrate that the integrity of the piRNA pathway is essential for the ongoing repression of mobile elements and argue against a model in which silent epigenetic states, once set by the action of piwi proteins on chromatin, can autonomously maintain transposon silencing.

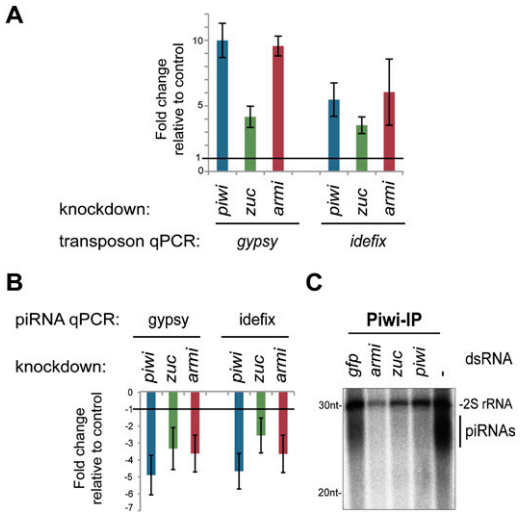


Figure 1. Effect of *piwi*, *zucchini*, and *armitage* knockdown on transposon silencing in OSS cells. OSS cells were treated with dsRNA against *piwi*, *zuc*, or *armi* for 6 d. qPCRs were normalized to internal controls *rp49* (*A*) or bantam (*B*). Fold changes relative to cells treated with gfp-dsRNA are shown on a linear scale. Error bars represent one standard deviation over three biological replicates. Transcripts (*A*) and two abundant piRNAs (*B*) corresponding to *gypsy* and *idefix* retroelements were detected by qPCR. (*C*) Small RNAs coimmunoprecipitating with Piwi in untreated cells and cells treated with dsRNA against *gfp*, *armi*, *zuci*, or *piwi* were labeled with ^{32}P at their 5′ termini.

Armitage *is a component of the somatic piRNA pathway*

Nearly a dozen proteins have been linked to the fully elaborated piRNA pathway that operates in germ cells (Malone et al. 2009). Many of these show germ cell-specific expression patterns consistent with their selective biological effects. Mutations in *armitage* (*armi*) result in coincident loss of the characteristic nuclear accumulation of Piwi protein and a reduction in Piwi-associated piRNAs (Malone et al. 2009). Unlike most germline-specific pathway components, an examination of RNA-seq data from OSS cells indicated substantial *armi* expression (Supplemental Fig. S1C). We therefore suppressed *armi* by RNAi and examined effects on transposon expression. Notably, *gypsy*, *blood*, and *idefix* were strongly derepressed, implying a role for *armi* in both the somatic and germline compartments (Fig. 1A; Supplemental Fig. S1B).

The *Drosophila* mutant *armi*^{1} represents a P-element insertion in the 5′ untranslated region (UTR) of *armitage*. A second allele, *armi*$^{72.1}$, was derived from *armi*^{1} by imprecise excision (Cook et al. 2004). RNA-seq data covered the *armi* ORF in OSS, but no reads were detected corresponding to the germ cell 5′ UTR (Supplemental Fig. S1D). This raises the possibility that *armi* expression might be driven by an alternative promoter in somatic cells, and that the *armi* alleles examined thus far may have spared the activity of that promoter.

Armitage and Zucchini function at the initiation phase

To investigate whether Armi and Zuc act at the initiation or effector phase of the piRNA pathway, we examined piRNAs. Silencing of *piwi* reduced levels of two abundant piRNAs, corresponding to *gypsy*, or *idefix* (Fig. 1B). Similar effects were noted upon silencing of *armi* or *zuc*. Aggregate OSS piRNA levels can be measured qualitatively by radioactive phosphate exchange of small RNAs in Piwi immunoprecipitates. As expected, RNAi against *piwi* virtually eliminated piRNAs in immunoprecipitates (Fig. 1C). Silencing of *armi* or *zuc* produced indistinguishable effects.

In germ cells, *armi* mutation causes loss of the prominent nuclear localization of Piwi (Malone et al. 2009). We observed a similar phenotype upon knockdown of *armi* in somatic OSS cells (Fig. 2A). Because of the mixed cell types present in ovaries, previous studies had not been able to distinguish whether Armi loss simply caused Piwi mislocalization or whether Armi influenced Piwi expression or stability. In OSS cells, knockdown of *armi* reduced Piwi protein levels by approximately fivefold, equivalent to a targeted knockdown of Piwi itself without affecting *piwi* mRNA (Fig. 2B,C). We noted a similar loss of Piwi protein from the nuclei in cells exposed to *zuc*-dsRNAs (Fig. 2A). In this case, Piwi protein but not mRNA levels also fell (Fig. 2B,C).

Considered together, these data strongly suggest roles of Armi and Zuc in the initiation phase of the piRNA pathway. A role for Armi, along with a previously unrecognized component, Yb, in the somatic pathway, is also supported by a recent report from Brennecke and colleagues (Olivieri et al. 2010). Either protein could play

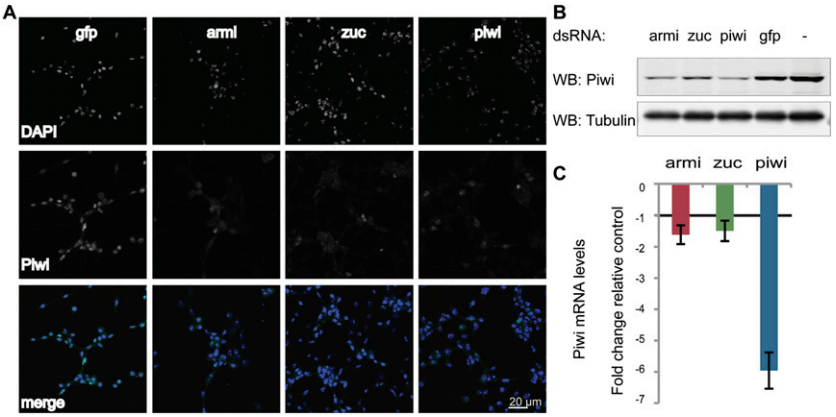


Figure 2. Piwi protein localization and levels upon knockdown of *armitage* or *zucchini*. (*A*) Piwi subcellular localization was determined by immunostaining in OSS cells treated with *zuc*-dsRNA, *armi*-dsRNA, or *piwi*-dsRNA. *Gfp*-dsRNA-treated cells were used as control. (*B*) Piwi protein levels in total cell extracts were determined by Western blotting. Tubulin was used as a loading control. (*D*) *Piwi* qPCRs were normalized to *rp49*. Fold changes relative to cell treated with gfp-dsRNA are shown on a linear scale. Error bars represent one standard deviation over three biological replicates.

a role in primary piRNA biogenesis, aiding piRNA production or loading, with this model resting on the presumption that association with mature piRNAs influences Piwi protein stability (Fig. 3A, steps 1, 2). Alternatively, Armi or Zuc could be core components of mature Piwi RISC, with loss of either subunit destabilizing associated components of the complex (Fig 3A, step 3).

Armitage is a component of Piwi RISC

To investigate these alternative models, we performed proteomic analysis of Piwi RNPs. Piwi immunoprecipitates contained a number of peptides from Armi, suggesting that this protein is present in Piwi RISC (Fig. 3B). Of note, we also detected association of both Piwi and Armi with Squash (Squ), another previously identified piRNA pathway component (Pane et al. 2007). Piwi could be also detected in Squ immunoprecipitates by Western blotting. Although no Zuc peptides were seen in multidimensional protein identification technology (MudPIT), Piwi could be detected to a low extent in Zuc immunoprecipitates (Fig. 3C; Supplemental Fig. S2). Overall, the emerging picture suggests that both Armi and Squ are components of Piwi RISC. Lower levels of Piwi associated with Zuc might indicate a weaker or more transient association of Zuc with Piwi RISC.

Squash impacts the piRNA effector step

Mutations in *squash* (*squ*) show little impact on piRNA populations in mutant ovaries (Malone et al. 2009). Similarly, upon sequencing of small RNAs in Piwi immunoprecipitates, we failed to detect any differences in associated piRNA populations upon comparison of *squ* homozygous mutant animals to heterozygous siblings (Fig. 4B). Animals harboring two *squ* alleles interrupted by early stop codons did, however, display an effect on transposon silencing.

As compared with heterozygous siblings, *squ* mutants showed significant derepression of *gypsy* (Fig. 5A). This occurred without any detectable change in an abundant *gypsy* piRNA or overall Piwi levels (Fig. 5B,C). In contrast, no substantial changes were detected in *idefix* or *ZAM* (Fig. 5A,B); however, *I-element* and *blood* were strongly derepressed (Supplemental Fig. S3).

Considered together, these results point to a role of squash in the effector phase of the piRNA pathway. We did note a slight but reproducible reduction in Piwi protein levels in homozygous squ mutants (Fig. 5C; Supplemental Fig. S4). However, this was well within the range observed in Piwi heterozygotes, where the piRNA pathway functions completely normally.

A possible role for Zucchini in piRNA biogenesis

In the initial screen that placed *zuc* within the piRNA pathway, two alleles were identified (Schupbach and Wieschaus 1991; Pane et al. 2007). *zuc*^{HM27} represents an early stop mutation resulting in a putative null allele (referred to as *zuc* mut). This mutant strongly affects piRNA silencing in both germline and somatic cells of the ovary (Pane et al. 2007; Malone et al. 2009). While somatic piRNAs are depleted in this mutant, ping-pong signatures remain intact (Malone et al. 2009). This places Zuc outside of the adaptive phase, consistent with our accumulating evidence for a role in the initiation phase.

While the biochemical properties of Zuc have yet to be analyzed, its protein sequence places it as a member of the phospholipase D (PLD) family of phosphodiesterases. These share a HxK(x)$_4$D motif, whose integrity is

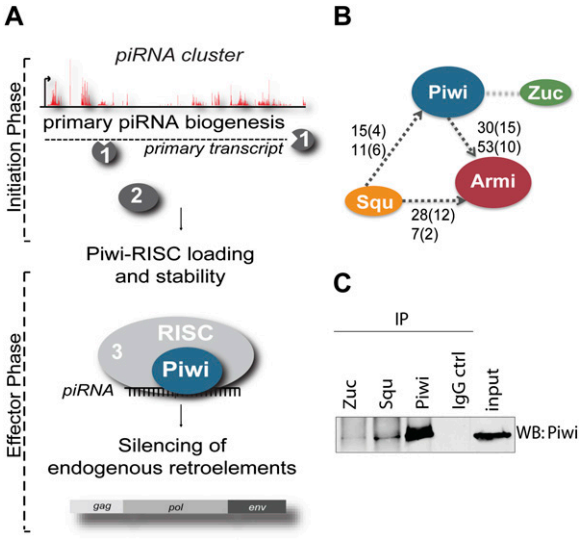


Figure 3. MudPIT analysis of Piwi, Zucchini, and Squash complexes. (*A*) A proposed model for the function of piRNA pathway components is shown. (*B*) Protein associations identified by MudPIT are depicted. Arrows point from the immunoprecipitated protein to the coimmunoprecipitated protein. Numbers of identified peptides and corresponding unique sequences (shown in parentheses) of two biological replicates are indicated. Only peptides above the significance threshold were considered (see the Materials and Methods). (*C*) Zucchini, Squash, and Piwi were immunoprecipitated (IP) from OSS cells. The presence of Piwi was assessed by Western blotting.

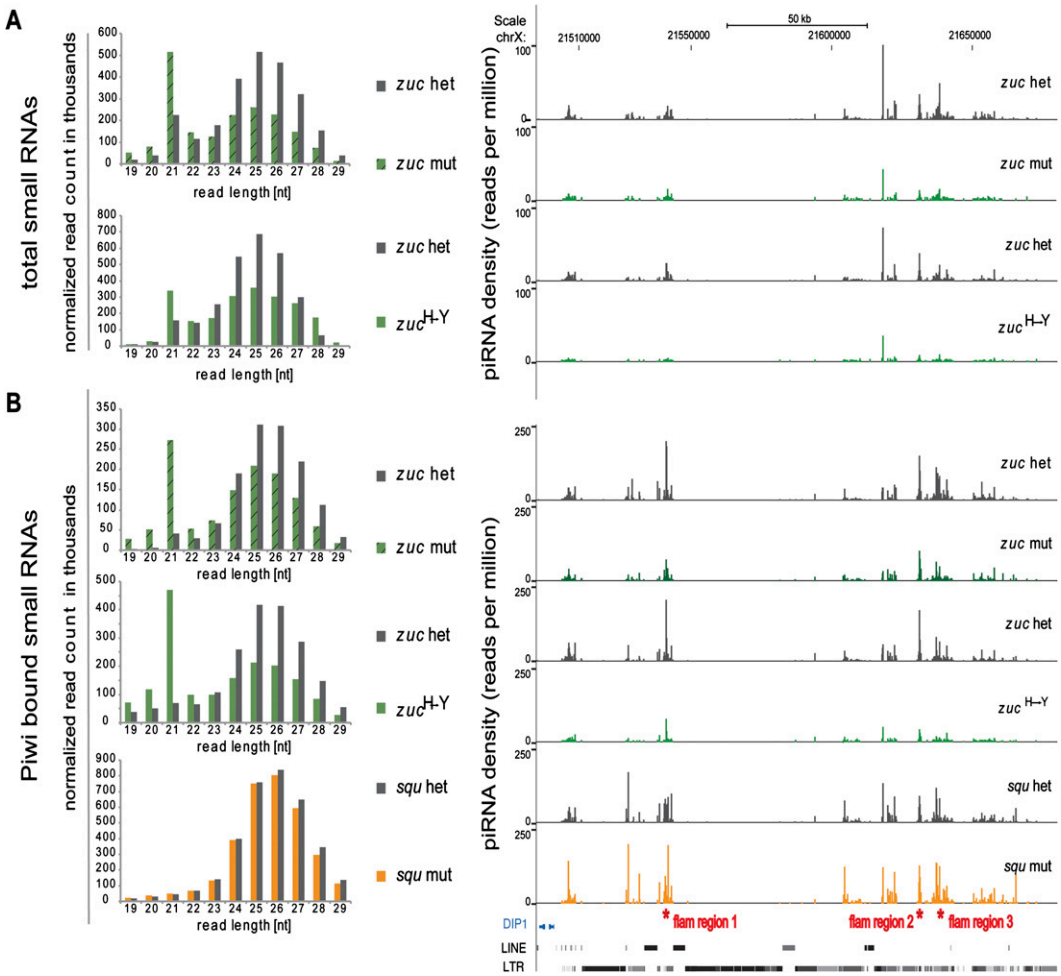


Figure 4. piRNA populations from *zucchini* and *squash* mutant ovaries. Heterozygous siblings serve as control. (*Left* panel) Size profiles of small RNAs mapping to repeats normalized to total read counts are shown. (*Right* panel) Densities of uniquely mapping piRNAs are plotted over the *flamenco* locus in reads per million (only small RNAs matching the plus strand are depicted). (*A*) Small RNAs cloned from total RNA. (*B*) Small RNAs from Piwi immunoprecipitates (IP). Regions of *flamenco* measured in Figure 5D are indicated by asterisks.

essential for catalytic activity (Zhao et al. 1997; Sasnauskas et al. 2010). The second *zuc* mutation that emerged in the original screen, zuc^{SG63}, contains a H → Y mutation within the phosphodiesterase motif that is predicted to render it catalytically inactive. To probe a role for Zuc catalytic activity in the piRNA pathway, we compared the presumed null (*zuc* mut) and catalytically dead (*zuc* H → Y) alleles for their effects on piRNAs and transposon silencing.

We analyzed total ovarian small RNAs from animals that were heterozygous or homozygous for the *zuc* H → Y allele and compared the resulting profiles to previously published analyses of the presumed *zuc* mut allele (Malone et al. 2009). In both cases, we saw strong reductions in total piRNAs and in populations that mapped uniquely to the *flamenco* locus (Fig. 4A), regardless of the normalization method used to compare libraries (Supplemental Fig. S5). Slightly stronger impacts were apparent when we compared profiles of Piwi immunoprecipitates (Fig. 4B). Here, piRNA populations corresponding to *flamenco* were almost completely lost. We did note an accumulation of 21-nt species in Piwi immunoprecipitates from both *zuc* mutant lines. These were enriched for a 5′ U, although not to the extent for longer piRNA species. The nature of these shorter, apparently Piwi-associated RNAs remains mysterious.

Both the presumed null and H → Y *zuc* alleles impacted transposon silencing (Fig. 5A). Between fivefold and 20-fold increases in *gypsy*, *ZAM*, and *idefix* were noted in comparison with heterozygous controls. Even stronger derepression could be observed for I-element, *HeT-A*, *1731*, and *blood* (Supplemental Fig. S3). The *zuc* H → Y and *zuc* mut alleles also showed similar impacts on piRNA populations (Fig. 5B) and the overall levels of Piwi protein (Fig. 5C; Supplemental Fig. S4).

