bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Dual-layer transposon repression in heads of Drosophila *melanogaster*.

Marius van den Beek[1,3], Bruno da Silva[1], Juliette Pouch[4], Mohammed el amine Ali[4] Chaouche, Clément Carré[1*] and Christophe Antoniewski[1,2,*]

[1]*Drosophila Genetics and Epigenetics* ; Sorbonne Université, CNRS, Biologie du développement - Institut de Biologie Paris Seine, 75005 Paris, France. [2]ARTbio Bioinformatics Analysis Facility, Sorbonne Université, CNRS, Institut de Biologie Paris Seine (IBPS), Paris, France. [3]present address: Institut Curie, PSL Research University, CNRS UMR 3215, INSERM U934, Stem Cells and Tissue Homeostasis, Paris, France. [4]Genomic facility, Institut de biologie de l'Ecole normale supérieure (IBENS), Ecole normale supérieure, CNRS, INSERM, PSL Université Paris 75005 Paris, France. *corresponding authors: christophe.antoniewski@upmc.fr; Phone: +33 1 44 27 70 05 & clement.carre@upmc.fr; Phone: +33 1 44 27 34 39.

Abstract

piRNA-mediated repression of transposable elements (TE) in the germline limits the accumulation of heritable mutations caused by their transposition in the genome. It is not clear whether the piRNA pathway plays a functional role in adult, non-gonadal tissues in *Drosophila melanogaster.* To address this question, we first analyzed the small RNA content of adult *Drosophila melanogaster* heads. We found that varying amount of piRNA-sized, ping-pong positive molecules in heads correlates with contamination by gonadal tissue during RNA extraction, suggesting that most of piRNAs detected in head sequencing library originate from gonads. We next sequenced the heads of wild type and *piwi* mutants to address whether *piwi* loss of function would affect the low amount of piRNA-sized, ping-pong negative molecules that are still detected in heads hand-checked to avoid gonadal contamination. We find that loss of *piwi* does not affect significantly these 24-28 RNA molecules. Instead, we observe increased siRNA levels against the majority of *Drosophila* transposable element families. To determine the effect of this siRNA level change on transposon expression, we sequenced the transcriptome of wild type, *piwi*, *dicer-2* and *piwi, dicer-2* double-mutant fly heads. We find that RNA expression levels of the majority of TE families in *piwi* and *dicer-2* mutants remain unchanged and that TE transcript abundance increases significantly only in *piwi, dicer-2* double-mutants. These results lead us to suggest a dual-layer model for TE repression in adult somatic tissues. Piwi-mediated transcriptional gene silencing (TGS) established during embryogenesis constitutes the first layer of TE repression whereas Dicer-2-dependent siRNA-mediated post-transcriptional gene silencing (PTGS) provide a backup mechanism to repress TEs that escape silencing by *piwi*-mediated TGS.

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Introduction

Transposable element (TEs) activity is thought to be an important force in genome evolution, as TE integration and excision can result in mutations that impact gene regulation networks (Tubio et al. 2014). However, these mutations may be detrimental to individuals, potentially decreasing lifespan and fertility (Wood et al. 2016). Therefore, limited TE mobilization is beneficial to both host and TE, whereas high TE activity decreases host fitness and adversely affects vertical transfer of the TE.

In *Drosophila melanogaster*, the siRNA (small-interfering RNA) and the piRNA (piwi-interacting RNA) pathways are important negative regulators of TE expression in somatic (Li et al. 2013; Ghildiyal et al. 2008a) and gonadal tissues (Vagin et al. 2006), respectively. Both pathways are active in the gonads, while the siRNA pathway is thought to be active in all somatic tissues. Dicer-2, the central endonuclease of the siRNA pathway operates on double-stranded RNA (dsRNA) molecules by processively translocating along the molecule and cutting every 21st nucleotide (Cenik et al. 2011; Welker et al. 2011). After the initial processing, these 21nt duplexes with 3'OH overhangs of 2 nt are rebound by Dicer-2 and r2d2 to be loaded into Argonaute-2 (Ago2). Ago2 can then engage in multiple rounds of endonucleolytic cleavage of transcripts with mature siRNA complementarity. Loss of functional siRNA pathway results in increased levels of TE expression and mobilization (Li et al. 2013; Xie, Donohue, and Birchler 2013; Czech et al. 2008; Ghildiyal et al. 2008a)–and compromised male fertility as well as defective sperm development (Wen et al. 2015).

The *Drosophila* piRNA pathway is well characterized for its role in maintaining germline stem cells (Cox et al. 1998; Jin, Flynt, and Lai 2013), genome integrity in the gonads and early embryo (Wang and Elgin 2011) and is therefore required for fertility. Its key components are 3 germline Argonaute-family proteins Piwi, Aubergine (Aub) and Argonaute-3 (Ago3). piRNA production relies on the processing of piRNA cluster transcripts, which are composed of TE fragments. The processing of these primary transcripts by the piRNA biogenesis machinery produces primary antisense piRNAs which are loaded into Piwi or Aub in a germline cells cytoplasmic structure called the *nuage* (Brennecke et al. 2007). Primary piRNA-loaded in Aubergine slices complementary sense TE transcripts between the 10th and 11th position, which will become the 5' end of a new TE sense piRNA. Sense piRNAs are then loaded onto

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Ago3, which in turn slices complementary antisense piRNA cluster transcripts between the 10th and 11th position. This cytoplasmic cyclic process referred to secondary piRNA amplification leaves a detectable "ping-pong" signature, in which 5' and 3' ends of piRNAs tend to overlap by 10 nucleotides (Brennecke et al. 2007; Gunawardane et al. 2007).

piRNA-loaded Piwi proteins can re-enter the nucleus, where piRNAs guide Piwi towards complementary nascent transcripts. Piwi then recruits factors (Maelstrom, SuVar3-9, dSETDB1, HP1 and silencio/Panoramix) that establish and maintain H3K9me3 at the surrounding genomic vicinity of TE insertion sites, possibly including protein coding genes and hence functions in TGS (Sienski, Dönertas, and Brennecke 2012; Brower-Toland et al. 2007; Wang and Elgin 2011; Rangan et al. 2011; Sienski et al. 2015).

Zygotic piwi expression has been detected ubiquitously in early embryos up to the 14th nuclear division (~2h after egg laying) (Mani, Megosh, and Lin 2014; Rouget et al. 2010), depletion of Piwi in nurse cells and oocytes results in early arrest of embryonic development (Mani, Megosh, and Lin 2014; Wang and Elgin 2011), and *piwi* acts as a suppressor of variegation in the eye (Gu and Elgin 2013; Pal-Bhadra et al. 2004), suggesting an important function for Piwi in germline cells maintenance and during early development of somatic tissues. In contrast the function of Piwi in larval or adult somatic tissue remains unclear. Piwi has both been reported to be present (Brower-Toland et al. 2007) or absent (Le Thomas et al. 2013) in 3rd instar larval salivary glands whereas Aubergine and Ago3 have been observed in non-overlapping cells of the adult central nervous system (Perrat et al. 2013).

In order to unravel the role of piRNA pathway in TE control of somatic adult tissues, we analyzed small RNA profile in wild type, *piwi* and *dicer-2* mutant heads fly. We provide evidence that previously reported ping-pong pairs in adult heads likely result from contamination with testicular RNA, suggesting that secondary piRNA amplification does not take place in adult heads. However, small RNA sequencing of *piwi* mutant heads reveal an increased levels of siRNAs against most TE families. RNA-sequencing of *piwi* single mutant heads and *dicer-2* single mutant heads showed only minor upregulation of transposable elements, whereas double-mutants of *piwi* and *dicer-2* showed increased TE levels. Our results suggest a dual-layer model of TE repression in somatic tissues. The first layer of TE repression is established by Piwi at the chromatin level during early development as shown in (Gu

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and Elgin 2013). When TEs escape the epigenetic Piwi silencing, PTGS triggered by dicer-2 and siRNAs mediates TE degradation to decrease TE burden.

Results

Secondary piRNA biogenesis is not detectable in adult heads

Ago3 and Aubergine were detected by immunofluorescence in the optic lobe of the *Drosophila* central nervous system (Perrat et al. 2013) and several teams sequenced small RNA molecules in head samples which had the size (24-28 nt), 3'-end 2'-O-methylation and ping-pong signature of *Drosophila* piRNAs (Ghildiyal et al. 2008b; Yan et al. 2011a; Mirkovic-Hösle and Förstemann 2014b). However, these putative piRNAs corresponded to a very minor fraction of the sequence datasets, which raises the possibility of sample contamination by gonadal tissues during RNA extraction. To address this issue, we analyzed small RNA matching TE sequences in 24 small RNA sequencing libraries prepared from adult male heads isolated by sieving after freezing (Reinhardt et al. 2012). The majority of reads were in the size range of siRNAs (21 nt). However, we observed a fraction of reads in each library within the 24-28 nt, size range of *Drosophila* piRNAs (mean 7.5%, interquartile range 5.6-10.4%; Supplemental Fig 1A, 1B).

We searched for ping-pong partners among 24-28 nt reads for each of the 24 small RNA sequencing libraries, as had been done in previous works (Ghildiyal et al. 2008b; Yan et al. 2011a; Mirkovic-Hösle and Förstemann 2014b). We detected a significant ping-pong signature (>20 pairs, z-score of the 10 nt overlap >= 2) in 5 out of the 24 analyzed libraries (hereafter referred to as *ping-pong positive* libraries, Supplemental Fig 1C, red asterisks), while 7 libraries (hereafter referred to as *ping-pong negative* libraries, Supplemental Fig 1C, black square) had a negative 10 nt overlap z-score.

The genotype of any of the analyzed libraries was not expected to differentially affect piRNA biogenesis, suggesting that the significant signatures we found might be due to stochastic contamination by gonadal material during the RNA extraction from heads prepared from adult male isolated by sieving after freezing (Reinhardt et al. 2012). To investigate this possibility, we performed differential expression testing of miRNAs between the 5 ping-pong positive libraries and the 7 ping-pong negative libraries, under the assumption that RNA introduced from contaminating tissues would include tissue-specific miRNAs that are not normally expressed in heads. Under our assumption, those putative tissue-specific miRNAs would be absent in

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ping-pong negative heads libraries (not contaminated) and present in ping-pong positive heads libraries (contaminated).

