# How germ line epigenetic regulators contribute to gene expression during genomic instability

By
© 2018
Alexandra A. Erwin
B.S, Biology, Missouri State University, 2012

Submitted to the graduate degree program in Ecology and Evolutionary Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Chair: Justin P. Blumenstiel

_____

Stuart J. Macdonald

_____

Maria E. Orive

_____

James R. Walters

_____

Lena C. Hileman

Date Defended: 29 March 2018

The dissertation committee for Alexandra A. Erwin certifies that this is the approved version of the following dissertation:

# How germ line epigenetic regulators contribute to gene expression during genomic instability

Chair: Justin P. Blumenstiel

Date Approved: 11 May 2018

## Abstract

Genetically identical individuals can have strikingly different phenotypes due to differences in gene expression. One reason for this is because epigenetic regulators can modulate mitotically or meiotically heritable changes in gene expression without changing DNA sequences. One way epigenetic factors can affect gene expression is by mediating how tightly DNA is packaged in the nucleus. When the DNA is tightly packed, in a structure known as heterochromatin, it is inaccessible for transcription, whereas lightly packed DNA, euchromatin, is accessible for transcription. The most well understood epigenetic regulators involve chemical modification of DNA like methylation and post-translational modifications of histone proteins, but small noncoding RNAs now are being recognized as important epigenetic regulators. Small RNAs can initiate the formation of heterochromatin as well as regulate genes post-transcriptionally by preventing the translation of mRNA.

Piwi-interacting RNAs (piRNAs) are maternally transmitted small RNAs that control mutagenic transposable elements (TEs). There is a striking interplay between heterochromatin formation and piRNA function as piRNAs target resident TE insertions for repression through heterochromatin formation. Although most DNA and chromatin based epigenetic marks are erased between generations, piRNA are directly deposited into the embryo to maintain TE repression across generations. Because piRNA can induce heterochromatin formation, the transgenerational transmission of piRNA is thought to be one of the ways that heterochromatin domains can be re-established in the embryo.

In chapters one and two, I examine how disparate maternal piRNA pools influence gene expression using a sterility syndrome of hybrid dysgenesis in *Drosophila virilis*. Chapter 1 outlines the effects of TE silencing small RNAs on gene expression in the germline and subsequent effects on future generations. Chapter 2 extends that study to investigate how disparate piRNA profiles and gene silencing in the genome affect somatic gene expression.

Misregulation of epigenetic regulators is associated with many diseases, ranging from kidney diseases and neurodegenerative diseases, to cancer. Recently, epigenetic regulation has been implicated as playing a key role in the aging process. In particular, the landscape of silent heterochromatin has been shown to redistribute in aged stem cells and cells of the soma, leading to aberrant gene and transposable element expression. Although important for understanding the biology of aging, these cells do not affect future generations. In fact, little is known about heritable epigenetic changes that occur in aged germline cells that do give rise to the next generation. In chapter 3, I investigate whether the epigenetic deregulation and TE derepression that has been shown to exist in the soma also exists in the *Drosophila melanogaster* germline.

# Acknowledgements

I would like to express my sincerest gratitude to my advisor, Dr. Justin Blumenstiel, for the guidance and mentorship he has provided throughout the pursuit of this degree. As both a rigorous scientist and an incredibly kind person, he has been a significant professional and personal role model. I feel beyond fortunate to have had the opportunity to pursue my degree under his guidance.

I would also like to thank my committee members, Drs. Lena Hileman, Stuart MacDonald, Maria Orive, and Jamie Walters for the guidance and valuable suggestions over the years. I would also like to recognize Dr. Laszlo Kovacs for the contributions that he made to my intellectual growth during my time as an undergraduate at Missouri State University. I must thank the Department of Ecology and Evolutionary Biology and the University of Kansas for graciously providing me with countless resources during my time here and I would especially like to thank my fellow graduate students for being a source of constant inspiration, motivation, and support.

I want to thank my grandparents Anna and Jack for their encouragement and generosity throughout my schooling. My parents, Dave and Natalya, for their immeasurable support throughout my life and for fostering a home environment where thoughtful dialogue flourished and scientific evidence was highly valued. I would also like to thank my parents-in-law, Tracy and Jim for their enthusiastic encouragement over the years. Finally, I would like to thank my husband, Mike, for the unwavering support he has shown throughout my undergraduate and graduate degree pursuits. I especially want to thank him for always being confident in my ability to accomplish and achieve. His assurance and encouragement was an integral source of inspiration and motivation.

Table of Contents

# List of Figures

# List of Tables

# Chapter 1
# piRNAs Are Associated with Diverse Transgenerational Effects on Gene and Transposon Expression in a Hybrid Dysgenic Syndrome of *Drosophila virilis*

## ABSTRACT

Sexual reproduction allows transposable elements (TEs) to proliferate, leading to rapid divergence between populations and species. A significant outcome of divergence in the TE landscape is evident in hybrid dysgenic syndromes, a strong form of genomic incompatibility that can arise when (TE) family abundance differs between two parents. When TEs inherited from the father are absent in the mother's genome, TEs can become activated in the progeny, causing germline damage and sterility. Studies in Drosophila indicate that dysgenesis can occur when TEs inherited paternally are not matched with a pool of corresponding TE silencing PIWI-interacting RNAs (piRNAs) provisioned by the female germline. Using the D. virilis syndrome of hybrid dysgenesis as a model, we characterize the effects that divergence in TE profile between parents has on offspring. Overall, we show that divergence in the TE landscape is associated with persisting differences in germline TE expression when comparing genetically identical females of reciprocal crosses and these differences are transmitted to the next generation. Moreover, chronic and persisting TE expression coincides with increased levels of genic piRNAs associated with reduced gene expression. Combined with these effects, we further demonstrate that gene expression is idiosyncratically influenced by differences in the genic piRNA profile of the parents that arise though polymorphic TE insertions. Overall, these results support a model in which early germline events in dysgenesis establish a chronic, stable state of both TE and gene expression in the germline that is maintained through adulthood and transmitted to the next generation. This work demonstrates that divergence in the TE profile is associated with diverse piRNA-mediated transgenerational effects on gene expression within populations.

# INTRODUCTION

In sexually reproducing species, two unique haploid genomes join together in syngamy to establish each generation. This mixing of genomes introduces potentially advantageous variation under changing environmental conditions, but also provides a condition ripe for exploitation by selfish elements (Hickey, 1982). Because syngamy can introduce selfish elements to new genomes and recombination can separate selfish elements from their harmful consequences, selfish elements such as transposable elements (TEs) can proliferate (Charlesworth and Charlesworth, 1983; Charlesworth and Langley, 1989). This is exemplified by the P element in Drosophila melanogaster. Through a likely horizontal transfer event from the distant species *D. willistoni*, the P element invaded D. melanogaster less than 100 years ago and is now found in D. melanogaster world-wide (Anxolabehere et al., 1988; Daniels et al., 1990).

Because TEs can be harmful and also drive a rapid accumulation of differences between species, they have been proposed to contribute to reproductive isolation. While their proliferative nature makes it very unlikely that TEs are drivers of speciation itself (Coyne, 1986, 1989), TE misregulation has been observed in a variety of interspecific hybrids. For example, increased TE expression is observed in malformed backcrosses between recently diverged species of lake whitefish (Dion-Cote et al., 2014). Similar observations have been made in species ranging from Arabidopsis (Josefsson et al., 2006) to wallaby (O'Neill et al., 1998, 2002). Studies in Drosophila using interspecific crosses have been especially important for our understanding of TE control in hybrids. The results so far have been idiosyncratic. Interspecific hybrids between closely related members of the *affinis*, *simulans*, *virilis* and *pseudoobscura* groups of Drosophila show little evidence for increased transposition (Coyne, 1986, 1989; Labrador et al., 1999; O'Neill et al., 2002; Vela et al., 2014). In contrast, increased transposition is observed in crosses between *D. buzzatii* and *D. koepferae* (Labrador et al., 1999; Vela et al., 2014). In the latter case, the increased rate of TE movement has been attributed to a form of genomic stress, though the

nature of this stress is not clear. Additionally, interspecific hybrids between *D. simulans* and *D. melanogaster* (which are more distantly related compared to those in previous crosses examining this question (Coyne, 1986, 1989) do show increased expression of TEs (Kelleher et al., 2012) and this is attributed to adaptive divergence in components of the TE regulatory machinery. Since species may differ significantly both in TE profile and regulatory machinery protein function, it is challenging to determine how divergence in TE profile alone contributes to TE activation in interspecific hybrids.

For this reason, intraspecific syndromes of hybrid dysgenesis provide critical insight into the role that divergence in TE profile can play in determining TE activity across generations. Hybrid dysgenesis is defined as a syndrome of hybrid sterility (Kidwell et al., 1977; Kidwell and Novy, 1979) and germline damage that occurs in intraspecific crosses when the male carries one or more TE families absent in the female (Bingham et al., 1982; Bucheton et al., 1984; Yannopoulos et al., 1987).The dysgenesis phenomenon in Drosophila has provided crucial insight into mechanisms of host genome defense by small RNAs. This is because activation of TEs inherited solely through the Drosophila male germline can be explained by the fact that the maternal germline is the primary agent of transgenerational TE repression via PIWI-interacting RNAs (piRNAs) maternally loaded into the egg (Aravin et al., 2003).

piRNAs are 23–30 nt RNAs found in complex with PIWI proteins and they play a crucial role in maintaining genome integrity via the repression of TEs. Many piRNAs are derived from TE fragments residing in distinct genomic regions known as piRNA clusters (Aravin et al., 2007; Brennecke et al., 2007; Brennecke et al., 2008). Anti-sense TE transcripts derived from these clusters are processed into piRNAs and, in complex with PIWI proteins, serve as guides to target resident TE transcripts for PIWI-mediated 'slicing' (Gunawardane et al., 2007). This system serves as a mechanism of genome defense because the proliferative nature of TEs can be inherently recognized by their tendency to transpose into

piRNA clusters, whereby they serve as guides to recognize mRNAs of the same TE family. In the absence of the maternally provisioned piRNAs that target TE mRNAs for PIWI-mediated slicing, paternally inherited TEs become activated in the progeny germline. This has been demonstrated for the P-M and I-Rsystems of hybrid dysgenesis in *D. melanogaster* (Chambeyron et al., 2008; Grentzinger et al., 2012; Khurana et al., 2011b). In the P-M system, P strains, but not M strains, carry active copies of the DNA transposon known as the P element. In the I-R system, I (Inducer) strains, but not R (Reactive) strains, carry active copies of the non-LTR retrotransposon I element. Dysgenesis arises when I or P strain males are mated, respectively, with R or M strain females lacking such elements. These females are unable to maternally provision matching piRNAs that target the activating element.

The P-M and I-R systems of dysgenesis represent cases in which the TE profile differs between strains with respect to only one family of inducing elements. In contrast, a third syndrome of hybrid dysgenesis in D. virilis represents a more complicated form of dysgenesis that appears to be caused by the mass action of multiple active elements abundant in one strain, but not another. Elements that likely contribute to this syndrome were first identified through direct analysis of induced lesions identified in the offspring of F1 progeny that escaped sterility from the dysgenic cross. The major driver of dysgenesis appears to be the Penelope element, the founding member of a clade of retroelements designated Penelope-like elements that are distinct from non-LTR and LTR retroelements (Evgen'ev and Arkhipova, 2005). Active copies of Penelope are abundant in the inducer strain (Strain 160) and only degenerate copies are present in the M/R-like reactive strain (Strain 9) (Lyozin et al., 2001). Furthermore, expression of the Penelope element is elevated in the ovaries and testes of F1 dysgenic progeny that have escaped ablation of the gonads (Evgenev et al., 1997). In addition to the Penelope element, three other elements (Helena, a non-LTR; Paris and Polyphemus, both DNA transposons (Blumenstiel, 2014; Vieira et al., 1998) are also more abundant in the 160 inducer strain, and these likely contribute to dysgenesis. A complex mode of hybrid dysgenesis, driven jointly by multiple elements, is supported by the fact that females of some

"neutral" strains—capable of preventing dysgenesis when crossed with inducer males but also incapable of induction (Rozhkov et al., 2013)–lack Penelope piRNAs in the their germline. If Penelope is the sole cause of paternal induction, it is difficult to explain how such strains could prevent induction when the female germline lacks Penelope piRNAs.

In light of this complexity, a second model for hybrid dysgenesis not directly driven by transposable elements in D. virilis has been proposed. A previous study showed that not only do inducer and reactive strains differ with respect to TE abundance, they also differ with respect to piRNA cluster activity (Rozhkov et al., 2010). In particular, small RNA sequencing in these two strains demonstrated that telomeric regions of the inducer strain exhibit uniquely strong piRNA cluster activity. Differences in telomeric cluster activity were proposed as potentially causative of dysgenesis. Here, we directly test this hypothesis by genetically assessing the contribution of inducer strain telomeres to hybrid sterility.

Because the 160 inducer strain and the 9 reactive strain of D. virilis differ with respect to multiple elements, the dysgenic syndrome in D. virilis may perhaps be more similar to that observed between species with respect to TE profiles, but with minimal divergence in protein coding function since it arises from an intraspecific cross. This syndrome may be considered an intermediate state between the P-M and I-R models and crosses between entirely different species that differ significantly at the genic level as well. Therefore, the dysgenic syndrome in D. virilis serves as a useful model for understanding the consequences of accumulating differences in TE profile between populations.

A fundamental question is how TE activation in dysgenic crosses influences the entire genomic TE landscape. Early studies of P element hybrid dysgenesis in D. melanogaster indicated downstream activation of additional TEs (Gerasimova et al., 1984), but this interpretation was soon called into doubt

(Eggleston et al., 1988). Nonetheless, the syndrome of hybrid dysgenesis in *D. virilis* provides strong support for cascading germline activation of TEs because multiple TEs were found to transpose in the germline of dysgenic progeny (Evgenev et al., 1997; Petrov et al., 1995). While different TE families may contribute to the initial induction of dysgenesis in D. virilis, germline co-mobilization has been demonstrated by the transposition of TEs that are evenly distributed between the two strains, in contrast to Penelope, Helena and Paris, which are more abundant in inducer strain 160 (Rozhkov et al., 2011; Scheinker et al., 1990; Vieira et al., 1998). Significantly, a recent study of the P element system indicates that the previous conclusion of no co-mobilization may have been premature (Khurana et al., 2011b). In the face of P element activation, DNA damage can perturb piRNA biogenesis and this defective piRNA biogenesis is presumed to drive the mobilization of additional TEs. Thus, global TE mobilization may also be observed in syndromes of hybrid dysgenesis that are driven by a single element. Whether a similar mechanism caused by DNA damage explains co-mobilization in the D. virilis system is unknown.

To fully understand the mechanisms underlying TE activation in dysgenic crosses, the developmental context must be considered since co-mobilization may be induced at any point in the developing or aging germline. A critical feature of the P-M system is that the germline crisis ameliorates with age. In particular, as flies age to 21 days, fertility is partially restored and piRNA levels that target the P element are restored to the same level as non-dysgenic reciprocal females. Thus, even though P element derived piRNAs are not maternally inherited, de novo piRNA production from paternally inherited P insertions is evident. This de novo piRNA production also coincides with restored silencing of P element mRNA. Rescue of germline crisis is also proposed to be enhanced by movement of other elements into piRNA clusters (Khurana et al., 2011b).

Here we use the unique system of hybrid dysgenesis in D. virilis to define the landscape of TE expression that coincides with the initial activation of multiple TE families. Because analysis of piRNA production and TE expression can be confounded in atrophied gonads, we focus solely on germline tissues that have escaped complete gonadal atrophy. Specifically, we examine piRNA and TE expression in 0–2 hour old embryos laid by F1 females from non-dysgenic crosses, and F1 escaper females of the dysgenic cross. This represents an endpoint of the dysgenic crisis and also provides insight into how the effects of hybrid dysgenesis in females that escape sterility can be passed on to further generations. In contrast to the P-M system, which may resolve within the germline as flies age (Khurana et al., 2011b), the effects of dysgenesis on TE expression in the D. virilis system persist through adulthood. Because this occurs within an intraspecific cross, increased levels of persisting TE expression are not explained by divergence in the piRNA machinery.

By comparing 0–2 hour old embryos laid by genetically identical females derived from dysgenic versus non-dysgenic crosses, we show that germline activation of TEs is driven by a multi-layered mechanism. Diverse elements are activated corresponding to TE copy number asymmetry between strains but there is also corresponding activation of some TEs that are evenly distributed between strains. This state of chronic increased TE expression is maintained as flies age, suggesting a different mechanism underlying co-mobilization compared to the P-M system. Interestingly, increased and persistent TE expression in the germline of females of the dysgenic cross coincides with a shift in piRNA pools. This shift in piRNA pools is associated with increased abundance of piRNAs that target genes outside of piRNA clusters, leading to significant effects on non-TE gene expression. Finally, differences in the TE profile between strains coincide with different modes of trans-generational gene regulation by genic piRNAs that arise from polymorphic TE insertions. Overall, this work identifies multiple modes by which differences in the TE landscape between strains can influence patterns of TE and gene expression across generations via piRNAs.

# MATERIALS AND METHODS

## Custom *D. virilis* TE library

Few annotated TEs are available for *D. virilis*. Therefore, we combined available annotated TE sequences with two computationally predicted libraries (generated with PILER (Smith et al., 2007) and REaS (Clark et al., 2007; Li et al., 2005) ( ftp://ftp.flybase.net/genomes/aaa/transposable_elements/) to generate a manually curated library. Several annotations (*Uvir*, *Helena*, *TART*, *Telemac*) were improved by manual curation. Portions of the PILER library were also manually curated. Additional sequence from the *Helena* element was obtained by interrogating a *de novo* assembly of the strain 160 genome. Redundancy was removed from this combined library first by removing repeats with significant blastn hits between and within the PILER and annotated library, with priority to annotated and longer sequences. With this filtered set, further redundancy was removed by blasting this library with and between the REaS library.

## Genome Sequencing and TE measurement from strain 9 and 160

Genome sequencing was performed on both reactive strains 9 (non-inducer) and strain 160 (inducer). As *D. virilis* has very high satellite content (more than 40%) (Bosco et al., 2007), we elected to used wandering third instar larvae for our DNA source as previously described (Blumenstiel et al., 2009). These tissues include polytene tissues that are underreplicated in heterochromatin, thereby reducing the number of satellite reads and enriching for euchromatic TE insertions that are expected to be the most active. Wandering 3rd larvae were collected from strain 9 and 160, rinsed with 50% bleach, and DNA was extracted. 100 bp paired end sequencing was performed on an Illumina GAII with 400 to 500 bp fragments. TE abundance estimation was performed with single ends that were trimmed by Sickle (https://github.com/najoshi/sickle), mapped to the TE library with BWA-MEM (LI, 2013) and normalized by read numbers mapping to the reference. Homogeneity within mapped reads was measured using

piledriver (github.com/arq5x/piledriver) and averaging the frequency of the major allele across all nucleotides within the mapping for each TE.

Estimation of zygotic effects of paternally inherited chromosomes

From the genome sequences of strain 9 and 160, restriction fragment length polymorphisms were identified that distinguish between the strains at two positions for each chromosome. F1 males were generated from a non-dysgenic cross and these males were crossed to strain 9. 96 F2 progeny were collected, (48 from each class, dysgenic or non-dysgenic) and genotyped for chromosomes inherited paternally with RFLPs. Log-odds ratios for the probability for being dysgenic with a given chromosome were estimated using a generalized linear model for logistic regression (binomial family with a logit link) in R. Some failed genotypes resulted in N = 92.

mRNA seq: Dysgenic and Non-dysgenic germline, Strains 9 and 160 germline, dysgenic and non-dysgenic soma

*Dysgenic vs. non-dysgenic germline.*

RNA for sequencing was collected from embryos laid by F1 mothers from the dysgenic and non-dysgenic directions of the cross. Ovaries were not selected because dysgenic ovaries are often atrophied. Therefore, the germline tissue represented in this experiment is derived from mothers that have escaped germline ablation. Paternal effects on embryos that might occur when dysgenic females are mated with sterile dysgenic brothers were minimized by equally mixing males from reciprocal directions of the cross and allocating them in mating cages between reciprocal F1 females. This also ensured improved egg laying from dysgenic females. Overall, we collected four pools of 0–2 hour old embryos, aggregated across several days, from large population cages containing many hundreds of adults grown up simultaneously from multiple bottles. Females were maintained as continuously laying with a constant supply of yeast and grape plates and eggs were collected after 0–2 hour durations and flash frozen in liquid nitrogen.

RNA was pooled from collections of 12 to 16 day old mothers and 19 to 21 day old mothers. From each age sample, two RNA-seq libraries were generated for single-end, 50 bp sequencing, for a total of four libraries per condition (dysgenic or non-dysgenic).

*160 vs. 9 germline*.

The same strategy was employed for collection of 0–2 hour old eggs from strain 9 and 160. Here, pools were also aggregated over multiple collections from large cages of pure strain 9 and 160 from multiple grow up bottles. Pool were aggregated over 7 to 15 days (young) and 15 to 25 days (old) and split for two RNA seq libraries per pool with reciprocal barcodes. Results presented are average RPKM per library and allele counts were pooled across all libraries.

*Dysgenic and Non-Dysgenic Soma*: RNA for analysis from somatic gene expression was obtained from 3 dysgenic and 3 non-dysgenic crosses. From each cross, 10 females were collected and aged 3 to 9 days. RNA was collected from these pools of 10 females, with abdomens removed.

Analysis of mRNA seq data

For estimates of gene expression level, RPKM was used. For statistical analysis, mapped count data was used. Reads were quality trimmed at the 3' end (up to 16 bp) and reads with 2 bp of quality less than 20 were excluded using the Galaxy server. TE RPKM estimates were obtained by directly mapping with BWA to the annotated TE library and normalizing with known TE length and the number of reads that mapped to the reference genome. mRNA RPKM estimates were obtained using the RNA-seq tool in CLC. Fold analysis was performed by calculating RPKM(+0.01) ratios. We used DEseq2 (Love et al., 2014) on RNAseq read count data to test for differential expression for both TEs and genes between dysgenic and non-dysgenic germ lines. Analysis of TEs and genes was performed separately. The model employed tested for treatment effects, age effects, and age x treatment interaction effects. While the same model was fit for genes and TEs, these models were run independently to account for possible differences

in the mean/variance relationship between these groups. Genes with fewer than 40 total reads mapped across all samples were removed from the analyses prior to calling differential expression in order to reduce the loss of power due to multiple testing. Statistical significance was assessed using FDR, focusing on treatment effect only. GO analysis was performed with GOrilla (Eden et al., 2009) using *D. melanogaster* orthologs genes sorted by FDR value for the test of treatment effect.

Additional germline mRNA seq was performed using the same protocol for pure strain 9 and strain 160 (RPKMs averaged for cluster analysis across 2 libraries each for 7–15 day old females and 15–25 day old females). Allele counts were determined by direct counting within the RNA-seq mappings (summed across all library mappings) for a SNP known to distinguish the two strains within the transcripts for *cdi* and *oysgedart*. Somatic mRNA seq analysis was likewise performed (3 libraries per condition, each library from 10 pooled female carcasses, RPKMs averaged across libraries and allele counts estimated as before).

Small RNA sequencing

All small RNA was size selected from 15% acrylamide, cut between an 18 bp oligo and the 30nt rRNA. Small RNA sequencing for dysgenic and non-dysgenic germline material was performed on embryos laid by the same mothers as for RNA seq, but at 15–16 days old, according to (Vigneault et al., 2012). Small RNAs from strain 9 and strain 160 pooled ovaries were sequenced according to (Li et al., 2009) with the oxidation reaction. Small RNAs from ovaries from individual F3 females was performed by Fasteris.

Small RNA Analysis

Reads were trimmed by removing adapters and filtered by size as piRNA (23-30nt) in CLC Genomics Workbench 7.0. Reads were then filtered by mapping to tRNA, ncRNA, miscellaneous RNA, and miRNA (including pre-miRNA) libraries from the *D. virilis* reference genome. The filtered 23–30 nt small RNA reads were mapped to our curated TE library with BWA.aln (Li and Durbin, 2010), using the default

parameters. Reads were normalized by non-unique mappers to the *D. virilis* reference genome using BWA.aln defaults. Calculations for ping-pong percent (Brennecke et al., 2008) and density of piRNA pairs were done with the R package viRome (http://www.ark-genomics.org/bioinformatics/virome), with some modifications. For genic small RNA analysis, reads were mapped uniquely with BWA.aln to the *D. virilis* reference genome, using default parameters. Reads were normalized by non-unique mappers to the genome. BEDTools intersect (Quinlan and Hall) was utilized to count piRNA hits on genes and CDS sequences. Fastq reads specific to genic piRNA hits were extracted using Enve-omics (https://github.com/lmrodriguezr/enveomics) and FASTX-Toolkit (hannonlab.cshl.edu/fastx_toolkit/index.html) was used to process and analyze nucleotide bias for genic piRNA subsets. We did not perform DEseq2 analysis because estimation of the dispersion parameter with DEseq2 is unlikely to be robust with about 200 TEs, compared to standard mRNA-seq analysis that estimates the dispersion parameter using thousands of genes (Love et al., 2014). Percent ping-pong was defined as the percent of 23 to 30 nt mapping reads that had a corresponding read on the opposite strand with a 10 bp 5'-5' overlap. We also measured piRNA biogenesis by determining ping-pong pair density. This measure was obtained by counting all non-redundant ping-pong pairs (counting each read only once) per kb.

Genetic analysis of genomic regions from strain 160 that maternally protect against dysgenesis

To identify regions of the 160 genome that protect against dysgenesis when present in females, an F3 mapping/QTL experiment was performed. F3 females were generated by crossing 160 females to strain 9 males (a non-dysgenic cross), followed by two rounds of backcrossing to strain 9 males. This resulted in the production of F3 mothers for which strain 160 was the great-grandmother. All but the final cross was performed *en masse*. Dysgenic crosses were performed with >160 single 4 to 5 day old tester F3 females mated with three 4 to 5 day old strain 160 males. Adults were transferred to new vials daily and dysgenesis was estimated by counting the number of dysgenic testes in progeny over all testes counted (2 per male) across three broods. Females were then collected, ovaries removed (for small RNA sequencing

by Fasteris, top protectors only) and carcasses retained for genomic DNA extraction. Genotyping was performed using the TaqMan Open Array platform on all females, with the exception of the top six females that had the strongest ability to protect against dysgenesis. SNPs distinguishing 160 and 9 chromosomes were chosen in pairs for redundancy, one pair at each telomere and pericentric region, as well as one or two euchromatic SNPs. Care was taken to avoid repeat sequences by screening with Blast against the reference and also using RepeatMasker. F3 Females were then genotyped for 160/9 heterozygosity, alongside pure strain 9 and 160 controls, by National Jewish Health. Single marker regression was carried out with RQTL after dropping individuals with missing genotype data from the analysis. 5000 permutations were done to find the significance threshold at alpha = 0.05. For the top 6 protectors, whole genome sequencing was performed (100 bp, paired-end) using the Nextera library prep protocol.

Genotyping by whole genome sequencing

Reference genome scaffolds from *D. virilis* were concatenated according to their supported positions and orientations on known Muller elements, with a large scaffold arbitrarily generated by concatenating scaffolds from unknown positions. Using this new "assembly" we mapped all strain 9 and strain 160 reads and generated two consensus genomes. A pseudo-diploid heterozygous genome was then assembled by placing these scaffolds into one file.

Paired-end reads from the top six protectors were mapped to the hybrid genome, using BWA's default parameters, with the goal of inferring spans of heterozygosity for strain 160 by identifying reads that map uniquely, under high stringency, to the strain 160 haploid reference. The mapping output was piped into SAMtools for filtering by quality score (-q 42) and post-alignment processing (SAM to BAM conversion and indexing). A relatively high cutoff quality score was used in order to remove reads that could have mapped promiscuously. We were able to remove all reads that mapped to more than one locus/allele leaving us with reads that are specific to either strain 9 or 160. Spans of heterozygosity for strain 160

were visualized with a sliding window for read density along the 160 chromosomes within the psuedo-diploid reference genome.

Chromatin immunoprecipitation sequencing

Chromatin isolation and immunoprecipitation was performed using 200 pairs of ovaries from strain 9 and strain 160 according to the protocol described in Du and Elgin (Gu and Elgin, 2013). Immunoprecipitation (IP) was performed using anti-H3K9me2 (Abcam 1220) at 1:100. Input and IP DNA was used for SR50-bp Illumina sequencing. H3K9-me2 enrichment was estimated using the IP:Input ratio of reads (with duplicates removed) uniquely mapping to the *cdi* or *oysgedart* locus (normalized to the total number of reads mapping in the library).

## RESULTS

Genome wide asymmetry in TE abundance in a dysgenic cross of *D. virilis*

Previous studies identified *Penelope* to be the primary driver of dysgenesis in the *D. virilis* system because multiple active copies reside in the inducer strain 160, but only degenerate copies reside in the reactive strain 9. *in situ* hybridization has identified more than 45 euchromatic *Penelope* insertions in strain 160 and none in strain 9 (Vieira et al., 1998). In addition, the *Helena* elements (euchromatic insertions Strain 160: 18; Strain 9: 0) and *Paris* elements (euchromatic insertions Strain 160: 26; Strain 9: 0) were shown to be more abundant in inducer strain 160 (Vieira et al., 1998). Recently a third element, *Polyphemus*, was identified as more abundant in strain 160. Using genome sequence reads from strains 160 and 9, mapped to a *D. virilis* TE/repeat library (Supp. Data 1.1), we identified additional factors more abundant in strain 160 and potentially contributing to dysgenesis. Consistent with previous results, *Penelope* and *Polyphemus* showed the largest excess in strain 160, validating this approach. Using a 3-fold cutoff as a threshold, we further validated our detection methods by confirming that Helena and Paris copy numbers are enriched in strain 160 (Vieira et al., 1998). Overall, we identified eleven elements

enriched in strain 160 and three elements enriched in strain 9. Of the eleven enriched in strain 160, two repeat sequences (258 and 1069; Fig 1.1A) show no apparent evidence of TE related coding capacity (Supp. Data 1.2). Likewise, the three repeat sequences enriched in strain 9 show no apparent evidence of TE related coding capacity. In addition to the four TEs known to be overrepresented in copy number in strain 160, and the two putatively non-TE repeat sequences, we identify five additional elements enriched in strain 160 (Fig 1.1A). These are candidates for contributing to the dysgenic syndrome. It is important to note that in this comparison there is a form of ascertainment bias. Because strain 160 is more closely related to the reference strain, repeats entirely absent from the reference strain (but possibly present in strain 9) will be excluded.

Coincident with the excess of multiple elements in strain 160, we found that the telomeric TART elements exhibit higher mapping abundance in strain 160, albeit below the 3-fold enrichment threshold (Fig 1.1A). TART elements have functioned as telomeres for millions of years in *Drosophila* (Casacuberta and Pardue, 2003), and our result demonstrates a strain-specific increase in bulk abundance of this long term resident. Telomeric TE content is under piRNA control (Khurana et al., 2010; Savitsky et al., 2006; Shpiz et al., 2011), and previous work has shown increased piRNA cluster activity in the telomeric regions of strain 160 compared to strain 9 (Rozhkov et al., 2010). Since reduced piRNA function can lead to increased telomeric TE activity, our observed TART excess in strain 160 may be a readout of compromised piRNA function in the inducer strain that is either a cause or consequence of TE excess. For these reasons, it was critical to test for a role of telomeric regions in the induction of dysgenesis. Overall, we found that diverse TE families are in excess in strain 160. This finding is consistent with the hypothesis that the invasion of the *Penelope* element itself into the reactive strain contributes to genome instability, possibly through the co-mobilization of other TEs within the strain (Evgen'ev, 2013).

TE age analysis identifies different modes for TE asymmetry between strains

Divergence in TE abundance between strains can result from different processes. For example, long-resident TEs may be in excess in one strain due to strain- or population-specific recent re-activation. By contrast, entirely new TEs may have invaded a species and have yet to spread equally throughout the genomes of the individuals within the population. The *P* element invasion in *D. melanogaster* is an example of the latter process (Daniels et al., 1990). It has recently invaded and is only present in natural populations/strains collected within the last seven decades.

To distinguish among alternative processes contributing to asymmetry in TE abundance between strains 160 and 9, we performed an age analysis of TE families using high sequence homogeneity within a TE family as an indicator of recent activation or invasion. A phylogenetic approach using full-length fragments is ideal for this purpose, but full-length TE assemblies are not available with short read sequencing technology. Therefore, we estimated relative TE family age by examining the sequence heterogeneity within mapping reads (Fig 1.1B) by considering the average frequency of the most common nucleotide variant, across all nucleotide variants within the mapping.

A young element that has recently invaded will show high similarity (higher homogeneity) among copies, nearing 1 for an average frequency of the major nucleotide variant. Older elements, with patterns of activation that occurred in the more distant past, accumulate sequence-level differences among insertions, which contribute to lower homogeneity. This accumulation of differences among multiple copies is evident by lower nucleotide frequencies of the most common variant. For a recent TE *re-activation* in only one strain, we expect higher sequence homogeneity within that strain but higher heterogeneity in the other strain, arising from degraded copies. For an element that has recently invaded a species and is

present in both strains, but achieves greater copy number in one strain, we expect a similar level of sequence homogeneity in both strains. Finally, for an element that has recently invaded a species, but is entirely absent in one strain (similar to the *P* element in *D. melanogaster*), we expect higher homogeneity of reads in the carrying strain, but much higher heterogeneity in the naive strain, arising from sequence heterogeneity within the marginally mapping reads.

Our age analysis of TE families revealed two classes of TEs that are enriched in inducer strain 160 (Fig 1.1B). Consistent with previous analyses (Blumenstiel, 2014; Lyozin et al., 2001; Morales-Hojas et al., 2006), one class includes Penelope, Polyphemus, Helena and Paris. In this class, we now also include Skippy and telomeric TART elements. Penelope and Polyphemus showed much higher homogeneity among copies in strain 160 compared to strain 9. We found the same pattern, albeit to a lesser extent, in Helena, Paris, Skippy and the telomeric TART elements. This pattern is highly consistent with recent activation of these elements from long-term resident status. A second class of elements exhibited a different pattern. Slicemaster, Uvir, 258, and 734 showed a pattern of nucleotide homogeneity consistent with similar age in both strains, while element 1069 appears to be slightly older in strain 160. In the case of elements like Slicemaster, which are very young (>99% nucleotide similarity), this can be explained by recent invasion of both strains but excess movement in strain 160, rather than re-activation in one strain from long-time resident status.

Increased germline TE expression in dysgenic females persists through adulthood

To determine the relationship between TE excess in strain 160 and TE expression in dysgenic progeny, we performed mRNA-seq from pooled 0–2 hour old embryos laid by F1 females of both the dysgenic (9 females X 160 males) and non-dysgenic (160 females X 9 males) directions of the cross. Notably, we did not measure TE expression in ovaries from F1 females because

dysgenic ovaries are atrophied and expression analysis from these tissues is confounded by altered ratios of somatic and germline tissue. 0–2 hour old embryos laid by F1 mothers represent a sample of pure germline tissue, albeit lacking piRNAs and mRNAs residing solely in nurse cells that are not loaded into the egg (Chambeyron et al., 2008). This is because zygotic transcription in *D. virilis*, as measured with the early, zygotic *fushi-tarazu* (*ftz*) gene, begins after 2 hours (Blumenstiel and Hartl, 2005). Confirmation that embryos in these samples were collected prior to the onset of zygotic transcription was obtained by examining *ftz* expression in our RNA-seq dataset.