Considered together, these data point to a requirement for the presumed catalytic center of Zuc in the initiation phase of the piRNA pathway. Other PLD family nucleases that have been characterized to date cleave nucleic acids leaving 5′ phosphate and 3′ hydroxyl termini (Pohlman et al. 1993; Zhao et al. 1997; Sasnauskas et al. 2010). These are the characteristics one might expect for a processing enzyme that catalyzed primary piRNA biogenesis. Previous studies have posited the requirement for several nucleolytic activities in the piRNA pathway. One is thought to form the 5′ ends of primary piRNAs. The

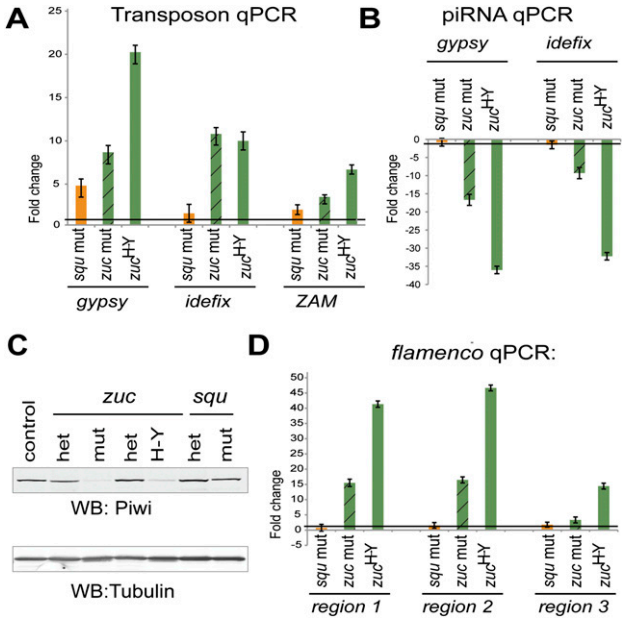


Figure 5. Transposons, piRNAs, and piRNA precursors in *zucchini* and *squash* mutant ovaries. (*A*) Transcripts of *gypsy*, *idefix*, and *ZAM* transposons were detected by qPCR. (*B*) Individual piRNAs targeting *gypsy* and *idefix* were detected by qPCR. (*C*) Piwi protein levels in mutant and heterozygous ovary extracts were measured by Western blotting. Tubulin serves as loading control. Celera sequencing strain (S-strain) is shown in addition. (*D*) Three ~100-nt regions of *flamenco* that are normally highly processed into piRNAs were detected by qPCR. The positions of these segments are indicated in Figure 4. qPCR data were normalized to internal controls *rp49* (*A*,*C*) or bantam (*B*). Fold changes relative to heterozygous siblings are shown on a linear scale. Error bars represent one standard deviation over three technical replicates.

3′ ends of these species could be formed prior to Piwi loading or could be coupled to protein binding, as is posited for the ping-pong cycle. The nucleolytic center of Piwi proteins themselves form the 5′ ends of secondary piRNAs, with their 3′ ends proposed to be created by a separate enzyme. Based on its impacts in the soma on Piwi complexes, we imagined that the Zuc catalytic center might form either the 5′ or 3′ ends of primary piRNAs.

To evaluate this hypothesis, we examined RNAs derived from the *flamenco* locus in control ovaries or in tissues from animals homozygous for either of the two *zuc* mutant alleles. The prevailing model holds that the *flamenco* locus is transcribed as a continuous, single-stranded precursor spanning >150 kb (Brennecke et al. 2007). We reasoned that a defect in primary processing might result in an accumulation of long RNAs from this locus, since they would not be effectively metabolized into piRNAs. By quantitative PCR (qPCR) using primer pairs spanning three different regions of *flamenco* (see Fig. 4B, right panel), we saw 15-fold to 45-fold increases in *flamenco*-derived long RNAs in *zuc* mutant ovaries (Fig. 5D; Supplemental Fig. S6).

Considered as a whole, our results strongly support a role for Zucchini in the primary processing of piRNAs from the *flamenco* locus. Given its size, it is virtually impossible to follow the fate of the intact *flamenco* transcript by Northern blotting. Three different segments of the locus do show an accumulation consistent with their failure to be parsed into piRNAs. However, several alternative explanations can also be envisioned. For example, if Zucchini impacts Piwi stability, feedback controls might operate to inhibit primary biogenesis. Without a direct, biochemical demonstration that Zucchini processes piRNA cluster transcripts, its assignment as a primary biogenesis enzyme must be viewed as provisional. However, any alternative model must account for the requirement for its phosphodiesterase active site, and, at present, a direct role in piRNA biogenesis seems the most parsimonious conclusion.

While our studies do not ascribe specific functions to Armitage and Squash, they do support their assignment to the initiator and effector phases, respectively. Armitage is a putative helicase, although no analyses as yet indicate whether this biochemical activity is required for its function. Our placement of this protein in the initiation phase and its intimate association with Piwi perhaps suggest a role in loading or stability of Piwi RISC. Squash, of all the components examined in this study, had the most variable effects on transposon control in somatic cells of the ovary (Figs. 1A, 5A), but both its physical association with Piwi RISC and its impact on transposons without an effect on piRNAs imply a role in the effector phase. While the studies reported herein can suggest roles for known pathway components at specific points in the piRNA pathway, a definitive conclusion regarding the part played by any of these proteins will require reconstitution of the pathway in vitro.

Materials and methods

OSS cell culture and knockdown

The OSS cell line was a kind gift from Yuzo Niki, and was cultured as described (Niki et al. 2006). dsRNA was prepared as described (http://www.flyrnai.org/DRSC-PRS.html). Cells were transfected with dsRNA using Xfect Transfection Reagent (Clontech). Six days after transfection, RNA and protein were extracted. Detailed information for RT-qPCR, Western blotting, and immunofluorescence is available in the Supplemental Material.

Fly stocks and allelic combinations

Fly stocks and allelic combinations used were *squ mut*: *squ*HE47/PP32 (Pane et al. 2007); *zuc mut*: *zuc*HM27/Df(2l)PRL (Pane et al. 2007); and *zuc H → Y*: *zuc*HM27/SG63 (Pane et al. 2007).

Preparation of small RNA libraries

Small RNA libraries were prepared as described (Brennecke et al. 2007). Detailed information is available in the Supplemental Material.

Acknowledgments

We thank Vasily Vagin, Anna Jankowska, Adam Rosebrock, Yaniv Erlich, Julius Brennecke, Alexei Aravin, Delphine Fagegaltier, and current and former members of the Hannon laboratory for helpful discussions and support. The OSS cell line was a gift of Yuzo Niki (Ibaraki University). Fly stocks were kindly provided by Gertrud Schupbach (Princeton). PMG is supported by Grant 5T32GM065094 from the NIH and by a William Randolph Hearst Foundation Scholarship. S.F. is supported by the Elizabeth Sloan Livingston Foundation. B.C. is supported by the Boehringer Ingelheim Fonds. G.J.H. is an investigator of the Howard Hughes Medical Institute. This work was supported by grants from the NIH (to G.J.H.). Sequences reported in this study can be found at GEO using accession number GSE24108 (which includes libraries GSM593297–593304).

References

Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, et al. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442:** 203–207.

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

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

Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322:** 1387–1392.

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

Cook HA, Koppetsch BS, Wu J, Theurkauf WE. 2004. The *Drosophila* SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. *Cell* **116:** 817–829.

Cox DN, Chao A, Lin H. 2000. *piwi* encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* **127:** 503–514.

Ghildiyal M, Zamore PD. 2009. Small silencing RNAs: An expanding universe. *Nat Rev Genet* **10:** 94–108.

Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442:** 199–202.

Grivna ST, Beyret E, Wang Z, Lin H. 2006. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev* **20:** 1709–1714.

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

Klattenhoff C, Xi H, Li C, Lee S, Xu J, Khurana JS, Zhang F, Schultz N, Koppetsch BS, Nowosielska A, et al. 2009. The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* **138:** 1137–1149.

Klenov MS, Lavrov SA, Stolyarenko AD, Ryazansky SS, Aravin AA, Tuschl T, Gvozdev VA. 2007. Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the *Drosophila melanogaster* germline. *Nucleic Acids Res* **35:** 5430–5438.

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE. 2006. Characterization of the piRNA complex from rat testes. *Science* **313:** 363–367.

Lau NC, Robine N, Martin R, Chung WJ, Niki Y, Berezikov E, Lai EC. 2009. Abundant primary piRNAs, endo-siRNAs, and microRNAs in a *Drosophila* ovary cell line. *Genome Res* **19:** 1776–1785.

Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, Seitz H, Horwich MD, Syrzycka M, Honda BM, et al. 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137:** 509–521.

Malone CD, Hannon GJ. 2009. Small RNAs as guardians of the genome. *Cell* **136:** 656–668.

Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ. 2009. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* **137:** 522–535.

Niki Y, Yamaguchi T, Mahowald AP. 2006. Establishment of stable cell lines of *Drosophila* germ-line stem cells. *Proc Natl Acad Sci* **103:** 16325–16330.

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* doi: 10.1038/emboj.2010.212.

Pal-Bhadra M, Leibovitch BA, Gandhi SG, Rao M, Bhadra U, Birchler JA, Elgin SC. 2004. Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science* **303:** 669–672.

Pane A, Wehr K, Schupbach T. 2007. zucchini and squash encode two putative nucleases required for rasiRNA production in the *Drosophila* germline. *Dev Cell* **12:** 851–862.

Pohlman RF, Liu F, Wang L, More MI, Winans SC. 1993. Genetic and biochemical analysis of an endonuclease encoded by the IncN plasmid pKM101. *Nucleic Acids Res* **21:** 4867–4872.

Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC. 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 MC. 2009. A regulatory circuit for piwi by the large Maf gene traffic jam in *Drosophila*. *Nature* **461:** 1296–1299.

Sarot E, Payen-Groschene G, Bucheton A, Pelisson A. 2004. Evidence for a piwi-dependent RNA silencing of the *gypsy* endogenous retrovirus by the *Drosophila melanogaster flamenco* gene. *Genetics* **166:** 1313–1321.

Sasnauskas G, Zakrys L, Zaremba M, Cosstick R, Gaynor JW, Halford SE, Siksnys V. 2010. A novel mechanism for the scission of double-stranded DNA: BfiI cuts both 3′–5′ and 5′–3′ strands by rotating a single active site. *Nucleic Acids Res* **38:** 2399–2410.

Schupbach T, Wieschaus E. 1991. Female sterile mutations on the second chromosome of *Drosophila melanogaster*. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* **129:** 1119–1136.

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

Zhao Y, Stuckey JA, Lohse DL, Dixon JE. 1997. Expression, characterization, and crystallization of a member of the novel phospholipase D family of phosphodiesterases. *Protein Sci* **6:** 2655–2658.

Appendix 2: Shutdown is a component of the *Drosophila* piRNA biogenesis machinery

Contributions to publication:

The initiative to begin experiments on the involvement of Shutdown in the piRNA pathway was due, in part, to shutdown being identified as a hit in the OSS genome-wide screen. I also contributed the OSS RNA-seq library.

Citation:

Preall JB, Czech B, Guzzardo PM, Muerdter F, Hannon GJ. 2012. Shutdown is a component of the Drosophila piRNA biogenesis machinery. RNA 18: 1446-1457.

shutdown is a component of the *Drosophila* piRNA biogenesis machinery

JONATHAN B. PREALL,[1] BENJAMIN CZECH,[1] PALOMA M. GUZZARDO, FELIX MUERDTER, and GREGORY J. HANNON[2]

Howard Hughes Medical Institute, Watson School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

ABSTRACT

In animals, the piRNA pathway preserves the integrity of gametic genomes, guarding them against the activity of mobile genetic elements. This innate immune mechanism relies on distinct genomic loci, termed piRNA clusters, to provide a molecular definition of transposons, enabling their discrimination from genes. piRNA clusters give rise to long, single-stranded precursors, which are processed into primary piRNAs through an unknown mechanism. These can engage in an adaptive amplification loop, the ping-pong cycle, to optimize the content of small RNA populations via the generation of secondary piRNAs. Many proteins have been ascribed functions in either primary biogenesis or the ping-pong cycle, though for the most part the molecular functions of proteins implicated in these pathways remain obscure. Here, we link *shutdown* (*shu*), a gene previously shown to be required for fertility in *Drosophila*, to the piRNA pathway. Analysis of knockdown phenotypes in both the germline and somatic compartments of the ovary demonstrate important roles for *shutdown* in both primary biogenesis and the ping-pong cycle. *shutdown* is a member of the FKBP family of immunophilins. Shu contains domains implicated in peptidyl-prolyl *cis-trans* isomerase activity and in the binding of HSP90-family chaperones, though the relevance of these domains to piRNA biogenesis is unknown.

Keywords: piRNAs; transposon silencing; RNAi; FKBP; germ cells

INTRODUCTION

Eukaryotic genomes are prone to the accumulation of repetitive sequences, including transposable elements, over evolutionary time (McClintock 1953; Kim et al. 1994; Brennecke et al. 2007; Chambeyron et al. 2008; Feschotte 2008). The genomic instability brought about by transposon activity is a double-edged sword. Low levels of transposition can drive evolution in the long term, but loss of control over mobile elements in any individual can threaten reproductive success. Mechanisms for suppressing transposon activation in the germline are therefore both potent and widely conserved (Grimson et al. 2008). In animals, the PIWI-interacting RNA (piRNA) pathway is key to transposon silencing in reproductive tissues (Aravin et al. 2006; Girard et al. 2006; Lau et al. 2006; Vagin et al. 2006; Malone and Hannon 2009; Khurana and Theurkauf 2010; Senti and Brennecke 2010). In *Drosophila*, piRNAs are active both in the germ cell lineage and in a particular somatic lineage that encysts the germ cells and provides growth and maturation signals (Malone et al. 2009).

piRNA clusters sit at the apex of the pathway and, based upon their sequence content, define transposon targets for repression (Brennecke et al. 2007). piRNA clusters give rise to long, single-stranded transcripts (Brennecke et al. 2007) that are thought to be exported to the cytoplasm and processed into primary piRNAs, most likely in specialized cytoplasmic structures (Saito et al. 2010; Handler et al. 2011). A number of proteins have been implicated in primary piRNA biogenesis and their loading into PIWI-family proteins, including Armitage, Zucchini, Vreteno, and the Yb family (Klattenhoff et al. 2007; Pane et al. 2007; Malone et al. 2009; Szakmary et al. 2009; Haase et al. 2010; Olivieri et al. 2010; Saito et al. 2010; Handler et al. 2011; Zamparini et al. 2011). Yet, almost nothing is known about how each of these promotes the production of primary piRNAs.

The soma relies on a single piRNA cluster, *flamenco* (*flam*) (Brennecke et al. 2007). This ~180 kb, centromere-proximal locus on the X chromosome produces a piRNA population that is strongly enriched for species antisense to the *gypsy* family elements. These elements are active in follicle cells and can propagate by infection of germ cells through

[1]These authors contributed equally to this work.

[2]**Corresponding author**
E-mail hannon@cshl.edu

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.034405.112.

their capability to form virus-like particles (Pelisson et al. 1994; Chalvet et al. 1999). Somatic piRNAs are produced solely through primary biogenesis (Brennecke et al. 2007; Malone et al. 2009). In the germline, a greater variety of clusters targets a broad spectrum of mobile elements and engages an adaptive cycle, termed ping-pong, through which transposon mRNAs help to shape piRNA populations (Brennecke et al. 2007; Gunawardane et al. 2007). Here, antisense-oriented piRNAs derived from genomic clusters are loaded into Aubergine (Aub) and cleave active transposable element transcripts in an RNAi-like reaction. Unlike classical RNAi, this triggers the production of a new small RNA, derived from the target mRNA and with its 5′ end formed by Aub-mediated cleavage. The new, secondary piRNA is loaded into Ago3, which can then use this sense-oriented species to recognize and cleave cluster-derived transcripts, producing more antisense piRNAs via a similar mechanism.

Mutations in the *Drosophila* piRNA pathway generally result in sterility with stereotypical phenotypes in the male and female germline (Schupbach and Wieschaus 1991; Wilson et al. 1996; Gonzalez-Reyes et al. 1997; Lin and Spradling 1997; Cox et al. 2000; Cook et al. 2004). In part, these are thought to result from DNA double-strand breaks induced by element activity (Chen et al. 2007; Klattenhoff et al. 2007). Such breaks trigger meiotic checkpoint activation mediated by the *Drosophila chk2* ortholog, *loki,* which in turn disrupts dorsal-ventral axis formation during oogenesis. Hence, mutations in secondary piRNA genes such as *aubergine* display fused dorsal appendages and other hallmarks of oocyte ventralization (Theurkauf et al. 2006). Transposon silencing is also critical for the maintenance of germline stem cells (Lin and Spradling 1997; Cox et al. 2000; Houwing et al. 2007). In the male germline, loss of *Su(ste)* piRNAs derepresses the repetitive *Stellate* locus, which disrupts spermiogenesis by causing the overproduction and eventual crystallization of Stellate protein within the testis (Bozzetti et al. 1995; Aravin et al. 2001). Several mutants that are now known to affect the *Drosophila* piRNA pathway—including *aubergine*, *zucchini*, *squash*, *vasa*, and *cutoff*—were first described in a female sterility screen by Schüpbach and Wieschaus over 20 yr ago (Schupbach and Wieschaus 1989, 1991). Of the genes identified in that study that would eventually come to be known as piRNA factors, all but *cutoff* were classified phenotypically as having defects in dorsal appendage formation (Schupbach and Wieschaus 1991).

