

Transposable element dynamics and PIWI regulation impacts lncRNA and gene expression diversity in *Drosophila* ovarian cell cultures

Yuliya A. Sytnikova,¹ Reazur Rahman,¹ Gung-wei Chirn,¹ Josef P. Clark, and Nelson C. Lau

Department of Biology and Rosenstiel Basic Medical Science Research Center, Brandeis University, Waltham, Massachusetts 02454, USA

Piwi proteins and Piwi-interacting RNAs (piRNAs) repress transposable elements (TEs) from mobilizing in gonadal cells. To determine the spectrum of piRNA-regulated targets that may extend beyond TEs, we conducted a genome-wide survey for transcripts associated with PIWI and for transcripts affected by PIWI knockdown in *Drosophila* ovarian somatic sheet (OSS) cells, a follicle cell line expressing the Piwi pathway. Despite the immense sequence diversity among OSS cell piRNAs, our analysis indicates that TE transcripts are the major transcripts associated with and directly regulated by PIWI. However, several coding genes were indirectly regulated by PIWI via an adjacent de novo TE insertion that generated a nascent TE transcript. Interestingly, we noticed that PIWI-regulated genes in OSS cells greatly differed from genes affected in a related follicle cell culture, ovarian somatic cells (OSCs). Therefore, we characterized the distinct genomic TE insertions across four OSS and OSC lines and discovered dynamic TE landscapes in gonadal cultures that were defined by a subset of active TEs. Particular de novo TEs appeared to stimulate the expression of novel candidate long noncoding RNAs (lncRNAs) in a cell lineage-specific manner, and some of these TE-associated lncRNAs were associated with PIWI and overlapped PIWI-regulated genes. Our analyses of OSCs and OSS cells demonstrate that despite having a Piwi pathway to suppress endogenous mobile elements, gonadal cell TE landscapes can still dramatically change and create transcriptome diversity.

[Supplemental material is available for this article.]

Transposable elements (TEs) are parasitic genetic entities found across different organisms and have the potential to severely damage host genomes. Important open questions are which mechanisms have animals evolved to limit TEs from mobilizing and disrupting essential genes, how TEs can evade this control to fulfill their own needs to replicate, and what is the impact on global gene expression from these competing events. While genetics can explore these questions in gonads of intact animals (for review, see Lau 2010; Siomi et al. 2011), biochemical approaches with somatic and stem cells have also yielded much insight in TE biology, such as in Han and Boeke (2004), Coufal et al. (2009), Garcia-Perez et al. (2010), and Quinlan et al. (2011). The *Drosophila* ovarian somatic sheet (OSS) cell line serves as a niche for examining TE control in a gonad-like context because these cells are derived from follicle cells of the *Drosophila* ovary and express the Piwi pathway—an important gonad-specific mechanism of TE repression (Lau et al. 2009; Robine et al. 2009; Saito et al. 2009; Haase et al. 2010).

The Piwi pathway is a conserved TE control mechanism in animal gonads that is adaptive to new TE invasions because animals encode large intergenic loci (also called master control loci) (Brennecke et al. 2007) that can take in TE sequence elements and express them as Piwi-interacting RNAs (piRNAs). These piRNAs incorporate into a complex with Piwi proteins and are thought to act via base-pairing to target TE loci (Siomi et al. 2011). These Piwi/piRNA complexes then trigger gene silencing mechanisms that are

still not fully understood. Since most animal genomes, including humans, are loaded with TE sequences, it is possible that the repressive mechanisms of the Piwi pathway are not absolute and that TEs may be retained for a useful function (Levin and Moran 2011; Cowley and Oakey 2013). However, the diversity of all piRNA sequences in gonadal cells is so immense that many piRNAs could theoretically target other transcripts beyond TEs, including many coding genes if multiple mismatches are tolerated between targets and Piwi/piRNA complexes (Fig. 1A).

To test the hypothesis that in gonadal cells there may be unidentified non-TE PIWI/piRNA targets, we conducted a genome-wide survey for transcripts associated with PIWI and for transcripts that were regulated by PIWI in OSS cells. We deployed a comprehensive suite of RNA deep-sequencing approaches on OSS cells such as PIWI crosslinking immunoprecipitation (PIWI CLIP-seq) (Fig. 1B) and profiling total mRNAs and mRNAs from cellular fractions, including nascent RNAs (Nascent-seq) (Fig. 1C). Comparison of steady-state mRNA and nascent RNAs was conducted between OSS cells treated with a control small interfering RNA (siRNA, siGFP) or a siRNA targeting PIWI (siPIWI). Surprisingly, the PIWI-associated and PIWI-regulated transcripts in OSS cells were quite distinct from a recent study of PIWI-regulated transcripts in ovarian somatic cells (OSCs) (Sienski et al. 2012). This compelled us to examine the TE landscapes in these two cultures as well as earlier passages of both cell lines. Our study reveals that TEs are the primary targets of repression by the Piwi pathway in OSCs and OSS

¹These authors contributed equally to this work.