A set of 27 miRNAs was significantly enriched in ping-pong positive libraries relative to ping-pong negative libraries at an adjusted P-value (Benjamini-Hochberg) of 0.01 (Supplemental Table S1). Since we analyzed male heads, the contaminating RNA giving rise to ping-pong signature might stem from the testis. To investigate this possibility we randomly added reads from testicular small RNA libraries (Toledano et al. 2012; Rozhkov et al. 2010) to ping-pong negative head libraries whose sizes were normalized by sampling to $2.5*10^6$ reads for comparison. We find that adding ~2% testes reads to mimic the contamination in the ping-pong negative head libraries is sufficient to detect a ping-pong signature (>20 pairs, z-score for 10 nt overlap >= 2; Supplemental Fig 1D).

In addition, we found a strong overlap of miRNAs differentially expressed between ping-pong positive and ping-pong negative libraries on the one hand, and miRNAs differentially expressed between ping-pong negative libraries with or without simulated testis contamination on the other hand: the 10 most significantly changed miRNAs between ping-pong negative libraries with or without 2% of testicular reads are also significantly changed between ping-pong positive and ping-ping negative libraries (9 miRNAs with p<0.01, 1 miRNA with p<0.025; Supplemental Table S1 and Supplemental Figure 1E). Altogether, these data strongly suggest that when detected in the head, the ping-pong signature is mostly due to contamination by gonadal abundant small RNAs during sample preparation, in particular if a sieving strategy after freezing is used to isolate the heads and not followed by eye check. However, we cannot exclude that Aubergine and Ago3 are producing secondary piRNAs in very low amounts or in a small set of head cells.

Ping pong negative piRNA-like small RNAs in adult heads are not *piwi*-dependent

In agreement with our previous observations, new libraries prepared from wild type (+/+), heterozygous (*piwi* +/-) or homozygous (*piwi* -/-) *piwi*, and heterozygous (*dicer-2* +/-) or homozygous (*dicer-2* -/-) *dicer-2* heads carefully checked to avoid contamination by gonadal debris were ping-pong negative (Supplemental Figure 1F). However, we still detected a small amount of piRNA-sized reads (24-28nt) in these library (Figure 1A insets). Since Piwi is sufficient for piRNA biogenesis without ping-pong signature in ovarian follicle cells (Lau et al. 2009; Robine et al. 2009; Saito

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

et al. 2009), we further explored the possibility that these 24-28nt piRNA-sized reads are Piwi-dependent primary piRNAs.

As the main function of Piwi has been proposed to induce transcriptional repression of TEs in ovaries (Sienski, Dönertas, and Brennecke 2012; Rozhkov, Hammell, and Hannon 2013), we focused on small RNA reads aligning to TEs.

We observed that for wild type, *piwi* heterozygous and *piwi* homozygous mutant heads, the 24-28 nucleotide reads show a slight bias for aligning to the sense strand of TEs (Figure 1A, inset), which is further amplified in *dicer-2* mutant flies. This is in contrast to piRNAs molecules in gonads that predominantly align to the antisense strand of TEs (Brennecke et al. 2007). Importantly, the fraction of 24-28 nucleotide antisense to TEs seemed to increase slightly in *piwi* homozygous mutant heads. Together, these results suggest that the small fraction of 24-28nt RNAs in ping-pong negative head libraries are not *piwi*-dependent RNAs (Figure 1B).

Piwi mutation unmasks a TE-specific siRNA response in adult somatic tissues

We next examined whether loss of *piwi* would affect siRNAs that align TEs sequences in the head. In both wild type and heterozygous *piwi* mutant heads we detect a substantial amount of 21-nt sense and antisense reads, which increases in homozygous *piwi* mutant (Figure 1A, 1E). Importantly, these reads are strongly reduced in *dicer-2* mutants, confirming that TE-aligned reads are siRNAs (Figure 1A, 1E). We consistently observed the increase of 21 nucleotide sense and antisense RNAs in *piwi* homozygous mutant heads when considering only TE families that had on average more than 20 aligned reads per 10 million matched reads ("cp10m"). To quantify siRNA variations we only considered the fraction of 21 nucleotide reads that aligns to the TE complementary strand, in order to minimize quantification of partially degraded TE transcripts, which would be expected to align to the sense strand with little size specificity (discussed in Malone et al. 2009). We further restricted analysis to TE families that had on average five or more 21-nt antisense reads per library after applying a size factor to correct for sequencing depth. By doing so, we determined that in *piwi* mutants, 27 out of 62 transposable element families show >2 fold increases of siRNAs, with an overall median fold change of 1.69 compared to a 0.92 fold change for piwi heterozygous mutant heads (Figure 1C, P-value 1.03e-05, Mann-Whitney U). While most TE siRNAs were hence moderately increased, gypsy siRNA expression was increased more than 45 fold.

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We also investigated if the observed changes would occur at earlier time points during development. We observed a 2.8 fold increase of gypsy-specific siRNA in *piwi*$^{-/-}$ mutant first instar larvae (L1) as compared to *piwi*$^{+/+}$ L1 , however the majority of TE family siRNAs remained unchanged at this stage (p-value 0.83, Mann-Whitney U, Figure 1D).

Together, these results suggested that increased TE siRNA levels upon Piwi loss is a response that is strongest at the adult stage. In addition to the previously reported suppression of variegation of wm^4 in adult eyes and mild decrease of HP1 occupancy at TEs loci in 3rd instar larvae (Gu and Elgin 2013), loss of Piwi at early stage of development also induces an increased production of TE-specific siRNAs later on in adult somatic tissues.

The siRNA response likely originates from Dicer-2-mediated TE transcript processing in adult heads

We next investigated the origin of the increase in TE derived siRNAs in *piwi* mutant heads. SiRNA response might be a direct consequence of increased TE transcription in the absence of Piwi, which in turn would result in increased processing of TE-derived dsRNAs by Dicer-2 into siRNAs. Alternatively, the elevated levels of TE-siRNAs might originate from piRNA clusters, which produce both piRNAs and siRNAs in the germline. If this were to be the case we should be able to detect an increase of specific 21-nucleotide RNAs originating from these piRNA clusters.

As most germinal piRNA clusters are transcribed bidirectionally we quantified both 21 nt sense and antisense RNAs that map exclusively to piRNA clusters (defined in Brennecke et al. 2007). We thereby excluded sequences shared with TE insertions elsewhere in the genome, allowing us to separate production of siRNAs originating from clusters and those originating from TE insertions. We observed a low quantity of 21nt reads derived from piRNA cluster in both *piwi* mutant and control wild type heads (between 2% and 3,5% of total TE reads, Supplemental Figure 2A). These cluster-derived 21 nt RNAs increased slightly in *piwi* mutant heads, although to a lesser extent that the increase of 21 nt antisense TE reads (1.44 fold vs 2.15 fold, Supplemental Figure 2B). The increase of TE-specific siRNAs in *piwi* mutant heads was thus unlikely to be caused by an increase in Dicer-2-mediated processing of piRNA cluster transcripts, favoring the hypothesis in which increased transcription of euchromatic TE insertions leads to an increase in TE-specific siRNA production through Dicer-2 activity. In principle, the increase of TE-specific siRNAs could be

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

maternally inherited or stably maintained from early development. We took advantage of the fact that mature Ago2-loaded siRNAs are single-stranded and that double-stranded RNAs, among which the substrates of Dicer-2, are frequently deaminated through the action of Adenosine Deaminase Acting on RNA (ADAR) enzyme, which converts adenosine (A) to inosine (I) (Keegan et al. 2005; Palladino et al. 2000; Wu, Lamm, and Fire 2011). This change manifests in a A to G mismatches in RNA sequencing datasets as compared to the DNA based reference genome. We therefore determined the frequency of all nucleotide mismatches for all FlyBase-listed TE insertions in our small RNA sequencing datasets (Figure 1G). The amount of A to G mismatches is not elevated over other mismatches in early embryos (1.5 % of all 21 nucleotide reads matching to TE insertions). We detect a similar frequency of A to G mismatches (1.2 to 1.4%) in first instar larvae, however no other mismatches were preferred, suggesting that ADAR might be active at low level in 1st instar larvae. In contrast, we detected a higher frequency of A to G mismatches in wild type heads (3.6 to 4.1%) that, despite a >2-fold increase of 21 nt RNA, further increases in *piwi* mutants (5.7%). Since this increase would not be observed if siRNA were inherited from earlier developmental stages, the data suggest that siRNA are actively produced by Dicer-2 from double-stranded TE RNA substrates in *piwi* mutant adult heads.

Neither loss of Piwi nor Dicer-2 leads to strong upregulation of TEs

To determine whether the observed increase of siRNA production efficiently counteracts any increased TEs transcription caused by a lack of Piwi-mediated TGS at earlier stage, we sequenced the head transcriptome of *piwi* mutants, *dicer-2* mutants and *piwi, dicer-2* double-mutants and compared these to wild type head transcriptome. In *piwi* mutants, transcript levels of most TEs remain unchanged in those heads, with the notable exception of *gypsy,* whose level increases about 5-fold (Figure 2A). *Gypsy* is also the TE against which we observed the strongest increase of siRNA levels (Figure 1C), suggesting that the transcription of TEs is indeed increased in *piwi* mutant heads and correlates with Dicer-2 dependant siRNA production through the dicing of double-stranded TE RNAs.

Contrary to sense transposon transcripts, we did detect a significant increase in antisense transcripts in *piwi* mutant heads (Figure 2B, +/+ vs Piwi -/- blue boxes), that might form transient duplexes with sense TE transcripts, serving as a substrate for Dicer-2-mediated siRNA processing. Similarly, most TEs are not upregulated in

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

dicer-2 mutants, except for *297*, which produces a significant amount of siRNA in wild type heads, perhaps indicating inefficient Piwi-mediated TGS for this TE family.

We conclude that Piwi and Dicer-2 are redundant for the maintenance of TEs repression for most TE families in adult heads and that both Piwi-mediated TGS and Dicer-2-mediated PTGS can efficiently repress TEs.