Full penetrance of dysgenesis, evidenced by fully ` gonads, is observed in approximately 50% of male and female progeny from 9 female X 160 male crosses. Therefore, embryos analyzed by mRNA-seq from the dysgenic cross were those laid by mothers that escaped full sterility. In contrast to other systems, hatch rates are normal in eggs laid by escaper females. For clarity, these tissues will be referred to as dysgenic, even though these tissues escaped complete atrophy. Sexual maturity in *D. virilis* occurs at about 5 days. To determine the dynamics of TE expression as flies aged, we analyzed mRNA-seq data from 0–2 hour old embryos laid by F1 mothers 12–16 days old, and 19–21 days old.

First, ignoring age effects, our mRNA-seq results indicated different modes of increased TE expression in the dysgenic germline (Fig 1.2A). Overall, we find 15 TEs that were differently expressed between dysgenic and non-dysgenic germlines (FDR<0.05). Of these, nine are

significantly up in dysgenic and six are significantly up in reciprocal females, but the magnitude of increased expression in the dysgenic germline is significantly greater (Mann-Whitney U test comparing magnitudes of expression difference among differentially expressed TEs: $U = 0$, $p<0.05$). This is shown by the fact that all nine of the TEs with significantly increased expression in the dysgenic germline show more than a two-fold increase, but none of the six that are higher in expression in non-dysgenic progeny show this level of expression difference.

TEs with excess abundance in the inducer strain showed increased expression in the dysgenic germline. There were high magnitude differences for some elements (*Skippy* and *Helena*; Fig 1.2A) but for others, the observed differences in expression between dysgenic and non-dysgenic germlines were more modest (*Paris*, *Polyphemus* and *TART;* Fig 1.2A) Importantly, while an assemblage of TEs are more highly expressed in the dysgenic germline, many TEs are expressed at equal levels in dysgenic and non-dysgenic crosses. This is also seen for TE expression in the *P-M* system (Khurana et al., 2011b).

Overall, there is not a general rule that *all* elements more highly expressed in dysgenesis are higher in copy number in the inducer strain (Fig 1.2B). Ten of eleven elements that are more abundant in the inducer strain were expressed at higher levels in dysgenic progeny. However, five of the nine elements with significantly higher expression in the dysgenic germline are in slightly higher copy number in strain 9 (Fig 1.2B). This finding is consistent with the observed co-mobilization that was originally identified through genetic approaches—TEs with similar copy number between strains can be co-mobilized by dysgenesis. For example, the *Ulysses* element is in similar copy number between strains, mobilizes in dysgenesis and shows about a two-fold increase in expression in the dysgenic germline, though this expression difference was not significant. In contrast, the *Telemac* element, at near-equal abundance

between strains 160 and 9, was one of the five expressed at slightly higher levels in the non-dysgenic germline.

Accounting for age of F1 mothers, we found that the observed patterns of increased TE expression in the dysgenic germline were maintained through adulthood (Fig 1.2C and 1.2D). For example, the *Helena* element, which is more abundant in the inducer strain, shows an approximate 30-fold higher expression in embryos laid by 12–16 day old mothers. And in embryos laid by 19–21 year old mothers, *Helena* maintains an approximately 15-fold higher expression. This persisting level of increased TE expression in the dysgenic germline stands in contrast to the *P-M* system, where by 21 days, *P* element expression equalizes between dysgenic and non-dysgenic females (Khurana et al., 2011b).

Maternal piRNA and siRNA deposition is an inconsistent predictor of TE expression in dysgenesis

To determine how maternal inheritance of piRNAs (defined as small RNAs, 23–30 nt, filtered against known non-piRNA classes) and also siRNAs (defined as small RNAs, 21 nt, filtered against known non-siRNA classes) might explain increased and persistent TE expression in the dysgenic germline we sequenced 18 to 30 nt RNAs from 0–2 hour old embryos laid by strain 9 and strain 160 mothers, and by F1 females from reciprocal crosses between the two strains. For F1 germline small RNAs, we collected embryos from the same pool of mothers used for mRNA-seq, but at intermediate maternal age (15–16 days old). This allowed us to determine whether the persistent differences in TE expression in the F1 germline of the dysgenic cross could be explained by a persistent defect in piRNA biogenesis.

A large number of TEs, including many with greater copy number in Strain 160, showed higher levels of maternally provisioned piRNA in strain 160 compared to strain 9 (Fig 1.3A). However, many TEs without large differences in copy number between strains also showed a more than 10-fold excess of maternally provisioned piRNA in the strain 160 genetic background (Fig 1.3A). Strikingly, despite the asymmetry in maternal provisioning observed in the strain 160 compared to the strain 9 background, piRNA differences are much less pronounced in the germlines of F1 individuals derived from the reciprocal 160 x 9 crosses (Fig 1.3B, non-overlapping 95% C.I.s for Pearson's correlation coefficient comparing to Fig 1.3A). This result is not consistent with global persistence of the maternally provisioned piRNA profile across generations. A significant exception to this is the *Helena* element, which maintains a higher level of piRNA in the non-dysgenic germline.

Fig 1.3C demonstrates the degree to which asymmetry in maternal provisioning predicts TE expression in reciprocal dysgenic and non-dysgenic progeny in the next generation. Many of the elements that are more abundant in strain 160 have greater piRNA abundance in the 160 female germline and also are expressed at higher levels in the germline of the dysgenic cross. This drives a significant positive correlation between the log 2 of the ratio of piRNA abundance between strains and relative TE expression levels. However, with Spearman's rho less than 0.2, this is not a strong relationship. Therefore, maternal provisioning of piRNA is only a modest predictor of TE expression in hybrid dysgenesis. For example, element 750 shows no difference in piRNA abundance between parental strains but is more highly expressed in the dysgenic germline. Additionally, a reduced level of TE piRNA abundance persisting in the dysgenic germline is positively associated with increased expression for many (Fig 1.3D, for example *Helena*), but not all elements. In contrast to *Helena*, *Skippy* and *Slicemaster* are both more highly expressed in the dysgenic germline even though both show *higher* levels of piRNA abundance in the dysgenic germline. Therefore, multiple mechanisms appear to explain increased TE expression that persists in the dysgenic germline.

We also examined the role that siRNAs had in predicting differences in TE expression between reciprocal progeny. Similar to the effect of piRNA provisioning, we found that differences in the maternal load of TE siRNAs were predictive of differential TE expression between dysgenic and non-dysgenic progeny (Fig 1.3E). Since differences in maternal piRNA and siRNA load are strongly correlated between the strains (Fig. 1.8), this is not surprising. Interestingly, in contrast to piRNA levels, zygotic siRNA levels were not predictive of differences in TE expression between reciprocal progeny (Fig 1.3F). Overall, maternal piRNA levels are predictive of F1 piRNA levels but maternal siRNA levels are not predictive of F1 siRNA levels (Fig 1.8). Interestingly, maternal piRNA levels are in fact predictive of F1 siRNA levels (Fig 1.8). These results suggest that while siRNA levels are poor predictors of TE expression differences, their biogenesis in F1 progeny may be coupled to piRNA abundance.

Increased and persisting dysgenic TE expression is not associated with a collapse of global piRNA biogenesis

Raw abundance measures of piRNAs ignore critical aspects of their biogenesis and recent studies have demonstrated that globally reduced signatures of robust piRNA biogenesis likely contribute to the mobilization of diverse TEs (Kelleher et al., 2012; Khurana et al., 2011b). In contrast to interspecific crosses that show near complete collapse of the 23–30 nt small RNA pool, we found no evidence that piRNA biogenesis is skewed away from the 23–30 nt expectation based on the size distribution of small RNA reads (Fig 1.4A). For each TE, we estimated the percent ping-pong (Brennecke et al., 2008) as well as the normalized density of ping-pong pairs in the dysgenic and non-dysgenic germline. When we compared metrics directly (first column of heatmaps, Fig 1.4B) we found little evidence that piRNA biogenesis is grossly perturbed in the dysgenic cross, though a more sensitive comparison using

normalized Z-scores indicated a modest reduction in piRNA abundance and density of ping-pong pairs (Fig 1.4B). This is observed in the Z-score heat maps (Fig 1.4B, second column of heatmaps) for abundance and ping-pong pair density. Both showed an excess of negative Z-scores for dysgenesis ($p<0.0001$, Wilcoxon Signed-Rank Test). Importantly, both ping-pong abundance and ping-pong pair density are normalized, proportional measures of abundance that are likely influenced by increases in the abundance of non-TE, genic piRNAs in the same library (see below).

In contrast to the piRNA abundance measures, we found no significant evidence for global ping-pong biogenesis disruption as measured by percent ping-pong. Note, for example, that many row Z-scores for percent ping-pong showed weaker Z-scores for non-dysgenic compared to dysgenic piRNA (compare upper and lower portions of the percent ping-pong Z-score heatmap, Fig 1.4B). A Wilcoxon Signed-Rank test also found no significant difference ($p = 0.06$) in percent ping-pong Z-score between dysgenesis and non-dysgenesis piRNA. If there is any tendency for perturbed piRNA biogenesis, it is not uniform across elements.

Overall, there is a significant relationship between the difference in the normalized density of ping-pong pairs between parents and the difference between dysgenic and non-dysgenic germlines (Fig 1.9), demonstrating a role for maternal provisioning in establishing piRNA biogenesis in the next generation. In light of this, we found that transposable elements more highly expressed in dysgenesis (red bars, Fig 1.4B) are in excess among elements with a reduced percent ping-pong signature. In particular, we found a significant correlation between difference in percent ping-pong Z-score and fold TE expression between dysgenic and non-dysgenic germlines ($p = 0.044$, Fig 1.4C). This trend is driven by the top eight elements that show higher expression in dysgenesis and all have lower percent ping-pong Z-scores in dysgenesis

(Fig 1.4C). These results do not support a model of global disruption in piRNA biogenesis maintained in adult flies. Rather, they support a model in which the persistence of higher expression for some TEs is driven by idiosyncratic defects in the restoration of piRNA biogenesis in aged females that occur on a TE-by-TE basis. Strikingly, for several TEs, signatures of piRNA biogenesis appear largely restored, despite increased expression in the dysgenic germline. For TEs that are increased in expression in the dysgenic germline, there appear to be multiple causes (Table 1.1), including, but not limited to, reduction in ping-pong pairs, suggesting multiple modes of TE derepression in hybrid dysgenesis.

To distinguish among the drivers of TE expression differences between the dysgenic and non-dysgenic germline, we used the *leaps* package in R to identify the single variable that was the best predictor within a multiple regression framework. We considered copy number difference between the two strains, differences in piRNA and siRNA abundance in parents and progeny, and differences in percent ping-pong. Differences in maternal piRNA abundance was selected as the single best predictor (*R-squared*: 0.054, $p = 0.0005$). Using an alternate approach to model selection using the AIC (step AIC from the *MASS* package), the best fit model included only two variables: maternal siRNA abundance and piRNA abundance in the progeny (Multiple *R-squared*: 0.0674, $p = 0.0005$). Both of these variables are correlated with maternal piRNA abundance. Nonetheless, the selection of these two variables without including the maternal piRNA variable suggests that the influence of maternal piRNA abundance in the single variable model is jointly mediated by maternal siRNA pools and zygotic piRNA pools. It should be noted, however, the amount of variance explained by these models is low.

Many genes are differentially expressed in dysgenic and non-dysgenic germlines

We found that 267 genes are significantly upregulated, and 300 significantly downregulated in the dysgenic compared to non-dysgenic germline (FDR = 0.05). We performed GO analysis using the

GOrilla program that identifies enriched ontologies in sorted lists (Eden et al., 2009). Sorting by FDR value, we identified several interesting terms for genes down-regulated in dysgenesis. These include: reproductive process (FDR = 0.00015), chromatin organization (FDR = 0.0015) and gene silencing (FDR = 0.037) (Supp. Table 1.1). Since chromatin marks and gene silencing are important for TE control, down-regulation of these genes may play a role in increased TE expression in the dysgenic germline. However, a large number of GO terms were identified in the set of down-regulated genes and GO analysis can be difficult to interpret (Pavlidis et al., 2012).

Increased genic piRNA production in the dysgenic germline

Recent studies have demonstrated that in addition to TEs, genes may also be the target of piRNA silencing. Genic targeting by piRNAs can arise from neighboring TE insertions that drag flanking sequences into piRNA biogenesis and gene silencing in *cis* (Gu and Elgin, 2013; Haynes et al., 2006; Olovnikov et al., 2013; Sentmanat and Elgin, 2012; Shpiz et al., 2014; Sienski et al., 2012). Previous work in the *D. virilis* system of dysgenesis identified a piRNA cluster overlapping the *center divider* (*cdi)*gene (Dvir\GJ14359) (Rozhkov et al., 2010). Global gene expression might be modulated if genic piRNA silencing were either attenuated or enhanced during dysgenesis. Therefore, we examined how the global landscape of genic piRNAs was influenced by the activation of diverse TEs in dysgenesis. For this analysis, we excluded genes that lacked hits to CDS regions. We also excluded genes that lacked orthologs in *D. melanogaster* to exclude TEs mis-annotated as genes.

In the dysgenic germline we found significant enrichment of piRNAs derived from genes. We identified 105 genes that had at least 5 piRNA per million mapping to genic CDS regions in either the parental strains or reciprocal F1 progeny. For these 105 genes, there was a significant excess of piRNAs in the

dysgenic germline (Fig 1.5A; p-value < 0.0001, Wilcoxon-signed rank test). Genic piRNA production, relative to parental strains, is also higher in non-dysgenic progeny, indicating that this may arise from crosses between strains with divergent piRNA profiles. For some genes, the genic piRNAs were predominantly anti-sense, but the majority of genes were associated with primarily sense strand piRNAs (Fig 1.5B). Comparing genic piRNA density across introns and exons, we found that piRNAs from these genes are enriched on exons (Paired T-test across genes contrasting intronic and exonic density: Library 1: $p = 0.0046$, Library 2: $p = 0.0024$). This suggests that genic piRNA processing may occur in the cytoplasm. We examined genic piRNAs for piRNA biogenesis signatures: first position U bias and 10th position A bias. Results indicate that the genic piRNAs are primary piRNAs (Supp. Table 1.2). From the entire set of 105 genes, focusing on 80 genes that primarily produce sense piRNAs, we found a signature of primary piRNA biogenesis and a very weak signature of secondary piRNA biogenesis (Library 1: U first position: 0.33, background U: 0.21, A 10th position: 0.25, background A: 0.22; Library 2: U first position: 0.33, background U: 0.21, A 10th position: 0.24, background A: 0.23).

In addition, we looked for TE insertions within 2 kb upstream and downstream of these piRNA targeted genes. For the several genes that produced both sense and antisense piRNA, we often found evidence of either a TE in the reference genome or an indication of a nearby insertion in genomic mappings of one of the two strains. Not only did these genes have sense and antisense piRNA, but unique piRNA also mapped to intergenic regions around these genes, indicative of cluster spreading in *cis*. In contrast, genes with sense piRNA mapping only to exons usually showed no evidence of proximal TE insertions and piRNA did not map to intergenic regions. This supports the idea that these sense and exon-only piRNA are generated from processing of genic mRNA in the cytoplasm.

The 105 genes were also more highly expressed above the genome-wide background (Fig 1.5C, Wilcoxon rank sum test, $p < 0.001$). Among the genes with the greatest excess of piRNA abundance in the dysgenic germline, the primary piRNA biogenesis signature was strongest for the genes with expression level higher than 1000 RPKM (Supp. Table 1.2). This indicates that these small RNAs are not simply degradation products of highly expressed genes. Importantly, the production of these genic sense piRNAs has an apparent effect on gene expression. Of the 105 genes, 89 were identified to show the highest piRNA abundance in the dysgenic germline. These 89 genes were more lowly expressed in the dysgenic germline compared to the non-dysgenic germline (Fig 1.5D, Wilcoxon rank sum test, $p < 0.001$). Among this set of 89 genes, there was a significant excess of genes more lowly expressed in dysgenesis compared to more highly expressed ($p = 0.03$, Sign Test). We attribute this to primary piRNA biogenesis from the sense strand since lower expression is observed for genes that are primarily targets of sense piRNA biogenesis (Fig. 1.10).

GO enrichment analysis using GOrilla indicated that these 105 genes are highly enriched for ribosomal proteins (FDR p-value = 1.9E-13, 19-fold enrichment; Supp. Table 1.3. Seventeen of these eighteen ribosomal genes produce more piRNA in the dysgenic samples. These ribosomal piRNA targets are highly expressed, show a strong signature of primary biogenesis (Supp. Table 1.2) and are not among the group that show differential mRNA expression between dysgenic and non-dysgenic samples. Notably, there is a gene with histone acetyltransferase activity, *nejire* (Dmel\nej), which has strong effects on TE expression upon knockdown in *D. melanogaster* (Czech et al., 2013) and is also orthologous to a piRNA target in our list (Dvir\GJ19060). *Nejire* produces more piRNA in the dysgenic samples than non-dysgenic, and has significantly lower mRNA levels in dysgenic versus non-dysgenic samples (p-adjusted value < 0.05).

Pericentric regions rather than telomeres influence dysgenesis

While the induction of hybrid dysgenesis in *D. virilis* has been attributed to TEs enriched in the inducer strain 160, it has been proposed that differences in telomeric cluster activity in the inducer strain may also contribute (Rozhkov et al., 2010). Therefore, we employed a genetic approach to determine whether identified telomeric (Brennecke et al., 2007) clusters were causal of dysgenesis or, perhaps, simply consequences of excess telomeric TART activity identified in the inducer strain. In the previous study (Rozhkov et al., 2010), two telomeric clusters specific to strain 160 were identified, one residing at the tip of the second chromosome and another residing at the tip of the sixth chromosome. We therefore tested whether chromosomes carrying telomeres from the inducer strain 160 were highly inductive, when transmitted paternally, or highly protective, when present in the female germline of the dysgenic cross.

We found that induction of dysgenesis is distributed across all chromosomes, with the exception of the dot sixth chromosome (Fig 1.6A). Therefore, strain 160 telomeres of the second and sixth chromosomes do not contribute uniquely to induction of dysgenesis. Using QTL analysis (Fig 1.6B) with special attention to telomeric and pericentric regions (motivated by the fact that these genomic compartments often contain TE-rich piRNA clusters (Brennecke et al., 2007; Grentzinger et al., 2012), we identified three genomic regions for which strain 160 variants at these positions significantly protected against F1 sterility when present in the mother (*i.e.*, dysgenesis; Fig 1.6C). The genomic region with the most significant effect corresponded to the pericentric region of chromosome 5. The pericentric region of the X chromosome also explained a significant proportion of variation in protective ability, followed by a euchromatic region in the proximal arm region of chromosome 4. We tested for interactions between these loci and saw no evidence for synergism (p-value for all interactions >0.2). Our previous work also found chromosome 5 to be the most protective, followed by the X and then chromosome 4 (Blumenstiel and Hartl, 2005). Together with the results from this study, there is strong genetic evidence that

pericentric, cluster-derived piRNAs play a role in the protection against dysgenesis in *D. virilis*. By contrast, variation in telomeric repeat abundance between strains does not explain variation in protection against dysgenesis. Thus, telomeric piRNA clusters and amplified TART elements are likely a result of TE destabilization in the inducer strain rather than a driver.

No single region or piRNA pool is protective against dysgenesis

To determine precise regions of the 160 genome critical for protection in a dysgenic cross, we performed whole genome sequencing of the six most protective F1 females for which DNA was available (Fig 1.6B and 1.6D). We found that among the most protective F1 females there was no single genomic region consistently derived from strain 160. In addition, across all six mothers, we identified at least one mother homozygous for strain 9 at each position of the genome (excluding unassembled regions). Therefore, no single genomic region appears critical for protection against dysgenesis.

To determine whether piRNA from any particular TE enriched in 160 was dispensable for protection, we sequenced small RNAs from the individual pairs of ovaries of the six most protective F1 females. We first reasoned that the only TEs that are candidate inducers of dysgenesis are those for which piRNA abundance is greater in strain 160 than in strain 9. Among the 221 repeats within the TE library (Supp. Data 1.1), there are 141 that meet this qualification. We further reasoned that if any female with full repressive ability had piRNA abundances for a TE that were similar to strain 9, we could rule out that TE as a driver of dysgenesis. Of the remaining 141 candidate TEs, 88 TEs have at least one protective female that has normalized piRNA abundance derived from that TE less than or equal to strain 9. Thus, we were left with 53 candidate contributing repeats. Using these criteria, we were unable to eliminate any of the

elements 3-fold enriched from strain 160. Therefore, we are unable to conclude that maternal loading of piRNA corresponding to a single TE can mediate protection against the induction of dysgenesis.

piRNA sequence data from the six most protective F1 females suggests a minimal role for the TART element piRNAs as mediators of maternal protection. This is because the vast majority of TART piRNAs in the protective mothers are derived from the tip of the X chromosome from 160 (Fig 1.6D) and three of the six females that are most protective against dysgenesis lack the X 160 telomere allele. Notably, the telomeric region of the X has special silencing properties in *D. melanogaster* that may explain the excess of TART piRNAs derived from this one genomic region (Marin et al., 2000; Niemi et al., 2004; Ronsseray et al., 2003; Ronsseray et al., 1998; Simmons et al., 2012), but in this system this region plays no role in protection against dysgenesis.

Using small RNA sequence data from the six most protective F1 females we characterized the genetic basis for *Penelope* endo-siRNA production that we had previously shown to be derived from the X-chromosome (Blumenstiel and Hartl, 2005). We confirmed that endo-siRNAs are not abundant when the X-chromosome from strain 160 is lacking (See individual 46; Fig 1.6D). These data also demonstrated that *Penelope* endo-siRNAs are contributed by several loci on the X chromosome (compare individual 98 to 43 and individual 43 to 50; Fig 1.6D). However, as one individual (46) lacking the X chromosome is protective against dysgenesis, we can confirm that *Penelope* endo-siRNAs play a minimal role in mediating maternal protection.

Genic piRNAs influence gene expression across generations in multiple ways

Together, the above results support a model of hybrid dysgenesis driven by the mass action of multiple transposable elements. However, divergence in the TE repertoire between strains can also lead to divergence in the genic piRNA profile. This is because genic piRNAs can be produced when TE inserts flank genes (Shpiz et al., 2014).

Rozhkov *et al*. showed that strain 160 possesses a number of piRNA clusters absent in strain 9 (Rozhkov et al., 2010). The most well-characterized cluster is a telomeric cluster at the tip of chromosome 2 encompassing the gene *center divider (cdi)*. This dual strand piRNA cluster was found to be present in strain 160, but absent in strain 9. We observed the same pattern in our divergent laboratory stocks (Fig 1.7). For a second cluster identified by Rozhkov *et al*., near the telomere of the 6th chromosome in strain 160, but absent in strain 9, we found the opposite pattern in our strains. This cluster in our strain 160 had 318 unique mappers per million mapped reads, whereas our strain 9 had 3,836 unique mappers per million mapped. The most parsimonious explanation for this is that this cluster was originally present in both lines, but independently lost in our strain 160 and Rozhkov *et al.'s* strain 9. Thus, some piRNA clusters may be prone to losing their activity over time.

In addition to *cdi*, we identified the *oysgedart* gene (Dvir\GJ17620) in strain 9 that displayed a novel mode of genic piRNA targeting. A *Ulysses* element insertion upstream of this gene in strain 9 is specifically associated with genic silencing and shunting *oysgedart* into piRNA biogenesis (Fig 1.7). 5' genic piRNAs derived from *oysgedart* are dual-strand in strain 9 ovaries (which includes somatic and germline tissue) but biased towards the sense strand in the germline.

Maternal deposition of piRNAs derived from a specific locus can mediate piRNA biogenesis at that locus in the next generation (de Vanssay et al., 2012; Le Thomas et al., 2014b). For *cdi*, a recent study demonstrated that maternal deposition leads to maintenance of piRNA biogenesis from the 160 *cdi* allele. In progeny of strain 160 mothers and strain 9 fathers, maternally deposited piRNAs derived from the *cdi* cluster also activate piRNA biogenesis and methylation of H3K9 from the strain 9 *cdi* allele (Le Thomas et al., 2014a).

We found that two modes of genic targeting by piRNAs result in contrasting effects on gene expression across generations. The first mode was found at the *cdi* locus (Fig 1.7). Here, maternal deposition of *cdi* piRNAs mediates silencing in the germline, but not the soma. *Cdi* gene expression is highly reduced in the germline of strain 160 and the germline silencing of both alleles of *cdi* is maintained when transmitted maternally and made heterozygous in combination with the wild-type strain 9 allele. In contrast, when the *cdi* cluster allele is transmitted paternally, piRNA cluster activity is reduced and germline expression levels are maintained near the level of strain 9, with similar contributions from each allele. Critically, this asymmetry in gene expression is not observed in the soma. Carcasses of females from reciprocal directions of the cross showed similar levels of expression from both alleles (Dysgenic Carcass: 17.1 RPKM [9 Allele: 7 counts; 160 Allele: 9 counts] vs. Non-dysgenic Carcass: 23.4 RPKM [9 Allele: 5 counts; 160 Allele: 12 counts]) (Fig 1.7).

The second mode was found at the *oysgedart* locus (Fig 1.7). The *oysgedart* allele in strain 9 shows a novel mode of genic targeting by piRNA. Similar to *cdi*, we identified sense and anti-sense piRNAs in the ovary. However, germline piRNAs derived from *oysgedart* are primarily sense derived. Strikingly, in

neither direction of the cross is this form of sense piRNA biogenesis maintained in progeny. Instead, the wild-type allele from strain 160 seems to function in *trans* to limit this mode of silencing. Thus, in both directions of the cross, the expression of *oysgedart* is maintained at equal levels, but only at about 60% of wildtype since the *Ulysses* insertion allele from strain 9 is expressed at a lower level. While sense piRNA biogenesis is turned off in reciprocal directions of the cross, expression of the 9 allele is reduced in *cis*. In neither direction of the cross does there appear a maternal effect on somatic expression for *oysgedart*. Even though expression of the 9 allele is reduced in the germline, it remains on in the soma of both dysgenic (RPKM: 62.0, [9 Allele: 21 counts; 160 Allele: 21 counts]) and non-dysgenic (RPKM: 80.9, [9 Allele: 29 counts; 160 Allele: 25 counts]) F1 females (Fig 1.7). This demonstrates that the allelic *cis* effects on local silencing by the *Ulysses* insertion are germline, not soma, specific (Fisher's exact test for difference in allele effects between soma and germline. Dysgenic: $p<0.0001$, Nondysgenic: $p<0.0001$). Interestingly, in contrast to *cdi* which has shown to be enriched for H3K9 methylation in strain 160 ovaries, we find no evidence that *oysgedart* is so enriched in strain 9 ovaries (Supp Table 1.4) (Le Thomas et al., 2014a).

We then tested how this mode of piRNA biogenesis is maintained in further generations. We found that the *cdi* piRNA cluster can cause heritable activation of piRNA biogenesis from the strain 9 allele, a process equivalent to paramutation. In F3 flies generated by backcrossing hybrid non-dysgenic females to strain 9 males (maintaining maternal transmission of the cluster) cluster behavior was maintained even when *cdi* was homozygous for the strain 9 allele (Fig 1.7). In one case, cluster behavior was lost. Because the cross scheme was maintained over two generations, we are unable to determine the generation in which the cluster activity was lost, however, it is clear that *cdi* can paramutate in one generation but that this paramutation is not robust across multiple generations (though see (de Vanssay et al., 2012). In the same F3 flies, we did not observe restoration of cluster maintenance for *oysgedart*.

# DISCUSSION

Due to their proliferative nature, differences in the TE landscape within populations and between species can accumulate rapidly. Since TEs are harmful, it has also been proposed that this divergence drives rapid evolution of the piRNA machinery itself (Kolaczkowski et al., 2010; Obbard et al., 2009; Simkin et al., 2013). Divergence at both levels is expected to greatly influence patterns of TE activity in crosses between individuals. Within a species, hybrid dysgenesis syndromes driven from a single element family reveal that maternally deposited piRNAs targeting the activating TE are critical for maintaining TE control and fertility. In contrast, crosses between species that differ with respect to both TE profile *and* the machinery of piRNA biogenesis show dramatic collapse of piRNA biogenesis and this can be attributed to divergence in the piRNA machinery rather than differences in the maternally deposited pool (Kelleher et al., 2012). Previous studies, combined with results presented here, indicate that the *D. virilis* system represents a complex form of intraspecific hybrid dysgenesis involving more divergent TE profiles compared to other syndromes of hybrid dysgenesis. However, because it is an intraspecific cross, there is minimal divergence in the TE regulatory machinery. Therefore, the dysgenic syndrome in *D. virilis* can be considered an important model for understanding the dynamics of TE control at an intermediate stage in the divergence of TE profile within a single species.

Here we show that differences between genetically identical dysgenic and non-dysgenic individuals are manifested in multiple ways. When genomes from two strains of *D. virilis* are brought together, TEs that are more abundant in one genome become more highly expressed in the germline of the next generation. This difference in TE expression persists in the germline as flies age. Coincident with this, there is also a persistent increase in TE expression for several TEs that are evenly distributed between strains.

We identify multiple modes by which piRNAs modulate gene expression. First, many genes become off-targets for piRNA biogenesis, and their expression levels are reduced. The mechanism for this is unclear, but may be driven by the same mechanism that leads to idiosyncratic defects in piRNA biogenesis for some TEs. One possibility is that compromised piRNA function in the cytoplasm leads to a shift in the targets of primary piRNA biogenesis. A similar increase in sense genic piRNA abundance has been observed in *rhino* and *uap56*mutants and this has been attributed to compromised specificity in piRNA processing (Zhang et al., 2012). Therefore, the increased genic piRNA abundance in the dysgenic germline may be a readout of compromised piRNA biogenesis and loss of specificity in the cytoplasm. Secondly, divergence in the TE profile between strains leads to differences in the pool of maternally deposited genic piRNAs. Depending on the nature of genic piRNAs, these can modulate gene expression in diverse ways across generations. For the *cdi* gene, maternally deposited piRNAs from both strands mediate gene silencing of both alleles in non-dysgenic progeny. Alleles of *cdi* share properties with imprinted genes since expression depends on which parent the allele is inherited from. However, *cdi* differs from canonical imprinting in that a silenced *cdi* allele is capable of silencing the other allele in *trans* when inherited maternally, similar to paramutations observed in maize, mice and recently in *Drosophila melanogaster* (Alleman et al., 2006; Arteaga-Vazquez et al., 2010; Chandler, 2007; Dalmay et al., 2000; de Vanssay et al., 2012).. Imprinted genes can have a significant downstream effect on patterns of gene expression (Mott et al., 2014). Even in *Drosophila*, where there is no evidence for DNA methylation, there are significant, albeit poorly understood, parent-of-origin allelic effects on global gene expression (Gibson et al., 2004; Wittkopp et al., 2006). Such transgenerational effects on gene expression may contribute to large numbers of genes being differentially expressed between the dysgenic and non-dysgenic germline. In contrast to *cdi*, a *Ulysses* insertion upstream of the *oysgedart* gene seems to mediate only deposition of sense piRNAs. In this case, the wild-type non-

insertion allele appears to resist transgenerational silencing and expression is maintained equally between reciprocal hybrids.

Since we identify a large number of differences in TE and genic expression that are mediated by piRNAs in diverse ways, it is difficult to distinguish between causal factors and downstream effects. However, our genetic analyses clearly demonstrate that both induction of and protection against dysgenesis is distributed across the genome. Since multiple TE families are in excess copy number in the inducer strain, the weight of evidence favors a complex mode of hybrid dysgenesis driven jointly by the mass action of multiple elements. This is supported by the fact that the regions of the genome that are most protective against dysgenesis are located in the pericentric regions, which are known to be critical sources of piRNAs (Brennecke et al., 2007; Grentzinger et al., 2012), further supporting a model in which pericentric regions play a unique role in TE control (Bergman et al., 2006).

Representing an intermediate state of TE divergence between single TE family dysgenesis syndromes and interspecific crosses, several observations are worth noting from the *D. virilis* system. First, in crosses between *D. melanogaster* and *D. simulans*, piRNA biogenesis is globally defective and this is attributed to the accumulation of incompatibilities that arise from rapid divergence in the protein sequence of the piRNA biogenesis machinery. In the *D. virilis* system where there is only divergence in the TE profile between strains, we see no such global collapse of piRNA biogenesis. In this sense, this is more similar to what is observed in the *P-M* system of dysgenesis.

However, there are some key distinctions between the *P-M* system, driven by a single element, and the *D. virilis* system of dysgenesis. In the *P-M* system, as flies age, piRNAs targeting the active element are restored and *P* element silencing is regained. We see very little restoration of silencing for many TEs that are most likely to contribute to dysgenesis. For example, the *Helena* element remains much more highly expressed in the dysgenic germline and this is also associated with failure to restore the piRNA pool targeting *Helena*. Therefore, despite the fact that these tissues (0–2 hour old embryos) have escaped the ablation event of dysgenesis, idiosyncratic defects in TE silencing persist and these are passed on to the next generation. Furthermore, increased TE expression is observed for some TEs that appear to have a restored level of piRNA biogenesis. Thus, a further defect in dysgenesis is a lack of effective silencing by a restored pool of piRNA. This suggests multiple mechanisms contribute to chronic increased TE expression in the dysgenic germline.

A long-standing question is what underlies TE co-mobilization in dysgenic syndromes in *Drosophila*. For many years, the *D. virilis* system was considered unique in that it was associated with increased movement of elements equally abundant between strains, the *Ulysses* element being the best example. Recent work by Theurkauf and colleagues suggest that co-mobilization in fact may be a general property of dysgenic syndromes (Khurana et al., 2011b). By what mechanism does the movement of one element activate the movement of others? The current working model is that DNA damage from a moving element activates the DNA damage response within the germline. This, in turn, drives Chk-2 mediated phosphorylation and degradation of Vasa. Vasa is critical for nuage assembly and germline piRNA biogenesis (Malone et al., 2009). In the absence of proper Vasa and germline piRNA function, resident transposons become activated, leading to co-mobilization.

This is a strong candidate for the mechanism of co-mobilization in the *D. virilis* system. A burst of TE mobilization in the germline at any time during development may drive movement of TEs found in equal

copy number between strains via the DNA damage response. However, it is difficult to explain the persisting and idiosyncratic pattern of increased expression for some TEs in the dysgenic germline that is observed in aged females that have escaped sterility. In the *P-M* system, Vasa degradation is relieved as the aging dysgenic flies recover partial fertility and *P*element silencing is restored. In *D. virilis*, full silencing is not restored, even in the germline that has escaped ablation. Coinciding with increased levels of expression that persist for some TEs, we also note that there is a persisting increase in piRNAs that target genes. This is similar to observations in flies defective in *rhino* and *uap56* (Zhang et al., 2012). We have shown that global defects in TE repression observed during hybrid dysgenesis share this feature. Strikingly, the genes that are piRNA targets are highly enriched for ribosomal proteins RNAs (FDR p-value: 1.9E-13, 19-fold enrichment; Supp. Table 1.3). The mechanism by which increased genic piRNAs target ribosomal protein RNAs is unclear, however there are multiple lines of evidence that disrupted levels of ribosomal proteins can lead to increased levels of repeat expression (Komili et al., 2007). For example, an early genome wide screen in *C. elegans* identified approximately 27 genes involved in transposon silencing, two of which were ribosomal proteins (Vastenhouw et al., 2003). Disruption of the ribosome is known to trigger nucleolar stress and p53 activation (Zhang and Lu, 2009). Thus, nucleolar stress mediated by genic piRNAs may also contribute to persisting TE activation in the dysgenic germline.