Subsequently, Munn and Steward (2000) mapped another of these female sterile mutations, *shutdown (shu, CG4735),* to an immunophilin gene of the FK506-binding protein (FKBP) family. Mutations in *shu* disrupt germ cell division, eventually causing the germline stem cells to fail entirely. Two strong alleles caused sterility in both males and females, while a third point mutant allele did not affect male fertility. In mutant females, stem cells that successfully divide generally produce faulty egg chambers that arrest mid-oogenesis. Germline clones for strong alleles of *shu* can produce mature oocytes, though they display typical patterning defects such as fused dorsal appendages. Considered together, these observations implicate *shu* as a component of the *Drosophila* piRNA pathway. This conjecture is supported by the presence of FKBP6, the mammalian protein most similar to Shutdown, in complexes with mammalian Piwi-family proteins, Miwi and Miwi2 (Vagin et al. 2009).

FKBPs play diverse biological roles ranging from facilitating protein folding to modulating transport (Ahearn et al. 2011), receptor signaling (Li et al. 2011), and meiotic recombination (Crackower et al. 2003; Kang et al. 2008). The FKBP domain is annotated as a peptidyl-prolyl *cis-trans* isomerase (PPIase), though there are many instances of well-conserved FKBP domains that lack PPIase activity (Gollan and Bhave 2010). The macrolide immunosuppressants FK506 (tacrolimus) and rapamycin (sirolimus) bind with sub-nanomolar affinity to the FKBP domain and block a key protein–protein interaction surface, but as is the case with PPIase activity, many family members display much reduced affinities for these compounds (DeCenzo et al. 1996; Gollan and Bhave 2010).

FKBP-class immunophilins display a variety of domain architectures. One arrangement, conserved from protozoa to humans, places a tetratricopeptide repeat (TPR) domain downstream from the FKBP domain (Pratt et al. 2004). The TPR domain is a protein–protein interaction module that binds heat shock proteins (HSPs), primarily of the HSP90 family in higher eukaryotes (Pratt 1998; Allan and Ratajczak 2011). Several crystal structures are available that highlight key conserved residues that participate in this interaction (Van Duyne et al. 1993; Ward et al. 2002). Connections between small RNA silencing pathways and HSP activity have been observed in several model systems (Smith et al. 2009). In particular, RNA-induced silencing complex (RISC) loading is facilitated by HSP90 and ATP hydrolysis (Iki et al. 2010; Iwasaki et al. 2010; Miyoshi et al. 2010; Iki et al. 2011).

Here, we report that *shutdown* is a critical element of the *Drosophila* piRNA pathway. Tissue-specific depletion of Shu results in derepression of transposon expression and a near-complete loss of mature piRNAs in both the somatic and germline lineages. Shu is cytoplasmically localized, and its loss disrupts the localization of all three piRNA effectors, Piwi, Aub, and Ago3. We hypothesize that Shu is an essential component of both primary and ping-pong–derived piRNA biogenesis, likely acting at a very early step that is shared between both piRNA systems.

RESULTS AND DISCUSSION

Clues to a role for FKBPs in the piRNA pathway

We previously carried out a proteomic analysis of mammalian PIWI proteins, Miwi and Miwi2 (Vagin et al. 2009). Among the components of these complexes were murine

FKBP6 and multiple HSPs. FKBP5 was also detected in Miwi immunoprecipitates with roughly half the coverage seen for FKBP6. Given the greater convenience of manipulating the piRNA pathway in *Drosophila*, we chose to examine potential roles for FKBP proteins in that model system.

The *Drosophila* genome encodes eight FKBP family members (Fig. 1A,B). Three, CG1847, CG5482, and FKBP59, are annotated to share the domain architecture of FKBP6, with their FKBP domains followed by a TPR. Shutdown is a potential fourth member of this group. Though its TPR

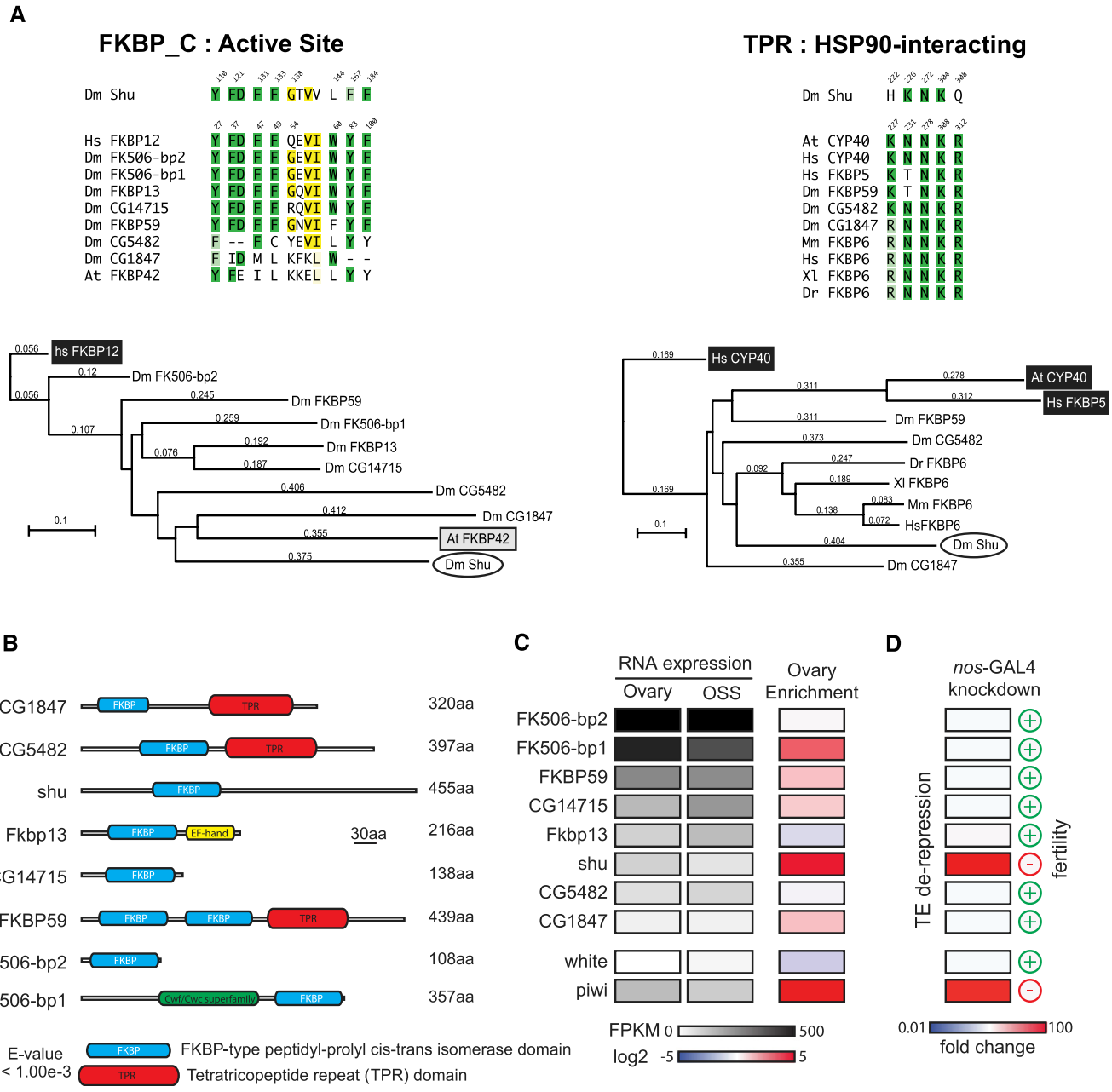


FIGURE 1. Shutdown is the only FKBP-family protein required for transposon silencing. (*A*) *Above* are shown the critical residues for the FKBP family peptidyl-prolyl *cis-trans* isomerase active site and the HSP90-interacting region in the TPR, as indicated, comparing the site in *Drosophila* Shutdown with those present in other family members. Residues in green indicate highly conserved residues with a known impact on PPIase activity, while those in yellow indicate a more poorly conserved region that has also been implicated. *Below* are evolutionary trees comparing each domain with family members present in other species. (*B*) The domain structures of the eight *Drosophila* FKBP family members are shown schematically. (*C*) Relative expression levels of *Drosophila* FKBP family members are shown for ovary and OSS RNAseq data sets. Relative enrichment in ovary versus other tissues is also shown. (*D*) Shown are relative *HetA* expression levels detected in ovaries from *Drosophila* engineered to express dsRNAs corresponding to each family member in the germline lineage. To the *right* is indicated whether dsRNA-expressing females are fertile (+) or sterile (−).

domain has substantially diverged in comparison to its paralogs (TPR_2, Pfam e-value = 0.0026), secondary structure predictions using the Phyre2 algorithm (www.sbg.bio.ic.ac.uk/phyre2) annotate the putative TPR as such with high confidence. Among *Drosophila* FKBPs, Shutdown is most closely related to FKBP6 ($E = 2 \times 10^{-36}$) overall, whereas FKBP59 is a potential FKBP5 ortholog ($E = 2 \times 10^{-46}$).

An examination of RNAseq both from the *Drosophila* ovarian somatic sheet (OSS) cell line and from a published ovarian data set revealed that several FKBPs are expressed in female reproductive tissues (Fig. 1C; Gan et al. 2010). A broader set of published microarray data (Chintapalli et al. 2007) suggested that the expression pattern of *shu* is much more biased to the ovary than is expression of other family members (Fig. 1C), a bias shared by many piRNA pathway components.

The FKBP_C domain is broadly conserved across evolution, though its PPIase activity is not (Kamphausen et al. 2002). Phylogenetic comparison of *Drosophila* FKBP_C domains to known active (*Homo sapiens* FKBP12) and inactive (*Arabidopsis thaliana* FKBP42) PPIase domains suggested that *shu* is more similar to inactive variants (Fig. 1A, bottom). Shutdown does retain more of the active site residues shown to be essential for PPIase activity in human FKBP12 (Fig. 1A, left) than does *AtFKBP42*. A conserved tryptophan residue (W60 in *Hs*Fkbp12) has been replaced by a leucine in Shutdown. Introduction of this change into Fkbp12 reduces PPIase activity by approximately eightfold and rapamycin and FK506 binding affinity by 10- and 75-fold, respectively. It is therefore likely that Shutdown does not represent an optimally active PPIase and may instead utilize the domain as a protein interaction interface, as do other FKBP family members (Gollan and Bhave 2010).

The Shu TPR domain is less well conserved and, in fact, shows little similarity to other TPRs known to bind HSP90 (Fig. 1A, right). In particular, nonconservative amino acid changes at two key residues suggested that the affinity of this domain for the C-terminal MEEVD of HSP90 is likely to be dramatically reduced compared with other family members (Ratajczak and Carrello 1996; Ward et al. 2002). Still, a *shu* allele (shu^{PB70}) bearing a point mutation at a nonconserved residue in the putative TPR is sufficient to cause female sterility, indicating that this region is essential for some aspects of Shu function.

Shutdown is implicated in transposon silencing

Recent work has suggested that Dcr-2 is a limiting factor that prevents conventional dsRNA triggers from inducing potent RNAi in *Drosophila* germ cells, but that this restriction could be overcome by enforced Dcr-2 expression (Handler et al. 2011; Wang and Elgin 2011). We took advantage of this observation by bringing UAS-driven dsRNA constructs from the Vienna *Drosophila* RNAi Center (VDRC) into a background containing a germline-specific GAL4-driver ({GAL4-nos.NGT}40; aka *nos*-GAL4) and a UAS-*Dcr-2* transgene. Among dsRNAs targeting all fly FKBP proteins, only those corresponding to *shu* had significant impacts on expression levels of the *HetA* transposon (Fig. 1D). Moreover, only dsRNAs targeting Shu caused female sterility, a property typical of piRNA mutants (Fig. 1D).

To validate *shu* as a novel piRNA pathway component, we compared the impact of its depletion to knockdowns of known piRNA pathway genes, *armi* and *piwi*. Germline silencing of each gene resulted in a similar level of derepression for 17 transposons, measured by quantitative PCR (qPCR) (Fig. 2A). The tissue specificity of our knockdown strategy was supported by the fact that germline-specific, telomeric transposons *TAHRE*, *HetA*, and *TART* were the most heavily derepressed (greater than 150-fold, $P < 0.01$), whereas RNA levels for primarily somatic elements, such as *ZAM,* remained unchanged (about 1.2- to 1.5-fold).

Shu RNAi also recapitulated the ventralized egg phenotype of shu^{PB70} germline clones, as evidenced by a high incidence of fused or abnormal dorsal appendages (Fig. 2B; Munn and Steward 2000). Surprisingly, the ventralization phenotype was not penetrant in *armi* and *piwi* knockdowns eggs, despite the eggs being nonviable (Fig. 2C). For *armi*, prior studies of mutants produce a clear expectation of ventralization upon potent knockdown (Klattenhoff et al. 2007; Orsi et al. 2010). For *piwi*, the prediction is less clear. Germline *piwi* clones were reported not to show this distinctive phenotype; however, RNAi-mediated *piwi* knockdown did produce eggs with a spindle morphology (Cox et al. 2000; Wang and Elgin 2011). In addition to causing sterility, *shu* depletion also reduced the number of nonviable eggs laid, suggesting that there may be additional requirements for *shu* function outside of piRNA-mediated transposon silencing.

Shu is required for Piwi, Aub, and Ago3 localization

In wild-type tissues, Piwi is localized to the nucleus of germline and somatic cells (Cox et al. 2000; Saito et al. 2006; Brennecke et al. 2007). Aub and Ago3 are expressed exclusively in the germline and are enriched in a perinuclear organelle called nuage (Lim and Kai 2007; Li et al. 2009). Proper localization depends upon normal piRNA production and loading into PIWI family proteins, and disruption of this pattern is an indicator of impaired biogenesis (Malone et al. 2009).

Depletion of *shu* using the *nos*-GAL4 driver resulted in redistribution of Piwi from nurse cell nuclei to the syncytial cytoplasm of the developing egg chamber, while neighboring somatic follicle cells retain proper nuclear Piwi localization (Fig. 2D). Similarly, the ping-pong factors Ago3 and Aub were redistributed from nuage to cytoplasmic foci, while the localization of the core nuage component Vasa was not altered (Fig. 2D). Driving the *shu* dsRNA using

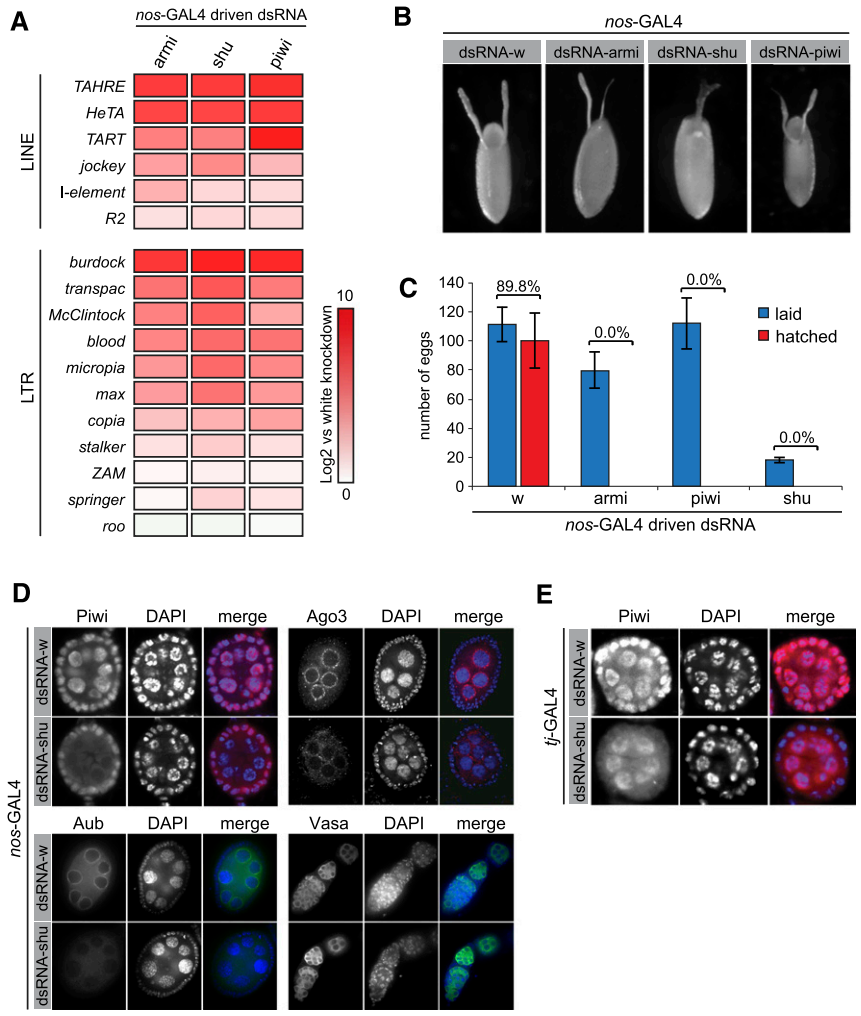


FIGURE 2. Phenotypes of *Drosophila* with germline-specific *shu* knockdown. (*A*) Depletion of *shu* in the germline results in derepression of multiple, unrelated transposons from the LINE and LTR families. Derepression, relative to *white* RNAi, is displayed as log2 fold change in heat map form. Analysis of flies with germline knockdown of Armi and Piwi, two known piRNA components, is displayed for comparison. (*B*) Germline-knockdown of *shu* causes patterning defects as indicated by the presence of fused dorsal appendages. (*C*) Depletion of *shu* causes female sterility. *shu* RNAi females lay fewer eggs compared with controls or animals depleted of other piRNA pathway factors. Hatching rates for all knockdown animals are zero, indicating complete sterility. (*D*) Depletion of *shu* in the germline using *nos*-GAL4 results in Piwi delocalization from nuclei and in Aub and Ago3 delocalization from nuage. Vasa localization is not changed. Depletion of *white* is shown as control. (*E*) *Tj*-GAL4–driven knockdown of *shu* in somatic follicle cells also causes Piwi delocalization. RNAi against *white* is shown as control.