Piwi and Dicer-2 compensatory mechanism revealed in double-mutant

To confirm our hypothesis that the siRNA response in adult somatic tissues compensates for the loss of Piwi earlier in development, we analyzed RNA libraries from *piwi, dicer-2* double-mutant heads. We found that the majority of TE families is significantly upregulated in *piwi, dicer-2* double-mutant heads compared to single *piwi* or *dicer-2* mutants (Figure 2 A, B, C).

To illustrate the impact of the increased siRNA levels in *piwi* mutants on TE transcripts, we plotted for each TE family the log2 of the fold change of siRNAs in *piwi* mutants on the X-axis, and the fold change in TE transcript level on the y-axis. We see a tendency for TE transcripts that are targeted by more siRNAs in *piwi* mutants to decrease in abundance when comparing *piwi* mutant to the wildtype (Figure 2E, top panel). If we perform the same analysis but focus on the change of TE levels between *piwi* mutant and *piwi*, *dicer-2* double mutants (Figure 2E, bottom panel), we see a tendency for TEs transcripts to increase in abundance in the double mutant. This further suggests that increased siRNA production in *piwi* mutants efficiently counteracts loss of the transcriptional repression (TGS) established by *piwi* earlier during embryogenesis. If this compensation is failing in *piwi*, *dicer-2* double mutants TE transcript levels increase globally.

Altogether our results suggests a dual layer expression control of Piwi-mediated TGS and Dicer-2-mediated PTGS to firmly repress TEs levels in the adult soma.

Piwi and Dicer2 do not restrict expression of genes containing TEs in their genomic loci

Piwi has previously been shown to silence genes adjacent to TEs insertions and genes that carry TEs sequences in their boundaries (Sienski, Dönertas, and Brennecke 2012). Our transcriptomic data does not support a significant trend of upregulation of genes that contain TE insertions in their genomic boundaries, whether we consider all TE families, or only those that are upregulated in *piwi*, *dicer-2* or *piwi*, *dicer-2* double-mutants (Figure 3A). Piwi has also been shown to repress genes that produce *traffic jam* class piRNAs from their 3'UTR (Robine et al.

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

2009). Again, we detect no expression bias for these genes in any of our heads mutant conditions (Figure 3C).

Discussion

Origin of piRNA-like molecules in somatic tissues

Through the analysis of a large number of small RNA sequencing libraries we revisited previous reports of putative secondary piRNA biogenesis in adult heads (Yan et al. 2011b; Ghildiyal et al. 2008a; Mirkovic-Hösle and Förstemann 2014a). Based on our analysis, it is likely that earlier observations resulted from small contamination with gonadal RNA, which can easily occur when collecting heads by vortexing frozen flies and filtering fly-parts by sieving. PiRNA-like molecules in adult *Drosophila* heads were previously observed in libraries that were beta-eliminated before sequencing, and/or that originated from siRNA pathway gene mutants (Ghildiyal et al. 2008; Mirkovic-Hösle and Förstemann 2014). Beta-elimination prevents sequencing of small RNAs with unprotected 3'OH groups thereby increasing the apparent sequencing frequency of 2'O-methylated contaminating piRNAs from gonadal origin. Likewise, decrease of siRNA sequencing in siRNA pathway mutant heads is expected to amplify the sequencing frequency of piRNAs, including contaminants from gonadal origin.

Our results highlight the importance of controlling the purity of RNA preparations. Considering the enormous increase of TE expression in piRNA mutant ovaries, often reaching more than 1000 fold upregulation, best practice for analysis should be a control that clearly shows the degree of tissue purity of the RNA preparation. We show that careful examination and hand-selection of fly heads eliminates most of artifactual detection of piRNA ping-pong patterns in head RNA preparations.

When contamination is limited, the marginal fraction of remaining 24-28 nt RNAs is not sensitive to Piwi loss. This implies that either these piRNA-like molecules are primary piRNAs produced independently of Piwi (Nagao et al. 2010) or that they are degradation products. Together with their sense bias, the absence of a marked peak in the piRNA-size range (Figure 1A) and the low transcriptional activity of piRNA clusters in heads, these observation strongly suggest that these 24-28 nt RNAs are not piRNAs.

Piwi likely exerts its function in adult heads through inherited transcriptional repression set in early development

Zygotic p*iwi* mutations are suppressors of variegation, revealing a link between the piRNA pathway and heterochromatin silencing (Pal-Bhadra et al. 2004). In contrast,

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

late inactivation of *piwi* in eye imaginal discs does not suppresses variegation (Gu and Elgin 2013), suggesting that Piwi silencing established during early embryogenesis is maintained in the absence of Piwi product. As we were unable to evidence a role of piRNAs in TE silencing in adult heads and found that zygotic loss of *piwi* alone has little effect on steady state RNA levels of most of TE families in heads, our work is in line with this model.

Repression of TEs in the absence of siRNAs

Dicer-2 mutations caused a strong reduction of TE-specific siRNAs. However, this did not result in major changes of TEs RNA expression for most TE families in adult heads, with the notable exception of *297*. This observation is in line with the results of Ghildiyal *et al* (Ghildiyal et al. 2008b) who found *297* expression to be strongly increased in heads of *dicer-2* mutants, and Xie et al., (Xie, Donohue, and Birchler 2013) who demonstrated increased somatic transposition of 297 in *dicer-2* mutants. How repression of specific TEs escapes from Piwi and mainly relies on siRNA pathways requires further investigation.

A dual-layer TEs repression by small RNAs

We show that in adult heads, the loss of only *piwi* or *Dicer-2* does not lead to significant change in expression of most of TE families. However, loss of piwi is associated to increased levels of antisense TE transcripts and of siRNAs, whereas double *piwi Dicer-2* mutants show and increased level of transcripts from a large panel of TE families. We propose that these observations reflect a dual-layer repression mechanism whereby residual expression of TE that results from incomplete piwi-mediated TGS feeds Dicer-2 for siRNA production that in turn reduces TE transcript levels through PTGS. This mechanism might be especially relevant for TEs that can escape Piwi-mediated repression, such as 297, or TEs that do not induce yet Piwi-mediated TGS because they are not integrated into piRNA clusters. It could also play a role in aging since an increase in TE expression and transposition was observed in aging adult flies, which was amplified in *Dicer-2* mutants and mitigated by *Dicer-2* overexpression (Wood et al. 2016).

In evolutionary terms, the “failsafe”, dual-layer repression mechanism could help during the early steps of a TE invasion of a fly population. Thus, siRNA-mediated PTGS would maintains a tolerable load of the TE transcripts, until the TE has integrated into a piRNA cluster and triggered a stable Piwi-mediated TGS.

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Material and Methods

Fly stocks

Flies were grown on standard *Drosophila* food at 25°C. All flies were brought into the w^{m4} background (Muller 1930). dicer-2^{R416X} and dicer-$2^{L811fsx}$ alleles were described in (Lee et al. 2004). piwi^2 and piwi^3 alleles were published in (Cox et al. 1998). Double-mutants were generated by crossing virgin female dcr2^{R416X}/CyO-GFP to male piwi^3/CyO-GFP flies. Offspring virgin dcr2^{R416X}/piwi^3 flies were then crossed to male wm4;Ln^{2R} Gla, wgGla^1, Bc^1/ CyO-GFP to establish wm4;piwi^2, dicer-2^{R416X}/CyO-GFP stocks. Stocks were then screened by PCR for the presence of the *piwi* mutation. The same procedure was applied to generate wm4;piwi^3, dicer-$2^{L811Fsx}$/CyO-GFP stocks. The table below provides the detailed genotype of all mutant combinations used.

Indicated name	Genotype	Maternal genotype	Paternal genotype
+/+	wm4;+/+	wm4;+/+	wm4;+/+
dicer-2 G31R/-	wm4;dcr2^{G31R}/dcr$2^{L811Fsx}$	wm4;dcr$2^{L811Fsx}$/CyO-GFP	wm4;dcr2^{G31R} /CyO-GFP
dicer-2 +/-	wm4;dcr$2^{L811Fsx}$/+	wm4;dcr$2^{L811Fsx}$/CyO-GFP	wm4;+/+
dicer-2 -/-	wm4;dcr$2^{L811Fsx}$/dcr2^{R416X}	wm4;dcr$2^{L811Fsx}$/CyO-GFP	wm4;dcr2^{R416X}/CyO-GFP
piwi +/-	wm4;piwi^2/+	wm4;piwi^2/CyO-GFP	wm4;+/+
piwi -/-	wm4;piwi^2/piwi^3	wm4;piwi^2/CyO-GFP	wm4;piwi^3/CyO-GFP
dicer-2-/-, piwi -/-	wm4;piwi^2, dicer-2^{R416X}/piwi^3, dicer-$2^{L811Fsx}$	wm4;piwi^2, dicer-2^{R416X}/CyO-GFP	wm4;piwi^3, dicer-$2^{L811Fsx}$/CyO-GFP

RNA extraction and sequencing

Head RNAs were prepared as followed. One- to two-day old flies were CO2 anesthetized, sorted by sex and genotype, transferred into 15 ml Falcon tubes and frozen in liquid nitrogen. The procedure was repeated multiple days until pools of 50 to 100 flies were obtained per biological replicate. Heads were separated from bodies by vortexing, followed by sieving and a careful selection of heads from thoraces and legs on a cooled metal plate. Heads were collected into 2 ml Precellys tubes for hard tissues and covered by 1ml Trizol. Heads were homogenized in two rounds of 5000 rpm for 30 seconds using a Precellys24 Tissue Homogenizer. Homogenate was centrifuged for 30 seconds at 13000 rpm and supernatant

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

transferred into a new 2 ml tube, 200µl of Chloroform was added and tubes were thoroughly vortexed. Further purification was as in (Rio et al. 2010). Remaining DNA was removed using Fermentas DNase I, RNase-free following the manufacturers instructions.

Small RNA library preparation and sequencing was performed on an Illumina HiSeq 2500 at Fasteris Life Sciences SA (Plan-les-Ouates, Switzerland) using the *Drosophila* small RNA track based on the Illumina TruSeq protocol.