# FIGURES

**Figure 1. 1 Multiple transposable elements are associated with induction of hybrid dysgenesis**. (A) Relative mapping abundance of single-end, 100 bp reads from strain 9 and strain 160 (normalized by reads mapping to the genome), to a consolidated repeat library. Eleven elements are in 3-fold excess in strain 160 and are indicated here and throughout with red. TART elements are about 1.7-fold in excess and are indicated here and throughout with blue. No apparent TEs were found in excess in strain 9. (B) Using piledriver (https://github.com/arq5x/piledriver) we assessed homogeneity within reads mapping to the TE library by determining the average frequency of the major variant in both strains. TEs in excess in strain 160 are either more homogenous in strain 160 or similarly aged between strains, with the exception of element 1069 which shows slightly more homogeneity in strain 9.

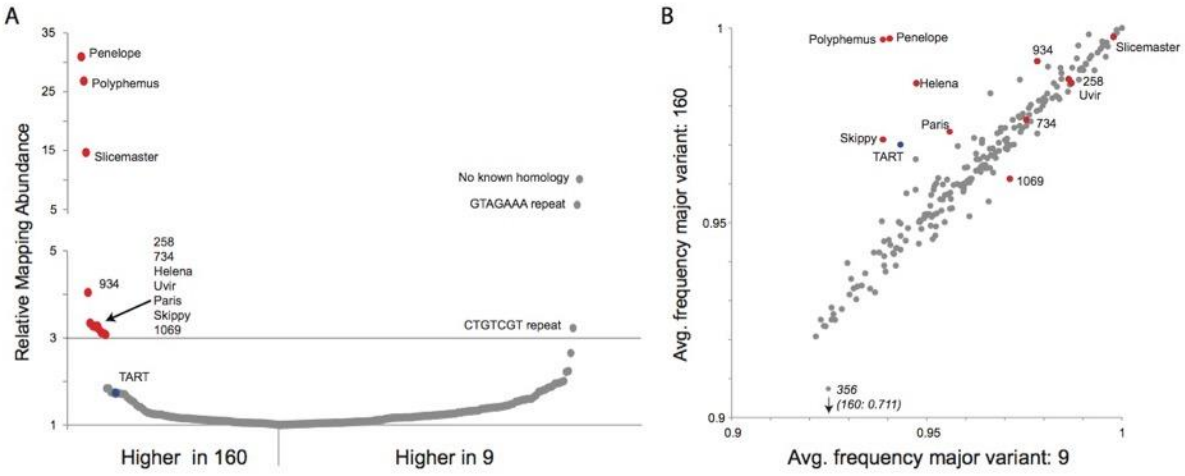


**Figure 1. 2. Increased TE expression in the dysgenic germline persists through adulthood.** (A) RPKM+0.01 (log 10, average across both ages) for TEs, Dysgenic vs. Non-dysgenic germline. TEs that are in excess in 160 are more highly expressed, as well as many TEs that are not in excess. (B) Fold excess in expression (RPKM+0.01, log 2, average across both ages) vs. fold excess in abundance in strain 160. Nearly all TEs that are in excess in 160 show increased expression in the dysgenic germline (11/12). But multiple TEs that are equivalent in abundance between strains are also increased in expression. (C,D) Increased expression in the dysgenic germline is maintained as flies age. Note: Log scale obscures magnitude of difference for some TEs that demonstrate significant differences in expression identified due to low variation across replicates.

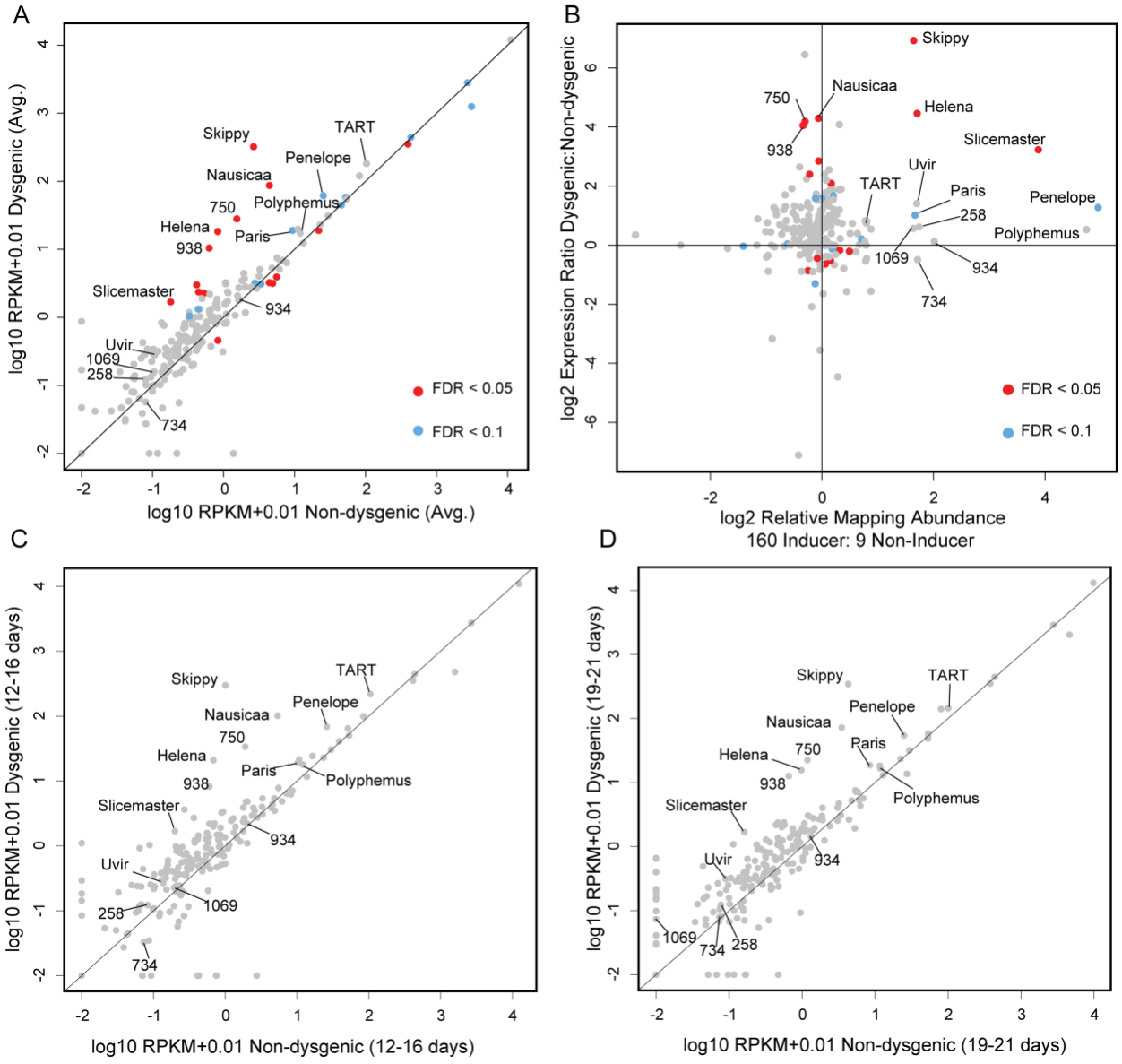


**Figure 1. 3**. **TE expression as a function of piRNA and siRNA abundance in parental strains and progeny**. (A). Normalized (per 1 million mappers) piRNA abundance +0.1 (log 10) in the strain 9 germline vs. the strain 160 germline. A large number of TEs show increased piRNA expression in the strain 160 germline, especially TART and others enriched in abundance in strain 160. Diagonal lines indicate 10-fold levels of difference (B) Normalized (per 1 million mappers) piRNA abundance +0.1 (log 10) in the dysgenic germline vs. the non-dysgenic germline. piRNA abundances for many TEs with greater excess in strain 160 become similar in the dysgenic germline. A significant exception to this is the Helena element. Diagonal lines indicate 10-fold levels of difference. (C) TE piRNA excess in strain 160 vs. relative expression level in dysgenesis. TE piRNA asymmetry between 160 and 9 is not the sole determinant of increased expression in dysgenesis. Some elements, such as 750, are increased in expression in dysgenesis, despite similar piRNA abundances in 9 and 160. (D) TE piRNA excess in the non-dysgenic germline vs. relative TE expression level in dysgenesis. Elements such as Skippy and Slicemaster show equilibrated piRNA abundances, but excess expression in the dysgenic germline. (E) TE siRNA excess in strain 160 vs. relative TE expression level in dysgenesis. (F) TE siRNA excess in the non-dysgenic germline vs. relative TE expression level in dysgenesis.

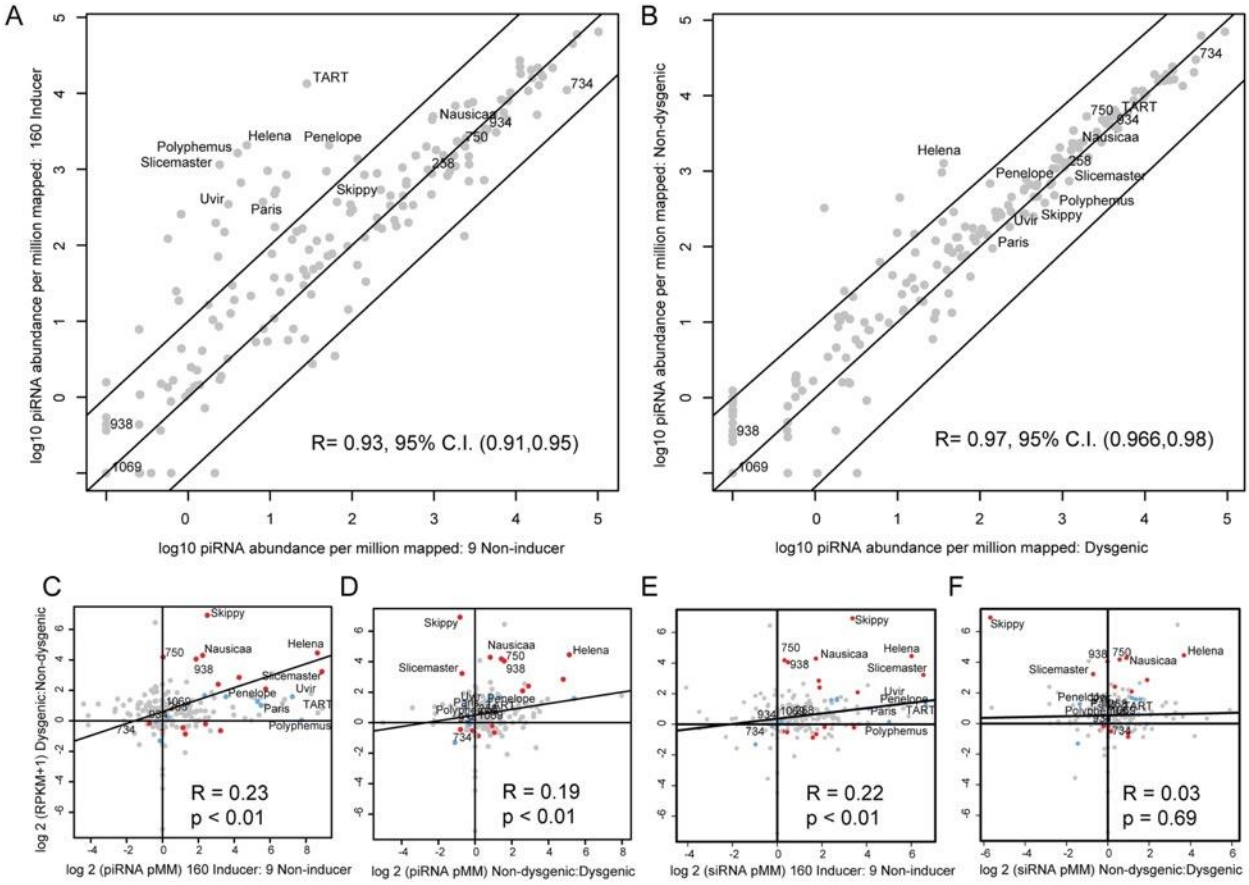


**Figure 1. 4. Signatures of piRNA biogenesis in the dysgenic germline show only modest defects.** (A) Size distributions of small RNAs are similar between dysgenic and non-dysgenic germlines. Distribution of all small RNAs (not normalized) from four germline libraries (2 dysgenic, 2 non-dysgenic) filtered for tRNA, rRNA and snoRNA. (B) piRNA biogenesis signature heatmaps. TEs upregulated in the dysgenic germline (a difference of 5 RPKM or higher) are indicated with red bars. TEs upregulated in the non-dysgenic germline (a difference of 5 RPKM or higher) are indicated with purple. On the left are heatmaps for raw measures of abundance, the density of ping-pong pairs and percent ping-pong. On the right are heatmaps for the same metrics, but by row z-score. For raw measures, there are no globally discernible effects of dysgenesis on piRNA biogenesis. Row z-scores in dysgenesis do show lower values for abundance measures (abundance and ping-pong pair density), but not percent ping-pong (see text). (C) Fold excess in expression in dysgenesis vs. the difference in percent ping-pong Z-score between dysgenic and non-dysgenic germline. Of the top eight that are most differently expressed in dysgenesis, all have lower ping-pong z-scores in dysgenesis.

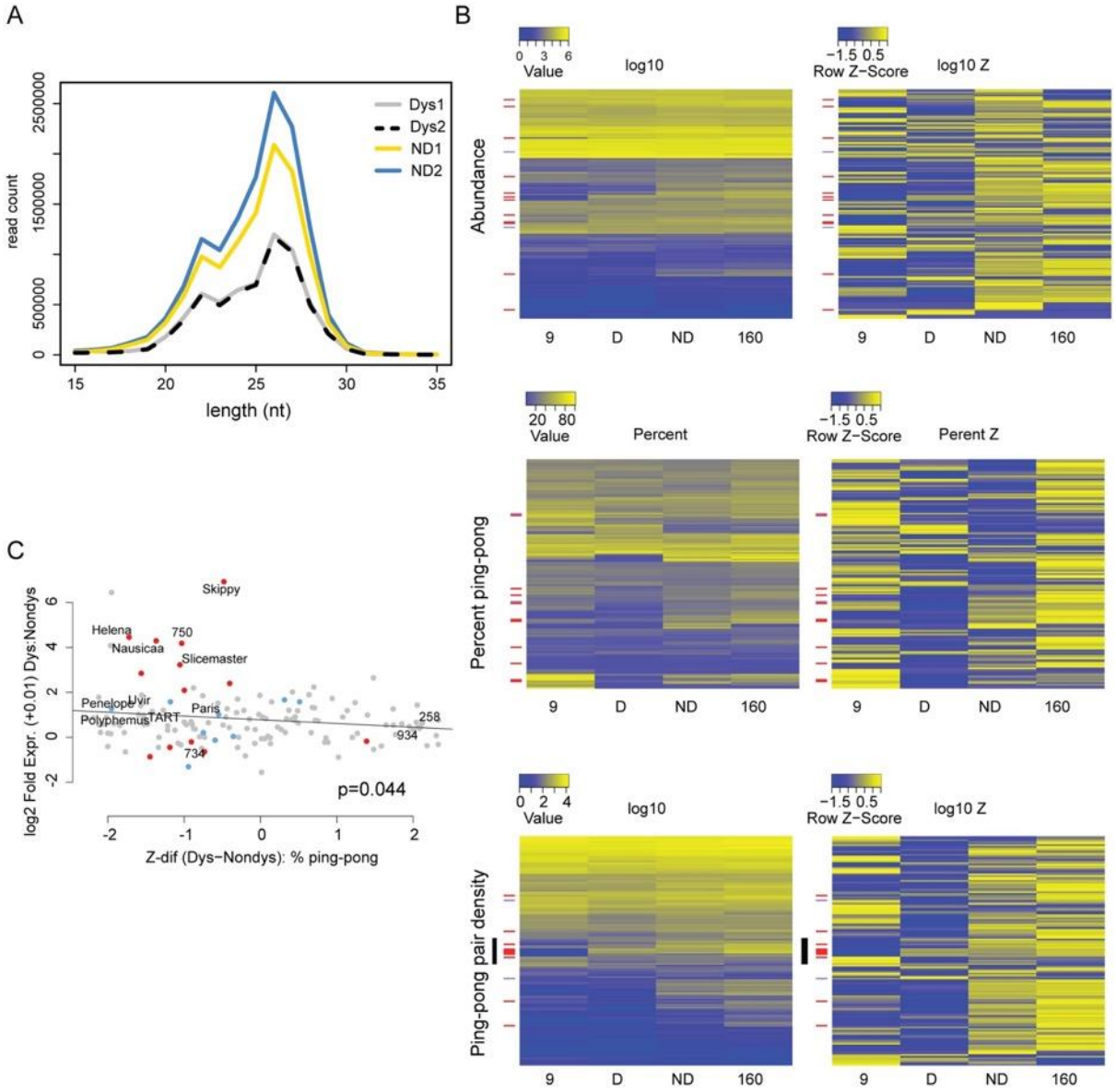


**Figure 1. 5**. **Genic piRNA targeting is increased in the dysgenic germline.** (A) log10 Z-score heat map of genic (CDS) piRNA density for D. melanogaster orthologs (above a threshold of 5 piRNAs per CDS per 1 million mapped in at least one of four columns). Of these 105 genes, there is an excess of genic piRNAs in the dysgenic germline (89 genes with greatest genic targeting in dysgenesis, P<0.001) (B) Sense vs. Anti-sense abundance for piRNAs in genic piRNA class for one library (Sample 1). Some CDS regions are predominantly the source of anti-sense piRNAs, but the majority are biased as a source of sense strand piRNA (C) Distribution of expression levels (log 10 RPKM+0.01) for all genes in the genome and piRNA target genes (expression levels from non-dysgenic germline). Genic piRNA targets are derived from more highly expressed genes ($p < 0.001$). (D) Of 105 genes, the 89 that show excess genic piRNA in dysgenesis are also more lowly expressed in dysgenesis. Shown is the distribution of expression ratios (dysgenic:non-dysgenic) for all genes and genes that are increased as a source of genic piRNAs in dysgenesis ($p < 0.001$).

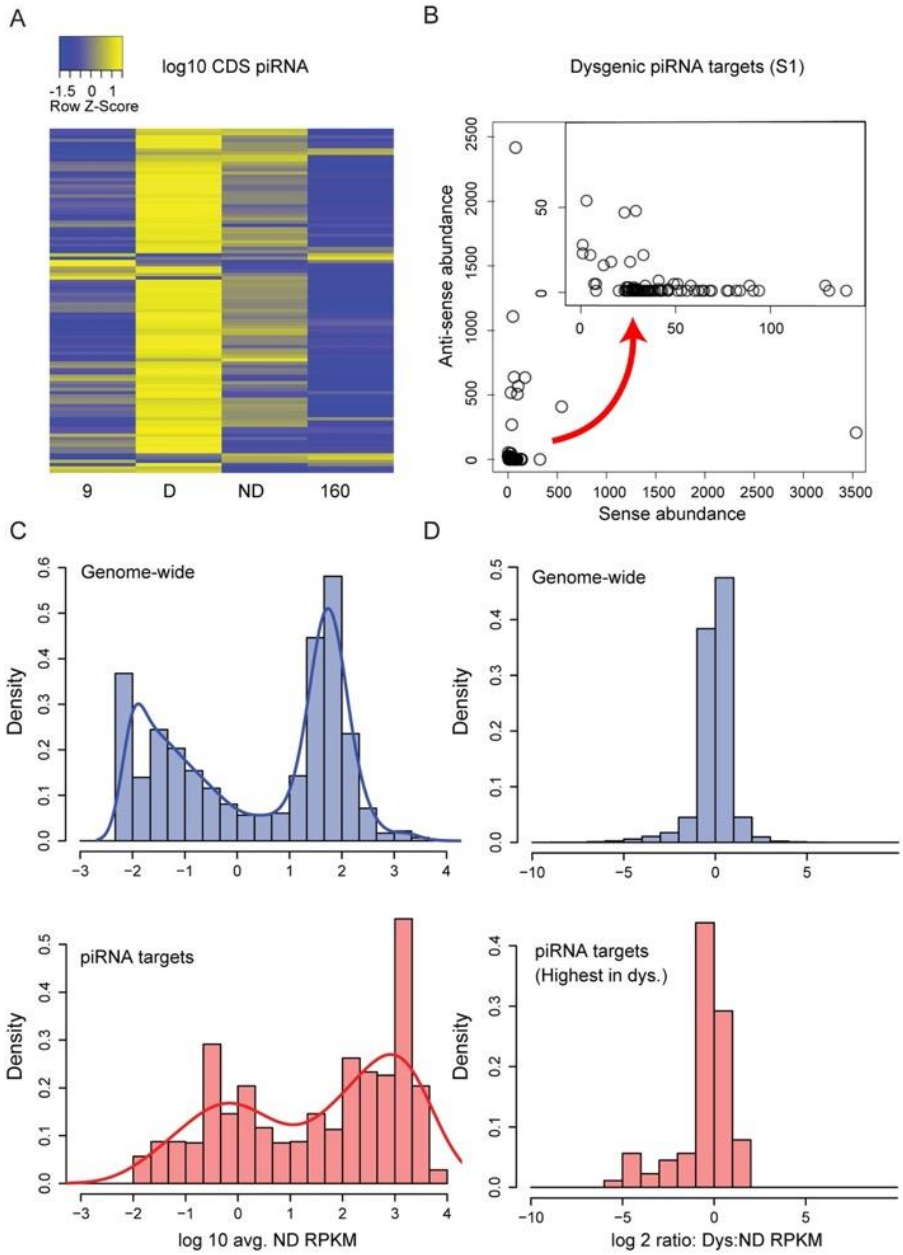
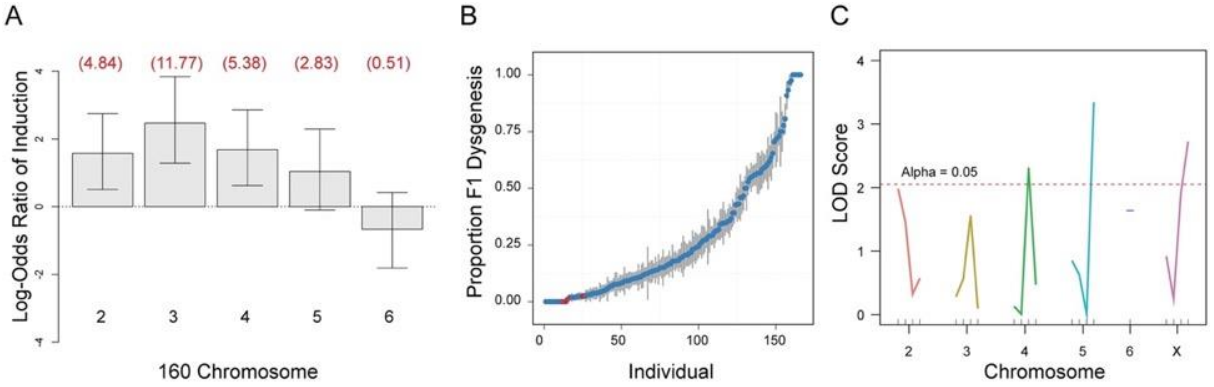


**Figure 1. 6. Genetic analysis of zygotic induction and maternal repression of gonadal atrophy.** (A) Induction of sterility by 160 is broadly distributed across the genome, with the exception of chromosome 6 (the dot chromosome). Log odds ratios for probability of induction were estimated by crossing F1 males to strain 9, determining whether F2s had male gonadal atrophy and genotyping F2s to determine the chromosomes inherited from the father. Estimates were determined using a generalized linear model for logistic regression (binomial family with a logit link). Values in red are actual odds ratios. Whiskers are 95% confidence intervals. Chromosome 5 is significant at 0.1 level only. X chromosome is not scored because dysgenesis is scored in males and males do not inherit the X from their fathers (N = 92). (B) Scatterplot showing proportion of dysgenic testes (y axis) observed in the progeny of each F3 female individual (x axis). Red dots indicate F3 females that were selected for whole genome sequencing. (C) Single marker QTL analysis identified 3 putative QTLs: one flanking the centromeres of the 5th and X chromosomes and one of the tested euchromatic regions of the 4th chromosome. (D) Top row: Results from the genotyping assay. Colored rectangles represent the presence of strain 160 SNPs in individuals, ranked from top to bottom (most protective individuals on top). Scatterplots: sequencing results. Each dot represents the average number of base pairs that uniquely mapped to every 10kb of the 160 genome. Valleys indicate regions of strain 9 homozygosity. Black dots above scatterplots show the location of each SNP used for our genotyping assay. Grey background demonstrates that no region of the genome from 160 is necessary to protect against dysgenesis. Right-most columns: Number of piRNAs mapped to TART sequences, per million reads, for each F3 female individual. Color intensity is representative of TART piRNA abundance. Number of 21 nt endo-siRNAs mapped to Penelope sequences, per million 21 nt reads, for each F3 female individual. Color intensity is representative of Penelope endo-siRNA abundance. Red bar indicates position of one of several Penelope endo-siRNA loci on the X.



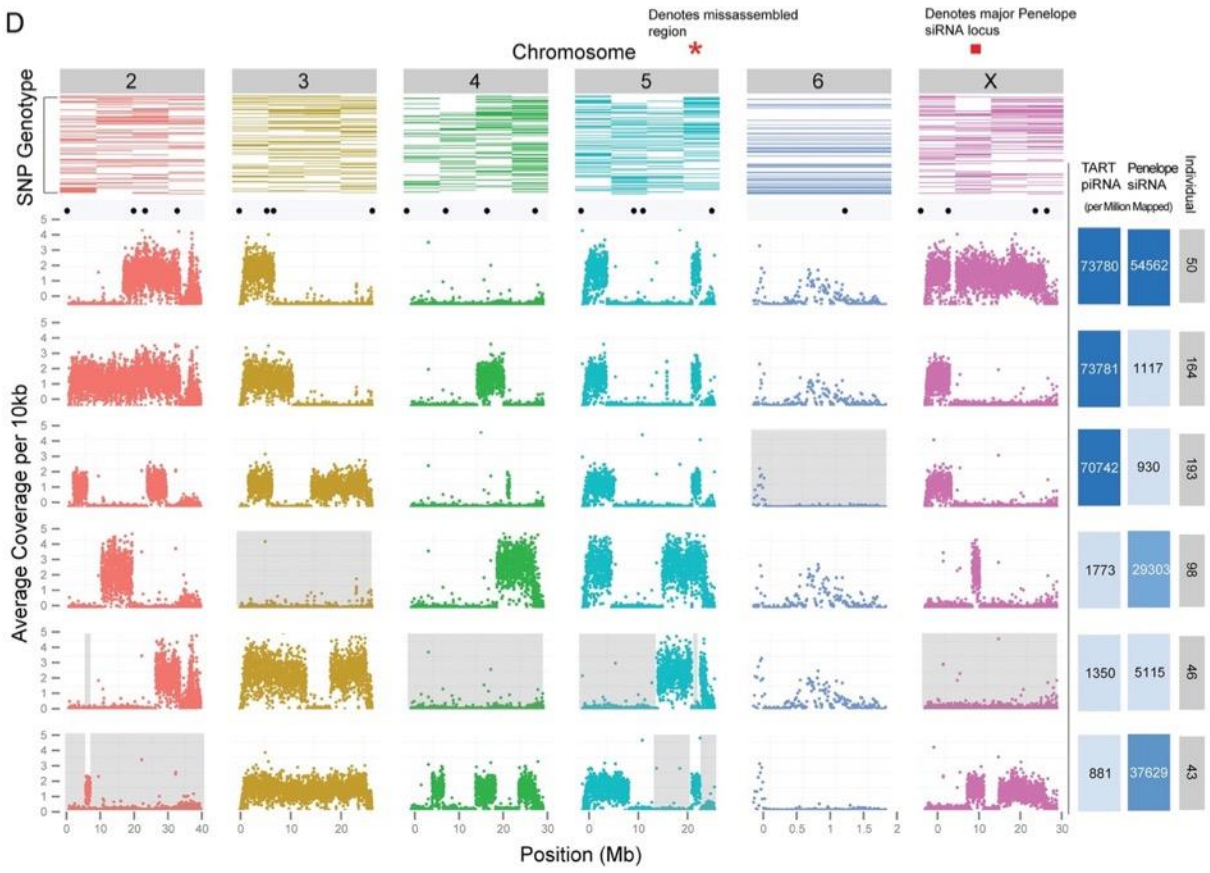

D
Denotes missassembled region
Denotes major Penelope siRNA locus
Chromosome
2
3
4
5
6
X
SNP Genotype
Average Coverage per 10kb
Position (Mb)
TART piRNA (per Million Mapped)
Penelope siRNA
Individual
73780 54562 50
73781 1117 164
70742 930 193
1773 29303 96
1350 5115 46
881 37629 43

**Figure 1. 7. Germline and ovary genic cluster behavior across generations for D. virilis orthologs of center divider and oysgedart from D. melanogaster.** piRNA mapping densities are indicated. mRNA-seq RPKM for germline (0–2 H embryo) is also indicated. Allelism was determined by counting mRNA-seq reads based on SNPs that distinguish strain 9 and 160. Strain 160 cluster identity is maintained for cdi in non-dysgenic progeny in which strain 160 is the mother. This is correlated with silencing of both alleles in the non-dysgenic germline. In contrast, the cluster is not maintained in the dysgenic germline and both alleles are expressed. Somatic expression is not affected. Germline cluster identity for oysgedart (which in the germline is predominantly sense) is lost in progeny. In this case, expression is even between reciprocal progeny, but germline expression is lower from the 9 allele in both directions of the cross. For cluster behavior in F3 backcrosses, heterozygosity or homozygosity of the respective allele is indicated. Notice how cluster identity is maintained for cdi to varying degrees in individuals homozygous for the 9 allele. In contrast, cluster activity is absent in all progeny for oysgedart.

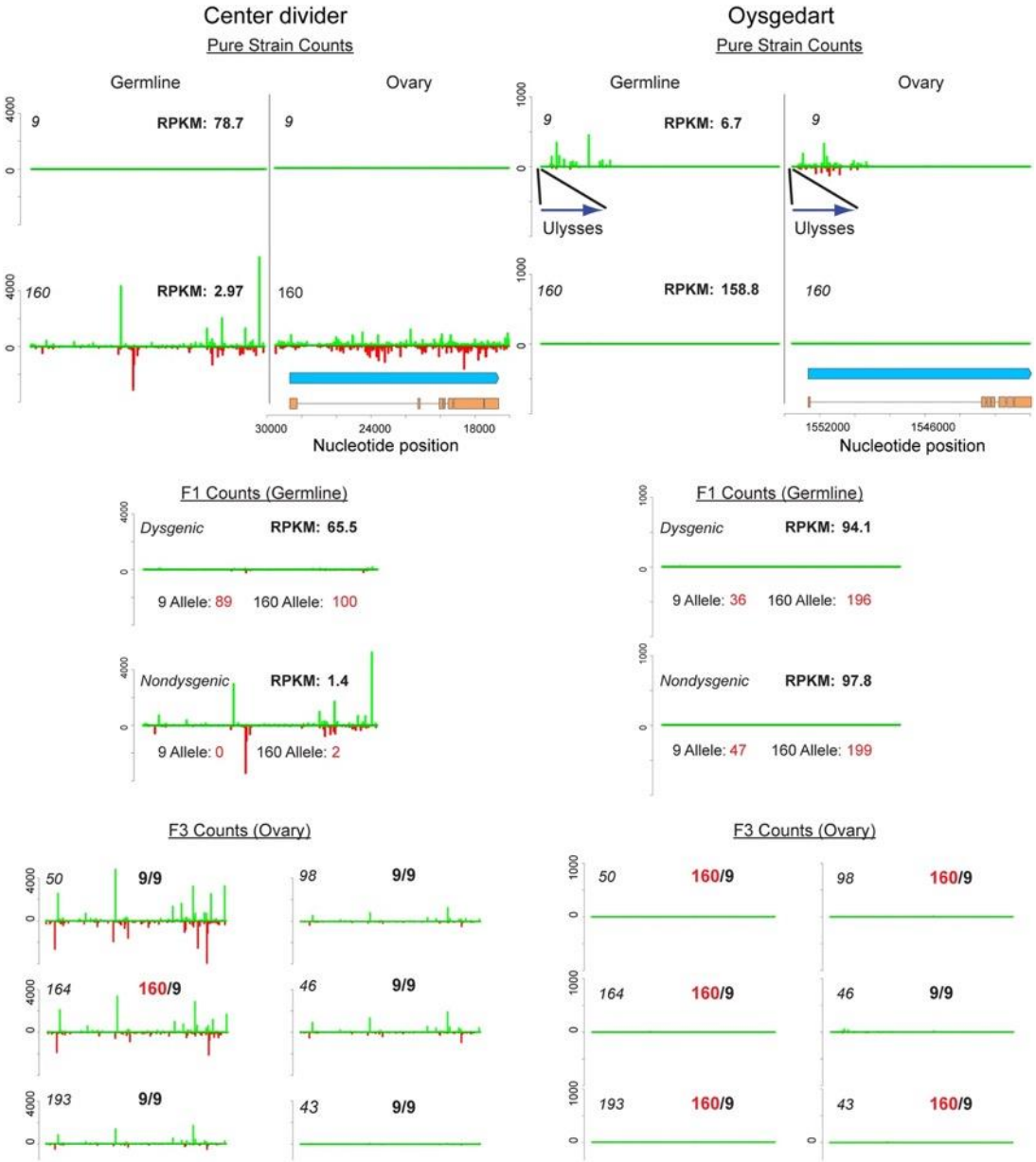
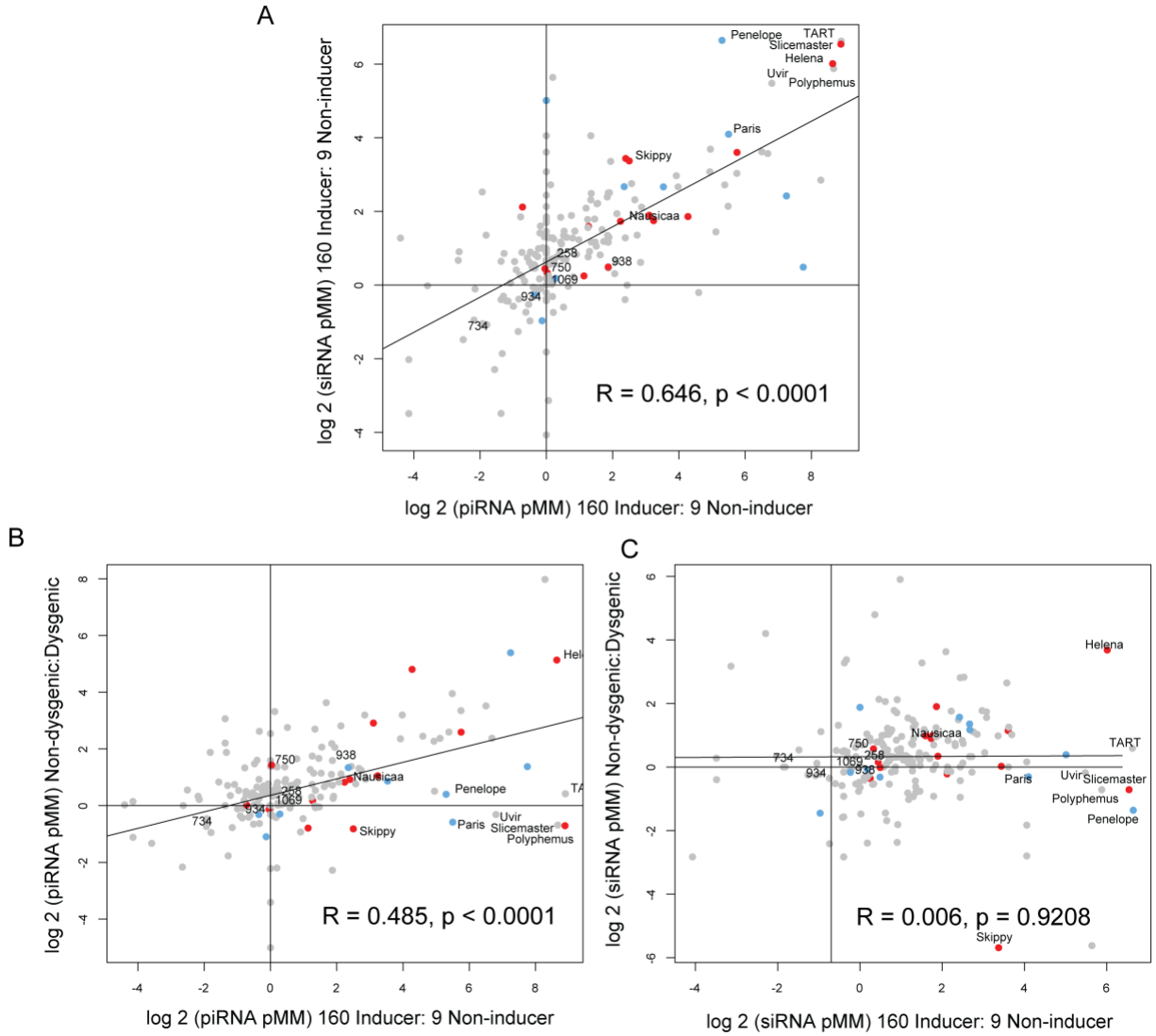


**Figure 1. 8. Relationships between maternal and zygotic abundance for piRNA and siRNA pools**. Red indicates TEs with significant differences in expression at FDR<0.05. Blue indicates TEs with significant differences in expression at FDR<0.1. A) Log 2 of piRNA abundance ratio (160:9, per million mapped) vs. Log 2 of siRNA abundance ratio (160:9, per million mapped). B) Log 2 of piRNA abundance ratio (160:9, per million mapped) vs. Log 2 of piRNA abundance ratio (non-dysgenic:dysgenic, per million mapped). C) Log 2 of siRNA abundance ratio (160:9, per million mapped) vs. Log 2 of siRNA abundance ratio (non-dysgenic:dysgenic, per million mapped). D) Log 2 of piRNA abundance ratio (160:9, per million mapped) vs. Log 2 of siRNA abundance ratio (non-dysgenic:dysgenic, per million mapped). E) Log 2 of piRNA abundance ratio (non-dysgenic:dysgenic, per million mapped) vs. Log 2 of siRNA abundance ratio (non-dysgenic:dysgenic, per million mapped).