GAL4 expressed from the soma-specific *traffic jam* promoter (*tj*-GAL4) caused delocalization of Piwi from the nuclei of follicle cells, while germline Piwi remained unaffected (Fig. 2E).

Despite its effects on the localization of PIWI-family proteins, we found that the bulk of Shutdown was not associated with domains characteristic of those piRNA pathway components. We generated N- and C-terminal GFP fusions of Shu expressed under the control of the ubiquitous *Actin5c* promoter. We examined the localization of Shutdown fusion proteins by transfection of OSS cells. Control constructs showed the expected localization with GFP-Piwi accumulating in nuclei and with GFP-Armi showing strong perinuclear localization consistent with its association with Yb-bodies. Zucchini features sequence homology with phospholipase D and was reported to localize to the outer membrane of mitochondria. In our studies, it displayed considerable overlap with the mitochondrial stain MitoTracker CMXRos (Supplemental Fig. S2). While cytoplasmic foci of GFP-tagged Shu were visible using both N- and C-terminal constructs, they did not overlap with the previously characterized localization patterns of other piRNA pathway proteins. Considered together, these data indicate that Shu is neither enriched in known structures associated with silencing nor required for assembly of a core nuage component.

Shu is essential for accumulation of both primary and secondary piRNAs

Strong derepression of germline and somatic transposons and the loss of characteristic localization patterns for Piwi-family proteins suggested that *shu* might function as a core piRNA biogenesis component, similar to *armi*. To address this possibility, we cloned and sequenced small RNAs from ovaries in which we drove the expression of *white* (*w*), *shu*, and *piwi* dsRNAs in the germline (*nos*-GAL4) or soma (*tj*-GAL4), as described above. Germline small RNAs libraries were normalized using the number of unique reads mapping to the *flam* locus, which is unaffected by germline-specific knockdowns. Germline-specific *shu* knockdown dramatically reduced the observed piRNA population compared with the *white* knockdown control. Small RNA reads with the characteristic piRNA size (23–29 nucleotides [nt]) mapping to the germline-specific, dual-strand *42AB* cluster were reduced 11.4-fold overall (8.2× on plus strand, 14.4× on minus strand). In contrast, *piwi* knockdown produced only a 2.8-fold overall reduction (2.8× on plus strand, 2.7× on minus strand) (Fig. 3A).

The incomplete loss of piRNAs in the *piwi* knockdown likely reflects the fact that piRNAs from *42AB* are normally loaded into each of the three *Drosophila* PIWI proteins

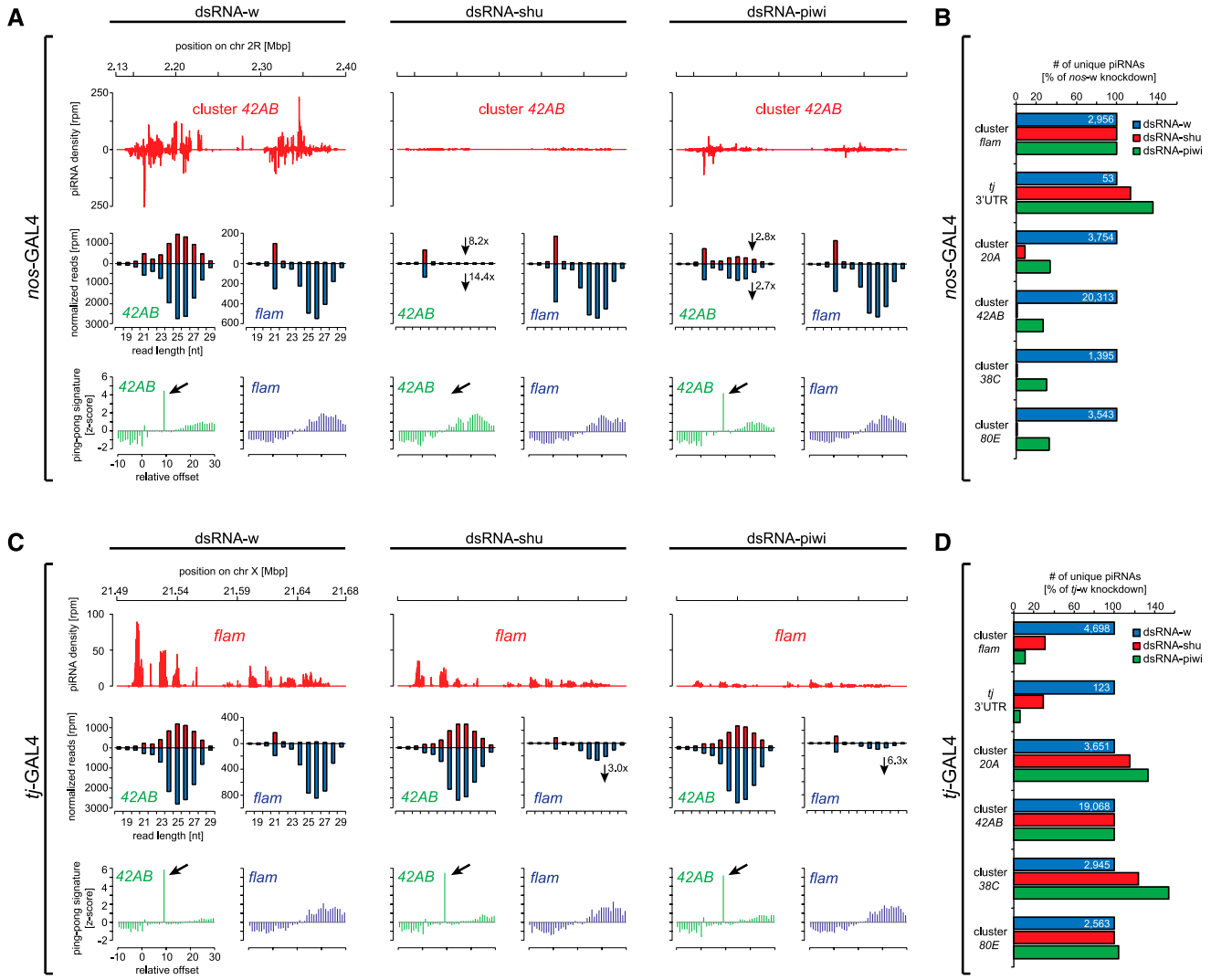


FIGURE 3. Knockdown of *shu* causes loss of cluster-derived piRNAs in both somatic and germline tissues. (*A*) At the *top* is shown a histogram of small RNAs mapping to the germline-specific *42AB* cluster in flies expressing the indicated dsRNAs specifically in germ cells. In the *middle*, the size distribution of RNAs derived from each strand of the *42AB* and *flamenco* clusters is shown as a histogram. At the *bottom* are histograms reflecting the relative enrichment of RNAs overlapping by the indicated number of nucleotides, plotted by *Z*-score, for the *42AB* and *flamenco* clusters in the indicated knockdown animals. The peak at position 9 (arrow) is indicative of a ping-pong interaction. (*B*) A histogram shows relative piRNA levels for a series of germline and somatic clusters. Total reads were normalized across libraries to piRNAs mapping to *flamenco*, which is unaffected in germline-specific knockdowns. For each cluster, changes in mapping piRNAs are shown with reference to the *white* control, which is set to 100%. *C* and *D* are similar to *A* and *B* except that dsRNA expression is driven by a follicle cell–specific *tj*-GAL4 driver. In *C*, at the *top*, reads are shown mapping to the soma-specific *flamenco* cluster. In *D*, reads are normalized across libraries to those derived from *42AB*, whose activity is not affected in the soma-specific knockdown.

(Brennecke et al. 2007), with loading of Aub and Ago3 occurring independently of Piwi function. Small RNAs mapping to this cluster in the *shu*-depleted germline also showed a clear reduction in ping-pong signatures, defined as the frequency of reads with a paired opposite strand read overlapping by 10 nt (Fig. 3A). In contrast, *piwi*, which does not participate significantly in ping-pong amplification, had no effect on ping-ping signatures upon knockdown. The effects of *shu* knockdown appear to be specific to the piRNA pathway. Reads corresponding to miRNAs were not reduced in *shu* knockdown animals (Supplemental Fig. S3). Though enforced Dicer-2 expression generally increased the endo-siRNA fraction, we did not note any further effect of *shu* knockdown, even on endo-siRNAs mapping to piRNA clusters (e.g., *42AB*) (Fig. 3A).

We also analyzed effects of *shu* or *piwi* depletion on other piRNA clusters. We compared reads that could be uniquely mapped to each annotated cluster to the *white* knockdown controls. Reads were set to 100% in the *white* library (normalized read number for the *white* knockdown library is show as a blue bar). piRNAs derived from the 3′ UTR of *traffic jam*, a genic locus that produces piRNAs only in

follicle cells, showed no impact of *shu* and *piwi* knockdowns (Fig. 3B), as expected. In contrast, all germline clusters analyzed showed a dramatic reduction of piRNA levels upon expression of *nos*-GAL4–driven *shu* dsRNAs (<10% remaining as compared to *white* RNAi). Depletion of Piwi had similar effects, although the reduction was less profound (~30% of *white* levels, as seen for *42AB*), probably due to intact Ago3 and Aub loading.

Primary and secondary piRNA biogenesis mechanisms in the germline exhibit some degree of interdependence. For example, disruption of ping-pong in *ago3* mutants or upon Aub knockdown feeds back and reduces the number of primary piRNAs loading into Piwi through unknown mechanisms (Li et al. 2009; Wang and Elgin 2011). Follicle cells, which are of somatic origin, express no detectable Aub or Ago3 and do not use ping-pong amplification. Thus, we directly tested the involvement of *shu* in primary piRNA production by sequencing small RNAs from *tj*-GAL4–driven dsRNA in ovaries. PiRNA-sized small RNAs were normalized using the number of unique reads mapping to the germline-specific *42AB* locus, which is unaffected by *tj*-GAL4–mediated knockdowns.

The sole somatic, unidirectional *flamenco* cluster produces abundant piRNAs that load only the Piwi protein. Thus, as expected, depletion of *piwi* caused a significant reduction in piRNAs derived from this locus (5.2-fold) (Fig. 3C,D). Follicular knockdown of *shu* also produced a marked reduction in *flam* piRNAs (2.9-fold) (Fig. 3C,D). As expected, piRNAs uniquely mapped to *flam* showed no ping-pong signature in any of the somatic knockdowns. Reads corresponding to germline clusters remained unchanged in piRNA abundance, with no shift in size profiles or ping-pong signatures, indicating that, as expected, the pathway remains fully functional in germ cells of animals that have lost *shu* expression only in the soma. As in germ cell–specific knockdowns, miRNA abundance was unaffected (Supplemental Fig. S3).

Mapping small RNAs from our germline-specific knockdown animals to a set of known *Drosophila* transposon consensus sequence further supported a general requirement for *shu* in piRNA accumulation. We retained in our analysis only the 75 transposons with the highest abundance of corresponding piRNAs. Previous reports have demonstrated substantial expression biases for many transposons, with some showing preferential expression in the somatic lineage and others being found predominantly in germline lineages (Malone et al. 2009). For the set of germline-enriched transposons, *nos*-GAL4–driven dsRNA-*shu* substantially affected all known elements, reducing overall piRNA levels (Fig. 4A). In general, sense and antisense piRNAs were depleted to roughly similar extents, suggesting that loading of all three PIWI clade proteins is affected by loss of *shu*. In contrast, only a subset of transposons showed depletion of piRNAs in the *nos*-GAL4–driven *piwi* knockdowns. Elements with a known somatic expression bias, including *ZAM*, *tabor*, *gypsy*, and others (indicated by red dots), show little or no reduction in piRNA levels upon germline knockdown of either *shu* or *piwi* (Fig. 4A). Transposable elements with strong germline signatures (green asterisks), like the LINE element *Rt1b* or the LTR transposon *roo* (pao family), not only showed a severe reduction of their corresponding piRNA levels but also demonstrated a dramatic loss of ping-pong signatures (Fig. 4B). In contrast, soma-specific elements retain their piRNA levels and generally lack ping-pong signatures. As an example of such an element, piRNA levels for the LTR element *ZAM* (gypsy family) are shown (Fig. 4B, bottom).

Summary

A combination of biochemical and genetic approaches are beginning to link a substantial number of proteins to functions in the piRNA pathway. Some act exclusively in primary piRNA biogenesis and affect small RNAs in both the germline and somatic compartments of the *Drosophila* ovary (Malone et al. 2009; Haase et al. 2010; Olivieri et al. 2010; Saito et al. 2010; Handler et al. 2011; Zamparini et al. 2011). Others function exclusively in the germline, and these tend to selectively affect the ping-pong cycle that hones piRNA populations in response to the expression of transposon mRNAs or factors implicated in germline cluster transcription (Klattenhoff et al. 2009; Li et al. 2009; Patil and Kai 2010; Pane et al. 2011; Zhang et al. 2011; Anand and Kai 2012). Here, we followed clues initially provided by proteomic analysis of Piwi-family protein complexes in mice to link *shutdown*, a gene previously shown to be required for fertility in *Drosophila* females (Schupbach and Wieschaus 1991; Munn and Steward 2000), to the piRNA pathway.

Analysis of transposon expression patterns and small RNA libraries in *shu* knockdown cells and animals suggests a role either in piRNA biogenesis or in piRNA stabilization, perhaps by fostering loading of piRNAs into PIWI-family proteins. Shutdown is a member of the FKBP family and its constituent domains have been ascribed PPIase activity and the ability to interact with the HSP90 family of chaperone proteins. Either of these activities could underlie the role of Shutdown in the piRNA pathway. In particular, studies of the Argonaute clade have implicated HSP family chaperones as critical cofactors for small RNA loading (Iki et al. 2010; Iwasaki et al. 2010; Miyoshi et al. 2010; Iki et al. 2011). However, evolutionary comparisons indicate that both the PPI and HSP90-binding domains harbor variations that reduce activity when introduced into other well-studied FKBP family members. Thus, understanding the true role of Shutdown in both primary biogenesis and the ping-pong cycle will await further genetic analysis and the development of biochemical systems that recapitulate aspects of the piRNA pathway in vitro.