RNA-seq was performed in biological triplicates for +/+, *piwi* -/+, *piwi* -/-, *dicer-2* -/+ and *dicer-2* -/-, with one replicate per condition sequenced in paired-end mode (2 * 101) and two replicates sequenced in single-read mode (1*51). *piwi* -/-, *dicer-2* -/- samples were sequenced in biological duplicates in single-read mode. Total RNA was depleted of rRNA using Ribo-Zero™ Gold Kit (Epicentre). Directional RNA-seq library preparation and sequencing was performed at the Genomic Paris Centre (Paris, France) using the Epicentre ScriptSeq™ v2 RNA-Seq Library Preparation Kit on an Illlumina HiSeq 2000 instrument.

All sequence files generated in this study are available from the EBI European Nucleotide Archive database (http://www.ebi.ac.uk/ena/) under study accession numbers PRJEB8519 (small RNA-Seq GZI-2) and PRJEB25033 (other sequence runs).

Computational Analysis

The complete computational analysis pipeline was run on our in-house Galaxy server. All necessary workflows and tools will be publicly available at http://mississippi.fr/

Histories and Figures can be reach using the following individual links.
Figures: 1A and 1B; 1C and 1E; 1D and 1F; 1G
Figures: 2A, 2B and 2C; 2D and 2E
Figures: 3A; 3B; 3C
Supplemental Figures: 1A and 1B; 1C; 1D; 1E; 1F
Supplemental Figures: 2A; 2B

All small RNA libraries were quality controlled, sequencing adapter-clipped and converted to fasta reads. All reads that aligned to ribosomal RNA were discarded. All small RNA alignments were done using bowtie 0.12.7, allowing 1 mismatch between sequenced read and reference sequence (Langmead et al. 2009). To produce Supplemental Figure 1, fasta reads were aligned to the *Drosophila* genome

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(FlyBase release 5.49) (St Pierre et al. 2014), randomly placing reads that align equally well in multiple genomic locations (multimapper) using the bowtie option "-M 1". Size distribution and ping-pong signature were calculated using the mississippi toolsuite (https://testtoolshed.g2.bx.psu.edu/view/drosofff/mississipi_toolsuite_beta). The ping-pong signature was calculated by counting the number of pairs that overlap between 5 to 15 nucleotides between sense- and antisense aligned reads and transforming the obtained counts into z-scores (each count subtracted by the mean and divided by the standard deviation). Ping-pong positive libraries were selected by having a z-score higher than 2 and more than 20 pairs overlapping by 10 nucleotides. Ping-pong negative libraries were selected by having a negative z-score. To obtain a list of differentially expressed miRNA between ping-pong positive and ping-pong negative libraries reads were matched to the *Drosophila* pre-miRNAs of the miRBase 20 release (Griffiths-Jones 2004; Griffiths-Jones et al. 2006, 2008; Kozomara and Griffiths-Jones 2011, 2014). Differential expression profiling between ping-pong positive and ping-pong negative libraries was performed using edgeR_3.8.2 (Robinson, McCarthy, and Smyth 2010; McCarthy, Chen, and Smyth 2012) with standard settings. For simulating contamination with testis RNA 2 testis-libraries (accessions SRX135547, SRX023726) were downsampled to 10 million reads, pooled and 50000 randomly selected reads were added to 2.45 $*10^6$ randomly selected reads from ping-pong negative libraries. piRNA signature was calculated as before. Differential miRNA expression was calculated between simulated libraries and ping-pong negative libraries of equal size (randomly downsampled to 2.5 * 10^6), with libraries that were sampled from the same initial ping-pong negative library paired as a blocking factor. This allows for obtaining an accurate list of miRNAs (contamination signature) that should be expected to be to significantly change in abundance if a contamination occurred. Size distribution for small RNAs that align to TEs (Fig. 1A, 1B) was calculated from reads that matched any of the canonical TE sequences with 1 mismatch allowed, excluding reads that matched to ribosomal RNA, tRNA or abundant insect viruses. Abundance of 21 nt antisense RNA for each TE family was calculated by filtering reads to 21 nt length and aligning reads to canonical TE sequences, allowing only unique reads using the bowtie option "-v 1". Only antisense reads were counted, and only TEs with on average 20 reads per library were analyzed. Between-library normalized 21 nucleotide antisense TE counts were obtained by pooling these with miRNA reads (obtained as before) and calculating a normalization factor using the DESeq (Anders

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

and Huber 2010) function estimateSizeFactors. log2 fold changes were calculated by dividing normalized reads of mutants by normalized reads of controls and taking the logarithm. Difference of the population of log2 fold changes were tested using a two-tailed Mann-Whitney-U test. To calculate mismatch frequencies for 21 nucleotide small RNA, ribosomal, non-coding RNA and viral reads were filtered out. Remaining reads were aligned to the collection of TE insertions (FlyBase version 5.49), allowing 1 mismatch. Each possible mismatch was counted and divided by the total number of 21 nucleotide reads aligned to the collection of TE insertions.

For gene expression profiling, reads were quality-filtered using the FASTX toolkit with a Quality cut-off of 30 for 90% of the read. For the paired-end libraries only the R1 read of the pair was used and trimmed to 51 nucleotides. Reads were then aligned to the *Drosophila* genome release 5 (dm3) using Tophat2 (Kim et al. 2013). Default parameters were used, except that we supplied Gene Model annotations from UCSC genome browser for dm3 (http://support.illumina.com/sequencing/sequencing_software/igenome.html). Read counting was performed using featureCounts (Liao, Smyth, and Shi 2014) guided by the aforementioned Gene Model file.

For TE expression profiling, reads were further trimmed to 30 nucleotides and aligned to canonical TEs using bowtie 0.12.7, allowing 2 mismatches and only uniquely matching reads. Sense and antisense reads were counted and merge with gene counts. Differential expression profiling was performed using edgeR (Robinson, McCarthy, and Smyth 2010; McCarthy, Chen, and Smyth 2012). Genes with less than 5 reads on average across libraries were discarded from the analysis. Diverging from the default, we used Full Quantile between-library normalization as implemented by the EDAseq package (Risso et al. 2011) and removed unwanted variation using replicate samples with the RUVs function (choosing k=2) implemented in the RUVseq package (Risso et al. 2014). Library sequencing method (paired-end vs. single-read) was introduced together with gene-wise Full Quantile normalization offsets and gene-wise RUVs offsets as covariates in the edgeR design formula. All libraries were tested for differential gene expression against the wild type, and in addition the double-mutant was also tested against the piwi -/- mutant. Proportional Venn Diagrams in Figure 2C and Figure 2D were drawn using the Vennerable package (https://github.com/js229/Vennerable). The spearman rank correlation and corresponding P-value between log2 fold changes in TE 21 nucleotide antisense RNA (log2 fold change calculated from data underlying

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 1C) and sense TE transcript expression was calculated with the rcorr function in the Hmisc R package (http://cran.r-project.org/web/packages/Hmisc/). All graphs were plotted using ggplot2 (http://ggplot2.org/). GO terms for DNA damage and genes with TE insertions were retrieved from FlyBase (St Pierre et al. 2014).

Acknowledgements

We thank the Galaxy community for their support. This work received financial support to CA from the Agence Nationale de la Recherche grants ANR-13-BSV2-0007 and ANR-10-BLAN-1210. MdvB had a PhD fellowship from the Ministère de la Recherche et de l'Enseignement Supérieur. The École normale supérieure genomic core facility was supported by the France Génomique national infrastructure, funded as part of the "Investissements d'Avenir" program managed by the Agence Nationale de la Recherche (contract ANR-10-INBS-09).

References

Anders, Simon, and Wolfgang Huber. 2010. "Differential Expression Analysis for Sequence Count Data." *Genome Biology* 11 (10):R106.

Bayona-Feliu, Aleix, Anna Casas-Lamesa, Albert Carbonell, Paula Climent-Cantó, Milos Tatarski, Salvador Pérez-Montero, Fernando Azorín, and Jordi Bernués. 2016. "Histone H1: Lessons from Drosophila." *Biochimica et Biophysica Acta* 1859 (3):526–32.

Brennecke, Julius, Alexei A. Aravin, Alexander Stark, Monica Dus, Manolis Kellis, Ravi Sachidanandam, and Gregory J. Hannon. 2007. "Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila." *Cell* 128 (6):1089–1103.

Brower-Toland, Brent, Seth D. Findley, Ling Jiang, Li Liu, Hang Yin, Monica Dus, Pei Zhou, Sarah C. R. Elgin, and Haifan Lin. 2007. "Drosophila PIWI Associates with Chromatin and Interacts Directly with HP1a." *Genes & Development* 21 (18):2300–2311.

Cenik, Elif Sarinay, Ryuya Fukunaga, Gang Lu, Robert Dutcher, Yeming Wang, Traci M. Tanaka Hall, and Phillip D. Zamore. 2011. "Phosphate and R2D2 Restrict the Substrate Specificity of Dicer-2, an ATP-Driven Ribonuclease." *Molecular Cell* 42 (2):172–84.

Cox, D. N., A. Chao, J. Baker, L. Chang, D. Qiao, and H. Lin. 1998. "A Novel Class of Evolutionarily Conserved Genes Defined by Piwi Are Essential for Stem Cell Self-Renewal." *Genes & Development* 12 (23):3715–27.

Czech, Benjamin, Colin D. Malone, Rui Zhou, Alexander Stark, Catherine Schlingeheyde, Monica Dus, Norbert Perrimon, et al. 2008. "An Endogenous Small Interfering RNA Pathway in Drosophila." *Nature* 453 (7196):798–802.

Ghildiyal, Megha, Herve Seitz, Michael D. Horwich, Chengjian Li, Tingting Du, Soohyun Lee, Jia Xu, et al. 2008a. "Endogenous siRNAs Derived from Transposons and mRNAs in Drosophila Somatic Cells." *Science* 320 (5879):1077–81.

Ghildiyal, Megha, Hervé Seitz, Michael D. Horwich, Chengjian Li, Tingting Du, Soohyun Lee, Jia Xu, et al. 2008b. "Endogenous siRNAs Derived from Transposons and mRNAs in Drosophila Somatic Cells." *Science* 320 (5879). American Association for the Advancement of Science:1077–81.

Griffiths-Jones, Sam. 2004. "The microRNA Registry." *Nucleic Acids Research* 32 (Database issue):D109–11.