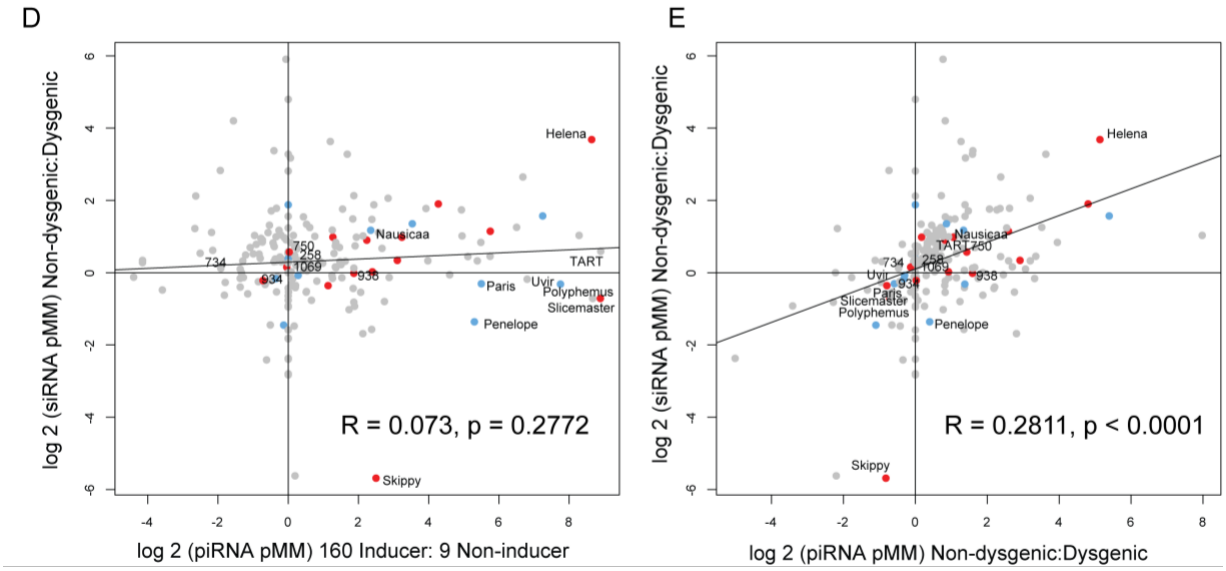

D
log 2 (siRNA pMM) Non-dysgenic:Dysgenic
log 2 (piRNA pMM) 160 Inducer: 9 Non-inducer
R = 0.073, p = 0.2772
Helena
Nausicaa
750
288
734
1069
934
938
TART
Paris
Uvir
Polyphemus
Slicemaster
Penelope
Skippy
E
log 2 (siRNA pMM) Non-dysgenic:Dysgenic
log 2 (piRNA pMM) Non-dysgenic:Dysgenic
R = 0.2811, p < 0.0001
Helena
Nausicaa
TART
750
734
288
1069
Uvir
934
938
Paris
Slicemaster
Polyphemus
Penelope
Skippy

**Figure 1. 9. Differences in TE ping-pong pair density are influenced by maternal differences in ping-pong pair density.** The difference in ping-pong pair density (per million mapped) between 160 and 9 (160 minus 9) vs. the difference in ping-pong pair density (per million mapped) between Non-dysgenic and Dysgenic (Non-dysgenic minus dysgenic). Larger differences in ping-pong pair density correspond to larger differences between dysgenic and non-dysgenic germline. However, many TEs differentially expressed show minimal differences in ping-pong pair density, either in parents or offspring. Red indicates TEs with significant differences in expression at FDR<0.05. Blue indicates TEs with significant differences in expression at FDR<0.1

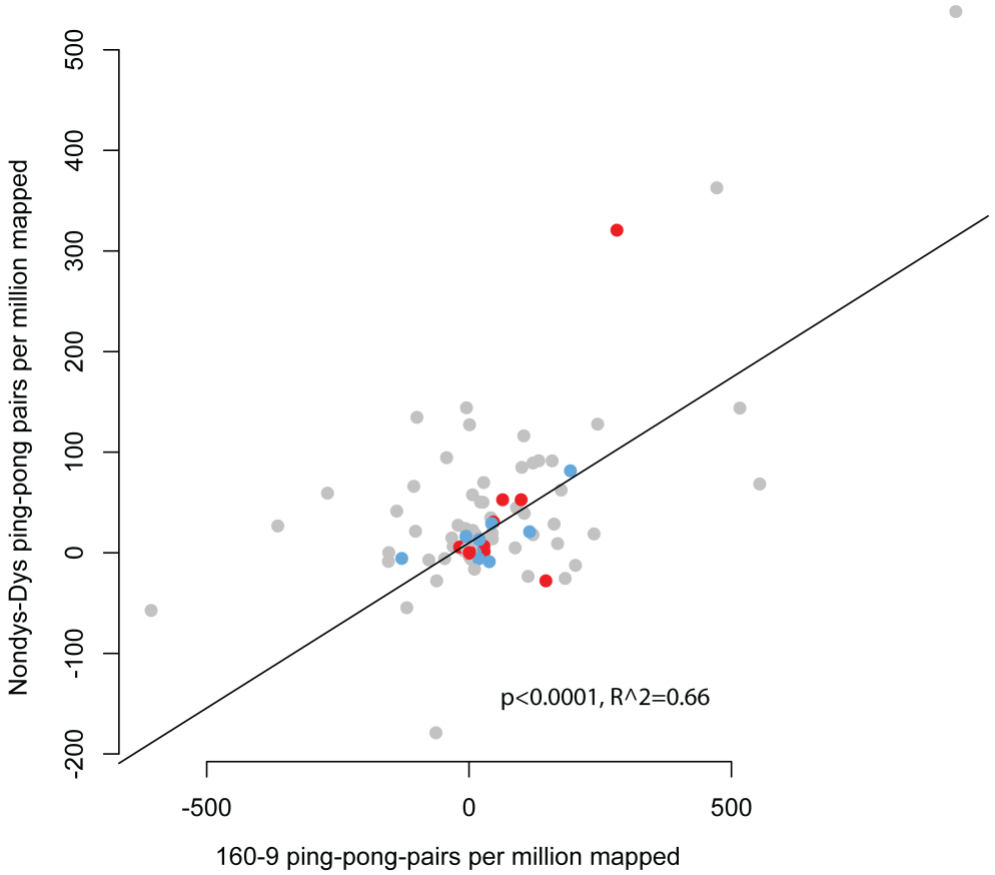
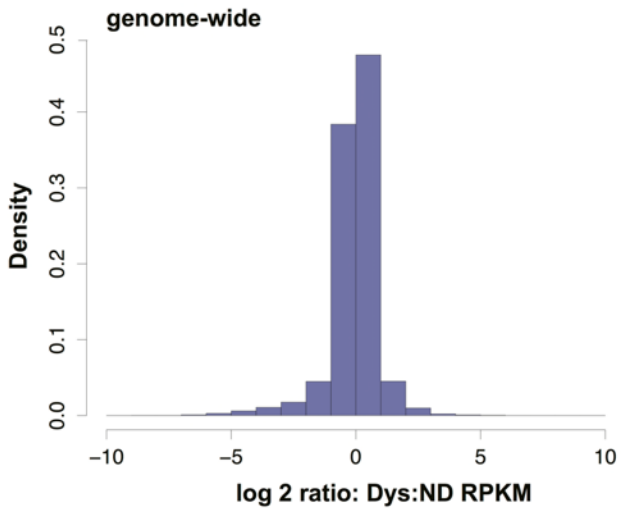
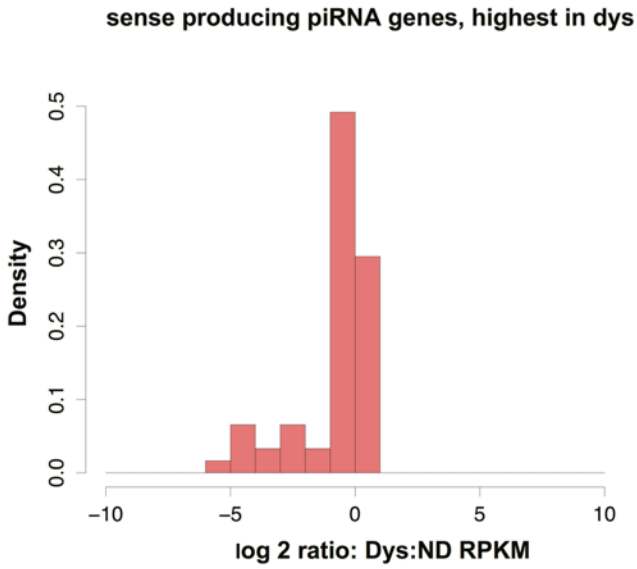


**Figure 1. 10. Average expression (RPKM) ratio for genes from 0–2 hour old embryos laid by dysgenic and non-dysgenic females**. A) Genome wide expression ratio distribution (Dysgenic:Non-Dysgenic) and B) Distribution of expression ratios (Dysgenic:Non-Dysgenic) for genes producing sense piRNAs, with highest piRNA abundance in the dysgenic treatment

A




B



# TABLES

**Table 1. 1. Properties of TEs significantly more expressed in the dysgenic germline by at last 2-fold (FDR<0.1).**

| TE | 160:9 Abund. (Log 2) | 160:9 piRNA (Log 2) | Expression: D (RPKM) | Expression: ND (RPKM) | PP pair density: D (pMM) | PP pair density: ND. (pMM) | TYPE |
|---|---|---|---|---|---|---|---|
| *Helena* | 1.71 | 8.64 | 18.25 | 0.83 | 0.31 | 53.04 | I |
| *495* | 0.21 | 2.35 | 1.05 | 0.33 | 37.79 | 119.30 | I |
| *967* | 0.17 | 5.75 | 2.28 | 0.53 | 0.21 | 6.72 | I |
| *Penelope* | 4.95 | 5.30 | 61.48 | 25.42 | 11.19 | 32.06 | I |
| *Slicemaster* | 3.88 | 8.89 | 1.69 | 0.18 | 12.63 | 13.89 | II |
| *Skippy* | 1.64 | 2.50 | 321.77 | 2.65 | 74.92 | 47.03 | III |
| *Paris* | 1.67 | 5.50 | 18.94 | 9.36 | 12.22 | 6.25 | III |
| *Nausicaa* | -0.06 | 2.24 | 86.90 | 4.43 | 14.19 | 44.83 | IV |
| *656* | -0.01 | 7.25 | 3.20 | 1.07 | 0.34 | 13.46 | IV |
| *1012* | -0.06 | 4.27 | 3.01 | 0.42 | 0.27 | 52.99 | IV |
| *190* | -0.12 | 3.53 | 1.32 | 0.44 | 16.66 | 45.73 | IV |
| *620* | -0.22 | 3.11 | 2.35 | 0.45 | 0.00 | 9.93 | IV |
| *938* | -0.34 | 1.87 | 10.42 | 0.63 | 0.00 | 0.00 | V |
| *750* | -0.30 | 0.03 | 28.03 | 1.54 | 7.17 | 12.84 | VI |

160:9 Abundance, 160:9 total piRNA, Expression: D (Dysgenic, RPKM), Expression: ND (Non-dysgenic, RPKM), Ping-pong pair density: D (Dysgenic, per Million piRNAs mapped), Ping-pong pair density: ND (Non-Dysgenic, per Million piRNAs mapped). Type I: Higher in copy number in 160, higher in piRNA abundance in 160, F1 piRNA ping-pong pair density defined by maternal loading. Type II: Higher in copy number in 160, higher in piRNA abundance in 160, F1 piRNA ping-pong density equilibrated. Type III. Higher in copy number in 160, higher in piRNA abundance in 160, F1 piRNA ping-pong density higher in dysgenic. Type IV: Higher in copy number in 9, higher in piRNA abundance in 160, F1 piRNA ping-pong pair density defined by maternal loading. Type V: Higher in copy number in 9, higher in piRNA abundance in 160, no ping-pong pairs detected. Type VI: Higher in copy number in 9, equivalent piRNA abundance in 9 and 160, F1 piRNA ping-pong density higher in non-dysgenic.

# Chapter 2:
# Effects of divergent piRNA profiles on somatic gene expression

## ABSTRACT

piRNAs are key regulators of transposable elements (TEs) that have been also shown to have strong effects on gene expression in the germline. The degree to which they contribute to intraspecific variation in gene expression in somatic tissues is poorly understood. Here, using a syndrome of hybrid dysgenesis in *D. virilis*, we evaluate how gene expression profiles differ in the soma of progeny whose mothers greatly differ in their piRNA profile. Moreover, because gonadal atrophy occurs in about 50% of progeny of the dysgenic cross, we can contrast the maternal effects of a divergent piRNA pool with the effects on gene expression that arise from loss of the germline. This allows us to determine whether variation in gene expression arising during hybrid dysgenesis could be explained by loss of germline, rather than contrasting maternal piRNA profiles. We have evaluated these contrasting effects by performing RNA sequencing on head and thorax tissue males and females of reciprocal crosses, with and without atrophied gonads. While there are substantial differences in gene expression in the heads and thorax of reciprocal progeny, we find that maternal piRNA contribution has little to no direct effect overall but may influence downstream expression of some genes through off-targeting in the germline. We also find that some reciprocal hybrid somatic expression differences are potentially caused by disparities in maternal mRNA loading. A comparison of expression profiles of carcasses with ovaries to genetically and epigenetically identical flies that lack ovaries, reveals changes in the expression of genes involved in metabolic processes and energy usage in tissues distant from the germline, with some gene expression differences equivalent to those previously seen in *D. melanogaster*.

# INTRODUCTION

While most genes exhibit Mendelian inheritance, parent-of-origin effects play a key role in determining variation in both gene expression and the phenotype. An extreme example of such parent-of-origin effects are the disparate fertility phenotypes in hybrid dysgenesis syndromes, where one direction of the cross leads to gonad ablation and sterility in the progeny, while the reciprocal cross maintains normal gonad development and function (Kidwell et al., 1977; Kidwell and Novy, 1979). This syndrome manifests from offspring inheriting transposable elements from the father that are absent in the mother (Bingham et al., 1982; Bucheton et al., 1984). Transposable elements are genetic sequences that replicate and can move autonomously, and with other repeat sequences make up approximately 14% of the *Drosophila virilis* genome (Drosophila 12 Genomes et al., 2007). Not only is their accumulation an important contributor to genome variation, but their movement can cause harmful genetic events such as chromosomal rearrangement and mutation. To limit their spread, an adaptive small-RNA immune system, PIWI-interacting RNA (piRNA), evolved in the animal germline (Aravin et al., 2007; Aravin et al., 2003; Brennecke et al., 2007). One interesting feature of the piRNA immune response in *Drosophila* is that piRNAs are transmitted exclusively through the maternal germline, deposited by the mother into the developing egg (Aravin et al., 2003). The reliance on this cytoplasmic contribution becomes problematic when piRNA pools are not complementary to the TE profile coming from the paternal genome. The lapse of immunity leads to an activation of TEs, a genome crisis that can result in apoptosis of germ line cells and subsequent sterility.

An essential question is how this genomic conflict that results in TE activation and sterility influences genomic expression of the organism as a whole. While the TE activation itself is limited to the germline,

there may be other cascading effects of divergent maternal piRNA pools that manifest in tissues of the soma.

For example, we and others have previously found that in addition to modulating TE expression, disparate piRNA profiles in strains can also influence gene expression through the off-targeting of genic sequences (Erwin et al., 2015; Le Thomas et al., 2014b; Shpiz et al., 2014). A common scenario of piRNA-mediated gene silencing is where a euchromatic TE insertion nucleates the formation of a piRNA-generating cluster, which spreads to the surrounding regions (Olovnikov et al., 2013; Shpiz et al., 2011). Notably, piRNA can silence at the transcriptional level by recruiting heterochromatin modifiers to establish silencing (Gu and Elgin, 2013; Sentmanat and Elgin, 2012; Sienski et al., 2012). This ability to nucleate heterochromatic silencing, paired with their epigenetic and transgenerational inheritance, may play a critical role in establishing the heterochromatin landscape in not only the adult but also in the next generation. To this point, we and others have shown that piRNA-mediated silencing is capable of paramutating naive alleles and that this robust form of silencing can be propagated across multiple generations (de Vanssay et al., 2012; Erwin et al., 2015). PIWI-interacting RNAs have also been found in other tissues beyond the germline (Ghildiyal et al., 2008; Jones et al., 2016; Lewis et al., 2018; Perrat et al., 2013; Rajasethupathy et al., 2012). For all of the reasons outlined above, it has been proposed that piRNA may be master regulators of gene expression. Thus, the dysgenesis syndrome can provide a unique insight into the role that divergence in piRNA profiles can play in determining somatic gene expression in progeny.

Defining the effects of maternal piRNA profile on somatic gene expression in progeny can be confounded by the effects of gonadal ablation that can occur during dysgenesis. This concern is further validated by the fact that other studies looking at *D. melanogaster* carcasses with and without germlines found dramatic differential expression in germline-distant tissues (Parisi et al., 2010; Parisi et al., 2004). The *Drosophila virilis* hybrid dysgenesis offers a fortuitous solution to this issue as hybrid sterility is not fully

penetrant in our system, leading to 50% of dysgenic progeny escaping the syndrome with intact and normal gonads. Thus, by comparing somatic tissue of dysgenic progeny with and without gonads against nondysgenic progeny that have gonads, we may identify the difference in gene expression attributed to gonad ablation rather than parental contribution of piRNAs. This not only allows us to compare reciprocal progeny without confounding factors from gonad tissue imbalance but also to investigate how gonadal ablation effects germline-distant tissue in *D. virilis* by comparing epigenetically identical hybrids within the dysgenic progeny that undergo versus escape sterility.

Here we use head and thorax mRNA data from dysgenic and nondysgenic flies, controlling for the amount of gonadal tissue, to tease apart the effects of sterility on gene expression from strict maternal effects. While we find that some of these expression differences may be due to downstream effects of genic off-targeting by piRNA, we find little evidence that overall expression differences are attributed to differences in maternal piRNA pools. Using carcasses from dysgenic progeny that differ in sterility phenotype also allows us to identify sets of genes that show variable patterns of gonad-dose dependence. We also find that *D. virilis* females exhibit similar responses to gonadal ablation as its 40-million-year-diverged relative, *D. melanogaster* (Obbard et al., 2012). Lastly, we report that *D. virilis* males show a milder and more variable response to gonadal ablation compared to females.

## METHODS

### Sample collection: non-dysgenic soma and dysgenic soma, varying in gonad amount

An overview of samples used for comparisons is given in Table 2.1

#### *Female dysgenic soma, varying in ovary amount*

For the dysgenic cross, a single strain 9 female was put in a vial with two strain 160 males. Samples for sequencing were obtained from F1 progeny of 8 dysgenic crosses. From each cross, F1 females were

collected and aged 12 to 15 days. Each sample was a pool of ~4 females from a single cross, with abdomens removed. For 3 of these crosses, we were able to collect pools from each ovary condition – carcasses from females with 2 normal ovaries, carcasses from females with 1 normal ovary and one atrophied ovary, and carcasses from females with two atrophied ovaries. This imbalance is due to the fact that there were not always enough replicates of a certain ovary condition within each cross. Additionally, two of the dysgenic atrophied ovaries samples (D18 and D12) had 2 flies instead of 4.

*Female non-dysgenic soma*

Samples for sequencing were obtained from 5 non-dysgenic crosses. From each cross, 4 females were collected and aged 12 to 15 days. In total, we collected 5 pools of 4 females with abdomens removed.

*Male dysgenic soma, varying in testis amount*

Samples for sequencing were obtained from 3 dysgenic crosses. F1 males were collected and aged 12 to 15 days. For each of the 3 crosses we collected a pool of 4 males with 2 normal testes and a pool of 4 males with 2 atrophied testes, abdomens removed.

*Male non-dysgenic soma*

Samples for sequencing were obtained from 3 dysgenic crosses. From each cross, 4 males were collected and aged 12 to 15 days. In total, 3 pools of 4 males with abdomens removed were collected.

RNA extractions and library prep

All samples were flash frozen in liquid nitrogen and RNA extracted using a standard Trizol extraction with 5PRIME Phase Lock Gel Heavy tubes to optimize recovery during phase separation. To prepare libraries, we used the Illumina TruSeq v2 RNA kit according to the instructions. We had a total of 29 libraries, 20 female libraries with 5 samples for each of the four conditions (nondysgenic, dysgenic 2 ovaries, dysgenic 1 ovary, dysgenic 0 ovaries) and 9 male libraries with 3 samples for each of the 3

conditions (nondysgenic, dysgenic 2 testes, dysgenic 0 testes). Libraries were pooled in groups of 9-10 per lane and run with single-end 100bp reads on an Illumina HiSeq 2500.

Analysis of mRNAseq data

For estimates of gene expression level, mRNA RPKM estimates were obtained using the RNA-seq tool in CLC, mapped against a masked *D. virilis* genome made with RepeatMasker. The genome was masked for individual repeats using the annotated *D. virilis* TE library from Erwin et al., 2015. The annotated TE library was included in the genome-wide analysis, allowing for expression estimates of TE families. Fold-analysis was performed by calculating RPKM(+.05) ratios. To test for differential expression analysis between conditions, CLC-generated mapped count data was used as input for the package DEseq (Love et al., 2014). GO analysis was performed with GOrilla (Eden et al., 2009) using *D. melanogaster* orthologs genes sorted by FDR p-value for the test of treatment effect. Clusters for gene expression patterns were generated using the R Package Mfuzz (Futschik and Carlisle, 2005), which employs a noise-robust soft clustering algorithm. Clusters were only generated for genes called significantly DE by DESeq for the ND versus D2 comparison and the D2 ovary versus D0 ovary comparison.

## RESULTS

Confounding effects of gonad dose on somatic gene expression in reciprocal hybrids

An overview of samples used for comparisons is given in Table 2.1 and a PCA for female samples is shown in Figure 2.1A. Here we define the comparison of the carcasses of females from the dysgenic cross that underwent gonadal ablation (dysgenic 0 ovaries; D0) versus those that escaped ablation (dysgenic 2 ovaries; D2) as the gonad effect contrast. The comparison between dysgenic females that escaped gonad ablation (D2) and nondysgenic females (ND) is the maternal effect contrast. We also compared the soma

of dysgenic females that underwent gonad ablation (D0) against nondysgenic females (ND) which incorporated both a maternal and gonad effect. An overview of the numbers of differentially expressed genes (FDR-adjusted p-values < .05) in female carcass comparisons is shown in Figure 2.1B.

The comparison that incorporated both a maternal and gonad effect (D0 vs ND) had the highest number of differentially expressed genes as this comparison included both differences in piRNA maternal profile and the most extreme differences in gonadal tissue. While other gonadal ablation studies have utilized genetic mutants or reciprocal hybrids (Parisi et al., 2010; Parisi et al., 2004), our study compares carcasses that are both genetically and epigenetically identical – a comparison that is only made possible with the stochastic and not fully penetrant property of the dysgenesis syndrome in *D. virilis*. When examining the effects of just gonadal ablation (D0 vs D2), we find that 371 genes are significantly differentially expressed (Fig. 2.1B, Fig. 2.2A). The majority (283/371) of genes in the gonad effect comparison are also present in the joint maternal and gonad effect comparison (Figure 2.1B), reflecting a clear gonadal ablation effect on somatic gene expression of reciprocal hybrids that are not matched for gonadal tissue.

To tease apart maternal effects in hybrids from effects of gonadal ablation in dysgenesis, we compared pools of reciprocal hybrid carcasses that were matched for gonadal tissue (D2 vs ND). This maternal effect analysis resulted in 889 significantly DE transcripts (Figure 2.1B, Figure 2.3A). We would expect a large portion of these maternal effect differences to also be significant in the joint maternal and gonad effect comparison (D0 vs ND) since both comparisons are between reciprocal hybrids. We find that the 889 genes are in fact correlated in their patterns of expression across the hybrid comparisons (Figure 2.1C), suggesting common maternal effects. However, only 335 of the 889 genes that were significant in the maternal effect comparison also showed significance in the joint maternal and gonad effect

comparison (Figure 2.1B). We wanted to assess if the genes that did not overlap were primarily a gonad-dose rather than maternal effect. We found a striking significant correlation between patterns of expression in the gonad effect comparison and the genes that did not overlap in the maternal effect and joint maternal and gonad effect analysis (R=0.92, p-value < 2.2E-16; Figure 2.1D), confirming that the expression of these genes was primarily influenced by gonad amount. Overall, these data strongly suggest that somatic gene expression comparisons in reciprocal hybrids can be dominated by gonadal ablation effects rather than maternal effects. It is interesting that the gonad effects were more apparent in the joint maternal effect and gonad effect comparison (D0 vs ND) than in just the gonad effect comparison (D0 vs D2). Even though these dysgenic females escaped sterility, there may still subtle differences in some components of the ovary like the occasional agametic ovarioles observed in some *D. virilis* crosses (Sokolova et al., 2013) that may become more apparent when compared to nondysgenic females that did not undergo a genomic crisis.

Concerted effects of germline ablation on germline-distant tissues across species

While there have been prior studies that have assessed the effects of germline ablation on germline-distant tissues in *D. melanogaster* (Parisi et al., 2010; Parisi et al., 2004), these comparisons were complicated by additional confounding factors because comparisons were either between genetically distinct mutants, or in the case of reciprocally crossed individuals, had different maternal effects. Here we are able to take advantage of the not fully penetrant dysgenesis syndrome and compare genetically and epigenetically identical carcasses differing only in gonad amount. It is worth emphasizing that the germline ablation that happens in *D. virilis* is an early event in embryogenesis, with no evidence of a cascading TE mobilization effect outside of the germline that may confound somatic tissue expression (Lozovskaya et al., 1990; Rozhkov et al., 2011; Sokolova et al., 2013). The gonad effect comparison, where carcasses of epigenetically and genetically identical flies differed in the amount of ovarian tissue (D2 versus D0)

showed significant differential expression (DE) of 371 genes (Fig. 2.2A). A gene ontology enrichment revealed that the processes these genes are involved in were primarily metabolic, with top terms associated with oxidation reduction (GO:0055114), small molecule metabolic process (GO:0044281), carbohydrate metabolic process (GO:0005975), cofactor metabolic process (GO:0051186), NADP metabolic process (GO:0006739). Other terms of interest include sterol, cholesterol, and lipid transport (GO:0015918; GO:0030301; GO:0006869) (full list of GO terms is in Supp. Table 2.1). Transcripts associated with metabolic processes being differentially expressed between carcasses that differ in gonad amount is not surprising since while much of gonad development is cell-autonomous, gene expression in other germline-distant tissues must be modulated to mediate overall energetics and behavior that coincide with reproduction. This enrichment for metabolic processes also aligns with the previous germline ablation studies in *D. melanogaster* (Parisi et al., 2010; Parisi et al., 2004).

The common gene ontology enrichment for metabolism in both *D. virilis* and previous studies in *D. melanogaster*, while not surprising, made us wonder exactly how granular the similarities in response to gonadal ablation was across species. To determine this, we compared the list of 449 genes that were differentially expressed in germline-present (flies that develop with a wild-type germline) and germline-naïve females (flies that develop in the absence of a functioning germline due to maternal effect sterile mutations) in Parisi et al., 2010 with our 371 differentially expressed genes that had orthologs in *D. melanagoster* (322 genes). We found 32 genes that showed up as differentially expressed in both studies which is significantly more than expected (hypergeometric probability test, p-value < .0001; http://nemates.org/MA/progs/overlap_stats.html) (Table 2.2). One of the most significant differentially expressed genes that appears across species is *larval serum protein 2* (GJ21616). This gene functions in protein storage, indicating that this pathway may be commonly utilized to alter energy storage in the absence of germline maintenance. Another interesting overlap was the downregulation of transcript *exuperantia* (*exu*) in the soma of flies that lacked germlines (GJ18765). *Exuperantia* plays a role in

localizing *bicoid* and *oskar* mRNA in the establishment of anterior/posterior axes in the developing oocyte (Hazelrigg et al., 1990; Riechmann and Ephrussi, 2004; Wang and Hazelrigg, 1994). It is unclear what its role is in somatic tissues, but it is worth noting that the expression of this gene may not be germline-autonomous.

It is also worth noting that the directionality of expression between germline-conditioned and germline-naïve carcasses across these two species was not the same for all genes. For example, there were also two chorion protein transcripts, *Cp36* and *Cp38*, that were differentially expressed in both studies, but in opposing directions. Chorion proteins build up the eggshell and are under tight temporal and spatial regulation during oogenesis. These chorion transcripts have not previously been reported to be expressed outside of the ovaries, although modENCODE data indicates that they are moderately expressed in *D. melanogaster* adult carcasses and have high expression in digestive systems (www.modencode.org). The fact that these transcripts show up as differentially expressed in carcasses varying in gonad amount across species underscores their tight association with reproduction and also calls for additional inquiry into how the germline-distant expression of these genes influences gametogenesis. It is also interesting that directionality of expression for these genes in the gonad-ablated soma differs between *D. virilis* and *D. melanogaster*. It is possible that some of this discrepancy is coming from species differences between *D. melanogaster* and *D. virilis* since a comparison across species found that chorion protein (Cp) regulation is evolving rapidly (Niepielko et al., 2014). It is also possible that while these are common targets to modulate during nutrient and energy allocation, their responses in the soma are variable.

Some gene expression is influenced by ovary-dose

The stochasticity of the germline ablation that occurs in the hybrid dysgenesis syndrome of *D. virilis* is highlighted by the fact that some flies show a unilateral gonad phenotype, exhibiting one ablated ovary and one normal ovary that has escaped dysgenesis. These intermediate phenotypes provide a unique opportunity to compare how variation in the amount of ovarian tissue affect gene expression. Specifically, we were curious if expression of genes that were altering metabolism and energy in response to germline ablation were sensitive to the dose of ovary tissue. To do this, we used a clustering R package Mfuzz (Futschik and Carlisle, 2005) to group the 322 differentially expressed genes that had orthologs between carcasses of ablated and non-ablated ovaries across dysgenesis progeny with 0, 1 and 2 ovaries. This allowed us to classify gene expression patterns. Cluster 1 and 2 consisted of 122 transcripts that were expressed in a pattern consistent with a dose-dependent effect of gonad tissue abundance (Fig. 2.2B). A gene ontology analysis for *D. melanogaster* orthologs of these genes revealed significant enrichment for oxidation-reduction process (GO:0055114), galactose metabolic process (GO:0006012), and co-translational protein targeting to membrane (GO:0006613). This suggests that the energy-related transcripts required to maintain reproductive tissue may be modulated based on the amount of tissue present.

We also identified 133 genes that demonstrated a response to ovary-dose in a binary way, showing consistent expression if ovarian tissue was present, regardless of whether that was in the form of 1 versus 2 ovaries (Fig. 2.2B). However, these genes did not show any significant enrichment for any gene ontology. Moreover, a similar somatic expression profile between carcasses of one versus two ovaries is underscored in the PCA, which did not categorically separate these samples (Figure 2.1A). Additionally, differential gene expression analysis comparing carcasses with two versus one ovary only found two genes, GJ18385 and GJ10579, to be significantly differentially expressed. While GJ10579 lacked an ortholog, GJ18385 was highly significant for DE (FDR adjusted p-value < 9.27E-11) and its ortholog was identified as having serine-type endopeptidase activity. Serine-type peptidases cleave peptide bonds in

proteins during digestion and some have been shown to respond in sex-dependent ways during times of nutrient stress, highlighting their potential role in resource allocation tradeoffs between reproduction and longevity (Bauer et al., 2006). Serine-type peptidases were also significantly DE in the Parisi et al., 2010 gonadal ablation study, indicating that this differential expression between carcasses of 1 and 2 ovaries could be a concerted response especially sensitive to gonad dose. The lack of differentially expressed genes in the soma of flies with one versus two ovaries, even though the clusters show a dose response, can be explained by the fact that carcasses of one ovary show intermediate expression which does not manifest in as large of a difference when compared to carcasses of flies with two ovaries.