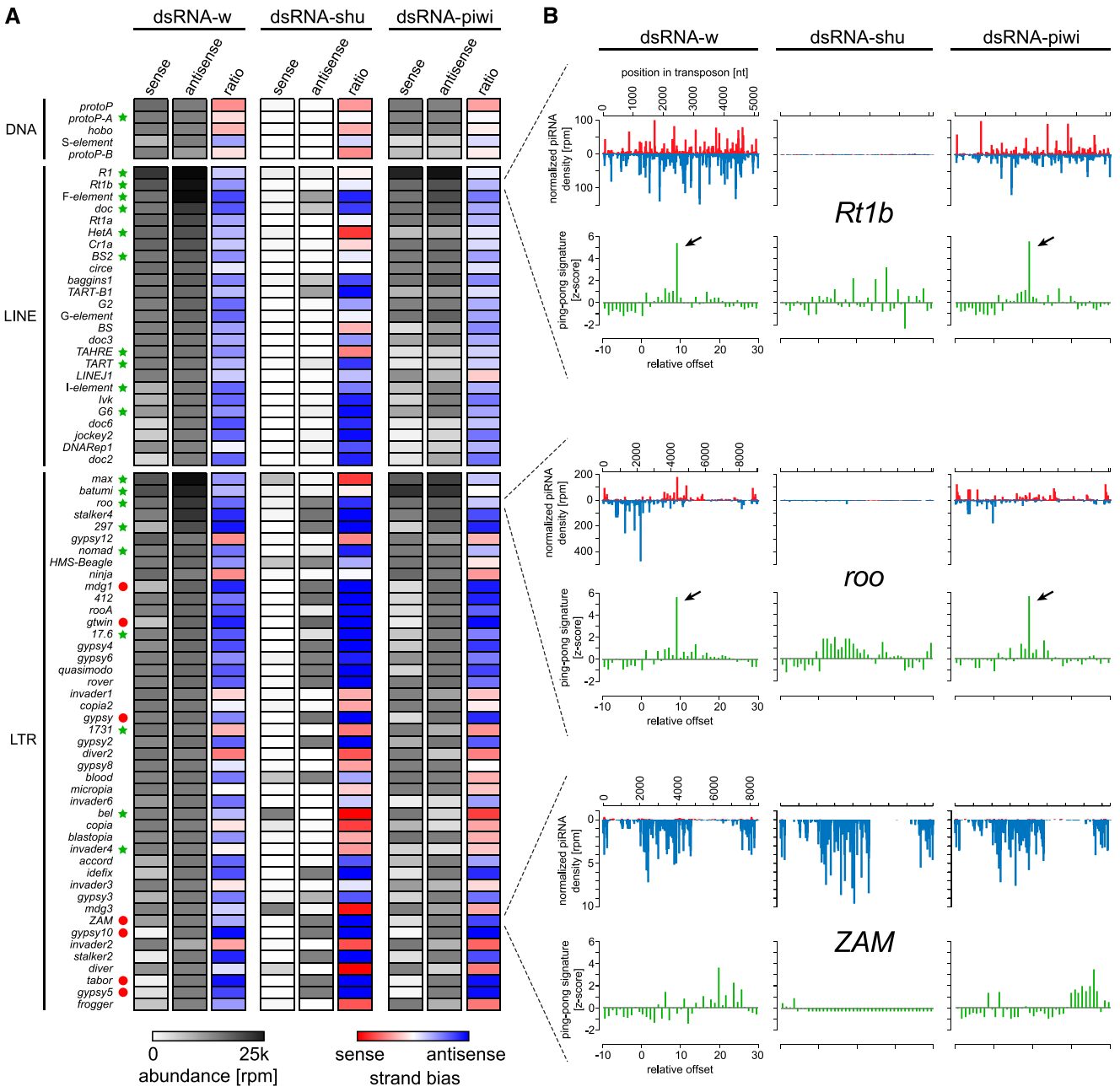


FIGURE 4. Loss of transposon control in *shu* knockdowns is a consequence of piRNA loss. (*A*) The heat map displays changes in piRNA abundance for each germline knockdown (as indicated) for the 75 elements most heavily targeted in our strain. Sense and antisense, with respect to the transposon coding strand, are quantified separately (gray heat maps), and their ratio is also indicated (red-blue heat map). (*B*) For three transposons, piRNAs are plotted along the length of the consensus sequence (*upper*) and a histogram of overlap between sense and antisense species (*lower*) is presented to indicate the degree of ping-pong (arrow highlights peak at position 9). Data are presented for *shu* and *piwi* knockdown and for a control (*white*). Two transposons with strong expression in the germline, *Rt1b* and *roo* (*top* and *middle*), are shown in comparison to a somatically biased element, *ZAM*. Since knockdown is germline specific, *ZAM* piRNAs are unaffected.

MATERIALS AND METHODS

Fly stocks and handling

Drosophila UAS-dsRNA strains were obtained from the VDRC. *nos*-GAL4 and *tj*-GAL4 driver lines were obtained from Bloomington and Kyoto, respectively (see Supplemental Table S1). For knockdown experiments, five males from dsRNA stocks were crossed with five virgin females expressing the desired GAL4 driver. Fertility of the FKBP-family F1 knockdown females was estimated by counting the number of eggs laid and crawling larvae 7 d post transfer to fresh media. Quantitative fertility measurements (shown in Fig. 1D) were obtained by transferring 3-d-old male and female F1 offspring (10 each) to grape-agar plates for

4 h and counting the eggs laid. Hatching frequencies were ascertained after 24 h (measurements were carried out in triplicate). For qPCR, small RNA libraries, and immunofluorescence experiments, ovaries were dissected from 2- to 3-d-old females fed with fresh yeast paste.

Expression of tagged transgenes in OSS

Full-length coding sequences of Shu, Piwi, Armi-RB, and Zuc were amplified from *Drosophila* ovary cDNA, cloned into pENTR/D-TOPO, and recombined into N- or C-terminal GFP destination vectors of the *Drosophila* Gateway collection (Terence Murphy, Carnegie Institute of Washington, Baltimore, MD). *Shu* was cloned into pAGW and pAWG, Zuc into pUWG, and Piwi and Armi into pUGW. Cells were transfected using Xfect reagent (Clontech) and costained with DAPI and MitoTracker Red CMXRos (Invitrogen).

RNA isolation, reverse transcription, qPCR

Ovaries were dissected into cold 1× PBS. Total RNA was extracted using TRIzol reagent (Invitrogen) following the manufacturer's instructions. One microgram RNA was treated with DNase I Amplification Grade (Invitrogen) according to the manufacturer's instructions. Complementary DNA was prepared by reverse transcription using oligo$(dT)_{20}$ primer and SuperScript III Reverse Transcriptase (Invitrogen). qPCR was carried out using SYBR Green PCR Master Mix (Applied Biosystems) and primers listed in Supplemental Table S2 on a Chromo4 Real-Time PCR Detector (BioRad). Transcripts were quantitated by the $\Delta\Delta C_t$ method (Livak and Schmittgen 2001), and normalized to transcript levels of *rp49*. Fold changes are expressed relative to control dsRNA-white knockdown RNA. Significance was calculated using a one-tailed heteroscedastic Student *t*-test of *rp49*-subtracted transposon c(t) values. All experiments were carried out in triplicates, with the average results shown.

Immunofluorescence

Ovaries were fixed in freshly prepared 4% paraformaldehyde for 20 min at room temperature. Blocking and permeabilization were carried out simultaneously in wash buffer (50 mM Tris at pH 6.8, 150 mM NaCl, 0.5% NP-40) supplemented with bovine serum albumin (5 mg/mL). All primary antibodies were diluted 1:1000 and incubated overnight at 4°C in wash buffer plus 1 mg/mL BSA. Anti-Ago3 and Anti-Piwi were generated in our laboratory (Brennecke et al. 2007); monoclonal mouse anti-Aub was provided by Mikiko Siomi (Nishida et al. 2007); and rabbit anti-Vasa (d-260) was purchased from Santa Cruz. Secondary AlexaFluor-488 and -568 antibodies were purchased from Invitrogen and used at 1:1000. Images were acquired on a Perkin Elmer UltraVIEW spinning disk confocal microscope.

RNAseq data analysis

For transcriptome libraries, 1 μg of total RNA from OSS cells transfected with GFP control dsRNA was used as input for the Illumina mRNA-Seq sample prep kit (catalog no. RS-930-1001). Libraries were made following the instructions by the manufacturer and sequenced on the Illumina GAII platform. RNAseq data were deposited to the Gene Expression Omnibus database (www.ncbi.nlm.nih.gov/geo/) under accession no. GSE38090. Publically available ovarian RNAseq data (GEO accession no. GSM424751) (Gan et al. 2010) were reanalyzed for this study. Raw sequence reads were iteratively mapped to the *Drosophila* genome (version dm3) using Bowtie (Langmead et al. 2009) with a tolerance of up to two mismatches. Remaining reads were also mapped to RefGene-annotated exon junctions with TopHat (Trapnell et al. 2009). Transcripts were quantitated using Cufflinks (Trapnell et al. 2010) and expressed as fragments per kilobase per million reads (fkpm) for relative comparison of FKBP family mRNA expression in the ovary.

Small RNA libraries and bioinformatic analysis

Small RNAs were cloned as described (Brennecke et al. 2007). For this study, the following small RNA libraries from total RNAs were prepared:

19–28 nt from *tj*-GAL4–driven dsRNA against white,
19–28 nt from *tj*-GAL4–driven dsRNA against shu,
19–28 nt from *tj*-GAL4–driven dsRNA against piwi,
19–28 nt from *nos*-GAL4–driven dsRNA against white,
19–28 nt from *nos*-GAL4–driven dsRNA against shu, and
19–28 nt from *nos*-GAL4–driven dsRNA against piwi.

Libraries were sequenced in-house using the Illumina GAII sequencing platform. Small RNA sequences were deposited in the Gene Expression Omnibus database (www.ncbi.nlm.nih.gov/geo/) under accession no. GSE38089. The analysis of small RNA libraries was performed similarly as described (Czech et al. 2008). In brief, Illumina reads were stripped of the 3′ linker and collapsed, and the resulting small RNA sequences were matched to the *Drosophila* release 5 genome (version dm3) without mismatches. Only reads that met these conditions were subjected to further analyses. For annotations we used a combination of UCSC (repeats/transposons; noncoding RNAs), miRBase (microRNAs), and FlyBase (protein coding genes; noncoding RNAs) tracks, as well as custom tracks (for synthetic markers, endo-siRNAs from structured loci, and miR and miR* strands) with different priorities. For comparison of small RNA counts between samples, libraries of dsRNA-white samples were set to 1 million reads. Next, all libraries were normalized based on unique piRNA-size mappers to the *flamenco* (for *nos*-GAL4 knockdowns) or *42AB* (for *tj*-GAL4 knockdowns) piRNA clusters. Heat maps were created by plotting the abundance of sequences (all piRNAs to a given transposable element or individual miRNA strands) as well as their strand bias within the indicated libraries.

Ping-pong analysis

For each piRNA, the relative frequency (*Z*-score) of an existing "neighbor" piRNA on the opposite strand within a certain window (10-nt upstream of and 30-nt downstream from each 5′ end of a piRNA) was calculated. In the case of germline and somatic piRNA clusters, this analysis was based on genomic mapping coordinates. For transposons, the 5′ coordinate of each mapping event to the respective transposon consensus sequence was used. Calculated frequencies were based on total cloning count. A spike at position 9 indicates more than average partners with a 10-nt overlap and is a signature of ping-pong amplified piRNAs.

DATA DEPOSITION

RNAseq data and small RNA sequences were deposited in the Gene Expression Omnibus database (www.ncbi.nlm.nih.gov/geo/) under accession no. GSE38098.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank members of the Hannon laboratory for helpful discussion, Yang Yu for providing DNA constructs, and Assaf Gordon for help with ping-pong analysis. We thank the VDRC, Kyoto, and Bloomington stock centers for fly stocks, and Julius Brennecke (IMBA, Austria) and Mikiko Siomi (Keio University, Japan) for antibodies. J.P. is supported by the American Cancer Society (award no. 121614-PF-11-277-01-RMC). B.C. was supported by the Boehringer Ingelheim Fonds. P.M.G. is supported by the NIH (grant 5T32GM065094) and by a William Randolph Hearst Foundation Scholarship. Work in the Hannon laboratory is supported by grants from the NIH and by a kind gift from Kathryn W. Davis. G.J.H. is an investigator of the HHMI.

Received May 14, 2012; accepted May 15, 2012.

REFERENCES

Ahearn IM, Tsai FD, Court H, Zhou M, Jennings BC, Ahmed M, Fehrenbacher N, Linder ME, Philips MR. 2011. FKBP12 binds to acylated H-ras and promotes depalmitoylation. *Mol Cell* **41:** 173–185.

Allan RK, Ratajczak T. 2011. Versatile TPR domains accommodate different modes of target protein recognition and function. *Cell Stress Chaperones* **16:** 353–367.

Anand A, Kai T. 2012. The tudor domain protein kumo is required to assemble the nuage and to generate germline piRNAs in *Drosophila*. *EMBO J* **31:** 870–882.

Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA. 2001. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr Biol* **11:** 1017–1027.

Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, et al. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442:** 203–207.

Bozzetti MP, Massari S, Finelli P, Meggio F, Pinna LA, Boldyreff B, Issinger OG, Palumbo G, Ciriaco C, Bonaccorsi S, et al. 1995. The *Ste* locus, a component of the parasitic *cry-Ste* system of *Drosophila melanogaster*, encodes a protein that forms crystals in primary spermatocytes and mimics properties of the β subunit of casein kinase 2. *Proc Natl Acad Sci* **92:** 6067–6071.

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

Chalvet F, Teysset L, Terzian C, Prud'homme N, Santamaria P, Bucheton A, Pelisson A. 1999. Proviral amplification of the Gypsy endogenous retrovirus of *Drosophila melanogaster* involves *env*-independent invasion of the female germline. *EMBO J* **18:** 2659–2669.

Chambeyron S, Popkova A, Payen-Groschene G, Brun C, Laouini D, Pelisson A, Bucheton A. 2008. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proc Natl Acad Sci* **105:** 14964–14969.

Chen Y, Pane A, Schupbach T. 2007. *Cutoff* and *aubergine* mutations result in retrotransposon upregulation and checkpoint activation in *Drosophila*. *Curr Biol* **17:** 637–642.

Chintapalli VR, Wang J, Dow JA. 2007. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet* **39:** 715–720.

Cook HA, Koppetsch BS, Wu J, Theurkauf WE. 2004. The *Drosophila* SDE3 homolog *armitage* is required for *oskar* mRNA silencing and embryonic axis specification. *Cell* **116:** 817–829.

Cox DN, Chao A, Lin H. 2000. *piwi* encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development* **127:** 503–514.

Crackower MA, Kolas NK, Noguchi J, Sarao R, Kikuchi K, Kaneko H, Kobayashi E, Kawai Y, Kozieradzki I, Landers R, et al. 2003. Essential role of Fkbp6 in male fertility and homologous chromosome pairing in meiosis. *Science* **300:** 1291–1295.

Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, et al. 2008. An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453:** 798–802.

DeCenzo MT, Park ST, Jarrett BP, Aldape RA, Futer O, Murcko MA, Livingston DJ. 1996. FK506-binding protein mutational analysis: defining the active-site residue contributions to catalysis and the stability of ligand complexes. *Protein Eng* **9:** 173–180.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* **9:** 397–405.

Gan Q, Chepelev I, Wei G, Tarayrah L, Cui K, Zhao K, Chen X. 2010. Dynamic regulation of alternative splicing and chromatin structure in *Drosophila* gonads revealed by RNA-seq. *Cell Res* **20:** 763–783.

Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442:** 199–202.

Gollan PJ, Bhave M. 2010. Genome-wide analysis of genes encoding FK506-binding proteins in rice. *Plant Mol Biol* **72:** 1–16.

Gonzalez-Reyes A, Elliott H, St Johnston D. 1997. Oocyte determination and the origin of polarity in *Drosophila*: the role of the spindle genes. *Development* **124:** 4927–4937.

Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP. 2008. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455:** 1193–1197.

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

Haase AD, Fenoglio S, Muerdter F, Guzzardo PM, Czech B, Pappin DJ, Chen C, Gordon A, Hannon GJ. 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 FS, 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.

Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, Filippov DV, Blaser H, Raz E, Moens CB, et al. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* **129:** 69–82.

Iki T, Yoshikawa M, Nishikiori M, Jaudal MC, Matsumoto-Yokoyama E, Mitsuhara I, Meshi T, Ishikawa M. 2010. In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. *Mol Cell* **39:** 282–291.

Iki T, Yoshikawa M, Meshi T, Ishikawa M. 2011. Cyclophilin 40 facilitates HSP90-mediated RISC assembly in plants. *EMBO J* **31:** 267–278.

Iwasaki S, Kobayashi M, Yoda M, Sakaguchi Y, Katsuma S, Suzuki T, Tomari Y. 2010. Hsc70/Hsp90 chaperone machinery mediates

ATP-dependent RISC loading of small RNA duplexes. *Mol Cell* **39:** 292–299.

Kamphausen T, Fanghanel J, Neumann D, Schulz B, Rahfeld JU. 2002. Characterization of *Arabidopsis thaliana At*FKBP42 that is membrane-bound and interacts with Hsp90. *Plant J* **32:** 263–276.

Kang CB, Hong Y, Dhe-Paganon S, Yoon HS. 2008. FKBP family proteins: immunophilins with versatile biological functions. *Neurosignals* **16:** 318–325.

Khurana JS, Theurkauf W. 2010. piRNAs, transposon silencing, and *Drosophila* germline development. *J Cell Biol* **191:** 905–913.

Kim A, Terzian C, Santamaria P, Pelisson A, Purd'homme N, Bucheton A. 1994. Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc Natl Acad Sci* **91:** 1285–1289.

Klattenhoff C, Bratu DP, McGinnis-Schultz N, Koppetsch BS, Cook HA, Theurkauf WE. 2007. *Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev Cell* **12:** 45–55.

Klattenhoff C, Xi H, Li C, Lee S, Xu J, Khurana JS, Zhang F, Schultz N, Koppetsch BS, Nowosielska A, et al. 2009. The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* **138:** 1137–1149.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25. doi: 10.1186/gb-2009-10-3-r25.

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE. 2006. Characterization of the piRNA complex from rat testes. *Science* **313:** 363–367.

Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, Seitz H, Horwich MD, Syrzycka M, Honda BM, et al. 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137:** 509–521.

Li L, Lou Z, Wang L. 2011. The role of FKBP5 in cancer aetiology and chemoresistance. *Br J Cancer* **104:** 19–23.

Lim AK, Kai T. 2007. Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. *Proc Natl Acad Sci* **104:** 6714–6719.

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

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}{}_{T}$ method. *Methods* **25:** 402–408.

Malone CD, Hannon GJ. 2009. Small RNAs as guardians of the genome. *Cell* **136:** 656–668.

Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ. 2009. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* **137:** 522–535.

McClintock B. 1953. Induction of instability at selected loci in maize. *Genetics* **38:** 579–599.

Miyoshi T, Takeuchi A, Siomi H, Siomi MC. 2010. A direct role for Hsp90 in pre-RISC formation in *Drosophila*. *Nat Struct Mol Biol* **17:** 1024–1026.