Griffiths-Jones, Sam, Russell J. Grocock, Stijn van Dongen, Alex Bateman, and Anton J.

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Enright. 2006. “miRBase: microRNA Sequences, Targets and Gene Nomenclature.” *Nucleic Acids Research* 34 (Database issue):D140–44.

Griffiths-Jones, Sam, Harpreet Kaur Saini, Stijn van Dongen, and Anton J. Enright. 2008. “miRBase: Tools for microRNA Genomics.” *Nucleic Acids Research* 36 (Database issue):D154–58.

Gunawardane, Lalith S., Kuniaki Saito, Kazumichi M. Nishida, Keita Miyoshi, Yoshinori Kawamura, Tomoko Nagami, Haruhiko Siomi, and Mikiko C. Siomi. 2007. “A Slicer-Mediated Mechanism for Repeat-Associated siRNA 5’ End Formation in Drosophila.” *Science* 315 (5818):1587–90.

Gu, Tingting, and Sarah C. R. Elgin. 2013. “Maternal Depletion of Piwi, a Component of the RNAi System, Impacts Heterochromatin Formation in Drosophila.” *PLoS Genetics* 9 (9):e1003780.

Jin, Zhigang, Alex S. Flynt, and Eric C. Lai. 2013. “Drosophila Piwi Mutants Exhibit Germline Stem Cell Tumors That Are Sustained by Elevated Dpp Signaling.” *Current Biology: CB* 23 (15):1442–48.

Keegan, Liam P., James Brindle, Angela Gallo, Anne Leroy, Robert A. Reenan, and Mary A. O’Connell. 2005. “Tuning of RNA Editing by ADAR Is Required in Drosophila.” *The EMBO Journal* 24 (12):2183–93.

Kim, Daehwan, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L. Salzberg. 2013. “TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions.” *Genome Biology* 14 (4):R36.

Klenov, Mikhail S., Sergey A. Lavrov, Alina P. Korbut, Anastasia D. Stolyarenko, Evgeny Y. Yakushev, Michael Reuter, Ramesh S. Pillai, and Vladimir A. Gvozdev. 2014. “Impact of Nuclear Piwi Elimination on Chromatin State in Drosophila Melanogaster Ovaries.” *Nucleic Acids Research* 42 (10):6208–18.

Kozomara, Ana, and Sam Griffiths-Jones. 2011. “miRBase: Integrating microRNA Annotation and Deep-Sequencing Data.” *Nucleic Acids Research* 39 (Database issue):D152–57.

———. 2014. “miRBase: Annotating High Confidence microRNAs Using Deep Sequencing Data.” *Nucleic Acids Research* 42 (Database issue):D68–73.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. 2009. “Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome.” *Genome Biology* 10 (3):R25.

Lau, Nelson C., Nicolas Robine, Raquel Martin, Wei-Jen Chung, Yuzo Niki, Eugene Berezikov, and Eric C. Lai. 2009. “Abundant Primary piRNAs, Endo-siRNAs, and microRNAs in a Drosophila Ovary Cell Line.” *Genome Research* 19 (10):1776–85.

Lee, Young Sik, Kenji Nakahara, John W. Pham, Kevin Kim, Zhengying He, Erik J. Sontheimer, and Richard W. Carthew. 2004. “Distinct Roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA Silencing Pathways.” *Cell* 117 (1):69–81.

Le Thomas, Adrien, Alicia K. Rogers, Alexandre Webster, Georgi K. Marinov, Susan E. Liao, Edward M. Perkins, Junho K. Hur, Alexei A. Aravin, and Katalin Fejes Tóth. 2013. “Piwi Induces piRNA-Guided Transcriptional Silencing and Establishment of a Repressive Chromatin State.” *Genes & Development* 27 (4):390–99.

Liao, Yang, Gordon K. Smyth, and Wei Shi. 2014. “featureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features.” *Bioinformatics* 30 (7):923–30.

Li, Wanhe, Lisa Prazak, Nabanita Chatterjee, Servan Grüninger, Lisa Krug, Delphine Theodorou, and Josh Dubnau. 2013. “Activation of Transposable Elements during Aging and Neuronal Decline in Drosophila.” *Nature Neuroscience* 16 (5):529–31.

Lu, Xingwu, Sandeep N. Wontakal, Harsh Kavi, Byung Ju Kim, Paloma M. Guzzardo, Alexander V. Emelyanov, Na Xu, et al. 2013. “Drosophila H1 Regulates the Genetic Activity of Heterochromatin by Recruitment of Su(var)3-9.” *Science* 340 (6128):78–81.

Malone, Colin D., Julius Brennecke, Monica Dus, Alexander Stark, W. Richard McCombie, Ravi Sachidanandam, and Gregory J. Hannon. 2009. “Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary.” *Cell* 137 (3):522–35.

Mani, Sneha Ramesh, Heather Megosh, and Haifan Lin. 2014. “PIWI Proteins Are Essential for Early Drosophila Embryogenesis.” *Developmental Biology* 385 (2):340–49.

McCarthy, Davis J., Yunshun Chen, and Gordon K. Smyth. 2012. “Differential Expression Analysis of Multifactor RNA-Seq Experiments with Respect to Biological Variation.”

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Nucleic Acids Research 40 (10):4288–97.

Mirkovic-Hösle, Milijana, and Klaus Förstemann. 2014a. "Transposon Defense by Endo-siRNAs, piRNAs and Somatic pilRNAs in Drosophila: Contributions of Loqs-PD and R2D2." *PloS One* 9 (1):e84994.

———. 2014b. "Transposon Defense by Endo-siRNAs, piRNAs and Somatic pilRNAs in Drosophila: Contributions of Loqs-PD and R2D2." *PloS One* 9 (1):e84994.

Muller, H. J. 1930. "Types of Visible Variations Induced by X-Rays inDrosophila." *Journal of Genetics* 22 (3). Springer India:299–334.

Nagao, Akihiro, Toutai Mituyama, Haidong Huang, Dahua Chen, Mikiko C. Siomi, and Haruhiko Siomi. 2010. "Biogenesis Pathways of piRNAs Loaded onto AGO3 in the Drosophila Testis." *RNA* 16 (12):2503–15.

Pal-Bhadra, Manika, Boris A. Leibovitch, Sumit G. Gandhi, Madhusudana Rao, Utpal Bhadra, James A. Birchler, and Sarah C. R. Elgin. 2004. "Heterochromatic Silencing and HP1 Localization in Drosophila Are Dependent on the RNAi Machinery." *Science* 303 (5658):669–72.

Palladino, M. J., L. P. Keegan, M. A. O'Connell, and R. A. Reenan. 2000. "dADAR, a Drosophila Double-Stranded RNA-Specific Adenosine Deaminase Is Highly Developmentally Regulated and Is Itself a Target for RNA Editing." *RNA* 6 (7):1004–18.

Perrat, Paola N., Shamik DasGupta, Jie Wang, William Theurkauf, Zhiping Weng, Michael Rosbash, and Scott Waddell. 2013. "Transposition-Driven Genomic Heterogeneity in the Drosophila Brain." *Science* 340 (6128):91–95.

Rangan, Prashanth, Colin D. Malone, Caryn Navarro, Sam P. Newbold, Patrick S. Hayes, Ravi Sachidanandam, Gregory J. Hannon, and Ruth Lehmann. 2011. "piRNA Production Requires Heterochromatin Formation in Drosophila." *Current Biology: CB* 21 (16):1373–79.

Reinhardt, Anita, Sébastien Feuillette, Marlène Cassar, Céline Callens, Hélène Thomassin, Serge Birman, Magalie Lecourtois, Christophe Antoniewski, and Hervé Tricoire. 2012. "Lack of miRNA Misregulation at Early Pathological Stages in Drosophila Neurodegenerative Disease Models." *Frontiers in Genetics* 3 (January). Frontiers:226.

Rio, Donald C., Manuel Ares Jr, Gregory J. Hannon, and Timothy W. Nilsen. 2010. "Purification of RNA Using TRIzol (TRI Reagent)." *Cold Spring Harbor Protocols* 2010 (6):db.prot5439.

Risso, Davide, John Ngai, Terence P. Speed, and Sandrine Dudoit. 2014. "Normalization of RNA-Seq Data Using Factor Analysis of Control Genes or Samples." *Nature Biotechnology* 32 (9):896–902.

Risso, Davide, Katja Schwartz, Gavin Sherlock, and Sandrine Dudoit. 2011. "GC-Content Normalization for RNA-Seq Data." *BMC Bioinformatics* 12 (December):480.

Robine, Nicolas, Nelson C. Lau, Sudha Balla, Zhigang Jin, Katsutomo Okamura, Satomi Kuramochi-Miyagawa, Michael D. Blower, and Eric C. Lai. 2009. "A Broadly Conserved Pathway Generates 3′UTR-Directed Primary piRNAs." *Current Biology: CB* 19 (24):2066–76.

Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. 2010. "edgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data." *Bioinformatics* 26 (1):139–40.

Rouget, Christel, Catherine Papin, Anthony Boureux, Anne-Cécile Meunier, Bénédicte Franco, Nicolas Robine, Eric C. Lai, Alain Pelisson, and Martine Simonelig. 2010. "Maternal mRNA Deadenylation and Decay by the piRNA Pathway in the Early Drosophila Embryo." *Nature* 467 (7319):1128–32.

Rozhkov, Nikolay V., Alexei A. Aravin, Elena S. Zelentsova, Natalia G. Schostak, Ravi Sachidanandam, W. Richard McCombie, Gregory J. Hannon, and Michael B. Evgen'ev. 2010. "Small RNA-Based Silencing Strategies for Transposons in the Process of Invading Drosophila Species." *RNA* 16 (8):1634–45.

Rozhkov, Nikolay V., Molly Hammell, and Gregory J. Hannon. 2013. "Multiple Roles for Piwi in Silencing Drosophila Transposons." *Genes & Development* 27 (4):400–412.