Carcasses of males that vary in testis number exhibit fewer gene expression effects in the soma

We also sequenced mRNA of dysgenic male carcasses that underwent gonadal ablation and carcasses of dysgenic males that had escaped ablation. In this case, we did not perform the analysis of males with only one atrophied gonad. In contrast to what was observed in females, males only showed 70 genes significantly differentially expressed between carcasses differing in gonadal state (Fig. 2.2C). Gene ontology analysis revealed that these genes were also primarily metabolic, with top terms including generation of precursor metabolites and energy (GO:0006091), electron transport chain (GO:0022900), respiratory electron transport chain (GO:0022904), oxidation-reduction process (GO:0055114), and ATP metabolic process (GO:0046034) (Supp. Table 2.2). While only 6 genes overlapped in the comparisons between females with and without ovaries and males with and without testes, DE genes in both gonadal ablation sets were positively correlated across sexes (Fig. 2.2D). One reason we may observe fewer genes differentially expressed in male carcasses with and without gonads is if males devote fewer energetic resources to maintaining gonadal tissue than females. The disparity between males and females could have also been due to a difference in power since 3 pooled replicates per treatment were collected for males compared to the 5 pooled replicates per treatment that we collected for females. To eliminate this as

a factor, we reperformed the female analysis with sets of only 3 samples. Even in the most extreme scenario where we sub-selected samples with the greatest amount of variation, females still showed 122 genes DE between treatments, almost double the DE seen in males. The 2010 Parisi et al., study also showed that germline-conditioned versus germline-naïve males had 2/3 less DE genes compared to the female comparison. Unlike in females, we did not find any genes that were common to our and Parisi's data sets. The fact that both our study and Parisi et al., show fewer DE genes in males than females supports the idea that males may devote fewer energetic resources to maintaining gonadal tissue than females.

Maternal effects drive extensive somatic gene expression variation in reciprocal hybrid carcasses

We found 889 transcripts differentially expressed in the maternal effect, which compared pools of nondysgenic and dysgenic female carcasses that were matched for gonadal tissue (ND vs D2) (Fig. 2.3A). We sought to test whether these DE genes could also be an extension of the gonadal-dose result. To evaluate this possibility, we first compared whether there was overlap between genes that were differentially expressed in the gonad effect analysis and the genes differentially expressed in the maternal effect comparison. We found that only 71 genes overlapped between the two sets (Figure 2.1B). Gene ontology enrichment for this set indicated that these genes were primarily metabolic, with top terms of galactose metabolic process and oxidation-reduction process (Supp. Table 2.3). If gonad-dose was the underlying factor in DE, we would expect the expression of these genes to be correlated across both comparisons. However, we did not find any significant correlation in expression (Pearson's product-moment correlation = 0.163, p-value < 0.17). We also did not find a positive correlation in the ratio of expression for the broader 889 genes DE between ND/D2 carcasses and the ratio of expression of D2/D0 again indicating that gonad-dose was likely not influencing expression differences in both comparisons (Pearson's product-moment correlation = −0.23, p−value < 7.954e−12; Fig. 2.3B).

We decided to visualize patterns of gene expression in a continuum from dysgenic flies with no ovaries to nondysgenic flies with ovaries to identify potential gonad dose effects. If gene expression depended on gonad dose, we would expect to see a somewhat linear pattern of expression consistent with our gonad continuum of 0, 1, and 2 ovaries in dysgenic flies to nondysgenic flies representing the greatest amount of gonad. We found 319 genes that were consistent with this pattern (cluster 2; Figure 2.3C), suggesting that even dysgenic flies with two ovaries seem to be affected by subtle dose effects of gonad.

The remaining genes fell into clusters that did not show this pattern (clusters 1 and 2; Figure 2.3C). These differentially expressed genes are thus likely due to maternal effects, not due to size or quality of gonad between nondysgenic and dysgenic flies. When we performed a gene ontology analysis for genes in the maternal effect clusters 1 and 3, some terms of interest that came up were regulation of membrane potential (GO:0042391), behavior-related terms (GO:0007610, GO:0007626, GO:0048512), circadian rhythm (GO:0007623, GO:0030431), axon and neuron guidance (GO:0007411, GO:0097485), regulation of growth (GO:0040008, GO:0001558, GO:0048589), and eye development (GO:0042051, GO:0045467, GO:0042462, GO:0009584, GO:0007602). These gene ontology enrichment terms suggest a potential relationship between maternal effects and neuron guidance and behavior. There are some maternally deposited mRNAs, like transcription factor *lola*, that promote axon growth. It is possible that differences in maternal deposition of factors like these might have long-term effects on brain development and behavior (Gates et al., 2011). Interspecific differences in maternal mRNA contribution have been reported, but how much maternal mRNA might vary within a species remains poorly understood (Paris et al., 2015). The *lola* ortholog in *D. virilis*, GJ20132, was significantly DE between nondysgenic and dysgenic carcasses (FDR-adjusted p-value < .003), with increased expression in nondysgenic offspring. Behavior may be especially sensitive to maternal effects as another study also identified behavior and

neurotransmission gene ontology enrichment in intraspecific *D. melanogaster* hybrids (Gibson et al., 2004).

Our TE-masked genome allowed us to incorporate transposable elements in our DE analyses. We did not find any TEs significantly differentially expressed in the soma of reciprocal hybrids, overall supporting the idea that TE activation in the dysgenesis syndrome is limited to the germline. However, we did find that in the comparison of ND to D0, the TE *Penelope* was borderline significant for expression differences in the soma. This is notable because *Penelope* is thought to be one of the major players in inducing dysgenesis as it varies in active and degenerative copy number between strains (Erwin et al., 2015; Evgen'ev, 2013; Evgen'ev and Arkhipova, 2005; Lyozin et al., 2001). This could be another downstream effect of piRNA loading where establishing chromatin silencing against *Penelope* may lapse in early development, and manifest into adult somatic tissues.

Maternal piRNA contributions do not directly determine somatic gene expression in reciprocal hybrids

When it was discovered that piRNA were not exclusively derived from transposable element sequences and could even be generated from genomic clusters, speculation followed about their possible role as master regulators of gene expression. In line with this idea, piRNA have been found to regulate the degradation of *nanos* mRNA in the developing embryo to facilitate axis patterning (Rouget et al., 2010) and in mouse spermiogenesis, piRNA instruct an even more extensive mRNA decay (Gou et al., 2014). These studies and others make it clear that piRNA have the potential to mediate long-range developmental effects by regulating mRNA in the developing embryo.

Because we have previously shown that the two parental strains involved in the *D. virilis* hybrid dysgenesis have disparate piRNA pools that regulate germline gene expression in a maternal-effect way, this syndrome provides a unique opportunity to measure how these differences may affect global gene expression in the adult. Specifically, we asked whether the somatic gene expression differences we saw between nondysgenic and dysgenic females could be attributed to maternal piRNA contribution. We found no correlation between differences in somatic gene expression in reciprocal hybrid carcasses and genic piRNA in the germline of parental strains (Fig. 2.3D). This supports our previous finding in which the piRNA-mediated silencing we observed for *cdi* and *oysgedart* in the germline did not extend to the somatic ovary tissue (Erwin et al., 2015).

Potential downstream effects of piRNA-regulated genes on somatic gene expression

While we do not find that piRNA are directly contributing to somatic gene expression differences in dysgenic and non-dysgenic progeny, it is possible that they may indirectly influence genes downstream of the ones they do target in the germline. Previously, we found that the maternal-effect repression of the gene *cdi* was caused by genic piRNA differences between strains (Erwin et al., 2015). Because *cdi* is an embryonic development gene, it is possible that its silencing may influence other downstream genes that it interacts with. The *D. melanogaster* ortholog of *cdi* is in fact reported to suppress the gene *sevenless,* which is required for proper cell fate of R7 cells in *Drosophila* eye development (Sese et al., 2006).

We checked to see if these genes that have been shown to genetically interact with *cdi* exhibited differential expression in the carcasses of reciprocal hybrid progeny. While only borderline significant (FDR-adjusted p-value <.06), *sevenless* expression is increased in the nondysgenic progeny, in the direction of the cross where *cdi,* the suppressor of s*evenless,* is silenced in the germline. *Brother of sevenless,* (*boss)* which is affected by *sevenless,* is also slightly increased, although not significantly, in

expression in the ND cross. It is worth noting that R7 cell development (GO:0045467), the cells in which the sevenless pathway is active, along with other eye development terms were enriched in the gene ontology analysis between nondysgenic and dysgenic carcasses. However, when we removed the sevenless pathway genes from the analysis, the GO terms remained significant, indicating that there may be other genes that are simultaneously altered, either through downstream effects or in alternate pathways. To this point, *lola* was also one of the genes that contributed to the enrichment of eye development terms, as it has also been found to contribute to determining cell fate in the eye. This underscores the difficulty of dissecting determinants of piRNA in the face of master regulators. Overall, we did not find strong evidence that the piRNA-mediated suppression of *cdi* in the germline contribute to long-range development effects in the soma.

Few similarities between female and male nondysgenic to dysgenic comparisons

*Drosophila virilis* dysgenic males also show the not fully penetrant sterility phenotype. Even though these reciprocal males are not genetically identical since they have different Y chromosomes, we were curious to see if we could identify overlaps by comparing our female ND/D comparisons to the male ND/D comparison to tease out common maternal effects. While we did not find any overlaps in DE expressed genes, we did find a slight correlation between female ND/D DE genes and expression in the male comparisons (Fig. 2.4). Notably, some of the most highly expressed genes that were correlated across the comparisons were genes from the mitochondrial genome. It is hard to know if or how these mitochondrial expression differences between reciprocal hybrids affect the differential gene expression we observe between nondysgenic and dysgenic females. It is worth pointing out however that any differences in mitochondrial genomes do not contribute to the dysgenesis syndrome itself as dysgenic F1 flies with mitochondria from the reactive strain are fully protective against dysgenesis when backcrossed to the inducer strain. The fact that there was not much similarity between nondysgenic and dysgenic carcass

gene expression across males and females could be due to the extensive variation that exists in sex-specific somatic expression (Arbeitman et al., 2016).

## DISCUSSION

piRNAs are critical in the regulation of transposable elements (TEs). Their ability to mediate heterochromatin formation and their transgenerational epigenetic inheritance make them candidates for master regulators of gene expression. While they have been shown to have strong effects on gene expression in the germline, the degree to which they contribute to variation in gene expression in somatic tissues remains poorly understood. Here we used a syndrome of hybrid dysgenesis in *D. virilis* to assess gene expression profiles in the soma of progeny whose mothers greatly differ in their piRNA profile. We additionally compared somatic tissues of reciprocal progeny with and without gonads to identify the difference in gene expression attributed to gonad ablation rather than the maternal effects of piRNAs.

We found that the majority of gene expression differences between reciprocal dysgenic hybrids are in fact caused by underlying differences in gonad amount. When we match reciprocal hybrids for gonad amount, we are able to tease out some maternal effects, but residual gonad dose effects still exist. This suggests that while some dysgenic hybrids escape sterility, there are still subtle somatic expression differences that can be attributed to subtle differences in gonad quality or performance. In teasing apart maternal-effect genes, one of our most striking findings is that the enrichment we see for genes involved in neuronal function, behavior, and eye development in our reciprocal hybrids may be a consequence of disparate piRNA or maternal mRNA contributions. While there is a lack of evidence for the direct role of piRNA as

master regulators of gene expression, we provide support for the possibility of downstream long-range effects from genic off-targeting that occurs in the germline. Some of the differences we observe in neuron development may be directly caused by piRNA acting directly in those tissues, since findings show that piRNA are not exclusive to the germline. Possible future studies include sequencing small RNA in the heads of these reciprocal hybrids and an investigation into whether the differences we observe in somatic gene expression lead to measurable behavioral differences in reciprocal hybrids.

We also report here that even when controlling for genetic and epigenetic effects, germline ablation has major consequences for somatic gene expression with regards to metabolism and overall energy allocation. There may also be commonalities in the genetic pathways that are modulated in response to gonadal ablation as we found overlaps in gene sets between our experiment and a similar experiment in *D. melanogaster*. While there is some gene expression overlap in responses between males and females with regards to gonad ablation, the maintenance of male sex organs in *D. virilis* seems to be more variable and have lesser consequence on gene expression of somatic tissues outside of the abdomen compared to maintenance of female sex organs.

FIGURES

**Figure 2. 1. Overview of comparisons between nondysgenic and dysgenic carcasses varying in gonad amount.** A) Principal component analysis of female carcass samples of log transformed normalized count data from DeSeq2. Dysgenic female carcasses, varying in the amount of reproductive tissue, tend to cluster more closely along principal component 2 to the exclusion of genetically identical samples with different maternal contributions. B) Overview of number of genes significantly DE (as called by DESeq) across female carcass comparisons. "D0" refers to dysgenic hybrids that undergo gonad ablation and have zero ovaries. "D2" and "ND" refer to reciprocal hybrids that have two ovaries. The D2 vs D0 comparison represents genetically and epigenetically identical dysgenic hybrids varying only in gonad amount, while the other two comparisons are between genetically identical but epigenetically different reciprocal hybrids varying in gonad amount (ND vs D2 and ND vs D0). C) Genes DE in reciprocal hybrids matched for gonad dose (ND vs D2) that don't overlap with significantly DE genes from reciprocal hybrids varying in gonad dose (ND vs D0), class A, show correlated expression patterns in both reciprocal hybrid comparisons (ND vs D2 and ND vs D0), suggesting commonality in maternal effects across treatments. D) Genes DE in reciprocal hybrids varying in gonad dose (ND vs D0) that don't overlap with significantly DE genes from reciprocal hybrids matched for gonad dose (ND vs D2) show a strong correlation with gene expression patterns in dysgenic hybrids varying only in gonad dose, suggesting that the majority of gene expression differences between ND and D0 can be attributed to effects of gonad ablation rather than maternal differences.

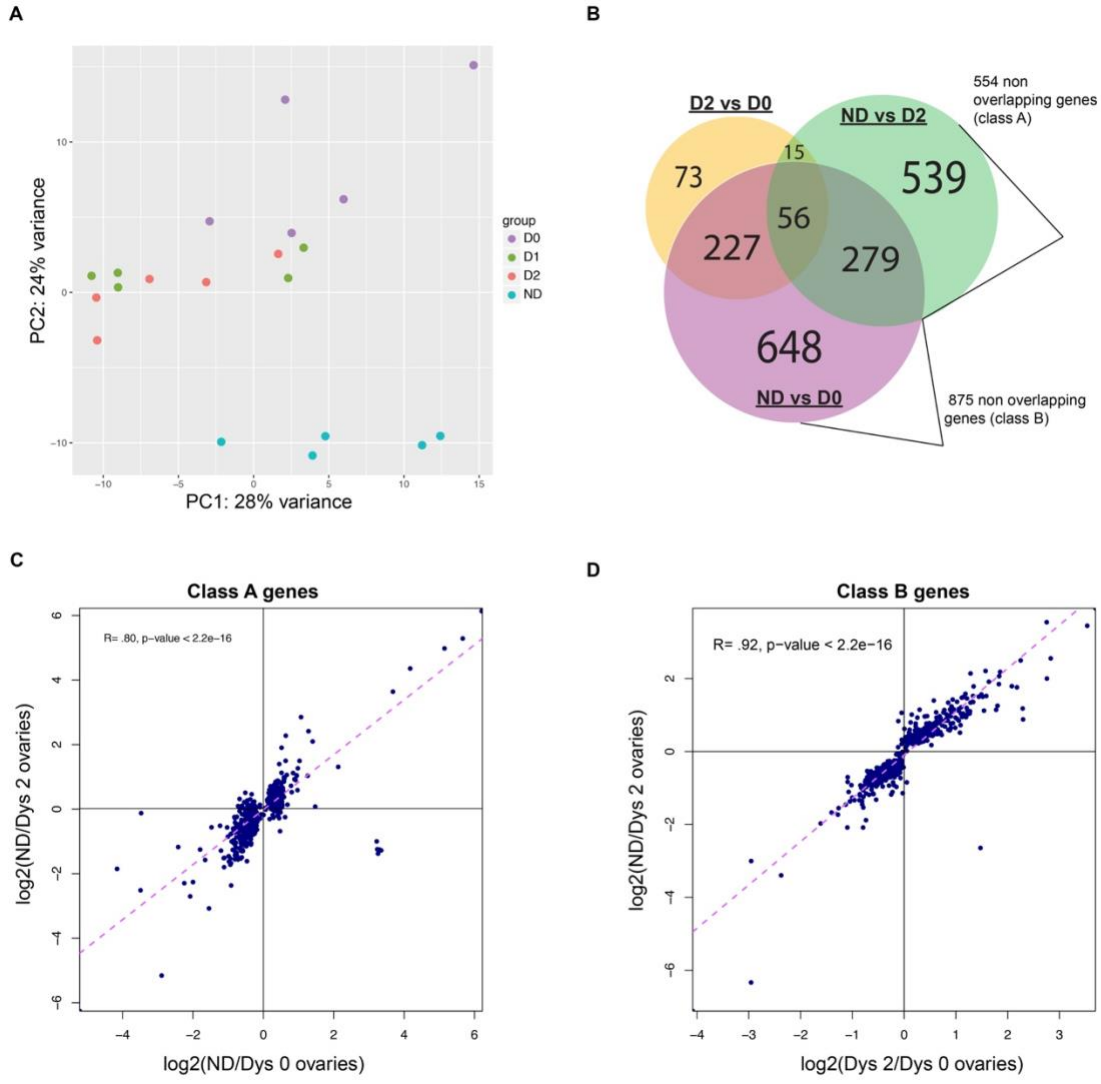


**Figure 2. 2. Germline-distant genes affected by gonadal ablation in females and males**. A) Average log2(RPKM+.5) of pooled dysgenic female carcasses differing in degree of gonad ablation. 371 transcripts (purple) are differentially expressed as called by DESeq, with a cutoff of an FDR adjusted p-value of .05. B) Gene expression patterns using the package Mfuzz of the DE genes between dysgenic females differing in number of ovaries present. Clusters A and B represent a gonad dose effect while clusters 3 and 4 show a dominant effect of gonad presence. C) Average log2(RPKM+.5) of pooled dysgenic male carcasses that differed in degree of gonad ablation. D) Ratios of gonad vs no gonad carcasses of genes DE across males and females.

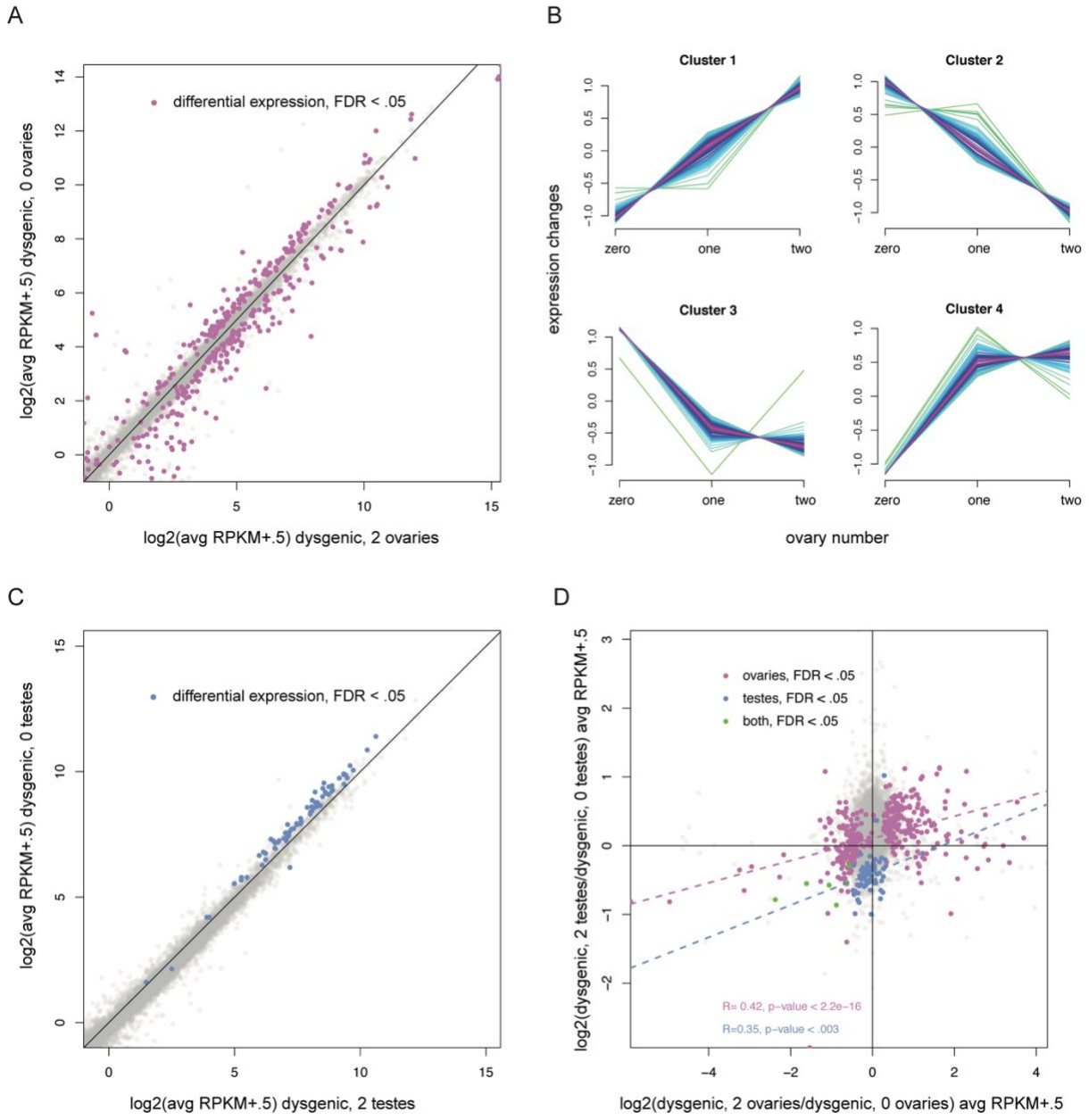


**Figure 2. 3. Expression differences between ND and D2 carcasses not attributed to maternal piRNA provisioning**. A) Average log2(RPKM+.5) of pooled nondysgenic and dysgenic female carcasses that were matched for gonadal tissue. 889 transcripts are differentially expressed as called by DESeq2, with a cutoff of an FDR adjusted p-value of .05. B) Ratios of ND/D plotted against dysgenic carcass w/2 ovaries vs 0 ovaries to see if DE transcripts were an extension of gonad dose. C) Gene expression patterns using Mfuzz of the 889 DE genes between ND and D2 across female samples. D) Ratio of piRNAs in parental strains (9/160) to ND/D2 progeny. piRNA differences between parental strains do not correlate to differences in gene expression in reciprocal progeny.

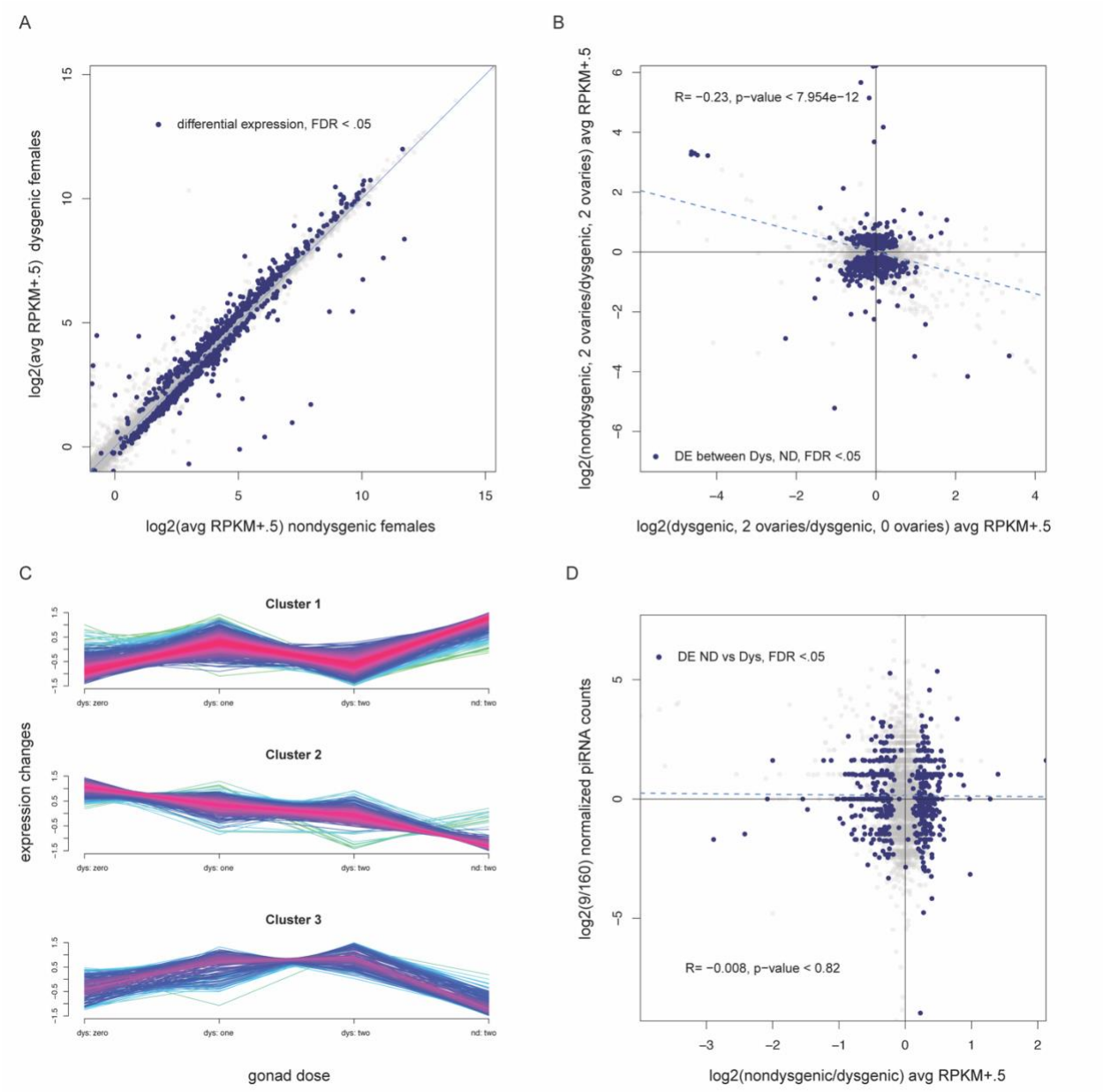
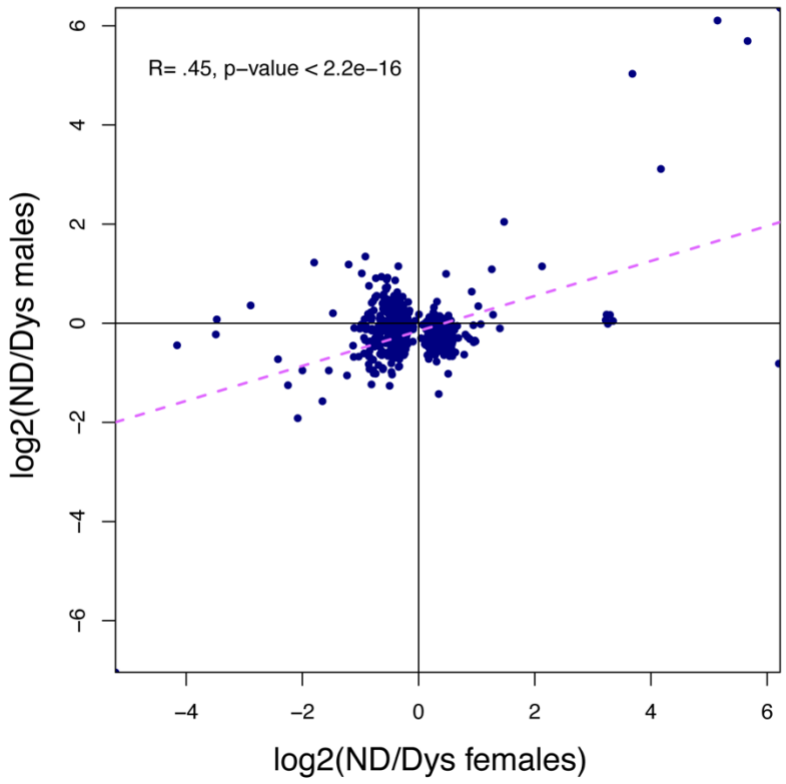


**Figure 2. 4. Reciprocal hybrid comparisons across males and females**. There is a slight correlation between genes DE in the ND/D female comparison and the ratio of expression between ND/D in males, mostly driven by a handful of highly expressed mitochondrial genes (4 highest points in top right quadrant). Many genes show opposite directionality between ND and D across sexes (upper left and lower right quadrants).



# TABLES

**Table 2. 1. Overview of samples**. Each sequencing library is a pool of ~4 F1 fly carcasses from a single cross, abdomens removed.

| | ***Female*** | | ***Male*** | |
|---|---|---|---|---|
| ***# Gonads*** | *ND* | *D* | *ND* | *D* |
| *2 (wt)* | 5 | 5 | 3 | 3 |
| *1* | | 5 | | |
| *0 (atrophied)* | | 5 | | 3 |

**Table 2. 2. Similarities in response to gonadal ablation across species.** Genes DE between D. melanogaster germline-conditioned and germline-naïve female carcasses in Parisi et al., 2010 and between dysgenic female carcasses with 2 versus 0 ovaries in D. virilis. Some gene ontology (GO) terms associated with orthologs in D. melanogaster are highlighted. FDR-adjusted p-value for our study using DESeq2, and the Parisi p-value is for an ANOVA. "+" and "-" indicate direction of fold-change.

| ***D. virilis* ID** | ***D. mel* common symbol** | **Selected GO terms** | **FDR-adj. p-val** | **Parisi ANOVA p-val** | **Germline condition vs germline naïve (Parisi)** | **Dysgenic (2 ovary) vs dysgenic (0 ovary)** |
|---|---|---|---|---|---|---|
| *GJ21616* | Lsp2 | GO:0005616 larval serum protein complex | 6.85E-08 | 0.0025 | -3.08 | -2.89 |
| *GJ18534* | CG9914 | | 5.63E-07 | 6.00E-04 | -2.18 | -1.84 |
| *GJ20385* | CG4847 | GO:0004216 cathepsin K | 5.63E-07 | 0.0037 | -1.89 | -1.69 |
| *GJ22204* | CG3609 | | 1.79E-05 | 0.0032 | -1.51 | -1.58 |
| *GJ11469* | Pdi | GO:0006457 protein folding, GO:0005783 endoplasmic reticulum, GO:0003756 protein disulfide isomerase | 1.88E-05 | 0.0016 | -1.25 | +1.39 |
| *GJ15366* | Tsf1 | GO:0006826 iron ion transport, GO:0006952 defense response | 2.79E-05 | 0.0019 | -2.18 | -1.64 |
| *GJ13795* | CG5618 | GO:0004782 sulfinoalanine decarboxylase | 3.11E-05 | 0.0053 | -1.25 | +2.29 |
| *GJ14566* | Cyp6d5 | GO:0015034 cytochrome P450 | 6.05E-04 | 4.00E-04 | -2.05 | -2.19 |
| *CP36* | Cp36 | GO:0007306 insect chorion formation, GO:0007292 oogenesis | 9.17E-04 | 0.0058 | +2.06 | -31.03 |
| *GJ10723* | CG12813 | | 1.09E-03 | 0.0026 | -2.54 | +2.34 |
| *GJ21384* | CG15674 | | 1.57E-03 | 0.0047 | -1.53 | +1.65 |
| *GJ22720* | CG3739 | GO:0008236 serine-type peptidase | 2.19E-03 | 9.00E-04 | -2.97 | -4.82 |
| *GJ23961* | CG5793 | | 2.36E-03 | 0.0035 | -1.35 | -1.81 |
| *GJ10508* | CG1544 | GO:0004591 oxoglutarate dehydrogenase | 2.36E-03 | 0.0043 | -1.19 | +1.99 |
| *GJ11518* | CG7529 | GO:0004091 carboxylesterase | 2.49E-03 | 0.0053 | -1.29 | +1.39 |
| *GJ21882* | Transaldolase | GO:0005737 cytoplasm, GO:0004801 transaldolase, GO:0009052 pentose-phosphate shunt | 3.93E-03 | 6.00E-04 | -1.65 | -1.59 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *GJ21687* | CG18067 | | 4.35E-03 | 0.0058 | -1.34 | -1.75 |
| *GJ18632* | Cp38 | GO:0007306 insect chorion formation | 9.18E-03 | 0.007 | +2.79 | -60.17 |
| *GJ24141* | CG6277 | GO:0004806 triacylglycerol lipase | 1.13E-02 | 0.0037 | -1.7 | +3.11 |
| *GJ18312* | CG3841 | GO:0004091 carboxylesterase | 1.48E-02 | 0.0088 | -1.59 | +1.47 |
| *GJ11812* | CG9466 | GO:0004559 alpha-mannosidase, GO:0005764 lysosome | 2.26E-02 | 0.008 | -1.98 | -1.53 |
| *GJ24063* | Lectin-galC1 | GO:0006952 defense response, GO:0005530 lectin, GO:0005531 galactose binding lectin | 2.32E-02 | 0.0061 | -1.45 | -1.37 |
| *GJ18704* | CG9672 | GO:0004252 serine-type endopeptidase | 2.35E-02 | 0.0039 | -1.96 | +2.16 |
| *GJ21836* | PGRP-SC2 | GO:0006955 immune response, GO:0005887 integral to plasma membrane, GO:0016019 peptidoglycan recognition, GO:0006952 defense response | 2.46E-02 | 6.00E-04 | -1.36 | +2.17 |
| *GJ15227* | CG9498 | | 3.00E-02 | 0.0057 | -1.96 | +1.39 |
| *GJ17181* | CG15406 | GO:0005353 fructose transporter | 3.03E-02 | 0.0017 | -1.61 | +4.89 |
| *GJ10176* | CG4462 | GO:0008513 organic cation porter | 3.03E-02 | 5.00E-04 | -1.21 | -1.25 |
| *GJ18538* | regucalcin | GO:0005509 calcium ion binding | 3.20E-02 | 0.001 | -1.55 | -1.39 |
| *GJ22429* | Jheh3 | GO:0006719 juvenile hormone catabolism, GO:0008096 juvenile hormone epoxide hydrolase, GO:0005792 microsome | 3.47E-02 | 0.0021 | -1.45 | -1.30 |
| *GJ22928* | CG17820 | | 4.06E-02 | 0.0049 | -2.6 | +2.03 |
| *GJ12938* | CG6602 | | 4.16E-02 | 0.0038 | -2.38 | +6.59 |
| *EXU* | exu | GO:0007283 spermatogenesis, GO:0008359 regulation of pole plasm bicoid mRNA localization, GO:0007300 nurse cell/oocyte transport | 4.52E-02 | 0.0082 | +1.77 | +1.28 |

# Chapter 3:
# Diverse changes in gene expression in the aging *Drosophila* ovary are not associated with a global release of transposable element expression

# ABSTRACT

Redistribution of heterochromatin during aging has been linked to the de-repression of transposable elements and an overall loss of gene regulation in the soma. Whether or not epigenetic factors such as heterochromatin marks are perturbed in reproductive and germline tissues is of particular interest because some epigenetic factors are known to transmit across generations. Additionally, the relative contribution of factors intrinsic or extrinsic to the germ line have in reproductive decline remains unknown. Using mRNA sequencing data from late stage egg chambers in *Drosophila melanogaster*, we show that age-related expression changes occur in genes residing in heterochromatin, particularly on the largely heterochromatic 4$^{th}$ chromosome. We further identify a striking age-related reduction in mitochondrial transcripts that we can attribute to the somatic tissues. Other than a modest increase in overall TE expression in the aging germline, we find no global TE derepression in reproductive tissues. Rather, the observed effects of aging on TEs are primarily strain and family specific. These results indicate unique responses in somatic versus germline tissue with regards to epigenetic aging effects and suggest that the global loss of TE control observed in other studies may be specific to certain tissues, genetic backgrounds and TE family. This study also demonstrates that while age-related effects can be maternally transmitted, the germline is generally robust to age-related changes.