Munn K, Steward R. 2000. The *shut-down* gene of *Drosophila melanogaster* encodes a novel FK506-binding protein essential for the formation of germline cysts during oogenesis. *Genetics* **156:** 245–256.

Nishida KM, Saito K, Mori T, Kawamura Y, Nagami-Okada T, Inagaki S, Siomi H, Siomi MC. 2007. Gene silencing mechanisms mediated by Aubergine–piRNA complexes in *Drosophila* male gonad. *RNA* **13:** 1911–1922.

Olivieri D, Sykora MM, 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.

Orsi GA, Joyce EF, Couble P, McKim KS, Loppin B. 2010. *Drosophila I-R* hybrid dysgenesis is associated with catastrophic meiosis and abnormal zygote formation. *J Cell Sci* **123:** 3515–3524.

Pane A, Wehr K, Schupbach T. 2007. *zucchini* and *squash* encode two putative nucleases required for rasiRNA production in the Drosophila germline. *Dev Cell* **12:** 851–862.

Pane A, Jiang P, Zhao DY, Singh M, Schupbach T. 2011. The Cutoff protein regulates piRNA cluster expression and piRNA production in the *Drosophila* germline. *EMBO J* **30:** 4601–4615.

Patil VS, Kai T. 2010. Repression of retroelements in *Drosophila* germline via piRNA pathway by the Tudor domain protein Tejas. *Current biology: CB* **20:** 724–730.

Pelisson A, Song SU, Prud'homme N, Smith PA, Bucheton A, Corces VG. 1994. Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila flamenco* gene. *EMBO J* **13:** 4401–4411.

Pratt WB. 1998. The hsp90-based chaperone system: involvement in signal transduction from a variety of hormone and growth factor receptors. *Proc Soc Exp Biol Med* **217:** 420–434.

Pratt WB, Galigniana MD, Harrell JM, DeFranco DB. 2004. Role of hsp90 and the hsp90-binding immunophilins in signalling protein movement. *Cell Signal* **16:** 857–872.

Ratajczak T, Carrello A. 1996. Cyclophilin 40 (CyP-40), mapping of its hsp90 binding domain and evidence that FKBP52 competes with CyP-40 for hsp90 binding. *J Biol Chem* **271:** 2961–2965.

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

Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, Siomi H, Siomi MC. 2010. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes Dev* **24:** 2493–2498.

Schupbach T, Wieschaus E. 1989. Female sterile mutations on the second chromosome of *Drosophila melanogaster*. I. Maternal effect mutations. *Genetics* **121:** 101–117.

Schupbach T, Wieschaus E. 1991. Female sterile mutations on the second chromosome of *Drosophila melanogaster*. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* **129:** 1119–1136.

Senti KA, Brennecke J. 2010. The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet* **26:** 499–509.

Smith MR, Willmann MR, Wu G, Berardini TZ, Moller B, Weijers D, Poethig RS. 2009. Cyclophilin 40 is required for microRNA activity in *Arabidopsis*. *Proc Natl Acad Sci* **106:** 5424–5429.

Szakmary A, Reedy M, Qi H, Lin H. 2009. The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in *Drosophila melanogaster*. *J Cell Biol* **185:** 613–627.

Theurkauf WE, Klattenhoff C, Bratu DP, McGinnis-Schultz N, Koppetsch BS, Cook HA. 2006. rasiRNAs, DNA damage, and embryonic axis specification. *Cold Spring Harb Symp Quant Biol* **71:** 171–180.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25:** 1105–1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28:** 511–515.

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

Vagin VV, Wohlschlegel J, Qu J, Jonsson Z, Huang X, Chuma S, Girard A, Sachidanandam R, Hannon GJ, Aravin AA. 2009. Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes Dev* **23:** 1749–1762.

Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J. 1993. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* **229:** 105–124.

Wang SH, Elgin SC. 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.

Ward BK, Allan RK, Mok D, Temple SE, Taylor P, Dornan J, Mark PJ, Shaw DJ, Kumar P, Walkinshaw MD, et al. 2002. A structure-based mutational analysis of cyclophilin 40 identifies key residues in the core tetratricopeptide repeat domain that mediate binding to Hsp90. *J Biol Chem* **277:** 40799–40809.

Wilson JE, Connell JE, Macdonald PM. 1996. *aubergine* enhances *oskar* translation in the *Drosophila* ovary. *Development* **122:** 1631–1639.

Zamparini AL, Davis MY, Malone CD, Vieira E, Zavadil J, Sachidanandam R, Hannon GJ, Lehmann R. 2011. Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in *Drosophila*. *Development* **138:** 4039–4050.

Zhang Z, Xu J, Koppetsch BS, Wang J, Tipping C, Ma S, Weng Z, Theurkauf WE, Zamore PD. 2011. Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. *Mol Cell* **44:** 572–584.

Appendix 3: Dnmt2-dependent methylomes lack biologically relevant DNA methylation patterns

Contributions to manuscript:

I was responsible for construction of bisulfite converted genomic DNA libraries from *Drosophila* embryos. Four libraries were constructed, three from wild-type flies and one from *dnmt2* mutant flies. This manuscript is being prepared for submission to *PNAS*.

Manuscript citation:

Raddatz G, Guzzardo PM, Olova N, Fantappié MR, Rampp M, Schaefer M, Reik W, Hannon GJ, and Lyko F. Dnmt2-dependent methylomes lack biologically relevant DNA methylation patterns. *In preparation.*

Dnmt2-dependent methylomes lack biologically relevant DNA methylation patterns

Günter Raddatz[1,*], Paloma M. Guzzardo[2,*], Nelly Olova[3], Marcelo Rosado Fantappié[4], Markus Rampp[5], Matthias Schaefer[1], Wolf Reik[3,6], Gregory J. Hannon[2,7], and Frank Lyko[1]

[1]Division of Epigenetics, DKFZ-ZMBH Alliance, German Cancer Research Center; Heidelberg, Germany; [2]Watson School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA; [3]Epigenetics Programme, The Babraham Institute, Cambridge, UK; [4]Instituto de Bioquímica Médica, Programa de Biologia Molecular e Biotecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; Department of Molecular Microbiology, [5]Computing Center (RZG) of the Max-Planck-Society, Max Planck Institute of Plasma Physics, Garching, Germany; [6]Wellcome Trust Sanger Institute, Cambridge, UK; [7]Howard Hughes Medical Institute

*G.R. and P.M.G. contributed equally to this study.

Corresponding author: Frank Lyko
Deutsches Krebsforschungszentrum
Im Neuenheimer Feld 580
69120 Heidelberg, Germany
phone: ++49-6221-423800
fax: ++49-6221-423802
e-mail: f.lyko@dkfz.de

Summary

Several organisms have retained *Dnmt2* as their only bona fide DNA methyltransferase gene. However, information about Dnmt2-dependent methylation patterns has been limited to a few isolated loci and the results have been discussed controversially. In addition, recent studies have shown that *Dnmt2* functions as a tRNA methyltransferase, which raised the possibility that *Dnmt2*-only genomes might be unmethylated. We have now used whole-genome bisulfite sequencing to analyze the methylomes of *Dnmt2*-only organisms at single-base resolution. Our results show that the genomes of *Schistosoma mansoni* and *Drosophila melanogaster* lack detectable DNA methylation patterns. Residual unconverted cytosine residues shared many attributes with bisulfite deamination artifacts and were observed at comparable levels in *Dnmt2*-deficient flies. Furthermore, genetically modified *Dnmt2*-only mouse embryonic stem cells lost the DNA methylation patterns found in wildtype cells. Our results thus uncover fundamental differences among animal methylomes and suggest that *Dnmt2*-only organisms lack biologically relevant DNA methylation patterns.

Introduction

DNA methylation is an epigenetic modification with important functions in cellular differentiation, organismal development and human disease (Feinberg, 2007; Mohn and Schubeler, 2009). DNA methylation is widely conserved in the animal and plant kingdoms and established and maintained by a family of enzymes termed DNA methyltransferases (Goll and Bestor, 2005; Law and Jacobsen, 2010). In animals, Dnmt3 enzymes act as de novo methyltransferases that establish DNA methylation patterns, most prominently during early stages of development. Established methylation patterns are then maintained by Dnmt1 enzymes, which copy methylation marks from methylated CpG dinucleotides on the parental DNA strand to complementary CpG dinucleotides on the daughter strand.

Dnmt2 (also known as Trdmt1) is the second member of the DNA methyltransferase family and shows strong sequence conservation to the catalytic motifs of established DNA methyltransferases, (Dong et al., 2001; Okano et al., 1998; Yoder and Bestor, 1998). However, the actual DNA methyltransferase activity of Dnmt2 was found to be substantially weaker than for other DNA methyltransferases (Goll et al., 2006; Hermann et al., 2003). It was later shown that Dnmt2 has a strong methyltransferase activity towards cytosine 38 in the anticodon loop of $tRNA^{Asp}$ and other tRNAs (Goll et al., 2006; Schaefer et al., 2010), which has been linked to the regulation of tRNA stability and protein synthesis (Schaefer et al., 2010; Tuorto et al., 2012). In agreement with this notion, several independent phylogenetic analyses have suggested that Dnmt2 is an ancient DNA methyltransferase that has switched its substrate specificity from DNA to tRNA (Iyer et al., 2011; Jurkowski and Jeltsch, 2011; Sunita et al., 2008).

The number of *Dnmt* genes can vary greatly between genomes and various organisms have been shown to encode different sets of Dnmt enzymes (Zemach and Zilberman, 2010). Mammalian genomes encode one *Dnmt1* gene and three paralogs of *Dnmt3*. This contrasts, for example, with the genome of the parasitic wasp *Nasonia vitripennis*, which encodes three paralogs of *Dnmt1* and a single *Dnmt3* homologue. These

variations have been interpreted to reflect multiple versions of a toolkit for phenotypic adaptation (Lyko and Maleszka, 2011). During evolution, specific parts of this toolkit could have been contracted or expanded to facilitate specific requirements for genome regulation (Lyko and Maleszka, 2011).

Interestingly, there is a diverse group of animal species that have retained *Dnmt2* as their only bona fide DNA methyltransferase. Global DNA methylation levels in these *Dnmt2*-only organisms have been found to be very low and are often discussed controversially (Jeltsch et al., 2006; Krauss and Reuter, 2011; Schaefer and Lyko, 2010). More recently, however, two prominent studies have provided support for a biologically important function of Dnmt2-dependent DNA methylation. For example, it has been suggested that Dnmt2-dependent DNA methylation regulates oviposition in *Schistosoma mansoni* (Geyer et al., 2011), the causative agent of bilharziosis. In addition, Dnmt2-dependent methylation of transposons has been linked to genome stability in *Drosophila melanogaster* (Phalke et al., 2009). These results have necessitated a more detailed analysis of genome methylation patterns in *Schistosoma* and *Drosophila*.

Over the past few years, whole-genome bisulfite sequencing has been established as a method to characterize genome-wide DNA methylation patterns at single-base resolution (Lister and Ecker, 2009). This method has been successfully used to characterize the methylomes of various animal organisms that are known to establish and maintain their methylation patterns by Dnmt1 and/or Dnmt3 enzymes. The results revealed a certain degree of diversity among animal methylomes, but also identified a number of conserved features, which include the specificity for CpG dinucleotides and an enrichment of methylation in defined genetic elements (Feng et al., 2010; Zemach et al., 2010). We have now used whole-genome bisulfite sequencing for an unbiased characterization of the *Schistosoma* and *Drosophila* methylomes. Our results fail to reveal any evidence for biologically relevant DNA methylation patterns in these organisms and thus uncover fundamental differences between Dnmt1/3-dependent and Dnmt2-dependent methylomes.

Results

To analyze the genomic DNA methylation pattern of *Schistosoma mansoni*, we isolated genomic DNA from the same strain (Puerto Rican) and developmental stage (adult worms) that was used for previous analyses (Fantappie et al., 2001; Geyer et al., 2011). After bisulfite deamination, DNA libraries were prepared and subjected to paired-end Illumina sequencing. Read pairs were subsequently mapped to the *S. mansoni* reference genome using BSMAP 2.0 (Xi and Li, 2009). This resulted in an average strand-specific genome coverage of 13x (Table 1). The conversion rate of mapped cytosine residues was 99.0 % (Table 1), which suggested that bisulfite deamination had been efficient and that the sequence information could be used for methylation analysis.

As an initial step towards a comprehensive methylome analysis, non-conversion ratios were determined for all cytosine residues with a sequence coverage >3. This revealed that 97% of the cytosines were completely converted (ratio <0.1). When the remaining cytosines were distributed into bins with increasing non-conversion ratios, the majority (93%) showed ratios <0.5, while only a minor fraction (7%) had ratios >0.5 (Figure 1A). This distribution strongly contrasts the results from honeybees, which have an established Dnmt1/3-dependent methylome with very low levels of DNA methylation (Lyko et al., 2010). Here, only 21% of all not completely converted cytosines showed a non-conversion (methylation) ratio <0.5, while 79% had a ratio of >0.5 (Figure 1A). These results suggest that the *Schistosoma* genome either contains extremely low levels or no DNA methylation at all.

A conserved feature of all known animal methylomes is their high degree of CpG-specificity (Feng et al., 2010; Zemach et al., 2010), which is related to the stable transmission of symmetric CpG methylation marks by maintenance methylation mechanisms. Therefore, we also determined the dinucleotide sequence context of non-converted cytosines in our *Schistosoma* dataset and could not detect any evidence for CpG specificity (Figure

1B). In contrast, unconverted cytosines from the honeybee methylome data set were strongly (>95%) enriched for CpG dinucleotides (Figure 1B). Finally, we also used our data for a detailed analysis of the *Schistosoma forkhead* gene that had previously been reported to be methylated in a *Dnmt2*-dependent manner (Geyer et al., 2011). Our results failed to reveal any evidence for DNA methylation at the previously analyzed region (Figure 1C). We did, however, observe dense clusters of incompletely converted cytosines outside of this region (Figure 1C), preferentially in regions of low sequence complexity. Similar patterns have not been described in any of the published methylomes, but correspond to known characteristics of bisulfite deamination artifacts (Warnecke et al., 2002). Taken together, these results strongly suggest that the *Schistosoma* genome is unmethylated.

To characterize the methylome of a second *Dnmt2*-only organism, we obtained genome-wide methylation profiles from *Drosophila melanogaster* embryos. A previous phylogenetic analysis of DNA methylation patterns had already suggested that DNA from *Drosophila* embryos is unmethylated (Zemach et al., 2010). However, the data was obtained from an unspecified strain and the sequencing coverage was comparably low (Table S1). To further characterize the *Drosophila* methylome, we therefore obtained an independent methylation profile from 0-2h w^{1118} embryos. This is the same strain and developmental stage that was used for a previous analysis describing DNA methylation of *Invader4* elements (Phalke et al., 2009). As a reference, 1% of human sperm DNA was spiked into the *Drosophila* sample prior to bisulfite conversion. Human sperm DNA is known to be highly methylated (Molaro et al., 2011) and the spiked-in DNA sample could thus serve as an important internal control. We obtained an average strand-specific genome coverage of 32x and a conversion rate of 99.7% for the *Drosophila* sample (Table 1), which suggested that the sequencing data was well-suited for further analysis.

A detailed analysis of the *Drosophila* data showed that the vast majority (99.7%) of cytosine residues appeared completely unmethylated (ratio <0.1), while only 0.003% showed a non-conversion ratio >0.5 (Figure 2A). This distribution was substantially different for the spiked-in human sperm DNA, which showed complete methylation (ratio >0.9) for 4.3 % of

the cytosine residues that were analyzed (Figure 2A). Pronounced differences between the *Drosophila* and the control sample were also detectable for the dinucleotide sequence context of non-converted cytosine residues. In the *Drosophila* dataset, only 11% of the non-converted cytosine residues were found in CpG dinucleotides (Figure 2B). This distribution strongly contrasted the control sample, which showed a high degree (98%) of CpG specificity (Figure 2B). Finally, we also used our data for a detailed analysis of *Drosophila Invader4* elements, which have previously been reported to be methylated in a Dnmt2-dependent manner (Phalke et al., 2009). Our results failed to reveal any evidence for DNA methylation at the previously analyzed region and at other *Invader4* sequences (Figure 2C). Together, these results strongly suggest that the *Drosophila* genome is unmethylated.

To investigate the significance of *Dnmt2* for the remaining non-converted cytosine residues in *Drosophila*, we generated genome-wide methylation profiles from homozygous *Dnmt2* mutant embryos. We isolated genomic DNA from the same strain and developmental stage that was used for previous analyses (Phalke et al., 2009). We obtained an average strand-specific genome coverage of 24x and a conversion rate of 99.4% (Table 1). Further data analysis showed that the vast majority (99.0%) of cytosine residues appeared completely unmethylated (Figure 2D). Compared to the wildtype *Drosophila* sample, non-converted cytosine residues appeared to be slightly more frequent in the *Dnmt2* mutant sample (Figure 2D), which can probably be attributed to a slight variation in the bisulfite conversion efficiency.