Saito, Kuniaki, Sachi Inagaki, Toutai Mituyama, Yoshinori Kawamura, Yukiteru Ono, Eri Sakota, Hazuki Kotani, Kiyoshi Asai, Haruhiko Siomi, and Mikiko C. Siomi. 2009. "A Regulatory Circuit for Piwi by the Large Maf Gene Traffic Jam in Drosophila." *Nature* 461 (7268):1296–99.

bioRxiv preprint doi: https://doi.org/10.1101/267039; this version posted February 21, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Sienski, Grzegorz, Julia Batki, Kirsten-André Senti, Derya Dönertas, Laszlo Tirian, Katharina Meixner, and Julius Brennecke. 2015. "Silencio/CG9754 Connects the Piwi--piRNA Complex to the Cellular Heterochromatin Machinery." *Genes & Development* 29 (21). Cold Spring Harbor Lab:2258–71.

Sienski, Grzegorz, Derya Dönertas, and Julius Brennecke. 2012. "Transcriptional Silencing of Transposons by Piwi and Maelstrom and Its Impact on Chromatin State and Gene Expression." *Cell* 151 (5):964–80.

St Pierre, Susan E., Laura Ponting, Raymund Stefancsik, Peter McQuilton, and FlyBase Consortium. 2014. "FlyBase 102--Advanced Approaches to Interrogating FlyBase." *Nucleic Acids Research* 42 (Database issue):D780–88.

Toledano, Hila, Cecilia D'Alterio, Benjamin Czech, Erel Levine, and D. Leanne Jones. 2012. "The Let-7-Imp Axis Regulates Ageing of the Drosophila Testis Stem-Cell Niche." *Nature* 485 (7400):605–10.

Tubio, Jose M. C., Yilong Li, Young Seok Ju, Inigo Martincorena, Susanna L. Cooke, Marta Tojo, Gunes Gundem, et al. 2014. "Extensive Transduction of Nonrepetitive DNA Mediated by L1 Retrotransposition in Cancer Genomes." *Science* 345 (6196):1251343.

Vagin, Vasily V., Alla Sigova, Chengjian Li, Hervé Seitz, Vladimir Gvozdev, and Phillip D. Zamore. 2006. "A Distinct Small RNA Pathway Silences Selfish Genetic Elements in the Germline." *Science* 313 (5785):320–24.

Vujatovic, Olivera, Katrin Zaragoza, Alejandro Vaquero, Oscar Reina, Jordi Bernués, and Fernando Azorín. 2012. "Drosophila Melanogaster Linker Histone dH1 Is Required for Transposon Silencing and to Preserve Genome Integrity." *Nucleic Acids Research* 40 (12):5402–14.

Wang, Sidney H., and Sarah C. R. Elgin. 2011. "Drosophila Piwi Functions Downstream of piRNA Production Mediating a Chromatin-Based Transposon Silencing Mechanism in Female Germ Line." *Proceedings of the National Academy of Sciences* 108 (52):21164–69.

Welker, Noah C., Tuhin S. Maity, Xuecheng Ye, P. Joseph Aruscavage, Ammie A. Krauchuk, Qinghua Liu, and Brenda L. Bass. 2011. "Dicer's Helicase Domain Discriminates dsRNA Termini to Promote an Altered Reaction Mode." *Molecular Cell* 41 (5):589–99.

Wen, Jiayu, Hong Duan, Fernando Bejarano, Katsutomo Okamura, Lacramioara Fabian, Julie A. Brill, Diane Bortolamiol-Becet, Raquel Martin, J. Graham Ruby, and Eric C. Lai. 2015. "Adaptive Regulation of Testis Gene Expression and Control of Male Fertility by the Drosophila Hairpin RNA Pathway. [Corrected]." *Molecular Cell* 57 (1):165–78.

Wood, Jason G., Brian C. Jones, Nan Jiang, Chengyi Chang, Suzanne Hosier, Priyan Wickremesinghe, Meyrolin Garcia, et al. 2016. "Chromatin-Modifying Genetic Interventions Suppress Age-Associated Transposable Element Activation and Extend Life Span in Drosophila." *Proceedings of the National Academy of Sciences of the United States of America* 113 (40):11277–82.

Wu, Diane, Ayelet T. Lamm, and Andrew Z. Fire. 2011. "Competition between ADAR and RNAi Pathways for an Extensive Class of RNA Targets." *Nature Structural & Molecular Biology* 18 (10):1094–1101.

Xie, Weiwu, Ryan C. Donohue, and James A. Birchler. 2013. "Quantitatively Increased Somatic Transposition of Transposable Elements in Drosophila Strains Compromised for RNAi." *PloS One* 8 (8):e72163.

Yan, Zheng, Hai Yang Hu, Xi Jiang, Vera Maierhofer, Elena Neb, Liu He, Yuhui Hu, et al. 2011a. "Widespread Expression of piRNA-like Molecules in Somatic Tissues." *Nucleic Acids Research* 39 (15):6596–6607.

———. 2011b. "Widespread Expression of piRNA-like Molecules in Somatic Tissues." *Nucleic Acids Research* 39 (15):6596–6607.

Zamudio, Jesse R., Timothy J. Kelly, and Phillip A. Sharp. 2014. "Argonaute-Bound Small RNAs from Promoter-Proximal RNA Polymerase II." *Cell* 156 (5):920–34.

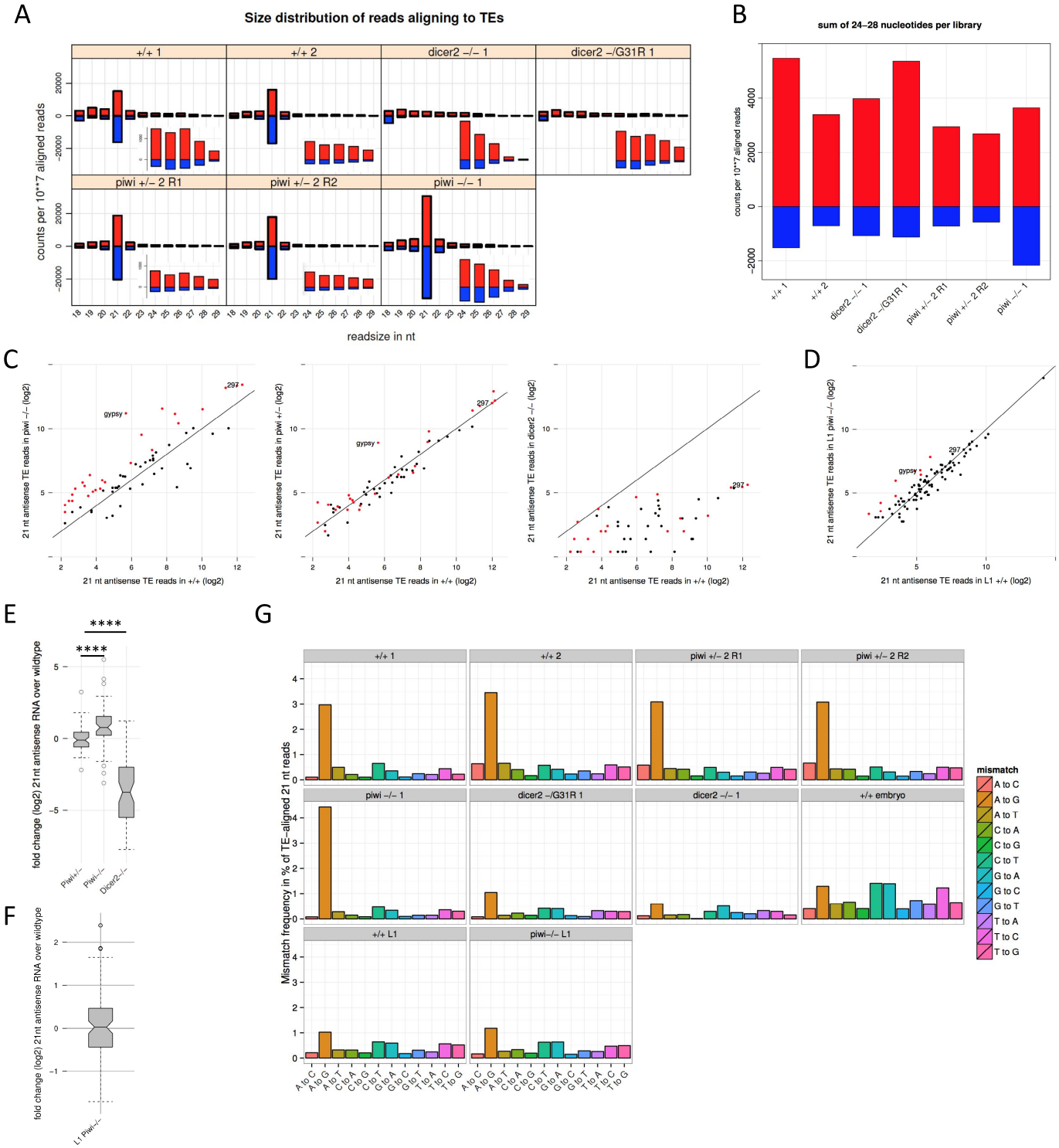


Figure 1. ***Loss of Piwi repression leads to an increase of TE-targeting siRNAs.*** (A) Overview of the size (x-axis) and amount (y-axis, in counts per 10 million mapped reads) of small RNA reads that align to TEs in adult heads of the indicated genotype. Zoom of the 24-28 nt fractions are shown as insets. (B) Sense or antisense reads of 24 to 28 nt were summed for each indicated genotype (x-axis). (C) Scatterplots displaying the abundance of 21 nt antisense reads in mutant (y-axis) and wild type (x-axis) heads. Red dots in the first panel indicate the transposon-specific 21 nt antisense reads that increased more than 2 fold in *piwi* homozygous mutant heads. These dots are shown for comparison in the second and third panel. (D) same as (C) but for wildtype 1st instar larvae (homozygous *piwi* mutant *vs* wild type). (E) Boxplots showing the distribution of 21 nt antisense fold changes (y-axis) between wild type and the indicated mutants (x-axis). Significance of differences between the distributions was assessed with Mann-Whitney U test. (F) as (E), but for wildtype 1st instar larvae (homozygous *piwi* mutant *vs* wild type). (G) Mismatches of 21 nt reads aligning to reference genome TE insertions with 1 mismatch allowed. Identity of mismatch is indicated on the x-axis, and the fraction of all reads with this mismatch identity over all TE matches is indicated on the y-axis. In all panels, the index of sample duplicate is indicated after the genotype when appropriate.