# INTRODUCTION

The age-related decline of the reproductive system has important consequences for evolution because reproductive success determines the fitness of an organism. Since the majority of aging studies focus on overall somatic decline, relatively little is known about the causes of reproductive aging. In humans, progressive delays in childbearing are leading more people to confront the reduced fertility and fecundity that accompanies advanced age (Billari et al., 2007; Dunson et al., 2002). Reproductive senescence is not unique to mammals, however. The invertebrate model *Drosophila melanogaster* shows a progressive decline in egg production at middle age, thought to be partially caused by a reduction in germline stem cell proliferation and decreased survival of developing eggs (Zhao et al., 2008). Possible mechanisms underlying these changes include reduced ovariole number, decreased rates in germline stem cell division, and apoptosis in egg chambers of older females (Pan et al., 2007; Zhao et al., 2008). Animals may have conserved mechanisms to regulate reproductive decline and control the relationship between reproduction and lifespan. Not only have mechanisms of gametogenesis been found to be similar across organisms, but the control of ovulation has also been shown to be conserved between *Drosophila* and humans (Sun and Spradling, 2013). Because *Drosophila* is an established model for studies of both reproductive and somatic aging, we used it here to examine age-related genome-wide expression changes in the germline and broader reproductive tissues.

While genetic causes have long been shown to determine longevity - through either inherited or somatic mutation, non-genetic contributions are also proving to be major factors. Epigenetic chromatin marks play an essential role in the maintenance of genome integrity through their repression of genes, repeat sequences, and transposable elements (reviewed by (Putiri and Robertson, 2011). The misregulation of epigenetic marks has been associated with many diseases, including kidney disease, nuerodegenerative diseases, and cancer (Figueroa-Romero et al., 2012; Muntean and Hess, 2009; Smyth et al., 2014).

Recently, epigenetic mis-regulation has been attributed to playing a key role in the aging process. In particular, the landscape of silent heterochromatin has been shown to redistribute in aged stem cells and cells of the soma, leading to aberrant gene expression (Bell et al., 2012; De Cecco, 2013a; Jiang, 2013; Larson et al., 2012; Shah et al., 2013; Wood et al., 2010). An additional consequence of this redistribution of heterochromatin is the observed de-repression of transposable elements in the soma during aging, notably in brains and fat body of *Drosophila*, and in a variety of other organisms including mammals (Chen et al., 2016; De Cecco, 2013b; Li et al., 2013; Maxwell et al., 2011; Patterson et al., 2015). Although interesting for the biology of aging, somatic cells do not affect future generations. Surprisingly, little is known about whether epigenetic changes occur in aged reproductive tissues and germline cells that may transmit these non-genetic but potentially heritable effects to the next generation.

The germ line is considered an immortal cell lineage. Thus, germ cells have unique strategies to faithfully transmit DNA indefinitely, such as greater telomerase maintenance (Wright et al., 1996) and greater resistance to genotoxic stress than somatic cells (Vinoth et al., 2008). However, age-related changes in the germline are known to occur. For example, some germ cells lose the ability to divide and differentiate normally (Zhao et al., 2008), the sperm of older human males are thought to be at risk for more de novo mutations based on parent-offspring mutations (Kong et al., 2012), and double strand break repair in oocytes in humans and mice declines with age (Titus et al., 2013). Additionally, age-dependent meiotic nondisjunction may be due to a loss of the protein complex that regulates the separation of sister chromatids over time (Subramanian and Bickel, 2008). However, some age-effects that have been observed in the germline may be due to extrinsic factors such as the microenvironment of the germ line stem cells (Boyle et al., 2007; Pan et al., 2007; Zhao et al., 2008). The relative roles of extrinsic versus intrinsic factors in contributing to germline aging are still being explored. In mammals, much of the current evidence points to a greater role of cell-extrinsic factors. Similar to flies, niche deterioration also may play a role in the mammalian system (Zhang et al., 2006). For example, it has been shown that

mammalian spermatagonial stem cells, when transplanted to a young environment, have extended functionality (Ryu et al., 2006; Schmidt et al., 2011). Signaling factors like insulin may also play a role in maintaining germline function in mammals (Hsu and Drummond-Barbosa, 2008; Yang et al., 2013). Thus, while the germline is generally considered to be immortal, components of the germline and its microenvironment are not immune to age-related changes.

Recent findings highlighting the large role of epigenetic changes in the aging process leads us to question whether similar mechanisms may also be at play in reproductive tissues. Although the majority of epigenetic marks are erased and re-established between generations, some epigenetic modifications are transmitted across generations through the germline. Longevity itself is a trait that has been shown to be epigenetically inherited in *C. elegans* (Greer et al., 2016; Greer et al., 2011; Spracklin et al., 2017). Of most relevance, *Drosophila* oocytes transmit the repressive histone mark H3K27me3 to their offspring (Zenk et al., 2017). This creates a potential for age-effects to be passed on to the next generation, an outcome that could pose new questions for traditional evolutionary aging theories that have been around for decades.

Few studies have characterized genome-wide, age-related expression in ovaries and we are not aware of any such studies in the *Drosophila* germline. We sought to determine whether age-related epigenetic changes occur in the germline and broader *Drosophila* reproductive tissues using mRNA expression as a proxy. Specifically, we asked whether the age-dependent transposable element release extends to the ovary by determining whether transposable elements were derepressed during aging. We further tested the heterochromatin aging hypothesis by testing whether genes in or near heterochromatin boundaries were aberrantly expressed, and if genes were globally misregulated in reproductive tissues. We find that gene expression changes are enriched in heterochromatic regions of the genome, but the direction of change is

not consistent with a global increase in expression of heterochromatin. Further, we only find idiosyncratic aging effects on TE expression and no global increase in expression. These results suggest that the age-related transposon release and the heterochromatin aging hypothesis do not extend to the *Drosophila* ovary in a simple manner.

## METHODS

### Fly stocks

*D. melanogaster* DGRP lines 237 and 321 were utilized for this study and maintained at 22 degrees Celsius and 12 hour light cycles.

### Egg Chamber Tissue Collection

Flies were maintained in bottles at controlled larval density (~100 per bottle) for two generations before tissue collections. Zero to one day old F2 females were transferred to individual vials for aging treatment and supplemented with two males ranging from 3-7 days old approximately every seven days to encourage egg production. Flies were moved to fresh vials weekly. Stage 14 egg chambers were dissected from ovaries of 3-4 and 32-34 day old females in PBS buffer. Using a thinly bristled paintbrush, 2-5 egg chambers from each female were added to single caps of .2mL tubes, stabbed with RNAse free needles in 30uL TRIzol, and flash frozen in liquid nitrogen.

### Embryo Tissue Collection

Embryos were collected from the Ral-321 strain only. Flies were maintained in bottles at controlled larval density (~100 per bottle) for two generations. F2 females were maintained continuously laying in bottles containing yeast paste and supplemented with younger males. For embryo collections, flies were moved to mating cages with petri dishes filled with fly food and ~5mL of yeast paste to acclimate overnight.

Embryos were plucked from food plates after approximately 45 minutes of laying. Embryos were rinsed with embryo wash (0.7%NaCl, 0.05% Triton X-100) and dipped into 50% bleach using a mesh net for 30s-1min followed by another rinse with embryo wash. Embryos were picked up with a thinly bristled brush and put into a TRIzol filled .2mL tube cap and flash frozen in liquid nitrogen.

RNA extraction and mRNA Sequencing

For RNA extraction, egg chambers from 5 females (~20 egg chambers total) were pooled. In total there were 5 pools for each age treatment across both strains. For embryos, ~20 embryos from each cage were pooled with 4 cages across two timepoints. Accounting for the TRIzol already in the samples from the collection stage, we added up to a total volume of 300uL TRIzol for RNA extractions. To improve recovery in the separation phase, we used 5PRIME Phase Lock Gel Heavy tubes. RNA was resuspended in 25uL of H20. Library preps were performed using the NEBNext Ultra Kit according to the manufacturer's instructions (New England Biolabs). NEBNext Ultra libraries were pooled in groups of 8-10 per lane, and run with single-end 100 bp reads on a HiSeq 2500.

Analysis of mRNA sequencing data

RNA-seq was performed in CLC Genomics Workbench 8 using release 6 of the *Drosophila melanogaster* reference genome. For expression values, RPKM estimates generated by the RNA-seq tool in CLC Genomics Workbench were used. FDR-adjusted p-values for significant differential expression were calculated with a CLC algorithm based on the DESeq2 package in Bioconductor (Love et al., 2014). To estimate TE family expression, an annotated TE library was included in the RNAseq analysis while the rest of the genome was masked for individual TE sequences. GO analysis was performed with GOrilla (Eden et al., 2009) using *D. melanogaster* orthologs genes sorted by FDR p-value for the test of treatment effect.

# RESULTS

## Genic transcripts differentially expressed with age in egg chambers of both strains

A number of studies have compared aging transcriptomes across tissues and even across species (Doroszuk, 2012; Lee, 1999; McCarroll et al., 2004; Pletcher, 2002; Zhan et al., 2007; Zou et al., 2000). Fewer studies, however, compare profiles in more than one natural strain (Highfill et al., 2016; Landis et al., 2004). Here we sought to determine how gene expression is modulated in the aging ovary in two different inbred Raleigh strains of *Drosophila melanogaster* obtained from the DGRP (Mackay et al., 2012). Since ovaries are highly heterogeneous, consisting of a mixture of somatic tissues, germline-stem cells and many different stages of oogenesis, we focused our RNAseq analysis using stage 14 egg chambers. This allowed us to minimize variation of cell type composition and to enrich for age-effects in the germline. Stage 14 egg chambers consist of an oocyte surrounded by a follicular sheath and represent the last stage of oogenesis before fertilization and oviposition. To measure differences in gene expression, we compared expression profiles in stage 14 egg chambers from mothers at 3-4 and 32-34 days post-eclosion (sample overview presented in Table 1). Overall, we identified 300 transcripts that were differentially expressed (DE) between young and old stage-14 egg chamber samples in a combined analysis with the two Raleigh lines (FDR adjusted p-value <.05), testing for age while controlling for strain in DESeq2.

Of the DE transcripts identified in the combined analysis, 106 transcripts show an average increase with age, while 194 show an average decrease with age across strains (Fig. 3.1A). Figure 3.1B demonstrates that the significantly differentially expressed transcripts are strongly correlated and show the same direction of expression changes between old and young egg chambers across the two strains (Pearson's product-moment correlation = 0.66, p-value < 2e-16). Seven of these genes have previously been

associated with regulation of lifespan. Notably, *hebe* (CG1623) overexpression increases both longevity and fecundity (Li and Tower, 2009) and *Hsp27* overexpression increases lifespan (Wang et al., 2004). Both of these transcripts showed average lower expression in older stage-14 egg chambers across the two strains (*hebe*: 4.11-fold decrease, FDR p-value < .1.28E-05; *Hsp27*: 1.6-fold decrease; FDR p-value <.006). *Hsp27* was also one of the most highly expressed genes (26th). Another gene, *POSH* (Plenty of SH3s, CG4909) has been shown to promote cell survival in both *Drosophila* and human cells when overexpressed (Tsuda et al., 2010). We find that this transcript shows a 1.46-fold increase with age in egg chambers (FDR adjusted p-value<4.05E-05). The other transcripts previously associated with regulation of lifespan include Thiolase (CG4581), Thor (CG8846), Coq2 (Coenzyme Q biosynthesis protein 2; CG9613), and Tpi (triose phosphate isomerase; CG2171). Other notable categories of gene ontology analysis using GOrilla (Eden et al., 2009) results for biological process by rank significance include terms pertaining to the electron transport chain (GO:0022900; GO:0022904), mitochondrial electron transport chain (GO:0006120), numerous metabolic processes, developmental and cellular processes involved in reproduction (GO:0003006; GO:0022412), eggshell chorion assembly (GO:0007306), many terms related to regulation of mitochondrial organization and fusion, determinant of adult life span (GO:0008340) and interestingly, miRNA metabolic process (GO:0010586). Full results from a gene ontology (GO) analysis for biological process, component, and function by rank significance is shown in Supplementary Table 3.1.

While these DE transcripts may provide a signature of senescence for egg chambers, the transcriptome, as a whole, shows only a very weak correlation in age-related patterns of expression across these two strains (Pearson's product-moment correlation = 0.04, p-value < 1.8e-06, Figure 3.1B). This demonstrates that many observed changes in gene expression in the aging ovary are likely to be strain specific. In fact, 59 genes show a significant strain by age effect in our analysis.

Egg chamber transcripts from the mitochondrial genome are significantly downregulated with age across both strains

Some sets of genes and gene pathways show consistent and concerted changes with age across various studies. Age-related changes in the expression of mitochondrial genes and genes associated with the electron transport chain have consistently been reported. This is most commonly observed as a decrease during aging (Andreu, 1998; Calleja, 1993; Fernandez-Silva SP, 1991; Girardot et al., 2006; Morel, 1995; Sohal et al., 2008). In particular, this pattern has been observed in transcripts associated with the mitochondrial electron transport chain in the gonads of mice (Sharov et al., 2008).

In stage 14 egg chambers, 11 transcripts from the mitochondrial genome significantly decreased with age in the DE analysis (Figure 3.1, Figure 3.2A). Nine of those transcripts also showed a significant strain by age effect, with greater age-related fold-changes observed in Ral_321 for seven transcripts, while two showed opposite age-related effects across the strains (Fig 3.2A). In addition to transcripts from the mitochondrial genome, we also found nuclear transcripts associated with the electron transport chain significantly enriched in a gene ontology analysis (Supp. Table 3.1). All of these nuclear transcripts were also downregulated with age in both strains (Fig. 3.2B). The downregulation of mitochondrial transcripts and those associated with the electron transport chain is in line with established mitochondrial dysfunction associated with age. Our finding lends support to decreased expression of mitochondrial transcripts being a general feature of aging across all tissue types but also highlights strain-specific discrepancies in the magnitude of mitochondrial age-related effects. The reduced expression of mitochondrial transcripts in reproductive tissues may be especially significant as this could contribute to the reduced oocyte quality seen in aged flies (Calleja, 1993; Girardot et al., 2006; Morel, 1995; Sohal et al., 2008) and humans (Johnson et al., 2007; Zhang et al., 2017).

Downregulation of egg shell chorion transcripts in aged egg chambers show both shared and strain-specific effects

We found a significant gene ontology (GO) enrichment for differentially expressed transcripts associated with eggshell chorion assembly (FDR q-value < 1.44E-04, 15.4-fold enrichment). All of these transcripts were downregulated with age in both strains (Figure 3.9). The downregulation of eggshell transcripts was especially striking in Ral_321, in which all but two eggshell transcripts showed a decrease with age (sign test: p-val < 1.60e-11; Fig 3.3). Ral_237 also showed more eggshell transcripts downregulated with age than expected by chance (sign test p-value <.04) but the effect was not as strong as in Ral_321 (Fig 3.9).

Somatic follicle cells work together to build the protective eggshell in oogenic stages 10-14. This process is dynamic, with transcript amounts changing rapidly between stages (Tootle et al., 2011; Yakoby et al., 2008). Due to the dynamism of expression in late stage oogenesis with regards to eggshell formation, we sought to verify that differential expression of chorion genes was not a consequence of different temporal snapshots in the collected samples. Tootle *et al.* (2011) performed a microarray analysis on 150 genes expressed in a stage-specific manner in the last 24 hours of follicle development, delineated by stages 9-10a, 10b, 12, and 14. This gene expression dataset included 30 previously known eggshell genes, 19 new candidate chorion genes, and other non-eggshell or chorion genes that showed 4-fold changes in expression at late stages of follicle development. Because this gene expression dataset provides an independent temporal profile of gene expression in late stage oogenesis, we cross-checked our young and old egg chamber expression data against the 49 eggshell-specific transcripts. Critically, gene expression in our samples is strongly correlated with expression in stage-14 egg chambers reported in Tootle 2008 (Pearson's product-moment correlation = 0.85, p-value < 7.80e-15) but not correlated in stages 9-10, 10b, or 12, confirming that we had captured stage 14 egg-chambers in our analysis (Fig. 3.8).

The decrease in chorion transcripts with age corroborates findings of numerous other studies (Carlson et al., 2015; Doroszuk, 2012; Pletcher, 2002) and here we demonstrated that this age-effect can also vary in effect between strains. The discrepancy between the strains could also be due to the fact that we used chronological age for sampling instead of physiological age. Doroszuk et al., 2012 finds that long-lived flies do not experience a typical decline of reproduction function in the later stages of life which may alternatively explain why we didn't detect as significant of chorion effects in the strain with slightly longer median lifespan (Doroszuk, 2012; Ivanov et al., 2015).

Differentially expressed genes in egg chambers enriched for residence in dispersed heterochromatin, but no global genome-wide relaxation of heterochromatic silencing

Previous studies have implicated aberrant gene expression changes with age to changes in the heterochromatin landscape in the soma (Bell et al., 2012; De Cecco, 2013a, b; Jiang, 2013; Larson et al., 2012; Shah et al., 2013; Wood et al., 2010). Genome-wide expression data can be utilized as a proxy for heterochromatic changes by assessing whether genes associated with regions of heterochromatin experience age-related changes in expression. Based on previous studies, we hypothesized that genes located near heterochromatin boundaries, specifically near telomeres and centromeres, may be enriched for differential expression in aging. Kharchenko et al., 2011 described a genome-wide chromatin landscape in Drosophila melanogaster based on 9 prevalent combinatorial patterns of 18 histone modifications (Kharchenko et al., 2011). Pericentromeric heterochromatin domains were characterized by high levels of H3K9me2/me3. We intersected locations of our gene set with the heterochromatin regions described in that study. Of the significantly differentially expressed egg chamber transcripts across both strains in age, we found enrichment for genes in locations of intercalary heterochromatin (Figure 3.3A, 47 differentially expressed genes from 1695 genes in heterochromatin, 300 genes differentially expressed

from 14289 total genes; Chi-squared with Yate's correction, two-tailed p-value = 0.034). We also found a striking enrichment for differentially expressed genes on the fourth or "dot" chromosome, which is primarily heterochromatic and carries only 84 genes (8 genes differentially expressed from 84 total genes on the dot, 300 genes differentially expressed overall from 14289 total genes; Chi-squared with Yate's correction, two-tailed p-value < .0001). Other than the enrichment for genes on the dot chromosome, there was no obvious signature of enrichment for differentially expressed genes specifically in pericentric heterochromatin (Fig. 3.3A). Critically, we find that the nature of expression change with genes associated with heterochromatin is not in one direction. Differentially expressed genes associated with heterochromatin both increase and decrease during aging (Fig. 3.3A). This is unexpected under the standard heterochromatic aging hypothesis where heterochromatin function becomes lessened and heterochromatic genes become derepressed. Therefore, while heterochromatic regions of the genome tend to be enriched for genes that change in expression during aging, this indicates a general release of regulation, but not release from silencing *per se*.

To test whether there was also a subtle derepression of genes located in heterochromatin genome-wide, we compared age-related expression of all genes which overlapped with heterochromatin in the genome. We found no obvious change in distributions of gene expression ratios between young and old egg chambers of genes located in described regions of heterochromatin compared to the rest of the genome (Figure 3.3B)

We also tested whether the strain specific age-related changes for genes in intercalary heterochromatic regions were due to euchromatic TE insertions that differed between strains. It has been shown that some euchromatic TE insertions can nucleate heterochromatin formation through piRNA targeting (Sentmanat and Elgin, 2012; Shpiz et al., 2014). We used the DGRP strain-specific TE insertion data from TIDAL-fly

(Rahman et al., 2015) to compare TE insertion locations across the two strains. However, we did not see strain-specific differences in TE insertions that correlated with aging effects that varied between the two strains.

Other studies have reported decreased expression in heterochromatin modifiers with age. We therefore determined whether genes associated with the gene ontology term for chromatin modifiers showed enrichment for a certain directionality change with age. In egg chambers of both strains, chromatin modifiers tended to increase in expression with age (Ral_237 exact binomial test, p-value < 0.0004; Ral_321 exact binomial test, p-value < .002). Chromatin modifiers in embryos, however, tended to decrease in expression with maternal age (exact binomial test, p-value < 0.03).

No global release of transposable element expression in aged egg-chambers

Previous studies have shown that transposable elements become derepressed in the soma during aging, notably in brains and fat body of *Drosophila*, and in a variety of other organisms including mammals (Chen et al., 2016; De Cecco, 2013a, b; Li et al., 2013; Maxwell et al., 2011; Patterson et al., 2015). However, a recent study on sequencing artifacts have called some of these results into question (Treiber and Waddell, 2017). Because TEs and small RNA mechanisms of genome defense are primarily expressed in the germline, we aimed to determine whether TE de-repression during aging occurs in reproductive tissues in which they are primarily active. In contrast to other studies, we found no global TE derepression in egg chambers. While one transposable element, *copia*, increased with age across both strains, the other four TEs that showed differential expression with age across strains decreased in expression (Fig. 3.1A, Fig. 3.4, Table 2). Additionally, two TEs, *pogo* and *Juan*, showed a significant strain-by-age effect, exhibiting opposing directions of expression with age across the strains (Fig 3.4C and 3.4D). Figure 3.4D also illustrates that the TEs that are significant in Ral 321 are dispersed

throughout the wider distribution of TE expression for Ral 237. There is also no correlation between the ratio of TE expression between young and old egg chambers across strains (Figure 3.4E).

piRNA pathway transcripts upregulated in aging egg chambers

TE control by piRNA in the germline has been shown to be sensitive to aging. This has been attributed to an increased capacity for TE fragments residing in heterochromatin to contribute to the piRNA pool in older flies (Grentzinger et al., 2012). Moreover, this effect of aging can be transmitted across generations since maternally transmitted piRNA pools establish piRNA biogenesis in offspring. Since some TEs did show significant differential expression with age, we sought to check whether genes in the piRNA pathway, which regulate TE expression in the *Drosophila* germline, showed any age-related-expression changes in egg chambers. Strikingly, 27 out of 31 piRNA pathway genes show an average transcriptional increase with age across the two strains in egg chambers (exact binomial test, p-value < 3.4E-05; Fig. 3.6). piRNA genes are also enriched in the top 10% of differentially expressed transcripts ranked significance (Chi squared with Yate's correction, p-value = .044). Notably, we did not see these age-effects carried over into the embryo (Fig 3.6C), indicating that this effect may primarily be happening in the follicle cells.

Differential expression in the aging egg chamber is driven by both somatic and germline changes

Stage-14 egg chambers consist of a mixture of somatic follicle cells and germline material. It is challenging to tease-apart intrinsic aging of the germline from extrinsic factors such as functional decay of the niche (Zhao et al., 2008). We found that *tsunagi* significantly decreased with age in egg chambers (1.75-fold decrease, FDR p-value < 0.0008) (Supp. Table 3.2). *Tsunagi* is required in the germline for proper oogenesis and plays a critical role in in oocyte fate (Mohr et al., 2001; Parma et al., 2007).

However, it's possible that transcript change observed in stage 14 egg-chambers could be coming from the somatic follicular sheath and therefore not necessarily indicative of possible age effects in the oocyte.

We sought to determine whether differential expression during aging was mostly driven by somatic or germline transcripts by performing RNAseq of 0-1 hour embryos of young and old mothers. Maternal germline transcripts can be sequenced because *Drosophila* embryos do not undergo zygotic transcription for approximately two hours. The age-related changes we had seen for *tsunagi* in egg chambers showed the same directionality changes between embryos of young and old mothers (Supp. Table 3.2) supporting the idea that this was an age-related effect occurring in the germline. We next checked whether the Differentially expressed transcripts in egg chambers showed similar directional changes in embryos of Ral_321 as egg chambers across age. It is important to note that many transcripts are not maternally deposited and therefore not expressed in 0-1 hour embryos. The genic transcripts that were differentially expressed in egg chambers and also expressed in 0-1 hr embryos (threshold expression set at above 1.0 RPKM average) did show a significant positive correlation (Pearson's Product-Moment Correlation, R= 0.31, p-value < 2.85e-06), indicating that some age-related gene expression changes were occurring in the germline rather than simply the follicle cells of stage-14 egg chambers (Fig. 3.7). However, the mitochondrial transcripts with detectable expression in embryos did not show any correlation with changes observed in egg chambers and showed opposite directionality of expression. Thus, we can attribute the observed decrease in mitochondrial transcripts in stage-14 egg chambers to effects in the somatic follicle cells. This also indicates opposing age-related effects occurring in the somatic follicle sheath versus the germline.

We also identified fewer genes with differentially expressed in 0-1 embryos of young versus old mothers when compared to the egg chamber data, though this may be attributed to lower power from fewer

replicates. The dynamic nature during early embryogenesis may have also contributed to high variability in gene expression in this stage (Supp. Table 3.2). In 0-1 hour embryos, zygotic transcription is low, but de-adenylation of maternal transcripts may contribute dynamically to rapid changes in apparent gene expression across 0-1 hours.

Our results indicate that age-effects occur in both the somatic cells surround the developing egg as well as in the germline. Additionally, some of transcriptional changes we identify, such as *tsunagi*, may be contributing factors to compromised germ cell division and differentiation that occurs with age.

Subtle de-repression of TEs in pre-zygotically active embryos of aged mothers

Because TEs are primarily active in the germline and maternal transcripts are deposited into the embryo, we may expect to see a correlation between TE age effects in the egg chambers and embryos. We find no such correlation in expression between late stage egg chambers and 0-1 hour embryos of the same strain (Fig. 3.5B and 3.5C). The TEs that were differentially expressed in egg chambers are for the most part in the middle of the distribution for TE expression in embryos (Figure 3.5B). We do however find a subtle, yet significant enrichment for TEs increasing in expression in embryos of old mothers (Figure 3.5A, Exact binomial test, $p < 1.462e-08$; Figure 3.5B). The differentially expressed TEs that we observed in egg chambers may be primarily driven by the somatic follicle cells, masking the subtle increase of expression in TEs of the oocyte. Alternatively, there may be an independent de-repression of TEs that occurs in embryos.

# DISCUSSION

With delays in childbearing on the rise, the study of reproductive decline grows increasingly relevant (Billari et al., 2007). Fruit flies are an excellent model organism to study because they experience a clear reproductive decline, existing age-related literature in flies is vast, and *Drosophila* share several mechanisms and pathways in ovulation and gametogenesis with mammals (Sun and Spradling, 2013).

Genome-wide RNAseq studies have shown that different tissues vary in age-related signatures, highlighting the importance of analyzing each tissue individually in each species (Sharov et al., 2008; Zhan et al., 2007). Reproductive tissues are unique in that they are a mix of interacting somatic and germline tissue. While the germline is widely recognized as being more resistant to aging than somatic cells, some age-related changes are known to occur. Critically, the relative contribution of factors intrinsic versus extrinsic to the germ line in reproductive decline remains poorly understood.

Here, we report a set of genes that show concerted changes across genetic backgrounds in aged egg-chambers. We additionally highlight the role genetic background plays in age-related effects. For example, while the decline we show in chorion-related transcripts with age parallels other studies, we propose that the severity of this age-effect depends on genetic background.

We also show that aging in late stage egg chambers mirrors that of other tissues, with a downregulation of transcripts from the mitochondria and nuclear transcripts associated with mitochondrial activity. Oocytes have significantly more mitochondria than any other cell, highlighting the incredible energy demands at stake in gametogenesis (May-Panloup et al., 2007). The dysfunction of oocyte mitochondria has been proposed as a possible mechanism involved in reduced competence of oocytes in older human infertility

patients (Zhang et al., 2017). One of the most well documented age-effects thought to reduce female fertility is chromosome abnormality in oocytes. There is evidence that reduced mitochondrial activity may contribute to this decline, as improper chromosome segregation has been induced in oocytes deficient in mitochondrial enzymes that metabolize pyruvate (Johnson et al., 2007). Our results support the idea that mitochondrial age effects could contribute to reproductive decline. Because mitochondria are maternally transmitted, the possible deposition of abnormal mitochondria with advanced age has been hypothesized to negatively contribute to offspring health. Here we find no evidence that mitochondrial transcript decline is propagated, as embryos of young and old mothers do not display the same expression patterns as seen in egg chambers. In contrast, maternally deposited mitochondrial transcripts in embryos increase with age. Thus, we propose that the effects of aging in the *Drosophila* ovary on mitochondrial gene expression are largely born out in somatic follicle cells.

Epigenetic changes have been implicated as playing an important role in the aging process in cells of the soma across model organisms. Specifically, genome-wide heterochromatin redistribution during aging has been linked to the de-repression of transposable elements and an overall loss of gene regulation. Whether or not epigenetic factors are perturbed in reproductive and germline tissues is of particular interest because some epigenetic factors are known to transmit across generations (Greer et al., 2016; Grentzinger et al., 2012; Zenk et al., 2017). Current theories of evolutionary aging depend on the assumption that age-related effects do not manifest in offspring. If age-related effects are in fact transgenerational, this could complicate current evolutionary explanations for aging.

While several studies have reported aberrant gene expression in aging on a genome-wide scale (De Cecco, 2013a; Jiang, 2013; Shah et al., 2013), we report no overall loss of gene regulation in aged egg chambers, consistent with another *Drosophila* study using whole bodies (Pletcher, 2002). Previously, it

was shown that reporter genes residing in heterochromatin regions of the fly experienced loss of silencing with age (Jiang, 2013). In line with these findings, we too show that genes that show significant age-related expression differences are enriched for regions of heterochromatin, providing evidence for age-related epigenetic changes occurring in late stage egg chambers of *Drosophila* oogenesis. However, we do not however find evidence that the landscape of heterochromatic silencing is relaxed in older egg chambers. Some studies have also reported a decrease in expression of transcripts involved with heterochromatin modification. Here we find that these transcripts do significantly change with age, although in opposite directionality between egg chambers and embryos. This opposing effect in the soma versus germline indicates that patterns of aging may not be universal across tissue types.

Of significant interest is the conserved age-related changes we found between egg chambers and embryos of aged females. These indicate changes in the aged germline *per se*, not simply in the gonad that is a mixture of somatic and germline tissues. These also indicate that aging effects on gene expression in older mothers can be deposited into embryos and transmitted across generations. Since many of the maternal RNA transcripts deposited in embryos are required for embryonic development, this raises the need for further studies of how the maternal transcript pool may change with age and how faithfully those transcripts are deposited into embryos.

A decline in repressive heterochromatin with age has been associated with TEs becoming active and mobile in aging somatic cells (Li et al., 2013; Patterson et al., 2015). Because increased transposition promotes DNA damage and increased mutagenesis, age-related transposable element de-repression has also been proposed to be an important component of genomic instability and a contributor to the prevalence of disease that accompanies advanced age. Here, we find no evidence that TEs are derepressed

as a general feature of aging in egg chambers. In contrast, we find that the handful of TEs that are differentially expressed with age tend to decrease in expression with age, in conflict with current TE aging theories, but in line with the idea of adaptive piRNA-mediated immunity with age (Khurana et al., 2011a). The increase in expression in piRNA pathway genes reported here also lends support to this hypothesis and suggests that, in contrast to non-reproductive tissues, mechanisms that limit the harm of TEs may be increased in aging reproductive tissues. It would be worth comparing relative piRNA levels complementary to these TEs in a future study. We also demonstrate that TE age-effects in egg chambers depend on both the genetic background and TE. It is also worth noting that a recent study demonstrates the role artifacts play in leading to incorrect transposition estimation in the soma, possibly throwing previous age-related results into question (Treiber and Waddell, 2017). One interesting finding in our study that deserves further investigation is the subtle increase in TE expression we found when comparing embryos of young and old mothers. In a future study, it would be worth repeating this experiment, paired with a comparison of piRNA profiles of embryos of young and old mothers.

In summary, here we show that there is evidence for age-related change within the reproductive tissues and germline of *Drosophila melanogaster*. However, these tissues are more robust to age-related change in gene expression than the soma, as we find no global TE derepression or global relaxation of heterochromatic silencing with age. We also report that some significant age-related changes in the egg chambers of ovaries persist in embryos. This study supports the conclusion that while there exists a potential to pass on age-related maternal effects, the germline is generally robust to age-related epigenetic changes.

FIGURES

**Figure 3. 1. Signature of age-related expression in egg chambers across genetic background.** (A) Average log 2(RPKM+.5) expression of stage 14 egg chamber transcripts of old 30 - 34 day old samples versus young 3-4 day old samples. Transcripts significantly expressed between young and old in a paired analysis (FDR<.05) are colored according to transcript type. Five TE transcripts are significantly differentially expressed across both strains with age, with only one, copia, showing an increase in expression. (B) Log2 ratios of old to young (RPKM+.5) expression between strains. The differentially expressed transcripts (FDR p <.05) are strongly and significantly correlated with age across strains.

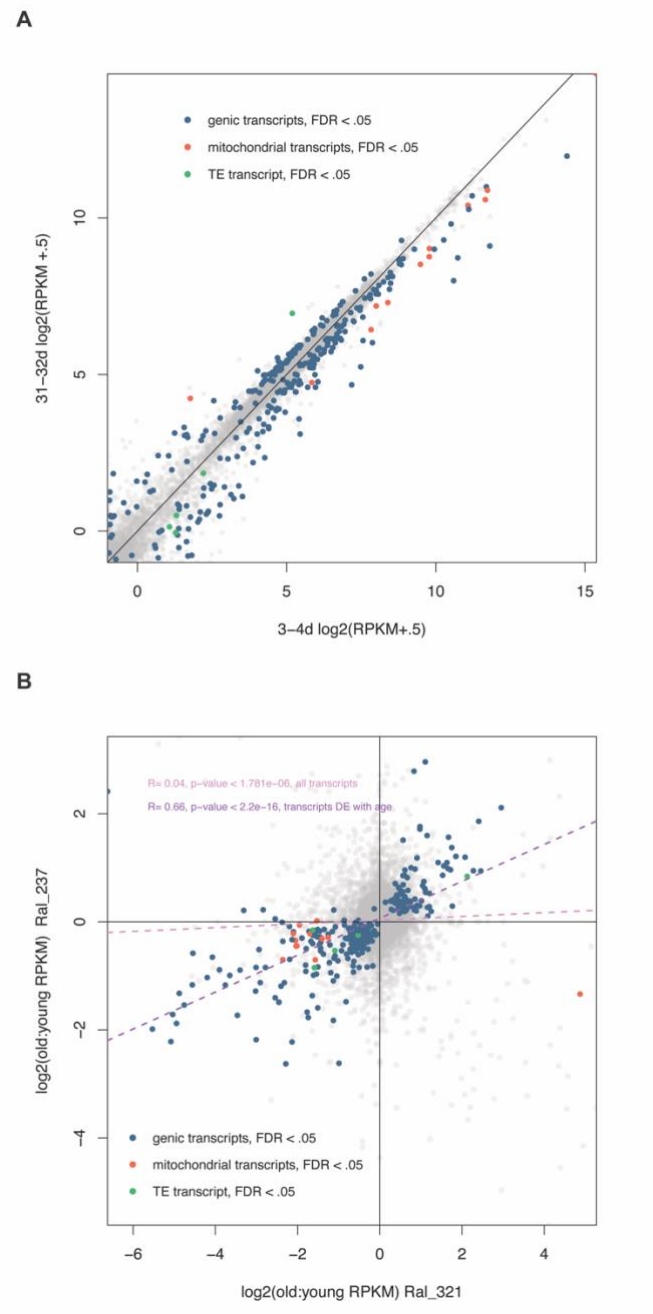


**Figure 3. 2. Majority of mitochondrial genome and nuclear mitochondrial transcripts decrease expression in egg chambers with age.** (A) There is an average reduction in mitochondrial genome transcript expression in stage 14 egg chambers across strains. Some transcripts are also significant for an age by strain interaction with greater age-related fold-changes (RPKM) in Ral_321. Gray color signifies no expression or no concerted change across strains. (B) Log2 ratios of old to young (RPKM+.5) expression between strains of mitochondrial transcripts from the nuclear genome.