To further investigate the significance of the remaining non-converted cytosine residues in the *Drosophila* datasets, we used a sliding window approach to identify sequences with an increased density of non-converted cytosines. The results showed that the vast majority (>99.9%) of 1 kb windows completely lacked non-converted cytosines (Figure 2E). Among the remaining windows, many contained a single non-converted cytosine residue (Figure 2E). Notably, a substantial fraction of windows with >2 non-converted cytosine residues in wildtype embryos also showed inefficient conversion in *Dnmt2* mutant embryos (Figure 2F). These results further argue against Dnmt2-mediated DNA methylation

and suggest that a large fraction of unconverted cytosines resides in sequences that are resistant to bisulfite deamination.

To further confirm the lack of DNA methylation in *Dnmt2*-only systems we also used bisulfite sequencing to analyze DNA methylation patterns in the mouse TKO ES cell line (Tsumura et al., 2006). This cell line is deficient for Dnmt1, Dnmt3a and Dnmt3b but has retained an intact *Dnmt2* gene (Fig. 3A). We obtained an average strand-specific genome coverage of 1x and a conversion rate of 98.3% (Table 1). A detailed analysis of the sequencing data showed that the vast majority (96.5%) of cytosine residues appeared completely unmethylated (ratio <0.1), while only 0.03% showed a non-conversion ratio >0.5 (Figure 3B). A substantially different distribution was observed in a published dataset from wildtype mouse ES cells (Stadler et al., 2011), which showed complete methylation (ratio >0.9) for 2.5 % of the cytosine residues (Figure 3B). A more detailed analysis showed a high proportion of CpG methylation and significant levels of non-CpG methylation in wildtype ES cells (Figure 3C), consistent with earlier observations (Lister et al., 2009; Ramsahoye et al., 2000). Even with a low-stringency cutoff (ratio >0.1), the TKO cells showed substantially reduced signals, which were similar to an unmethylated PCR fragment that had been spiked into the TKO sample prior to bisulfite conversion (Figure 3C). We also used the spiked-in PCR fragment to determine the significance of the observed non-converted cytosine residues by a statistical approach. The difference in the conversion efficiencies between the TKO dataset and the spiked-in control appeared small (0.12%, Table S2), but achieved borderline statistical significance (P=0.012, Fisher's exact test). It is possible that this result reflects a carry-over of residual amounts of DNA methylation from the parental ES cell line (Tsumura et al., 2006). It is also conceivable that the genomic DNA is intrinsically more resistant to bisulfite conversion than the PCR fragment. Similar, confounding factors did not exist for the analysis of the wildtype *Drosophila* dataset and the spiked-in human sperm DNA. Indeed, the difference in the deamination rate between the *Drosophila* sample and the spiked-in control was very small (0.04%, Table S2) and not statistically significant (P=1.0, Fisher's exact test). These results provide important confirmation for the absence of biologically relevant DNA

methylation patterns in *Dnmt2*-only organisms.

Discussion

The DNA methylation status of *Dnmt2*-only organisms has been a controversial topic for a long time. This may be related to the fact that the reported methylation levels were often close to the detection limits of the various methods that were used for DNA methylation analysis (reviewed in Krauss and Reuter, 2011). The results from chromatographical analyses (e.g., Geyer et al., 2011; Gowher et al., 2000; Lyko et al., 2000), may also have been affected by contamination with methylated DNA from other organisms, including bacteria. Similarly, immunodetection approaches of 5-methyl-cytosine (e.g., Kunert et al., 2003) could have been affected by low antibody specificity. Also, many previous bisulfite sequencing analyses were limited to isolated genomic loci (e.g., Geyer et al., 2011; Phalke et al., 2009), which made them more susceptible to false positive results. Finally, it is also possible that the conserved catalytic mechanism of Dnmt2 (Jurkowski et al., 2008) permits a limited „star activity“, i.e. a low enzymatic activity with relaxed substrate specificity, on DNA substrates. This “star activity” could be responsible for residual amounts of genuine DNA methylation and might become increased under certain experimental conditions (Hermann et al., 2003; Kunert et al., 2003). However, since we could not detect any relevant DNA methylation patterns in our analyses, we would interprete these spurious methylation marks as biological artifacts.

The comprehensive nature of whole-genome bisulfite sequencing datasets allows additional quality control steps during data analysis and permits the identification of false positives with higher sensitivity (Lister and Ecker, 2009). Also, whole-genome bisulfite sequencing protocols use substantially fewer PCR amplification cycles than locus-specific bisulfite sequencing protocols. This reduces the impact of PCR bias, another prominent source for false positive results in bisulfite sequencing (Warnecke et al., 1997). Our whole-

genome bisulfite sequencing approach therefore allowed an unbiased characterization of the *Schistosoma* and *Drosophila* methylomes and revealed that both organisms lack biologically relevant DNA methylation patterns. Our findings thus establish fundamental differences between *Dnmt2*-dependent and *Dnmt1/3*-dependent methylomes. We cannot presently exclude the possibility that Dnmt2 expression and, consequently, DNA methylation vary according to unknown environmental cues (Krauss and Reuter, 2011). However, our analyses used the same strains and culture conditions as previously reported (Geyer et al., 2011; Phalke et al., 2009). Furthermore, our results were derived from common laboratory strains of *Schistosoma mansoni* and *Drosophila melanogaster* and can therefore be considered as a reference for future studies.

In conclusion, our results directly contradict previous reports (Geyer et al., 2011; Phalke et al., 2009), which had suggested the presence of biologically relevant DNA methylation patterns in *Schistosoma* and *Drosophila*. These models are important representatives from a phylogenetically diverse group of organisms that do not encode a canonical DNA methyltransferase enzyme (Dnmt1 or Dnmt3), but have retained a *Dnmt2* gene. Our results establish fundamental differences between Dnmt2-dependent and Dnmt1/3-dependent methylomes and suggest that the genomes of *Dnmt2*-only organisms lack DNA methylation.

Experimental Procedures

DNA samples

For *Schistosoma* genomic DNA, adult male worms from the *S. mansoni* Puerto Rican strain were homogenized and DNA was isolated using the Qiagen DNeasy Blood & Tissue Kit, according to the manufacturer's instructions.

For *Drosophila* genomic DNA, 0-2 hour embryos were collected and dechorionated. To extract genomic DNA, embryos were homogenized in 2x PK Buffer (200 mM Tris-HCl, pH

7.5, 25 mM EDTA, pH 8, 300 mM NaCl, 2% SDS), followed by protein digestion with Proteinase K for 2 hours at 65ºC. DNA was extracted with an equal volume of phenol-chloroform, followed by ethanol precipitation. Samples were then treated with RNAse A for 15 minutes at 37ºC and genomic DNA was purified by phenol-chloroform extraction. For human sperm DNA, sperm cells were purified and then sonicated. Genomic DNA was purified by double phenol-chloroform extraction and ethanol precipitation.

For mouse TKO ES DNA, cells derived from the Dnmt1$^{-/-}$;Dnmt3a$^{-/-}$;Dnmt3b$^{-/-}$ clone #19 (Tsumura et al., 2006, a kind gift from Masaki Okano), were grown under feeder-free conditions on gelatine in complete ES medium (Ficz et al., 2011). Before harvesting, the cells were stained by immunofluorescence against 5-methylcytosine and Dnmt1 for potential parental cell line contaminations. Genomic DNA was prepared using the Qiagen DNeasy Blood & Tissue Kit, according to the manufacturer's instructions.

Sequencing

The *Schistosoma* bisulfite sequencing library was prepared as described previously (Lyko et al., 2010). Briefly, 5µg of high molecular weight DNA were used for fragmentation using the Covaris S2 AFA System in a total volume of 100µl. End repair of fragmented DNA was carried out in a total volume of 100µl using the Paired End DNA Sample Prep Kit (Illumina). For the ligation of the adaptors the Illumina Early Access Methylation Adaptor Oligo Kit and the Paired End DNA Sample Prep Kit (Illumina) were used. Adaptor-ligated fragments were purified using the E-Gel Electrophoresis System and directly transferred to bisulfite treatment using the EZ-DNA Methylation Kit (Zymo). Paired-end sequencing was performed on an Illumina HiSeq system.

The *Drosophila* bisulfite converted DNA libraries were generated using methods described previously (Hodges et al., 2011). In brief, 1 µg of genomic DNA was fragmented by sonication using a Bioruptor (Diagenode). For the w^{1118} sample, 10 ng of human sperm genomic DNA were spiked in before sonication. These fragments were incubated with a mixture of T4 DNA polymerase (NEB), T4 Polynucleotide kinase (NEB) and standard Taq

Polymerase (Roche), to repair ends, phosphorylate and adenylate the 3' ends. Paired-end Illumina adaptors synthesized with methylated cytosine were ligated to the fragments using Rapid T4 DNA Ligase (Roche). Ligated fragments were recovered using Zymo DNA Clean and Concentrator -5 columns and subjected to bisulfite conversion using EZ DNA Methylation kit (Zymo). Finally, a minimal PCR amplification of 15 cycles was performed using Roche High Fidelity Plus enzyme. Following this, samples were purified using Zymo DNA Clean and Concentrator -25 and prepared for paired-end sequencing for 76 cycles on an Illumina HiSeq system.

For the TKO ES bisulfite sequencing library, genomic DNA (spiked with a 2kb PCR fragment from M13mp18, 1:10,000) was fragmented via sonication with a Covaris E220 instrument in a total volume of 70 µl. Methylated adapters (Illumina) were ligated to 250ng of the fragmented DNA with the NEB Next DNA Library Prep Master Mix Set for Illumina, according to the manufacturer's instructions, and subsequently converted using the Imprint DNA Modification kit (Sigma). The ligated fragments were amplified (15 cycles) using Pfu Turbo Cx Hotstart DNA Polymerase (Agilent) and indexed adapter-specific primers for Illumina – iPCRtagT5 (Quail et al., 2012), followed by purification and size-selection using AMPure XP beads (Agencourt). Paired-end 100 bp sequencing was performed on an Illumina HiSeq system.

Reference sequences

Assembly v5.0 of the *S. mansoni* genome project (www.sanger.ac.uk/resources/downloads/helminths/schistosoma-mansoni.html), the dm3 genome assembly version of the *D. melanogaster* genome (www.flybase.org) and the mm9 assembly of the mouse genome (available at www.genome.ucsc.edu) were used as reference sequences.

Sequence mapping

Unique reads were trimmed to a maximal length of 80 bp and stretches of bases having a

quality score <30 at read ends were removed. Reads were mapped using BSMAP 2.0 (Xi and Li, 2009). Only reads mapping uniquely and with both read pairs having the correct distance were used in further analyses. Methylation rates were determined using a Python script distributed with the BSMAP package. To reduce the effects of sequencing errors, methylated Cs were only called when covered by >3 reads.

Accession numbers

Sequencing data have been deposited in the GEO database under the accession numbers GSE39996 (*D. melanogaster*, together with human sperm spike-in), GSE39997 (*S. mansoni*) and GSE42170 (mouse TKO ES cell line, together with M13 PCR fragment spike-in).
Reviewer access:
www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=tvoxlymekkkycry&acc=GSE39996
www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=zpujdmogeusycjq&acc=GSE39997
www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=dpurnewcksycqtg&acc=GSE42170

Acknowledgements

We thank Stephan Wolf, Andre Leischwitz and Kristina Tabbada for the Illumina sequencing services, Felix Krueger for bioinformatics support, Tobias Reber for IT support and Brigitte Engelhardt for graphical support. This work was supported by grants from the BBSRC, MRC and The Wellcome Trust to W.R., the NIH (5R01GM062534-12) to G.J.H., and from the Deutsche Forschungsgemeinschaft to M.S. and F.L. (FOR1082). P.M.G. is a NIH trainee on a CSHL WSBS NIH Kirschstein-NRSA pre-doctoral award (T32 GM065094), a William Randolph Hearst Scholar and a Leslie Quick Junior Fellow. N.O. is a recipient of a BBSRC CASE Studentship and a BI/EU Studentship.

References

Dong, A., Yoder, J. A., Zhang, X., Zhou, L., Bestor, T. H., and Cheng, X. (2001). Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. Nucleic Acids Res *29*, 439-448.

Fantappie, M. R., Gimba, E. R., and Rumjanek, F. D. (2001). Lack of DNA methylation in Schistosoma mansoni. Exp Parasitol *98*, 162-166.

Feinberg, A. P. (2007). Phenotypic plasticity and the epigenetics of human disease. Nature *447*, 433-440.

Feng, S., Cokus, S. J., Zhang, X., Chen, P. Y., Bostick, M., Goll, M. G., Hetzel, J., Jain, J., Strauss, S. H., Halpern, M. E.*, et al.* (2010). Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A *107*, 8689-8694.

Ficz, G., Branco, M. R., Seisenberger, S., Santos, F., Krueger, F., Hore, T. A., Marques, C. J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature *473*, 398-402.

Geyer, K. K., Rodriguez Lopez, C. M., Chalmers, I. W., Munshi, S. E., Truscott, M., Heald, J., Wilkinson, M. J., and Hoffmann, K. F. (2011). Cytosine methylation regulates oviposition in the pathogenic blood fluke Schistosoma mansoni. Nat Commun *2*, 424.

Goll, M. G., and Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. Annu Rev Biochem *74*, 481-514.

Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C. L., Zhang, X., Golic, K. G., Jacobsen, S. E., and Bestor, T. H. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science *311*, 395-398.

Gowher, H., Leismann, O., and Jeltsch, A. (2000). DNA of *Drosophila melanogaster* contains 5-methylcytosine. EMBO J *19*, 6918-6923.

Hermann, A., Schmitt, S., and Jeltsch, A. (2003). The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. J Biol Chem *278*, 31717-31721.

Hodges, E., Molaro, A., Dos Santos, C. O., Thekkat, P., Song, Q., Uren, P. J., Park, J., Butler, J., Rafii, S., McCombie, W. R.*, et al.* (2011). Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. Mol Cell *44*, 17-28.

Iyer, L. M., Abhiman, S., and Aravind, L. (2011). Natural history of eukaryotic DNA methylation systems. Prog Mol Biol Transl Sci *101*, 25-104.

Jeltsch, A., Nellen, W., and Lyko, F. (2006). Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. Trends Biochem Sci *31*, 306-308.

Jurkowski, T. P., and Jeltsch, A. (2011). On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. PLoS ONE *6*, e28104.

Jurkowski, T. P., Meusburger, M., Phalke, S., Helm, M., Nellen, W., Reuter, G., and Jeltsch, A. (2008). Human DNMT2 methylates tRNAAsp molecules using a DNA methyltransferase-like catalytic mechanism. RNA *14*, 1663-1670.

Krauss, V., and Reuter, G. (2011). DNA methylation in Drosophila--a critical evaluation. Prog Mol Biol Transl Sci *101*, 177-191.

Kunert, N., Marhold, J., Stanke, J., Stach, D., and Lyko, F. (2003). A Dnmt2-like protein mediates DNA methylation in Drosophila. Development *130*, 5083-5090.

Law, J. A., and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet *11*, 204-220.

Lister, R., and Ecker, J. R. (2009). Finding the fifth base: genome-wide sequencing of cytosine methylation. Genome Res *19*, 959-966.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M.*, et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315-322.

Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., and Maleszka, R. (2010). The honey bee epigenomes: differential methylation of brain DNA in queens and workers. PLoS Biol *8*, e1000506.

Lyko, F., and Maleszka, R. (2011). Insects as innovative models for functional studies of

DNA methylation. Trends Genet *27*, 127-131.

Lyko, F., Ramsahoye, B. H., and Jaenisch, R. (2000). DNA methylation in *Drosophila melanogaster*. Nature *408*, 538-540.

Mohn, F., and Schubeler, D. (2009). Genetics and epigenetics: stability and plasticity during cellular differentiation. Trends Genet *25*, 129-136.

Molaro, A., Hodges, E., Fang, F., Song, Q., McCombie, W. R., Hannon, G. J., and Smith, A. D. (2011). Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. Cell *146*, 1029-1041.

Okano, M., Xie, S., and Li, E. (1998). Dnmt2 is not required for *de novo* and maintenance methylation of viral DNA in embryonic stem cells. Nucleic Acids Res *26*, 2536-2540.

Phalke, S., Nickel, O., Walluscheck, D., Hortig, F., Onorati, M. C., and Reuter, G. (2009). Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine 5 methyltransferase DNMT2. Nat Genet *41*, 696-702.

Quail, M. A., Otto, T. D., Gu, Y., Harris, S. R., Skelly, T. F., McQuillan, J. A., Swerdlow, H. P., and Oyola, S. O. (2012). Optimal enzymes for amplifying sequencing libraries. Nat Methods *9*, 10-11.

Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci USA *97*, 5237-5242.

Schaefer, M., and Lyko, F. (2010). Solving the Dnmt2 enigma. Chromosoma *119*, 35-40.

Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M., and Lyko, F. (2010). RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. Genes Dev *24*, 1590-1595.

Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D.*, et al.* (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490-495.

Sunita, S., Tkaczuk, K. L., Purta, E., Kasprzak, J. M., Douthwaite, S., Bujnicki, J. M., and Sivaraman, J. (2008). Crystal structure of the Escherichia coli 23S rRNA:m5C

methyltransferase RlmI (YccW) reveals evolutionary links between RNA modification enzymes. J Mol Biol *383*, 652-666.

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H. R.*, et al.* (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells *11*, 805-814.

Tuorto, F., Liebers, R., Musch, T., Schaefer, M., Hofmann, S., Kellner, S., Frye, M., Helm, M., Stoecklin, G., and Lyko, F. (2012). RNA cytosine methylation by Dnmt2 and NSun2 promotes tRNA stability and protein synthesis. Nat Struct Mol Biol *19*, 900-905.