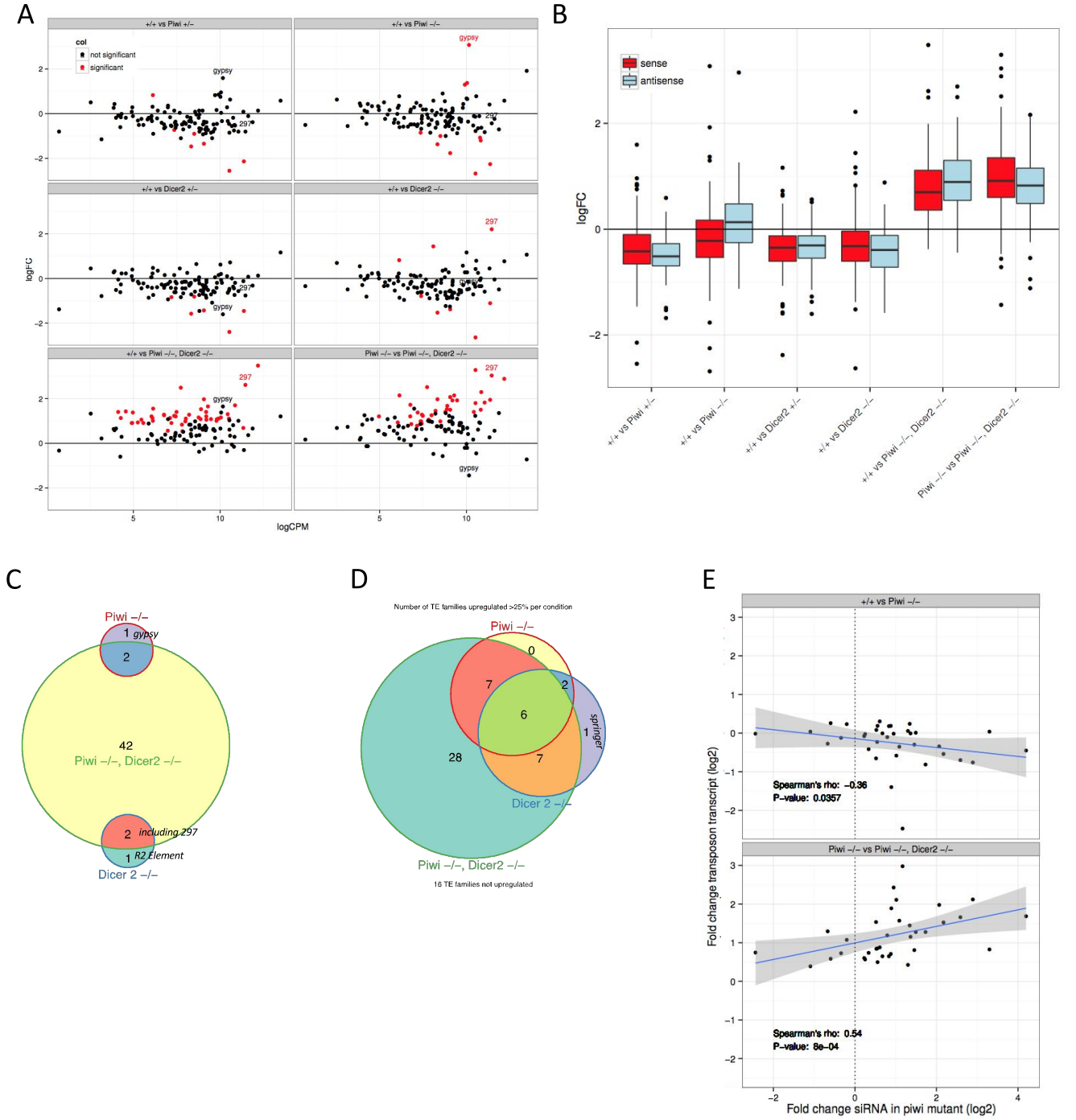


Figure 2. ***Piwi and Dicer-2 are complementary factors for the repression of TEs in adult heads.*** (A) Scatterplots displaying the log2 fold changes of sense TE transcript expression on the y-axis and the mean expression strength on the x-axis. log2 fold changes were calculated between the indicated mutant and the wild type TE levels, except in the last panel where we calculate changes between the double mutant and the *piwi* mutant. (B) Boxplots showing the distribution of log2 fold changes as in (A), but considering changes of sense (red) and antisense TE transcripts (blue) separately for each comparison (x-axis). (C) Venn Diagram showing the overlap of significantly (p.adj <0.05) up regulated TEs for the indicated mutants as compared to the wild type control. (D) as (C), but taking TEs whose transcript abundance increases more than 25% over the wild type. (E) Scatterplot displaying the correlation between log2 fold changes of 21 nt antisense RNA (siRNA) in *piwi* homozygous mutant heads compared to wild type heads on the x-axis and log2 fold changes of sense TE transcripts for the genotype comparisons indicated above each panel. Only TEs that passed a threshold of on average five 21 nucleotide antisense reads (after library-wise normalization, Material and Methods) over all small RNA libraries were analysed. Further, only TEs whose sense transcript level increased in *piwi*, *dicer-2* while insensitive to *piwi* loss are shown. A Scatterplot including all tested genotype that depicts all TEs that passed the siRNA threshold can be found in Supplemental Figure 2B. The blue line is a fit produced by the lmfit function and the grey area delimits the corresponding confidence interval.

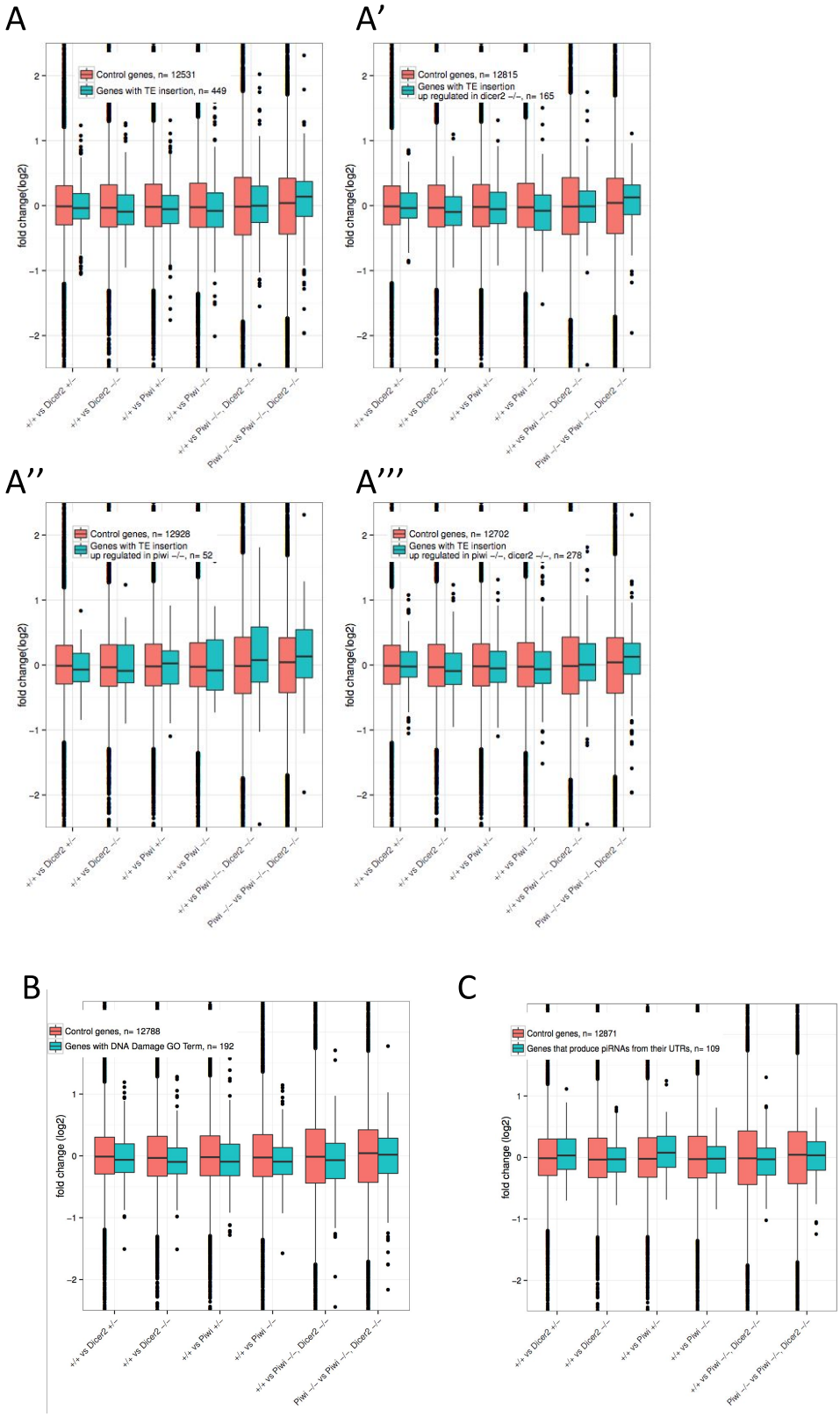
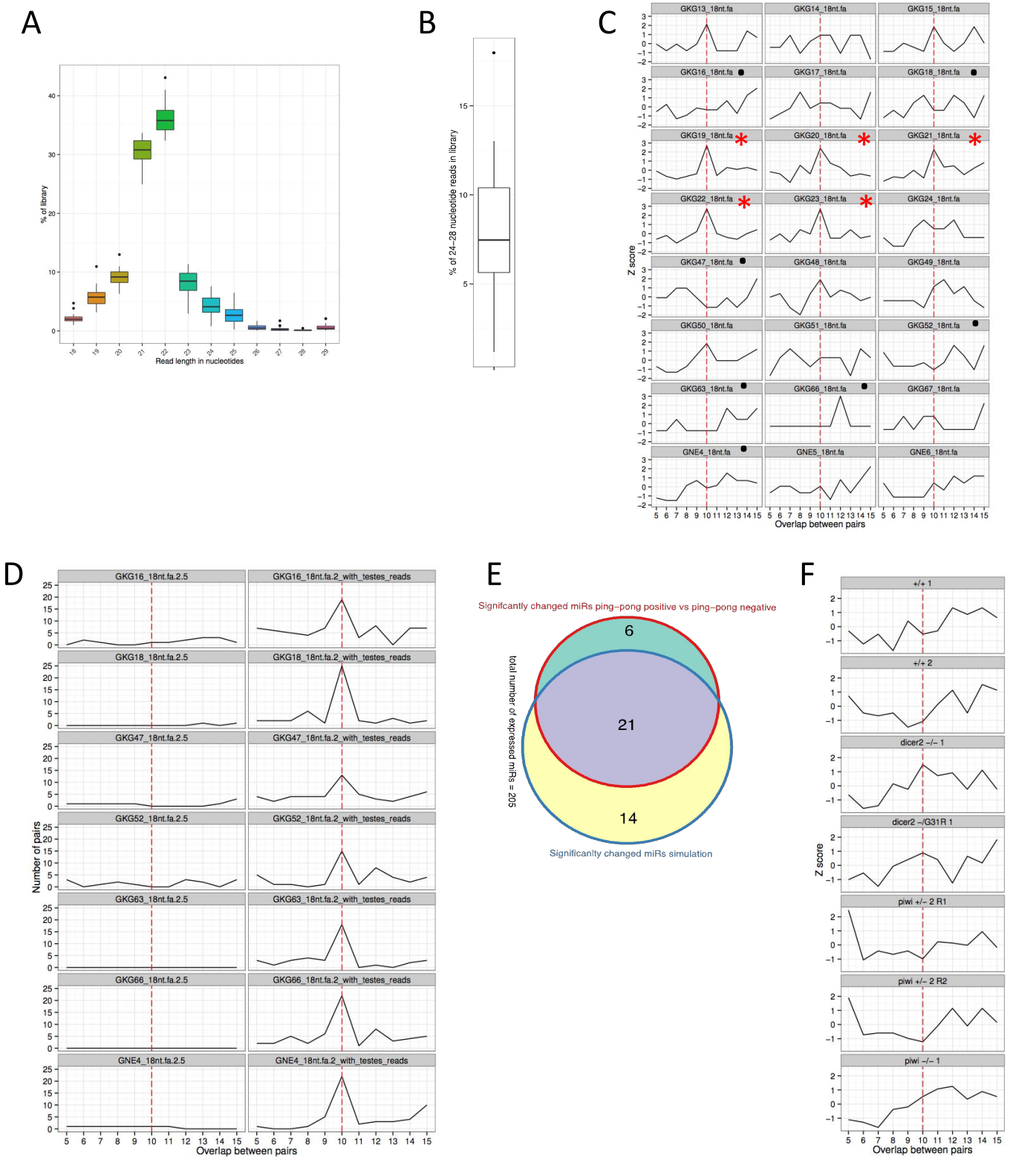
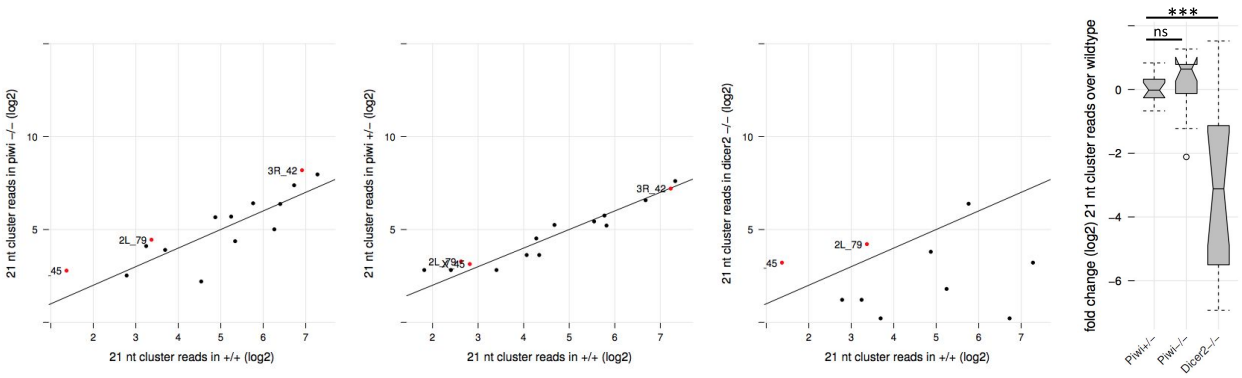