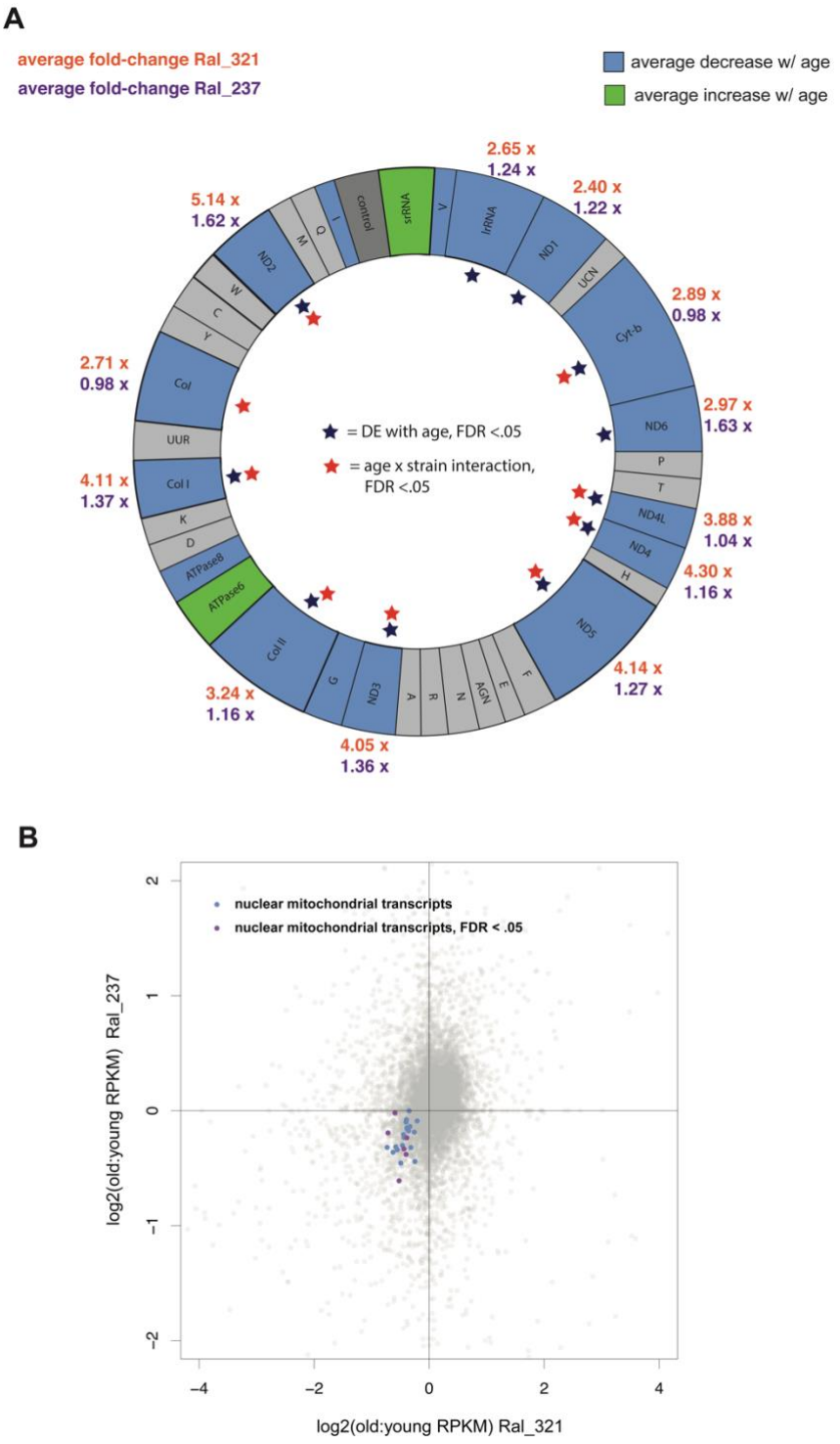


**Figure 3. 3. DE transcripts enriched for intercalary heterochromatin and the 4th chromosome.** (A) Positional information of differentially expressed genic transcripts across both strains. The notation "het" indicates that the genic location intersects with heterochromatin-associated proteins, H3K9me2/me3, as reported in Kharchenko et al 2011. DE (differentially expressed) genes located in regions of intercalary heterochromatin do not show a concerted directionality change of expression with age (Chi-squared with Yate's correction, two-tailed p-value <.034). The 4th chromosome is highly enriched for DE genes considering its limited gene composition Chi-squared with Yate's correction, two-tailed p-value < .0001. (B) Log2(young/old RPKM) of all genes located in heterochromatin versus Log2(young/old RPKM) genome-wide expression change with age. Genes in heterochromatin show similar age-related pattern of expression change as the rest of the genome.

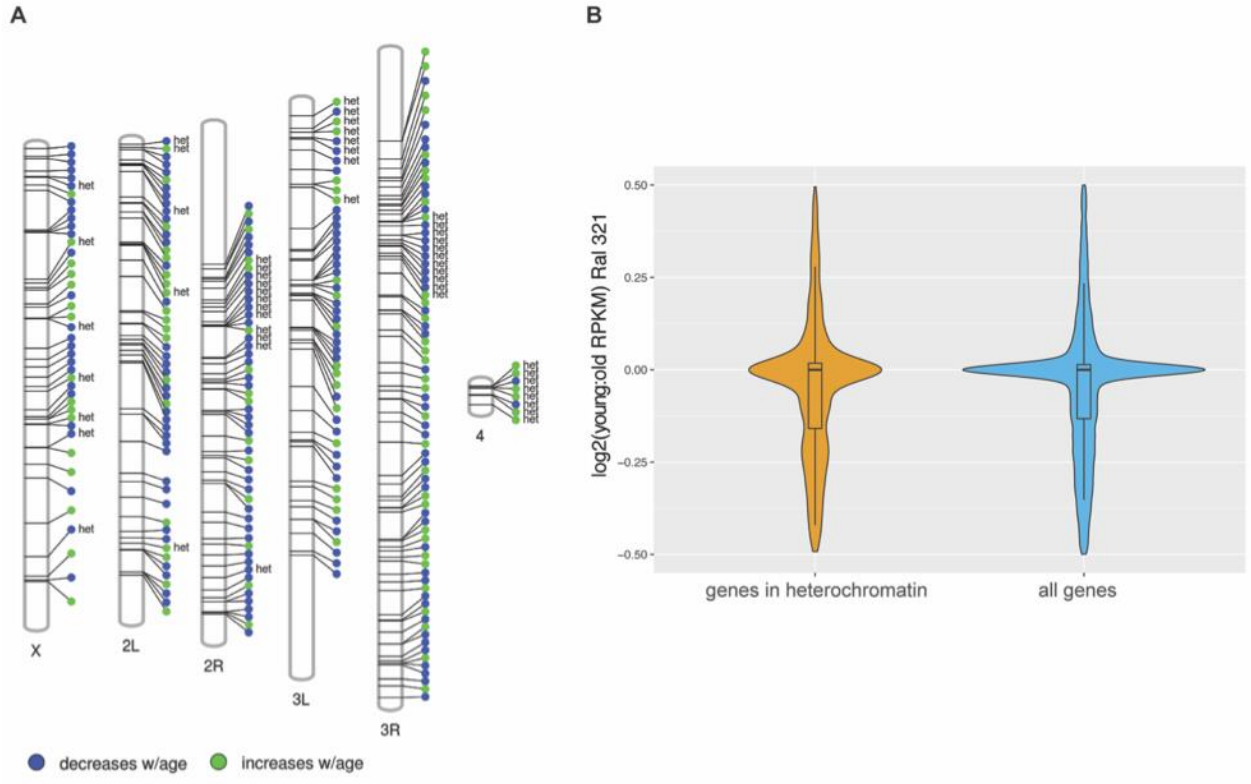


**Figure 3. 4. No global derepression of TEs in egg chambers from aged females.** A and B ) Heatmap of transposable element log2 (RPKM+1) expression in egg chambers normalized by Row Z-Score."O"= 30-34day old egg chambers, "Y" = 3-4 day old egg chambers. No clear patterns of TE expression occur with age. (C) TEs ordered by ratios of expression from old to young egg chambers in Ral_321. TEs significantly differentially expressed with age in Ral_321 tend to decrease with age. (D) TEs ordered by ratio of expression in Ral_237. Ral_237 shows differentially expressed TEs intercalated through a broader distribution of TE expression. (E) Log2 ratios of old to young RPKM+.5 of TE expression do not show a correlation with age across strains. Two TEs, pogo and Juan show significant age by strain interactions.

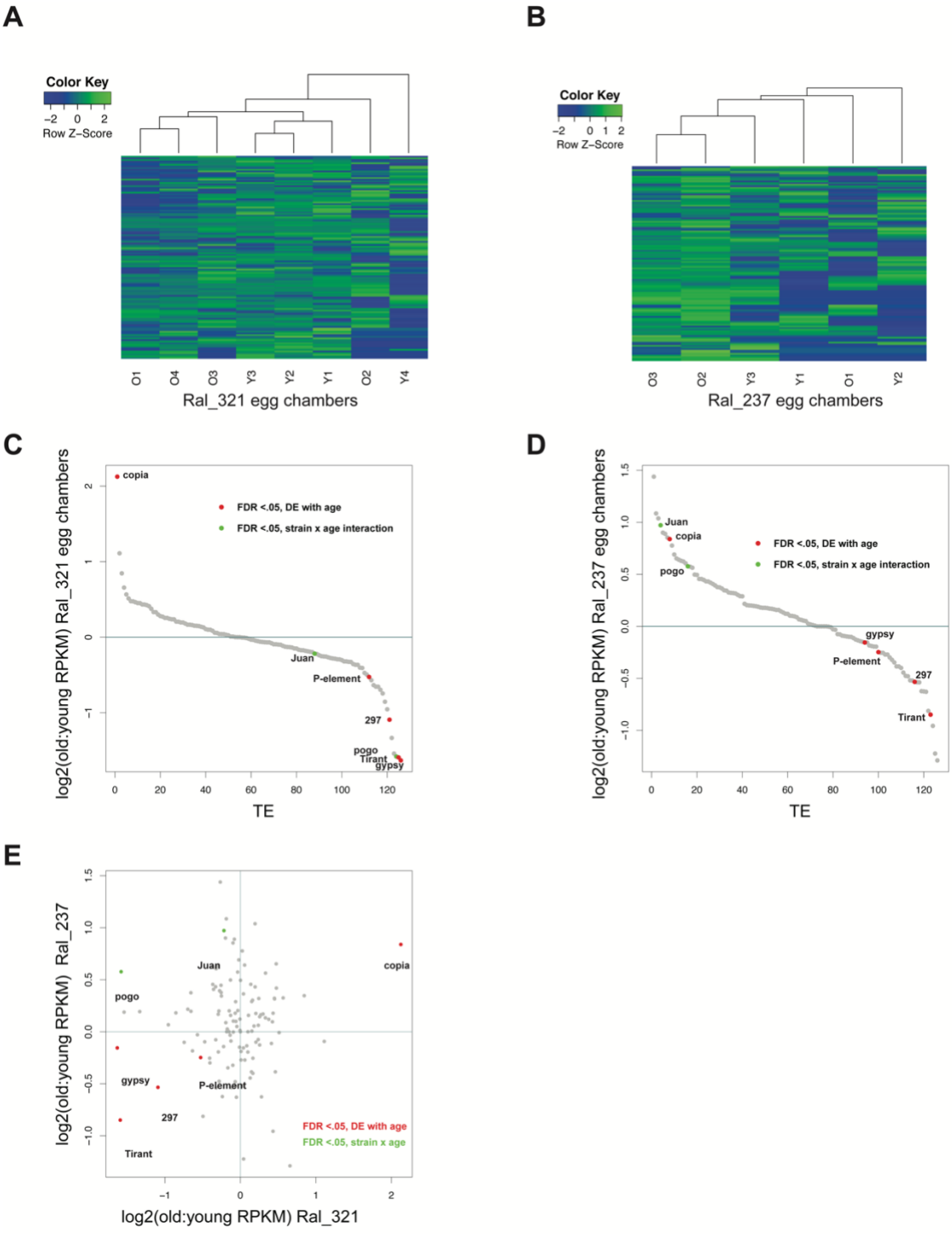


**Figure 3. 5. TE age effects differ from egg chamber to embryo.** (A) Heatmap of transposable element log2 (RPKM+1) expression in Ral_321 embryos normalized by Row Z-Score. "O"= embryos of 30-34day old mothers, "Y" = embryos of 3-4 day old mothers. Old samples show subtle increase of expression with age. (B) TEs ordered by ratios of expression from embryos of old versus young mothers in Ral_321. Transcripts that were differentially expressed in egg chambers of the same strain are interspersed within the broader distribution of TE expression. The majority of TE transcripts show increased expression with age. (C) Log2 ratios of old to young RPKM+.5 expression between egg chambers and embryos. TE transcripts change in egg chambers are not predictive of TE transcript changes in embryos.

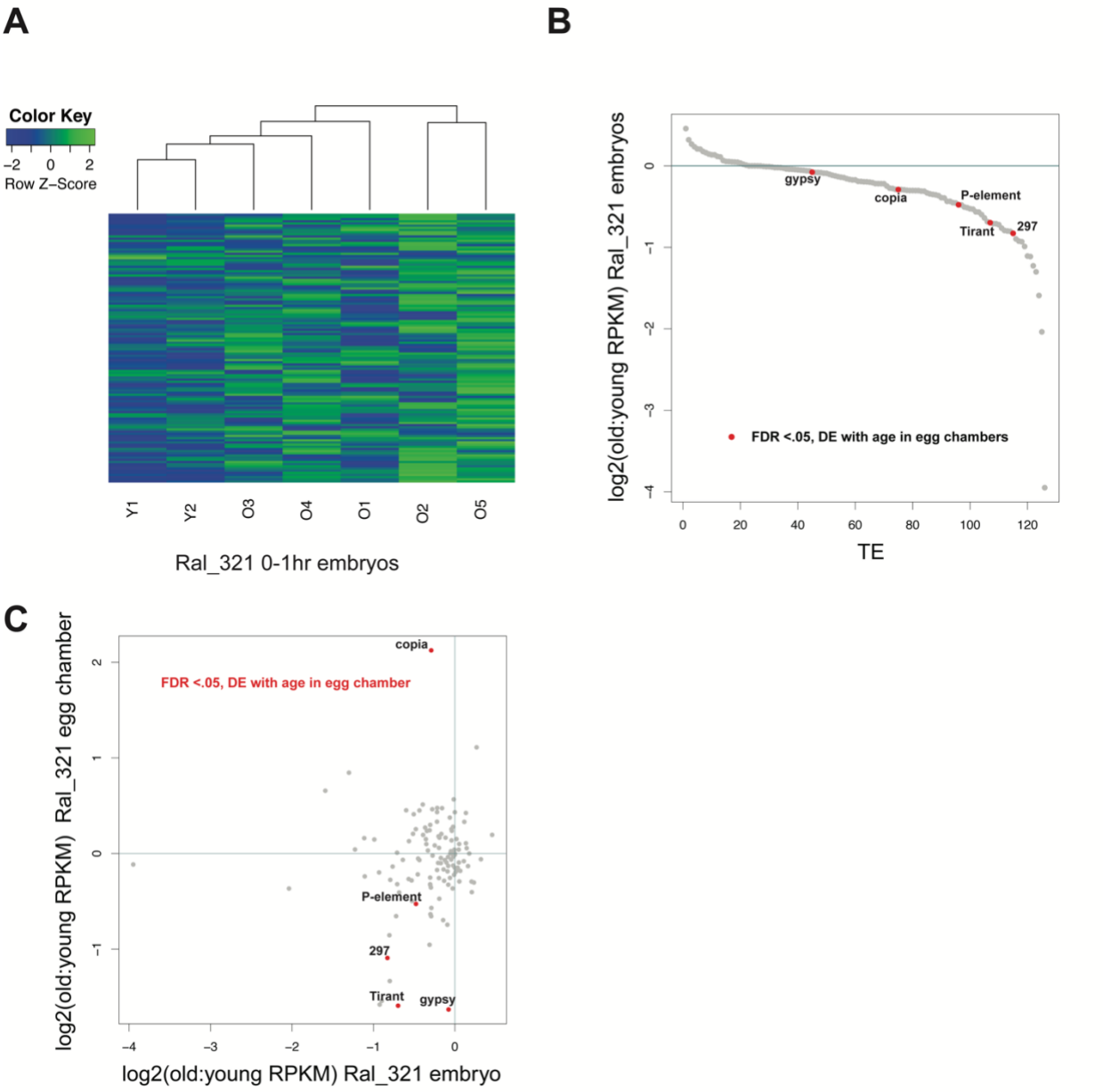


**Figure 3. 6. piRNA transcripts increase with age in egg chambers.** A) Expression (RPKM+.5) of piRNA pathway transcripts between egg chambers of young and old females. Red dots indicate transcripts that were in the top 10% of significant FDR-adjusted p-values. B) Log2 ratios of old to young piRNA pathway transcript expression (RPKM+.5) in egg chambers across strains. Both strains show a that a majority of piRNA transcripts increase with age. C) Log2 ratios of old to young piRNA pathway expression between egg chambers and embryos. Embryos of old mothers do not show increased transcript expression of piRNA genes.

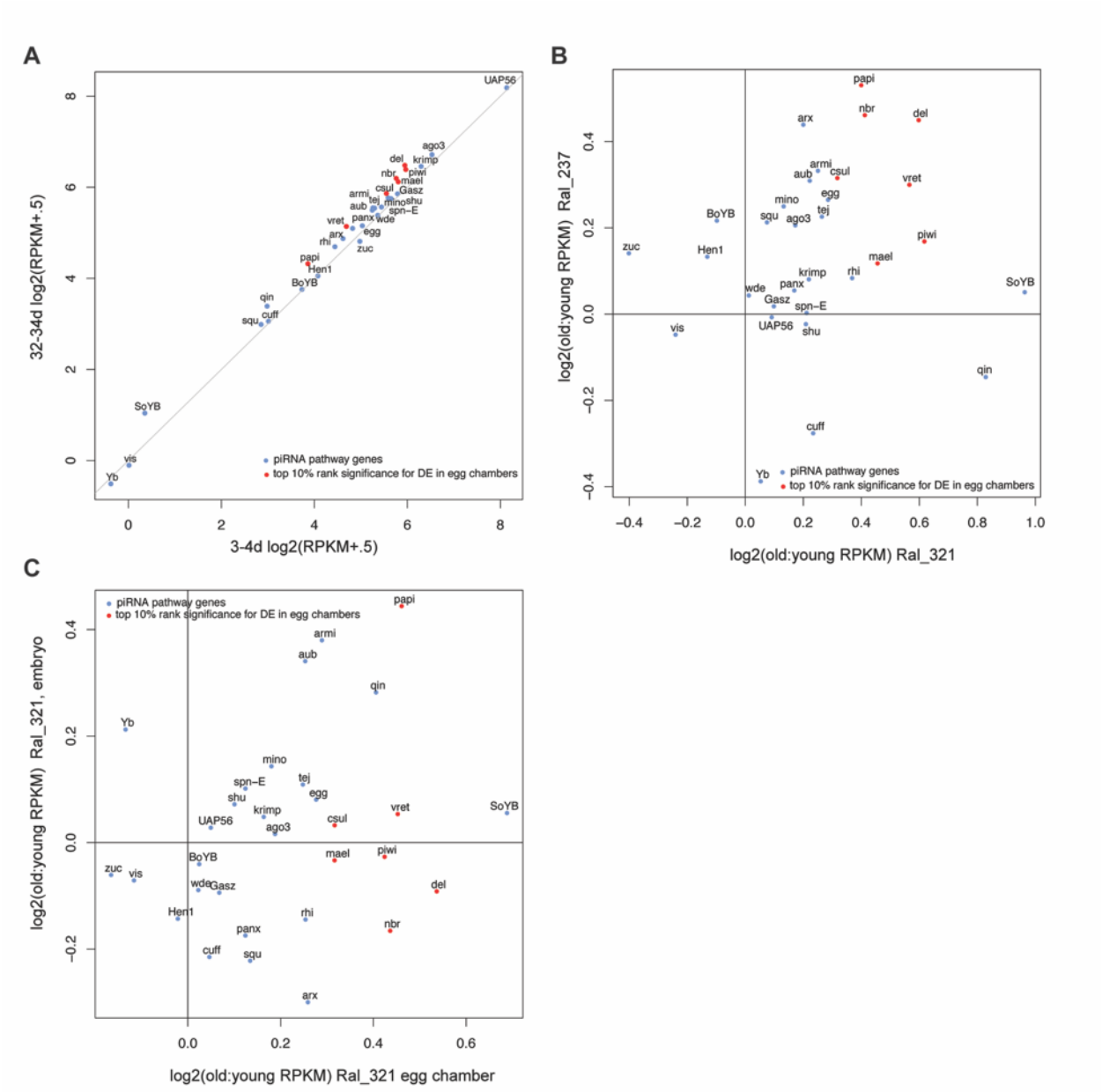


**Figure 3. 7. Some age-effects maternally deposited through germline.** Log2 ratios of old to young (RPKM+.5) expression between egg chambers and embryos of the same strain. Transcripts that have expression above an RPKM expression threshold of 0.5 in embryos are mildly correlated in age-related change. Transcripts from the mitochondrial genome do not show correlated age-related change between egg chambers and embryos.

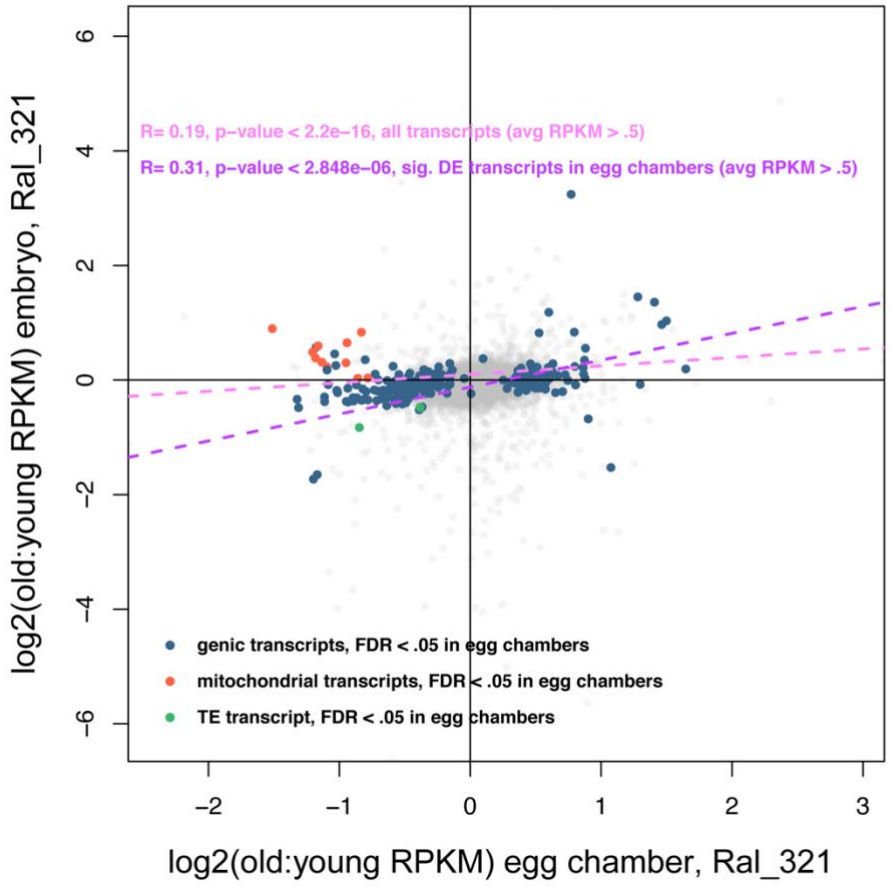


**Figure 3. 8. Verification of stage 14 transcript expression.** Transcripts that show stage-specific expression in final stages of oogenesis as defined by Tootle et al 2011. Transcript expression from stage 14 egg chambers is strongly correlated with stage 14 oogenic-specific transcript expression but not with the other stages in Tootle et al., 2011.

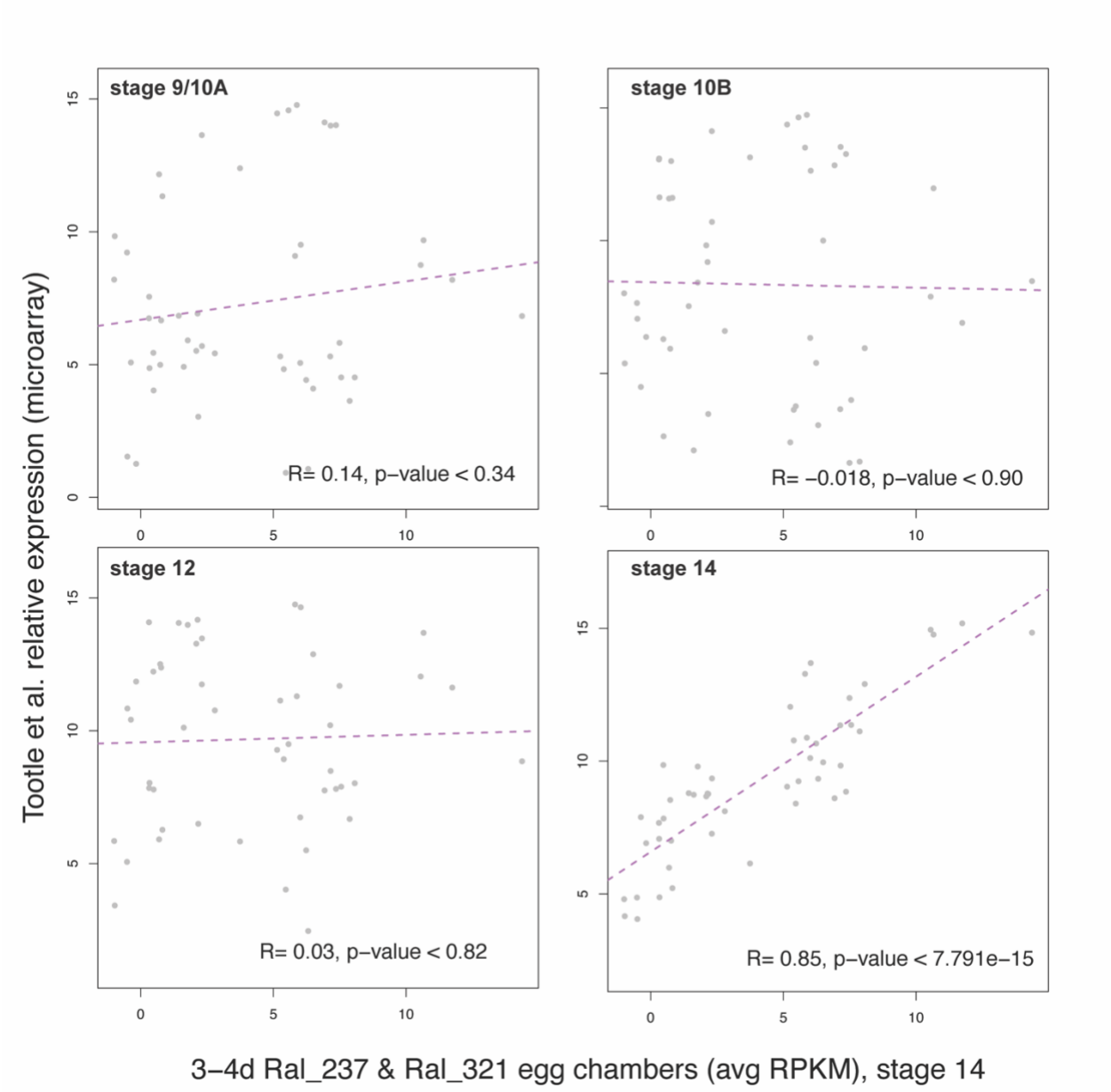
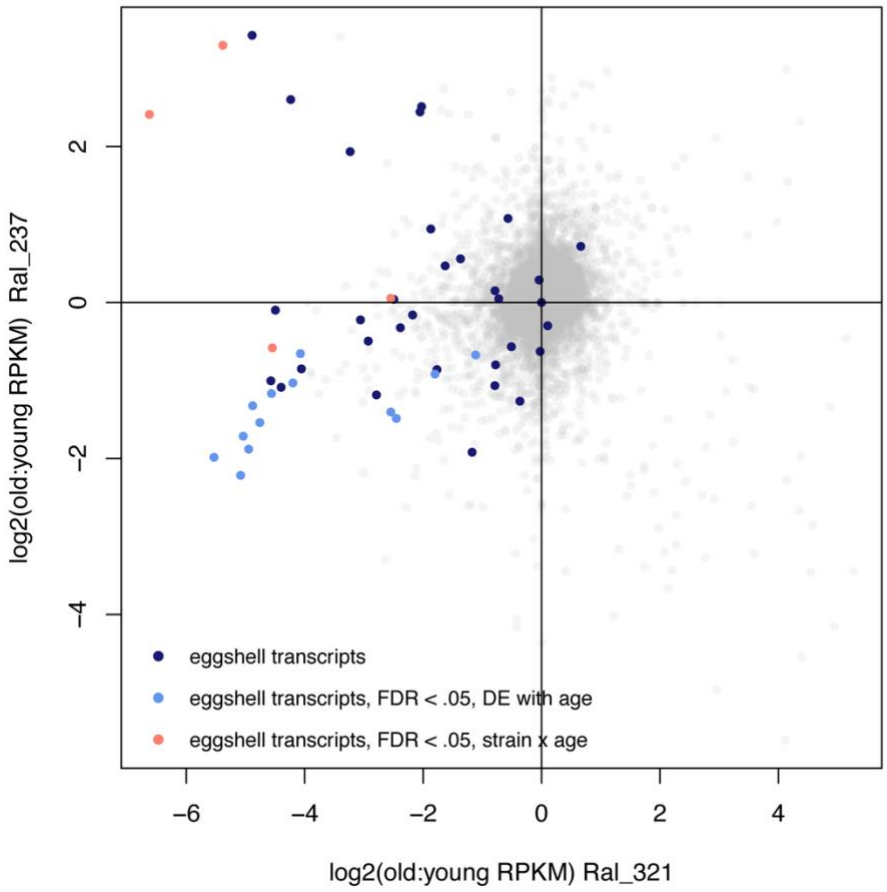


**Figure 3. 9. Transcripts associated with the eggshell are downregulated with age in both strains but show stronger age effects in Ral_321.** Log2 ratios of expression (RRKM + .5) of transcripts associated with the eggshell between young and old egg chambers across strains.



# TABLES

**Table 3. 1. Sample overview of stage 14 egg chambers.** Ral 237 and Ral 321 were the two DGRP strains utilized for RNA sequencing analysis. D=days post-eclosion. Each biological replicate is a pool of egg chambers from five females.

| DGRP STRAINS | # BIOLOGICAL REPLICATES (3-4D) | # BIOLOGICAL REPLICATES (32-34D) |
|---|---|---|
| RAL_321 | 4 | 4 |
| RAL_237 | 3 | 3 |

**Table 3. 2. Differential expression results for TEs.** TEs that show significant differential expression with age in egg chambers. Two TEs show a strain by age interaction. Fold change refers to fold change differences in RPKM levels. TEs significantly differentially expressed in egg chambers decrease with age in embryos but are not statistically significant.

| | ***EGG CHAMBER*** | | | | | ***EMBRYO*** | |
|---|---|---|---|---|---|---|---|
| ***TE*** | fold change | | | FDR adj. p-value | | fold decrease | FDR adj. p-value |
| | *Ral_321* | *Ral_237* | *drxn w/ age* | *age* | *strain x age* | *Ral_321* | *age* |
| *Tirant* | 3.02 | 1.80 | down | **0.007** | 0.940 | 1.62 | 1.0 |
| *297* | 2.13 | 1.45 | down | **0.010** | 0.869 | 1.78 | 1.0 |
| *Gypsy* | 3.10 | 1.11 | down | **0.011** | 0.125 | 1.06 | 1.0 |
| *Copia* | 4.36 | 1.79 | up | **0.034** | 0.849 | 1.22 | 1.0 |
| *P-element* | 1.44 | 1.19 | down | **0.041** | 0.939 | 1.39 | 1.0 |
| *Pogo* | 2.99 | 1.49 | down, up | 0.125 | **0.027** | 1.90 | 1.0 |
| *Juan* | 1.16 | 1.96 | down, up | 0.597 | **0.039** | 1.37 | 1.0 |

# REFERENCES

Alleman, M., Sidorenko, L., McGinnis, K., Seshadri, V., Dorweiler, J.E., White, J., Sikkink, K., and Chandler, V.L. (2006). An RNA-dependent RNA polymerase is required for paramutation in maize. Nature 442, 295-298.

Andreu, A.L., Arbos, M. A., Perez-Martos, A., Lopez-Perez, M. J., Asin, J., Lopez, N., Montoya, J. and Schwartz, S. (1998). Reduced mitochondrial DNA transcription in senescent rat heart. Biochemical and Biophysical Research Communications 252, 577-581.

Anxolabehere, D., Kidwell, M.G., and Periquet, G. (1988). Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of Drosophila melanogaster by mobile P elements. Mol Biol Evol 5, 252-269.

Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318, 761-764.

Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during Drosophila melanogaster development. Developmental Cell 5, 337-350.

Arbeitman, M.N., New, F.N., Fear, J.M., Howard, T.S., Dalton, J.E., and Graze, R.M. (2016). Sex Differences in Drosophila Somatic Gene Expression: Variation and Regulation by doublesex. G3 (Bethesda) 6, 1799-1808.

Arteaga-Vazquez, M., Sidorenko, L., Rabanal, F.A., Shrivistava, R., Nobuta, K., Green, P.J., Meyers, B.C., and Chandler, V.L. (2010). RNA-mediated trans-communication can establish paramutation at the b1 locus in maize. Proc Natl Acad Sci U S A 107, 12986-12991.

Bauer, M., Katzenberger, J.D., Hamm, A.C., Bonaus, M., Zinke, I., Jaekel, J., and Pankratz, M.J. (2006). Purine and folate metabolism as a potential target of sex-specific nutrient allocation in Drosophila and its implication for lifespan-reproduction tradeoff. Physiol Genomics 25, 393-404.

Bell, J.T., Tsai, P.C., Yang, T.P., Pidsley, R., Nisbet, J., Glass, D., Mangino, M., Zhai, G., Zhang, F., Valdes, A., et al. (2012). Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS Genet 8, e1002629.

Bergman, C.M., Quesneville, H., Anxolabehere, D., and Ashburner, M. (2006). Recurrent insertion and duplication generate networks of transposable element sequences in the Drosophila melanogaster genome. Genome Biol 7, R112.

Billari, F., Kohler, H.-P., Andersson, G., and Lundstrom, H. (2007). Approaching the Limit: Long-Term Trends in Late and Very Late Fertility. Population and Development Review, 149-170.

Bingham, P.M., Kidwell, M.G., and Rubin, G.M. (1982). The molecular basis of P-M dysgenesis - The role of the P-element, a P-Strain-Specific Transposon Family. Cell 29, 995-1004.

Blumenstiel, J.P. (2014). Whole genome sequencing in Drosophila virilis identifies Polyphemus, a recently activated Tc1-like transposon with a possible role in hybrid dysgenesis. Mob DNA 5, 6.