Warnecke, P. M., Stirzaker, C., Melki, J. R., Millar, D. S., Paul, C. L., and Clark, S. J. (1997). Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. Nucleic Acids Res *25*, 4422-4426.

Warnecke, P. M., Stirzaker, C., Song, J., Grunau, C., Melki, J. R., and Clark, S. J. (2002). Identification and resolution of artifacts in bisulfite sequencing. Methods *27*, 101-107.

Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinformatics *10*, 232.

Yoder, J. A., and Bestor, T. H. (1998). A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. Hum Mol Genet *7*, 279-284.

Zemach, A., McDaniel, I. E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916-919.

Zemach, A., and Zilberman, D. (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. Curr Biol *20*, R780-785.

Figure legends

Figure 1. Characterization of the *Schistosoma mansoni* methylome. (**A**) Average methylation levels were determined for all cytosine residues with a methylation ratio >0.1 and then distributed into bins with increasing methylation ratios (red bars). For comparison, the corresponding data is also shown for honeybee worker brains (grey bars), an established *Dnmt1/3*-dependent methylome with a very low DNA methylation level (Lyko et al., 2010). (**B**) Dinucleotide sequence contexts of unconverted cytosines in *Schistosoma* (red) and in honeybees (grey). (**C**) Position-specific non-conversion ratios (red) and coverages (grey) of the *Schistosoma forkhead* gene. The specific region previously reported to be methylated (Geyer et al., 2011) is indicated as a green bar. Sequence position numbers refer to GenBank accession number JF781495.

Figure 2. Characterization of the *Drosophila melanogaster* methylome. (**A**) Average methylation levels were determined for all cytosine residues and then distributed into bins with increasing methylation ratios (blue bars). For comparison, the corresponding data is also shown for human sperm DNA that was spiked (1%) into the *Drosophila* sample prior to bisulfite conversion (black bars). The actual numerical values of the first bins are 99.7% (*Drosophila*) and 92.9% (human sperm). (**B**) Dinucleotide sequence context of unconverted cytosines in *Drosophila* (blue) and in human sperm (black). (**C**) Position-specific non-conversion ratios (red) and coverages (grey) of the *Drosophila Invader4* element. Results are shown for the sequence with the lowest conversion rate among genomic *Invader4* elements. The specific region previously reported to be methylated (Phalke et al., 2009) is indicated as a green bar. Sequence position numbers refer to GenBank accession number AE014135.3. (**D**) Average methylation levels were determined for all cytosines from *Dnmt2* mutant embyos (green bars). The actual numerical value of the first bin is 99.0%. (**E**) Histograms showing the number of non-converted (ratio > 0.5) cytosine residues in 1 kb-windows in Dnmt2 mutant (green) and in wildtype (blue) *Drosophila* embryos. (**F**) Venn diagram showing overlapping

windows with >2 (left panel) or >20 (right panel) non-converted cytosine residues in wildtype (blue) and *Dnmt2* mutant (green) embryos.

Figure 3. Characterization of DNA methylation in the TKO mouse ES cell model. (A) Schematic illustration of Dnmt genotypes in wildtype and TKO mouse ES cells (**B**) Average methylation levels were determined for all cytosine residues and then distributed into bins with increasing methylation ratios (orange bars). For comparison, the corresponding data is also shown for wildtype mouse ES cells (grey bars). The actual numerical values of the first bins are 86.9% (wildtype) and 96.5% (TKO). (**C**) Fractions of non-converted (ratio >0.1) CpN dinucleotides in wildtype cells (grey bars), TKO cells (orange bars) and an unmethylated PCR fragment (black bars).

Table 1. Sequencing data.

organism	strain	no. of reads (pairs)	mapping efficiency	coverage	conversion
S. mansoni	Puerto Rican	254,347,320	73%	13.4x	99.02%
D. melanogaster	w^{1118}, E0-2	174,821,591	57%	32.0x	99.75%
D. melanogaster	$Dnmt2^{149}$, E0-2	129,094,743	52%	23.6x	99.42%
M. musculus	TKO ES cells	43,583,131	74%	0.8x	98.26%

E0-2 indicates 0-2 h old *Drosophila* embryos. Sequencing coverages are indicated per strand. For *Schistosoma* and *Drosophila*, the conversion rates were calculated as the average conversion ratio of all cytosine residues that were covered by the data set. The conversion rate of the mouse TKO ES cell sample was calculated as the average conversion ratio of all non-CpG dinucleotides that were covered by the data set.

Figure 1
Click here to download high resolution image

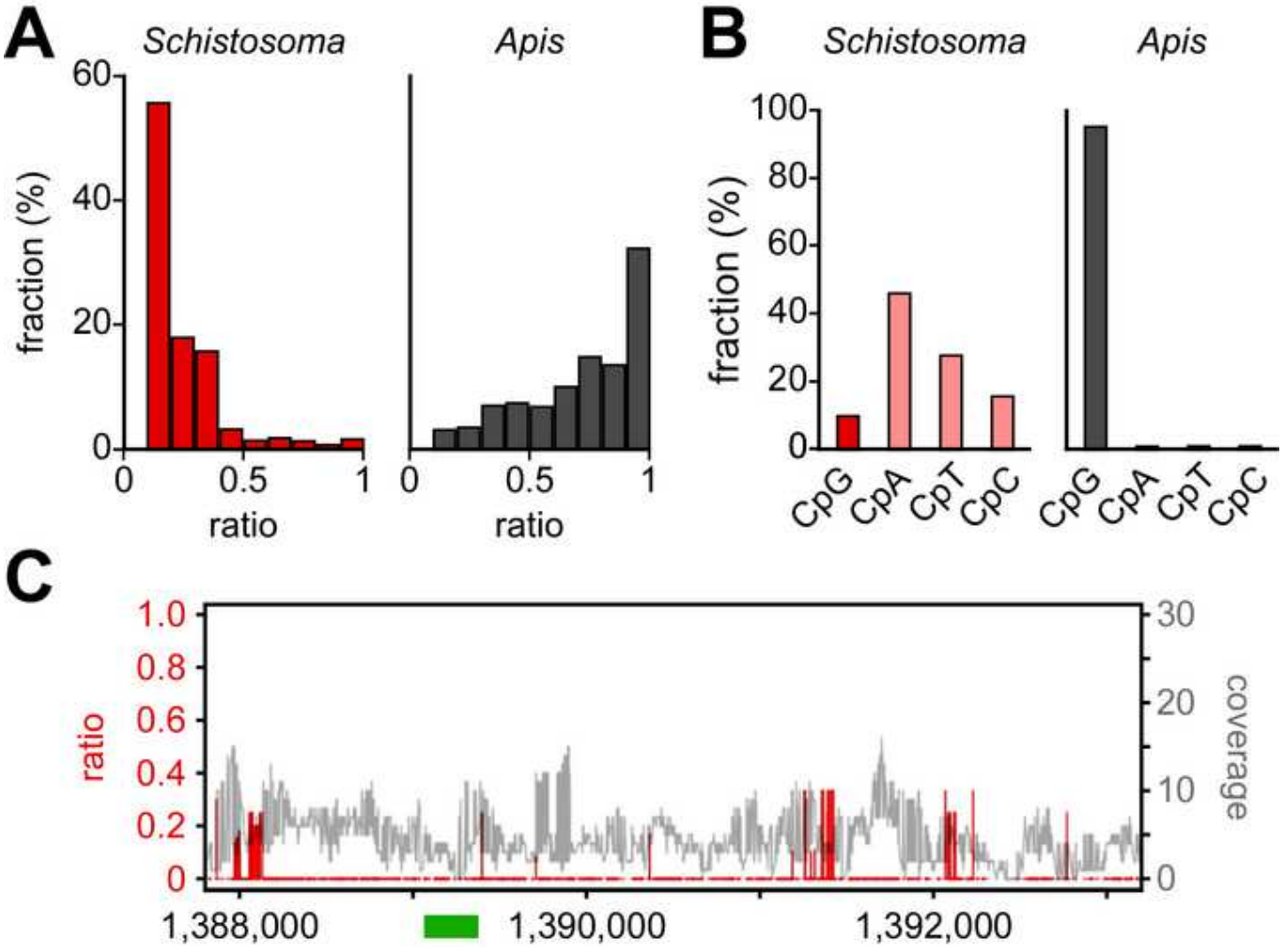


Figure 2
Click here to download high resolution image

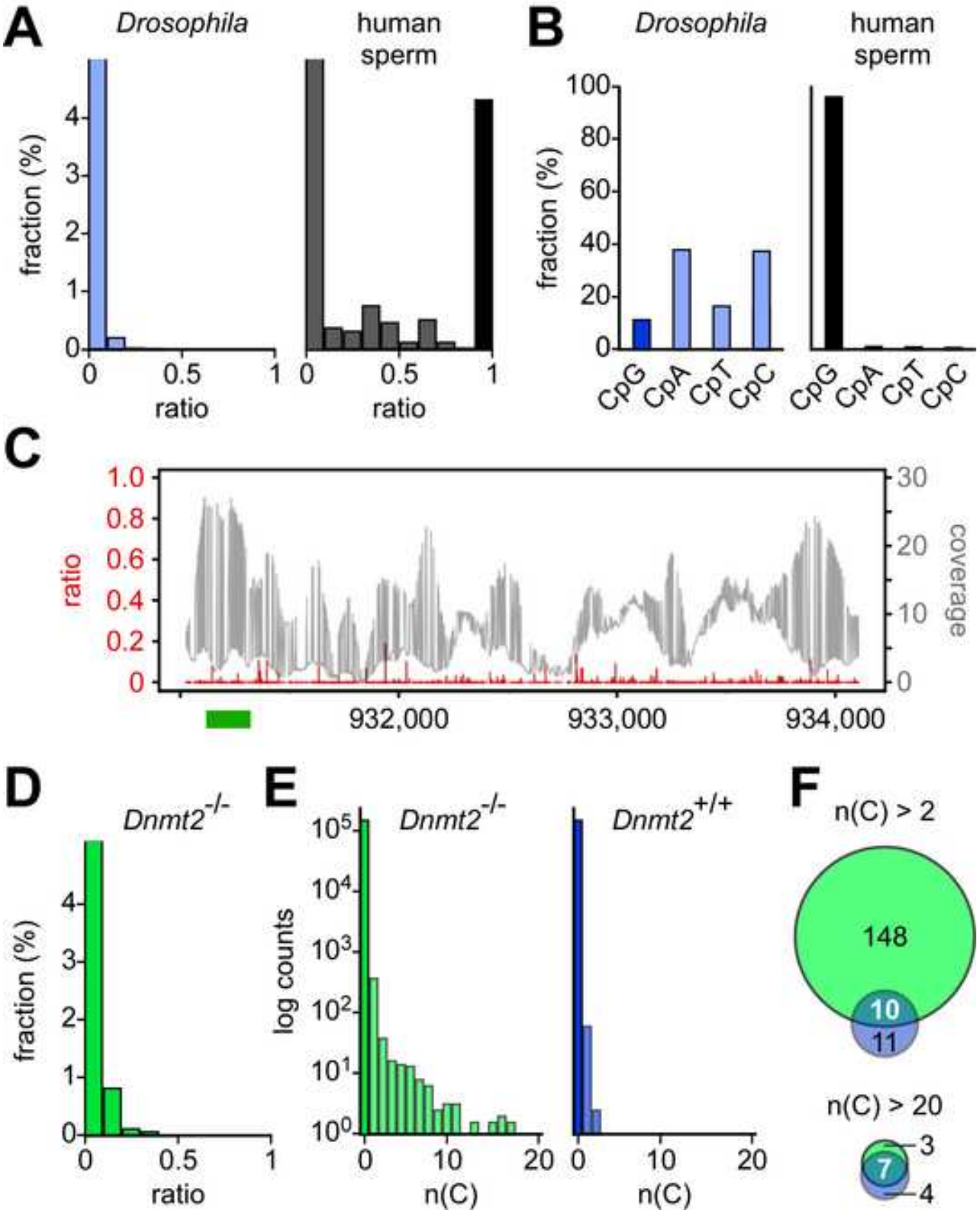
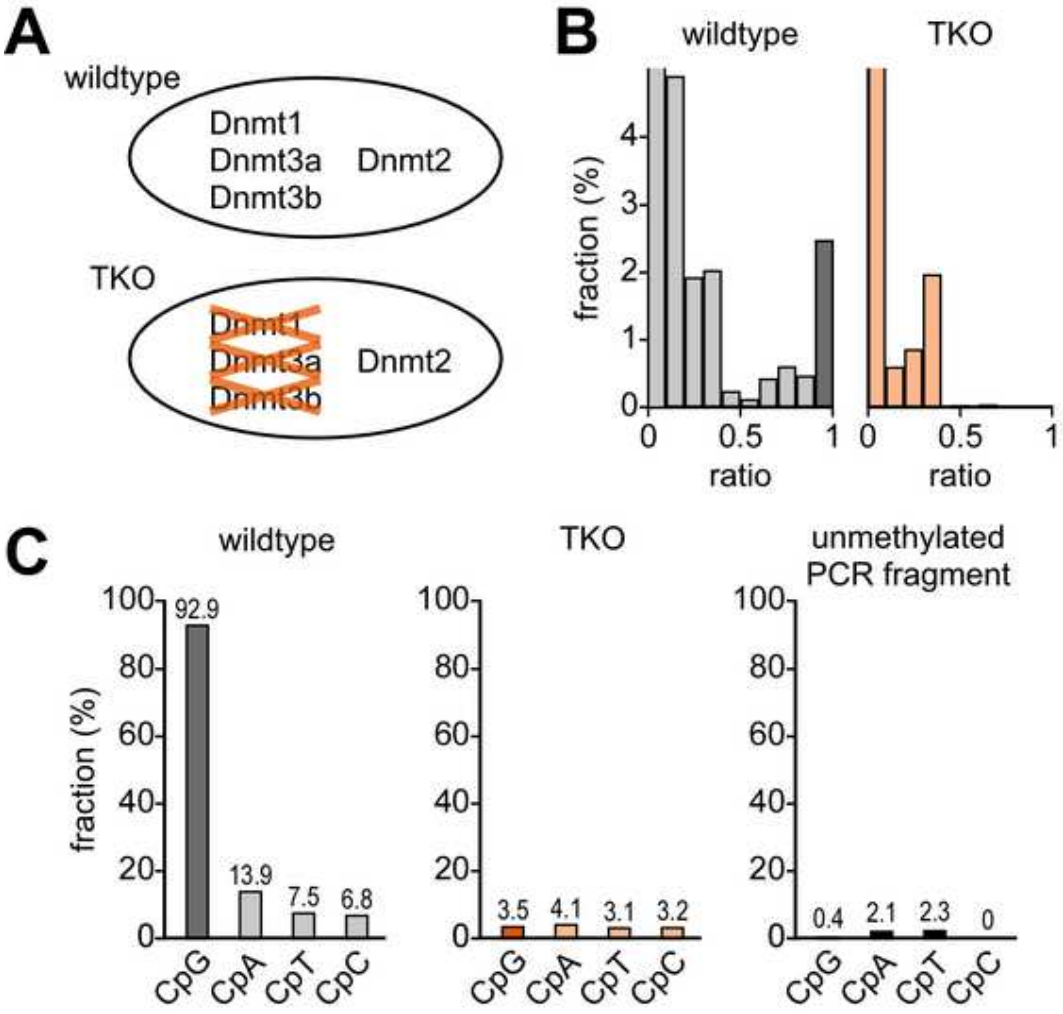


Figure 3
Click here to download high resolution image



Supplemental Inventory

Table S1. Comparison of *Drosophila* datasets, related to Figure 2. Compares the published *Drosophila* dataset (Zemach et al., 2010) to our newly generated dataset. As requested by reviewer #1.

Table S2. Bisulfite conversion rates of internal controls, related to Table 1. Allows the determination of false discovery rates and the statistical analysis of bisulfite conversion. As requested by reviewer #1.

Table S1. Comparison of *Drosophila* datasets, related to Figure 2.

dataset	strain	no. of reads	mapping efficiency	coverage	conversion
Zemach et al., 2010	n.s., E0-3	22,010,452 (SE)	96%	3.5x	99.62%
this study	w^{1118}, E0-2	174,821,591 (PE)	57%	32.0x	99.75%

n.s.: not specified. E0-3 indicates 0-3 h old embryos, E0-2 indicates 0-2 h old embryos. SE denotes single-end reads, PE denotes paired-end reads. Sequencing coverages are indicated per strand. The conversion rates were calculated as the average conversion ratio of all cytosine residues that were covered by the data set.

Table S2. Bisulfite conversion rates of internal controls, related to Table 1.

spike-in	source dataset	no. of reads (pairs)	conversion
2 kb PCR fragment	mouse TKO cell line	1,593	98.38%
human sperm DNA	*Drosophila* w^{1118}, E0-2	1,637,070	99.71%

The conversion rate of the human sperm DNA sample was calculated as the average conversion ratio of all non-CpG dinucleotides that were covered by the data set. The conversion rate of the unmethylated PCR fragment was calculated as the average conversion ratio of all cytosine residues that were covered by the data set.