Figure 3. ***piwi, dicer-2 double mutant in heads is not affecting drastically previously reported piwi-regulated genes in gonads.*** Boxplots of fold change for genes with reference genome TE insertions in their genomic boundaries do not suggest a trend for these genes to become de-repressed across any of the tested mutant condition (x-axis). We compared genes without TE insertions (red boxes) with genes with insertion of TE of any family (blue boxes, A), or with TE insertions that are up regulated in *dicer2* homozygous (blue boxes, A'), *piwi* homozygous (A'') or *piwi*, *dicer-2* double homozygous mutants (A'''). Boxplots of the distribution of fold changes for genes with DNA Damage Gene ontology terms (B) and genes that have been reported to host piRNA production from their UTRs (C).

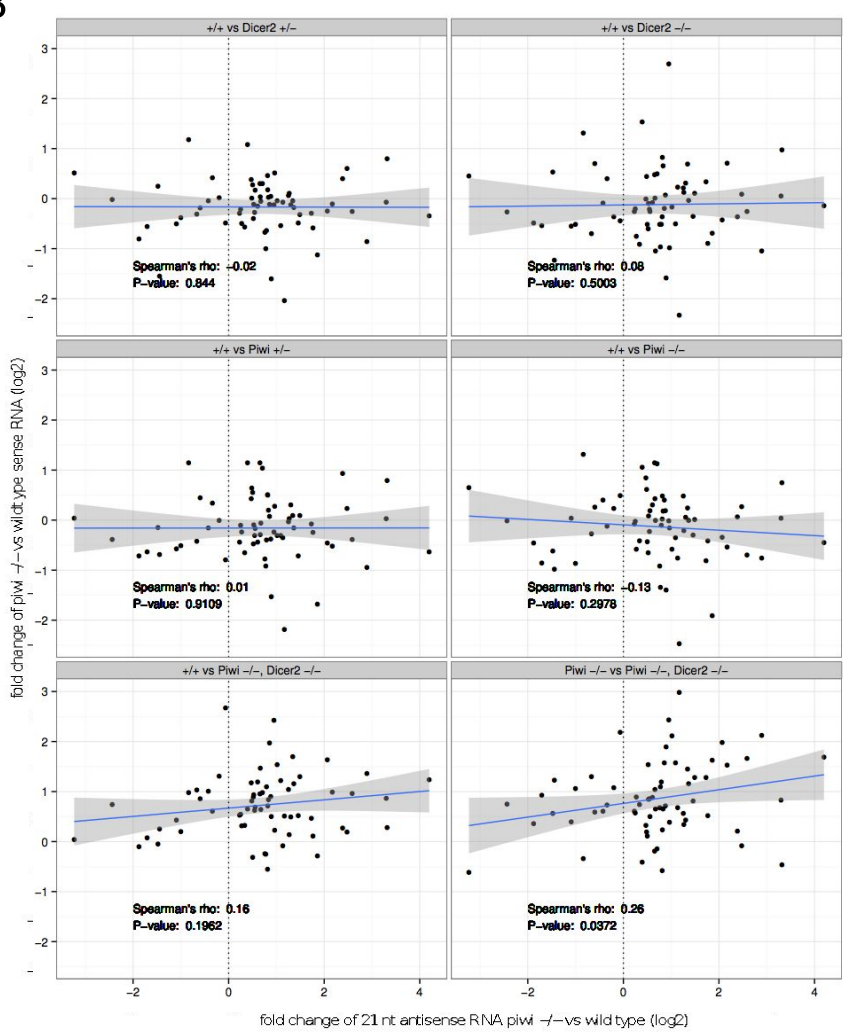


Supplemental Figure 1. ***Drosophila small RNA head libraries have variable amount of piRNA-sized reads and ping-pong signature in heads, which correlates with a contamination signature.*** (A) Boxplot showing the distribution of read lengths for the libraries analyzed in (C). (B) similar to (A), but for the fraction of 24-28 nucleotide length reads. (C) Ping-pong signature - tendency for small RNAs to overlap. The number of overlaps was transformed to Z scores to take into account the "cleanness" of the 10 nucleotide overlap as compared to other lengths of overlaps. Libraries with red asterisks were selected as ping-pong positive, and black dots indicate libraries selected as ping-pong negative libraries. (D) Ping-pong signature for $2.5*10^6$ ping-pong negative libraries with (right group of panels) and without (left panels) the addition of 2% of a testis-library. Number of pairs are shown instead of Z scores, as some down-sampled libraries had 0 overlapping pairs. (E) Venn Diagram showing the overlap of differentially expressed miRNAs between ping-pong positive compared to ping-pong negative libraries (red circle) and ping-pong negative libraries compared to ping-pong negative libraries with the addition of 2% of a testis library (blue circle). (F) ping-pong signature for small RNA libraries (as in (C)) of indicated genotypes of heads closely dissected by hands.

A



B



Supplemental Figure 2. (A) Scatterplots displaying the abundance of cluster-derived 21 nt reads in mutant (y-axis) and wild type (x-axis) heads. Red dots in the first panel indicate the cluster-specific 21 nt antisense reads that increased more than 2 fold in *piwi* -/- mutant heads. These dots are shown for comparison in the second and third panel. (A-right) Boxplots showing the distribution of 21 nt read fold changes (y-axis) between wild type and the indicated mutants (x-axis). Significance of differences between the distributions was assessed with Mann-Whitney U test. (B) Related to Figure 2E. Scatterplot displaying the correlation between log2 fold changes of 21 nt antisense RNA in *piwi* homozygous mutant heads compared to wild type heads on the x-axis and log2 fold changes of sense TE transcripts for the genotype comparisons indicated above each panel. All TE families that passed a threshold of on average five 21 nucleotide antisense reads over all small RNA libraries were analysed. The blue line is a fit produced by the lmfit function, and the grey area delimits the corresponding confidence interval.