Blumenstiel, J.P., and Hartl, D.L. (2005). Evidence for maternally transmitted small interfering RNA in the repression of transposition in Drosophila virilis. Proceedings of the National Academy of Sciences of the United States of America 102, 15965-15970.

Blumenstiel, J.P., Noll, A.C., Griffiths, J.A., Perera, A.G., Walton, K.N., Gilliland, W.D., Hawley, R.S., and Staehling-Hampton, K. (2009). Identification of EMS-Induced Mutations in Drosophila melanogaster by Whole-Genome Sequencing. Genetics 182, 25-32.

Bosco, G., Campbell, P., Leiva-Neto, J.T., and Markow, T.A. (2007). Analysis of Drosophila species genome size and satellite DNA content reveals significant differences among strains as well as between species. Genetics 177, 1277-1290.

Boyle, M., Wong, C., Rocha, M., and Jones, D.L. (2007). Decline in self-renewal factors contributes to aging of the stem cell niche in the Drosophila testis. Cell Stem Cell 1, 470-478.

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

Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., and Hannon, G.J. (2008). An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. Science 322, 1387-1392.

Bucheton, A., Paro, R., Sang, H.M., Pelisson, A., and Finnegan, D.J. (1984). The molecular basis of the I-R hybrid dysgenesis syndreom in Drosophila melanogaster - Identification, cloning and properties of the I-Factor. Cell 38, 153-163.

Calleja, M., Pilar Pena, Cristina Ugalde, Carmen Ferreiro, Roberto Marco, and Rafael Garesse (1993). Mitochondrial DNA Remains Intact during Droosphila Aging, but the Levels of Mitochondrial Transcripts Are Signficantly Reduced. The Journal of Biological Chemistry 268, 18891-18897.

Carlson, K.A., Gardner, K., Pashaj, A., Carlson, D.J., Yu, F., Eudy, J.D., Zhang, C., and Harshman, L.G. (2015). Genome-Wide Gene Expression in relation to Age in Large Laboratory Cohorts of Drosophila melanogaster. Genet Res Int 2015, 835624.

Casacuberta, E., and Pardue, M.L. (2003). Transposon telomeres are widely distributed in the Drosophila genus: TART elements in the virilis group. Proc Natl Acad Sci U S A 100, 3363-3368.

Chambeyron, S., Popkova, A., Payen-Groschene, G., Brun, C., Laouini, D., Pelisson, A., and Bucheton, A. (2008). piRNA-mediated nuclear accumulation of retrotransposon transcripts in the Drosophila female germline. Proceedings of the National Academy of Sciences of the United States of America 105, 14964-14969.

Chandler, V.L. (2007). Paramutation: from maize to mice. Cell 128, 641-645.

Charlesworth, B., and Charlesworth, D. (1983). The population dynamics of transposable elements. Genetical Research 42, 1-27.

Charlesworth, B., and Langley, C.H. (1989). The population genetics of Drosophila transposable elements. Annual Review of Genetics 23, 251-287.

Chen, H., Zheng, X., Xiao, D., and Zheng, X. (2016). Age-associated de-repression of retrotransposons in the Drosophila fat body, its potential cause and consequence. Aging Cell 15, 542-552.

Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., Iyer, V.N., et al. (2007). Evolution of genes and genomes on the Drosophila phylogeny. Nature 450, 203-218.

Coyne, J.A. (1986). Meiotic segregation and male recombination in interspecific hybrids of Drosophila Genetics 114, 485-494.

Coyne, J.A. (1989). Mutation rates in hybrids between sibling species of Drosophila. Heredity (Edinb) 63 ( Pt 2), 155-162.

Czech, B., Preall, J.B., McGinn, J., and Hannon, G.J. (2013). A transcriptome-wide RNAi screen in the Drosophila ovary reveals factors of the germline piRNA pathway. Mol Cell 50, 749-761.

Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D.C. (2000). An RNA-Dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. Cell 101, 543-553.

Daniels, S.B., Peterson, K.R., Strausbaugh, L.D., Kidwell, M.G., and Chovnick, A. (1990). Evidence for Horizontal Transmission of the P-Transposable Element between Drosophila Species. Genetics 124, 339-355.

De Cecco, M., Steve W. Criscione, Abigal L. Peterson, Nicola Neretti, John M. Sedivy, and Jill A. Kreiling (2013a). Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. Aging Cell 12, 247-256.

De Cecco, M., Steve W. Criscione, Abigal L. Peterson, Nicola Neretti, John M. Sedivy, and Jill A. Kreiling (2013b). Transposable elements become activate and mobile in the genomes of aging mammalian somatic tissues. AGING 5, 17.

de Vanssay, A., Bouge, A.L., Boivin, A., Hermant, C., Teysset, L., Delmarre, V., Antoniewski, C., and Ronsseray, S. (2012). Paramutation in Drosophila linked to emergence of a piRNA-producing locus. Nature 490, 112-115.

Dion-Cote, A.M., Renaut, S., Normandeau, E., and Bernatchez, L. (2014). RNA-seq reveals transcriptomic shock involving transposable elements reactivation in hybrids of young lake whitefish species. Mol Biol Evol 31, 1188-1199.

Doroszuk, A. (2012). Transcriptome analysis of a long-lived natural Drosophila variant: a prominent role of stress and reproduction genes in lifespan extension. BMC Genomics 13.

Drosophila 12 Genomes, C., Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., et al. (2007). Evolution of genes and genomes on the Drosophila phylogeny. Nature 450, 203-218.

Dunson, D., Colombo, B., and Baird, D. (2002). Changes with age in the level and duration of fertility in the menstrual cycle. Human Reproduction 17, 1399-1403.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10, 48.

Eggleston, W.B., Johnson-Schlitz, D.M., and Engels, W.R. (1988). P-M hybrid dysgenesis does not mobilize other transposable element families in D. melanogaster. Nature 331, 368-370.

Erwin, A.A., Galdos, M.A., Wickersheim, M.L., Harrison, C.C., Marr, K.D., Colicchio, J.M., and Blumenstiel, J.P. (2015). piRNAs Are Associated with Diverse Transgenerational Effects on Gene and Transposon Expression in a Hybrid Dysgenic Syndrome of D. virilis. PLoS Genet 11, e1005332.

Evgen'ev, M.B. (2013). What happens when Penelope comes?: An unusual retroelement invades a host species genome exploring different strategies. Mob Genet Elements 3, e24542.

Evgen'ev, M.B., and Arkhipova, I.R. (2005). Penelope-like elements--a new class of retroelements: distribution, function and possible evolutionary significance. Cytogenet Genome Res 110, 510-521.

Evgenev, M.B., Zelentsova, H., Shostak, N., Kozitsina, M., Barskyi, V., Lankenau, D.H., and Corces, V.G. (1997). Penelope, a new family of transposable elements and its possible role in hybrid

dysgenesis in Drosophila virilis. Proceedings of the National Academy of Sciences of the United States of America 94, 196-201.

Fernandez-Silva SP, P.V., Fracasso F, Gadaleta MN, Cantatore P (1991). Reduced synthesis of mtRNA in isolated mitochondria of senescent rat brain. Biochemical and Biophysical Research Communications 176, 645-653.

Figueroa-Romero, C., Hur, J., Bender, D.E., Delaney, C.E., Cataldo, M.D., Smith, A.L., Yung, R., Ruden, D.M., Callaghan, B.C., and Feldman, E.L. (2012). Identification of epigenetically altered genes in sporadic amyotrophic lateral sclerosis. PLoS One 7, e52672.

Futschik, M., and Carlisle, B. (2005). Noise-robust soft clustering of gene expression time-course data. Journal of Bioinformatics and Computational Biology 3.

Gates, M.A., Kannan, R., and Giniger, E. (2011). A genome-wide analysis reveals that the Drosophila transcription factor Lola promotes axon growth in part by suppressing expression of the actin nucleation factor Spire. Neural Dev 6, 37.

Gerasimova, T., Mizrokhi, L., and Georgiev, G. (1984). Transposition bursts in genetically unstable Drosophila melanogaster. Nature 309, 714-716.

Ghildiyal, M., Setiz, H., Horwich, M., Li, C., Du, T., Lee, S., Xu, J., Kittler, E., Zapp, M., Weng, Z., et al. (2008). Endogenous siRNAs Derived from Transposons and mRNAs in Drosophila Somatic Cells. Science 320

.

Gibson, G., Riley-Berger, R., Harshman, L., Kopp, A., Vacha, S., Nuzhdin, S., and Wayne, M. (2004). Extensive sex-specific nonadditivity of gene expression in Drosophila melanogaster. Genetics 167, 1791-1799.

Girardot, F., Lasbleiz, C., Monnier, V., and Tricoire, H. (2006). Specific age-related signatures in Drosophila body parts transcriptome. BMC Genomics 7, 69.

Gou, L.T., Dai, P., Yang, J.H., Xue, Y., Hu, Y.P., Zhou, Y., Kang, J.Y., Wang, X., Li, H., Hua, M.M., et al. (2014). Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. Cell Res 24, 680-700.

Greer, E.L., Becker, B., Latza, C., Antebi, A., and Shi, Y. (2016). Mutation of C. elegans demethylase spr-5 extends transgenerational longevity. Cell Res 26, 229-238.

Greer, E.L., Maures, T.J., Ucar, D., Hauswirth, A.G., Mancini, E., Lim, J.P., Benayoun, B.A., Shi, Y., and Brunet, A. (2011). Transgenerational epigenetic inheritance of longevity in Caenorhabditis elegans. Nature 479, 365-371.

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

Gu, T., and Elgin, S.C. (2013). Maternal depletion of Piwi, a component of the RNAi system, impacts heterochromatin formation in Drosophila. PLoS Genet 9, e1003780.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5 ' end formation in Drosophila. Science 315, 1587-1590.

Haynes, K.A., Caudy, A.A., Collins, L., and Elgin, S.C.R. (2006). Element 1360 and RNAi components contribute to HP1-dependent silencing of a pericentric reporter. Current Biology 16, 2222-2227.

Hazelrigg, T., Watkins, W., Marcey, D., Tu, C., Karow, M., and Xiaorong, L. (1990). The exuperantia gene is required for Drosophila spermatogenesis as well as anteroposterior polarity of the developing oocyte, and encodes overlapping sex-specific transcripts. Genetics, 607-617.

Hickey, D.A. (1982). Selfish DNA: A sexually-transmitted nuclear parasite. Genetics 101, 519-531.

Highfill, C.A., Reeves, G.A., and Macdonald, S.J. (2016). Genetic analysis of variation in lifespan using a multiparental advanced intercross Drosophila mapping population. BMC Genet 17, 113.

Hsu, H., and Drummond-Barbosa, D. (2008). Insulin levels control female germline stem cell maintenance via the niche in Drosophila. PNAS 106, 1117-1121.

Ivanov, D.K., Escott-Price, V., Ziehm, M., Magwire, M.M., Mackay, T.F., Partridge, L., and Thornton, J.M. (2015). Longevity GWAS Using the Drosophila Genetic Reference Panel. J Gerontol A Biol Sci Med Sci 70, 1470-1478.

Jiang, N., Guyu Du, Ethan Tobias, Jason G. Wood, Rachel Whitaker, Nicola Neretti, and Stephen L. Helfand (2013). Dietary and genetic effects on age-related loss of gene silencing reveal epigenetic plasticity of chromatin repression during aging. Aging 5.

Johnson, M.T., Freeman, E.A., Gardner, D.K., and Hunt, P.A. (2007). Oxidative metabolism of pyruvate is required for meiotic maturation of murine oocytes in vivo. Biol Reprod 77, 2-8.

Jones, B.C., Wood, J.G., Chang, C., Tam, A.D., Franklin, M.J., Siegel, E.R., and Helfand, S.L. (2016). A somatic piRNA pathway in the Drosophila fat body ensures metabolic homeostasis and normal lifespan. Nat Commun 7, 13856.

Josefsson, C., Dilkes, B., and Comai, L. (2006). Parent-dependent loss of gene silencing during interspecies hybridization. Current Biology 16, 1322-1328.

Kelleher, E.S., Edelman, N.B., and Barbash, D.A. (2012). Drosophila interspecific hybrids phenocopy piRNA-pathway mutants. PLoS Biol 10, e1001428.

Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Minoda, A., Riddle, N.C., Ernst, J., Sabo, P.J., Larschan, E., Gorchakov, A.A., Gu, T., et al. (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature 471, 480-485.

Khurana, J.S., Wang, J., Xu, J., Koppetsch, B.S., Thomson, T.C., Nowosielska, A., Li, C., Zamore, P.D., Weng, Z., and Theurkauf, W.E. (2011a). Adaptation to P element transposon invasion in Drosophila melanogaster. Cell 147, 1551-1563.

Khurana, J.S., Wang, J., Xu, J., Koppetsch, B.S., Thomson, T.C., Nowosielska, A., Li, C., Zamore, P.D., Weng, Z., and Theurkauf, W.E. (2011b). Adaptation to P Element Transposon Invasion in Drosophila melanogaster. Cell 147, 1551-1563.

Khurana, J.S., Xu, J., Weng, Z., and Theurkauf, W.E. (2010). Distinct functions for the Drosophila piRNA pathway in genome maintenance and telomere protection. PLoS Genet 6, e1001246.

Kidwell, M.G., Kidwell, J.F., and Sved, J.A. (1977). Hybrid Dysgenesis in Drosophila melanogaster: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. Genetics 86, 813-833.

Kidwell, M.G., and Novy, J.B. (1979). Hybrid dysgenesis in Drosophila melanogaster - Sterility resulting from gonadal-dysgenesis in the P-M system. Genetics 92, 1127-1140.

Kolaczkowski, B., Hupalo, D.N., and Kern, A.D. (2010). Recurrent Adaptation in RNA Interference Genes Across the Drosophila Phylogeny. Molecular Biology and Evolution 28, 1033-1042.

Komili, S., Farny, N.G., Roth, F.P., and Silver, P.A. (2007). Functional specificity among ribosomal proteins regulates gene expression. Cell 131, 557-571.

Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., et al. (2012). Rate of de novo mutations and the importance of father's age to disease risk. Nature 488, 471-475.

Labrador, M., Farre, M., Utzet, F., and Fontdevila, A. (1999). Interspecific hybridization increases transposition rates of Osvaldo. Mol Biol Evol 16, 931-937.

Landis, G.N., Abdueva, D., Skvortsov, D., Yang, J., Rabin, B.E., Carrick, J., Tavare, S., and Tower, J. (2004). Similar gene expression patterns characterize aging and oxidative stress in Drosophila melanogaster. Proc Natl Acad Sci U S A 101, 7663-7668.

Larson, K., Yan, S.J., Tsurumi, A., Liu, J., Zhou, J., Gaur, K., Guo, D., Eickbush, T.H., and Li, W.X. (2012). Heterochromatin formation promotes longevity and represses ribosomal RNA synthesis. PLoS Genet 8, e1002473.

Le Thomas, A., Marinov, G.K., and Aravin, A.A. (2014a). A transgenerational process defines piRNA biogenesis in Drosophila virilis. Cell Rep 8, 1617-1623.

Le Thomas, A., Stuwe, E., Li, S., Du, J., Marinov, G., Rozhkov, N., Chen, Y.C., Luo, Y., Sachidanandam, R., Toth, K.F., et al. (2014b). Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. Genes Dev 28, 1667-1680.

Lee, C. (1999). Gene Expression Profile of Aging and its Retardation by Caloric Restriction. Science.

Lewis, S.H., Quarles, K.A., Yang, Y., Tanguy, M., Frezal, L., Smith, S.A., Sharma, P.P., Cordaux, R., Gilbert, C., Giraud, I., et al. (2018). Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. Nat Ecol Evol 2, 174-181.

Li, C.J., Vagin, V.V., Lee, S.H., Xu, J., Ma, S.M., Xi, H.L., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., et al. (2009). Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. Cell 137, 509-521.

LI, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXivorg arXiv:1303.3997v2.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26, 589-595.

Li, R., Ye, J., Li, S., Wang, J., Han, Y., Ye, C., Wang, J., Yang, H., Yu, J., Wong, G.K., et al. (2005). ReAS: Recovery of ancestral sequences for transposable elements from the unassembled reads of a whole genome shotgun. PLoS Comput Biol 1, e43.

Li, W., Prazak, L., Chatterjee, N., Gruninger, S., Krug, L., Theodorou, D., and Dubnau, J. (2013). Activation of transposable elements during aging and neuronal decline in Drosophila. Nat Neurosci 16, 529-531.

Li, Y., and Tower, J. (2009). Adult-specific over-expression of the Drosophila genes magu and hebe increases life span and modulates late-age female fecundity. Mol Genet Genomics 281, 147-162.

Love, M.I., Huber, W., and Anders, S. (2014). Moderate estimation of fold change and dispersion for RNA-Seq data with DESeq2. bioRxiv.

Lozovskaya, E.R., Scheinker, V.S., and Evgenev, M.B. (1990). A hybrid dysgenesis syndrome in Drosophila virilis. Genetics 126, 619-623.

Lyozin, G.T., Makarova, K.S., Velikodvorskaja, V.V., Zelentsova, H.S., Khechumian, R.R., Kidwell, M.G., Koonin, E.V., and Evgen'ev, M.B. (2001). The structure and evolution of Penelope in the

virilis species group of Drosophila: an ancient lineage of retroelements. Journal of Molecular Evolution 52, 445-456.

Mackay, T.F., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., Casillas, S., Han, Y., Magwire, M.M., Cridland, J.M., et al. (2012). The Drosophila melanogaster Genetic Reference Panel. Nature 482, 173-178.

Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. Cell 137, 522-535.

Marin, L., Lehmann, M., Nouaud, D., Izaabel, H., Anxolabehere, D., and Ronsseray, S. (2000). P-element repression in Drosophila melanogaster by a naturally occurring defective telomeric P copy. Genetics 155, 1841-1854.

Maxwell, P.H., Burhans, W.C., and Curcio, M.J. (2011). Retrotransposition is associated with genome instability during chronological aging. Proc Natl Acad Sci U S A 108, 20376-20381.

May-Panloup, P., Chretien, M., Malthiery, Y., and Reynier, P. (2007). Mitochondrial DNA in the Oocyte and the Developing Embryo. Current topics in developmental biology 77C, 51-83.

McCarroll, S.A., Murphy, C.T., Zou, S., Pletcher, S.D., Chin, C.S., Jan, Y.N., Kenyon, C., Bargmann, C.I., and Li, H. (2004). Comparing genomic expression patterns across species identifies shared transcriptional profile in aging. Nat Genet 36, 197-204.

Mohr, S., Dillon, S., and Boswell, R. (2001). The RNA-binding protein Tsunagi interacts with Mago Nashi to establish polarity and localize oskar mRNA during Drosophila oogenesis. Genes Dev 15, 2886-2899.

Morales-Hojas, R., Vieira, C.P., and Vieira, J. (2006). The evolutionary history of the transposable element Penelope in the Drosophila virilis group of species. J Mol Evol 63, 262-273.

Morel, F., Mazet, F., Touraille, S., Alziari, S. (1995). Changes in the respiratory chain complexes activities and in the mitochondrial DNA content during ageing in D. subobscura. Mech Ageing Dev 84, 171-181.

Mott, R., Yuan, W., Kaisaki, P., Gan, X., Cleak, J., Edwards, A., Baud, A., and Flint, J. (2014). The architecture of parent-of-origin effects in mice. Cell 156, 332-342.

Muntean, A., and Hess, J. (2009). Epigenetic Dysregulation in Cancer. The American Journal of Pathology 175, 1353-1361.

Niemi, J.B., Raymond, J.D., Patrek, R., and Simmons, M.J. (2004). Establishment and maintenance of the P cytotype associated with telomeric P elements in Drosophila melanogaster. Genetics 166, 255-264.

Niepielko, M.G., Marmion, R.A., Kim, K., Luor, D., Ray, C., and Yakoby, N. (2014). Chorion patterning: a window into gene regulation and Drosophila species' relatedness. Mol Biol Evol 31, 154-164.

O'Neill, R.J., O'Neill, M.J., and Graves, J.A. (1998). Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. Nature 393, 68-72.

O'Neill, R.J., O'Neill, M.J., and Graves, J.A. (2002). Corrigendum:Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. Nature 420, 106.

Obbard, D.J., Gordon, K.H.J., Buck, A.H., and Jiggins, F.M. (2009). The evolution of RNAi as a defence against viruses and transposable elements. Philosophical Transactions of the Royal Society B-Biological Sciences 364, 99-115.

Obbard, D.J., Maclennan, J., Kim, K.W., Rambaut, A., O'Grady, P.M., and Jiggins, F.M. (2012). Estimating divergence dates and substitution rates in the Drosophila phylogeny. Mol Biol Evol 29, 3459-3473.

Olovnikov, I., Ryazansky, S., Shpiz, S., Lavrov, S., Abramov, Y., Vaury, C., Jensen, S., and Kalmykova, A. (2013). De novo piRNA cluster formation in the Drosophila germ line triggered by transgenes containing a transcribed transposon fragment. Nucleic Acids Res 41, 5757-5768.

Pan, L., Chen, S., Weng, C., Call, G., Zhu, D., Tang, H., Zhang, N., and Xie, T. (2007). Stem cell aging is controlled both intrinsically and extrinsically in the Drosophila ovary. Cell Stem Cell 1, 458-469.

Paris, M., Villalta, J.E., Eisen, M.B., and Lott, S.E. (2015). Sex Bias and Maternal Contribution to Gene Expression Divergence in Drosophila Blastoderm Embryos. PLoS Genet 11, e1005592.

Parisi, M.J., Gupta, V., Sturgill, D., Warren, J.T., Jallon, J.M., Malone, J.H., Zhang, Y., Gilbert, L.I., and Oliver, B. (2010). Germline-dependent gene expression in distant non-gonadal somatic tissues of Drosophila. BMC Genomics 11, 346.

Parisi, M.J., Nutall, R., Edwards, P., Minor, J., Naiman, D., Lu, J., Doctolero, M., Vainer, M., Chan, C., Malley, J., et al. (2004). A survey ovary-, testis-, and soma-biased gene expression in Drosophila melanogaster adults. Genome Biology 5.

Parma, D.H., Bennett, P.E., Jr., and Boswell, R.E. (2007). Mago Nashi and Tsunagi/Y14, respectively, regulate Drosophila germline stem cell differentiation and oocyte specification. Dev Biol 308, 507-519.

Patterson, M.N., Scannapieco, A.E., Au, P.H., Dorsey, S., Royer, C.A., and Maxwell, P.H. (2015). Preferential retrotransposition in aging yeast mother cells is correlated with increased genome instability. DNA Repair (Amst) 34, 18-27.

Pavlidis, P., Jensen, J.D., Stephan, W., and Stamatakis, A. (2012). A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. Mol Biol Evol 29, 3237-3248.

Perrat, P., DasGupta, S., Wang, J., Theurkauf, W., Weng, Z., Rosbash, M., and Waddell, S. (2013). Transposition-Driven Genomic Heterogeneity in the Drosophila Brain. Science 340.

Petrov, D.A., Schutzman, J.L., Hartl, D.L., and Lozovskaya, E.R. (1995). Diverse transposable elements are mobilized in hybrid dysgenesis in Drosophila virilis Proceedings of the National Academy of Sciences of the United States of America 92, 8050-8054.

Pletcher, S.D. (2002). Genome-Wide Transcript Profiles in Aging and Calorically Restricted Drosophila melanogaster. Current Biology 12, 712-723.

Putiri, E.L., and Robertson, K.D. (2011). Epigenetic mechanisms and genome stability. Clin Epigenetics 2, 299-314.

Quinlan, A.R., and Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842.

Rahman, R., Chirn, G.W., Kanodia, A., Sytnikova, Y.A., Brembs, B., Bergman, C.M., and Lau, N.C. (2015). Unique transposon landscapes are pervasive across Drosophila melanogaster genomes. Nucleic Acids Res 43, 10655-10672.

Rajasethupathy, P., Antonov, I., Sheridan, R., Frey, S., Sander, C., Tuschl, T., and Kandel, E.R. (2012). A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. Cell 149, 693-707.

Riechmann, V., and Ephrussi, A. (2004). Par-1 regulates bicoid mRNA localisation by phosphorylating Exuperantia. Development 131, 5897-5907.

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

Ronsseray, S., Marin, L., Lehmann, M., and Anxolabehere, D. (1998). Repression of hybrid dysgenesis in Drosophila melanogaster by combinations of telomeric P-element reporters and naturally occurring P elements. Genetics 149, 1857-1866.

Rouget, C., Papin, C., Boureux, A., Meunier, A.C., Franco, B., Robine, N., Lai, E.C., Pelisson, A., and Simonelig, M. (2010). Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. Nature 467, 1128-1132.

Rozhkov, N.V., Aravin, A.A., Zelentsova, E.S., Schostak, N.G., Sachidanandam, R., McCombie, W.R., Hannon, G.J., and Evgen'ev, M.B. (2010). Small RNA-based silencing strategies for transposons in the process of invading Drosophila species. Rna-a Publication of the Rna Society 16, 1634-1645.

Rozhkov, N.V., Schostak, N.G., Zelentsova, E.S., Yushenova, I.A., Zatsepina, O.G., and Evgen'ev, M.B. (2013). Evolution and dynamics of small RNA response to a retroelement invasion in Drosophila. Mol Biol Evol 30, 397-408.

Rozhkov, N.V., Zelentsova, E.S., Shostak, N.G., and Evgen'ev, M.B. (2011). Expression of Drosophila virilis retroelements and role of small RNAs in their intrastrain transposition. PLoS One 6, e21883.

Ryu, B.Y., Orwig, K.E., Oatley, J.M., Avarbock, M.R., and Brinster, R.L. (2006). Effects of aging and niche microenvironment on spermatogonial stem cell self-renewal. Stem Cells 24, 1505-1511.

Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A., and Gvozdev, V. (2006). Telomere elongation is under the control of the RNAi-based mechanism in the Drosophila germline. Genes & Development 20, 345-354.

Scheinker, V.S., Lozovskaya, E.R., Bishop, J.G., Corces, V.G., and Evgenev, M.B. (1990). A long terminal repeat-containing retrotransposon in mobilized during hybrid dysgenesis in Drosophila virilis. . Proceedings of the National Academy of Sciences of the United States of America 87, 9615-9619.

Schmidt, J.A., Abramowitz, L.K., Kubota, H., Wu, X., Niu, Z., Avarbock, M.R., Tobias, J.W., Bartolomei, M.S., and Brinster, R.L. (2011). In vivo and in vitro aging is detrimental to mouse spermatogonial stem cell function. Biol Reprod 84, 698-706.

Sentmanat, M.F., and Elgin, S.C. (2012). Ectopic assembly of heterochromatin in Drosophila melanogaster triggered by transposable elements. Proc Natl Acad Sci U S A 109, 14104-14109.

Sese, M., Corominas, M., Stocker, H., Heino, T.I., Hafen, E., and Serras, F. (2006). The Cdi/TESK1 kinase is required for Sevenless signaling and epithelial organization in the Drosophila eye. J Cell Sci 119, 5047-5056.

Shah, P.P., Donahue, G., Otte, G.L., Capell, B.C., Nelson, D.M., Cao, K., Aggarwala, V., Cruickshanks, H.A., Rai, T.S., McBryan, T., et al. (2013). Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. Genes Dev 27, 1787-1799.

Sharov, A.A., Falco, G., Piao, Y., Poosala, S., Becker, K.G., Zonderman, A.B., Longo, D.L., Schlessinger, D., and Ko, M. (2008). Effects of aging and calorie restriction on the global gene expression profiles of mouse testis and ovary. BMC Biol 6, 24.

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

Shpiz, S., Ryazansky, S., Olovnikov, I., Abramov, Y., and Kalmykova, A. (2014). Euchromatic transposon insertions trigger production of novel Pi- and endo-siRNAs at the target sites in the drosophila germline. PLoS Genet 10, e1004138.

Sienski, G., Donertas, D., and Brennecke, J. (2012). Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell 151, 964-980.

Simkin, A., Wong, A., Poh, Y.P., Theurkauf, W.E., and Jensen, J.D. (2013). Recurrent and recent selective sweeps in the piRNA pathway. Evolution 67, 1081-1090.

Simmons, M.J., Ragatz, L.M., Sinclair, I.R., Thorp, M.W., Buschette, J.T., and Grimes, C.D. (2012). Maternal enhancement of cytotype regulation in Drosophila melanogaster by genetic interactions between telomeric P elements and non-telomeric transgenic P elements. Genet Res (Camb) 94, 339-351.

Smith, C.D., Edgar, R.C., Yandell, M.D., Smith, D.R., Celniker, S.E., Myers, E.W., and Karpen, G.H. (2007). Improved repeat identification and masking in Dipterans. Gene 389, 1-9.

Smyth, L.J., McKay, G.J., Maxwell, A.P., and McKnight, A.J. (2014). DNA hypermethylation and DNA hypomethylation is present at different loci in chronic kidney disease. Epigenetics 9, 366-376.

Sohal, R.S., Toroser, D., Bregere, C., Mockett, R.J., and Orr, W.C. (2008). Age-related decrease in expression of mitochondrial DNA encoded subunits of cytochrome c oxidase in Drosophila melanogaster. Mech Ageing Dev 129, 558-561.

Sokolova, M.I., Zelentsova, E.S., Shostak, N.G., Rozhkov, N.V., and Evgen'ev, M.B. (2013). Ontogenetic consequences of dysgenic crosses in Drosophila virilis. Int J Dev Biol 57, 731-739.

Spracklin, G., Fields, B., Wan, G., Becker, D., Wallig, A., Shukla, A., and Kennedy, S. (2017). The RNAi Inheritance Machinery of Caenorhabditis elegans. Genetics 206, 1403-1416.

Subramanian, V.V., and Bickel, S.E. (2008). Aging predisposes oocytes to meiotic nondisjunction when the cohesin subunit SMC1 is reduced. PLoS Genet 4, e1000263.

Sun, J., and Spradling, A.C. (2013). Ovulation in Drosophila is controlled by secretory cells of the female reproductive tract. Elife 2, e00415.

Titus, S., Li, F., Stobezki, R., Akula, K., Unsal, E., Jeong, K., Dickler, M., Robson, M., Moy, F., Goswami, S., et al. (2013). Impairment of BRCA1-related DNA double-strand break repair leads to ovarian aging in mice and humans. Sci Transl Med 5, 172ra121.

Tootle, T.L., Williams, D., Hubb, A., Frederick, R., and Spradling, A. (2011). Drosophila eggshell production: identification of new genes and coordination by Pxt. PLoS One 6, e19943.

Treiber, C.D., and Waddell, S. (2017). Resolving the prevalence of somatic transposition in Drosophila. Elife 6.

Tsuda, M., Kawaida, R., Kobayashi, K., Shinagawa, A., Sawada, T., Yamada, R., Yamamoto, K., and Aigaki, T. (2010). POSH promotes cell survival in Drosophila and in human RASF cells. FEBS Lett 584, 4689-4694.

Vastenhouw, N.L., Fischer, S.E., Robert, V.J., Thijssen, K.L., Fraser, A.G., Kamath, R.S., Ahringer, J., and Plasterk, R.H. (2003). A genome-wide screen identifies 27 genes involved in transposon silencing in C. elegans. Curr Biol 13, 1311-1316.

Vela, D., Fontdevila, A., Vieira, C., and Garcia Guerreiro, M.P. (2014). A genome-wide survey of genetic instability by transposition in Drosophila hybrids. PLoS One 9, e88992.

Vieira, J., Vieira, C.P., Hartl, D.L., and Lozovskaya, E.R. (1998). Factors contributing to the hybrid dysgenesis syndrome in Drosophila virilis. Genetical Research 71, 109-117.

Vigneault, F., Ter-Ovanesyan, D., Alon, S., Eminaga, S., D, C.C., Seidman, J.G., Eisenberg, E., and G, M.C. (2012). High-throughput multiplex sequencing of miRNA. Curr Protoc Hum Genet Chapter 11, Unit 11 12 11-10.

Vinoth, K., Heng, B., Poonepalli, A., Banergee, B., Balakrishnan, L., Lu, K., Hande, M., and Cao, T. (2008). Human Embryonic Stem Cells May Display Higher Resistance to Genotoxic Stress As Compared to Primary Explanted Somatic Cells. Stem Cells and Development 17, 599-607.

Wang, H.D., Kazemi-Esfarjani, P., and Benzer, S. (2004). Multiple-stress analysis for isolation of Drosophila longevity genes. Proc Natl Acad Sci U S A 101, 12610-12615.

Wang, S., and Hazelrigg, T. (1994). Implications for bcd mRNA localization from spatial distribution of exu protein in Drosophila oogenesis. Nature 369, 400-403.

Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2006). Parent-of-origin effects on mRNA expression in Drosophila melanogaster not caused by genomic imprinting. Genetics 173, 1817-1821.

Wood, J.G., Hillenmeyer, S., Lawrence, C., Chang, C., Hosier, S., Lightfoot, W., Mukherjee, E., Jiang, N., Schorl, C., Brodsky, A.S., et al. (2010). Chromatin remodeling in the aging genome of Drosophila. Aging Cell 9, 971-978.

Wright, W., Piatyszek, M., Rainey, W., Byrd, W., and Shay, J. (1996). Telomerase Activity in Human Germline and Embryonic Tissues and Cells. Developmental Genetics 18, 173-179.

Yakoby, N., Bristow, C.A., Gong, D., Schafer, X., Lembong, J., Zartman, J.J., Halfon, M.S., Schupbach, T., and Shvartsman, S.Y. (2008). A combinatorial code for pattern formation in Drosophila oogenesis. Dev Cell 15, 725-737.

Yang, S.A., Wang, W.D., Chen, C.T., Tseng, C.Y., Chen, Y.N., and Hsu, H.J. (2013). FOXO/Fringe is necessary for maintenance of the germline stem cell niche in response to insulin insufficiency. Dev Biol 382, 124-135.

Yannopoulos, G., Stamatis, N., Monastirioti, M., Hatzopoulos, P., and Louis, C. (1987). hobo is responsible for the induction of hybrid dysgenesis by strains of Drosophila melanogater bearing the male recombination factor 23.5MRF. Cell 49, 487-495.

Zenk, F., Loeser, E., Schiavo, R., Kilpert, F., Bogdanovic, O., and Iovino, N. (2017). Germ line-inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition. Developmental Biology 357, 212-216.

Zhan, M., Yamaza, H., Sun, Y., Sinclair, J., Li, H., and Zou, S. (2007). Temporal and spatial transcriptional profiles of aging in Drosophila melanogaster. Genome Res 17, 1236-1243.

Zhang, D., Keilty, D., Zhang, Z., and Chian, R. (2017). Mitochondria in oocyte aging: current understanding. Facts Views Vis OBGYN 9, 29-38.

Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B.S., Schultz, N., Vreven, T., Meignin, C., Davis, I., Zamore, P.D., et al. (2012). UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell 151, 871-884.

Zhang, X., Ebata, K.T., Robaire, B., and Nagano, M.C. (2006). Aging of male germ line stem cells in mice. Biol Reprod 74, 119-124.

Zhang, Y., and Lu, H. (2009). Signaling to p53: ribosomal proteins find their way. Cancer Cell 16, 369-377.

Zhao, R., Xuan, Y., Li, X., and Xi, R. (2008). Age-related changes of germline stem cell activity, niche signaling activity and egg production in Drosophila. Aging Cell 7, 344-354.

Zou, S., Meadows, S., Sharp, L., Jan, L.Y., and Jan, Y.N. (2000). Genome-wide study of aging and oxidative stress response in Drosophila melanogaster. Proc Natl Acad Sci U S A 97, 13726-13731.