Corresponding author: nlau@brandeis.edu

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.178129.114>.

© 2014 Sytnikova et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

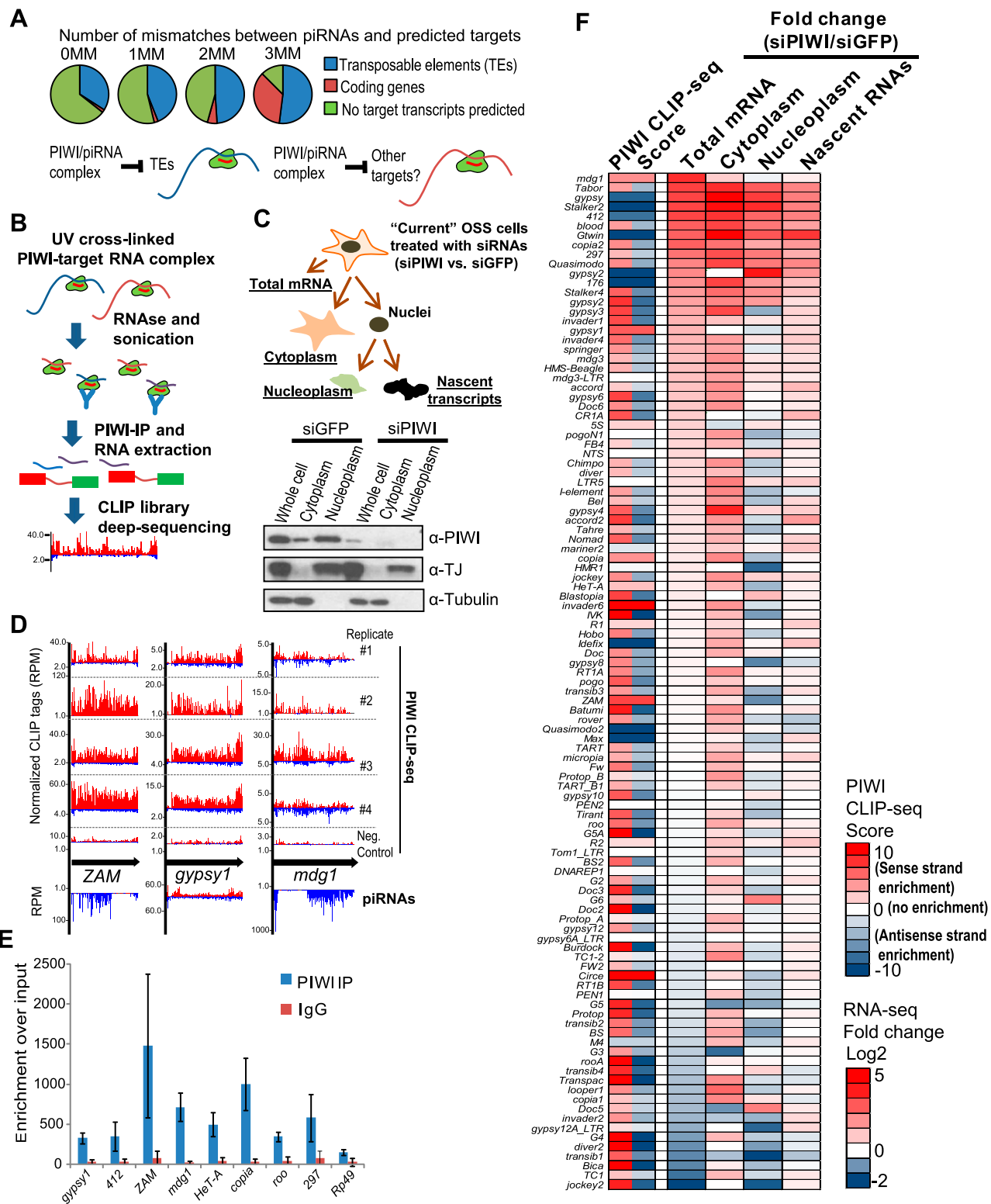


Figure 1. Transcriptome profiling and CLIP-seq confirm that transposable elements (TEs) are the main direct targets of PIWI-mediated regulation. (A) Bioinformatic prediction of PIWI/piRNA targets in OSS cells based on complementarity to piRNAs (Lau et al. 2009). With more mismatches (#MM), more coding genes are predicted to pair with piRNAs. These data suggest possible PIWI targets beyond TEs. (B) Scheme for PIWI CLIP-seq approach to identify PIWI associated transcripts. (C) Scheme for cellular fractionation and mRNA and nascent RNA isolation from OSS cells treated with siRNAs. Western blots confirm successful PIWI knockdown and separation of cytoplasmic from nucleoplasmic fractions. TJ is a transcription factor marking nuclear fractions, while tubulin is mainly cytoplasmic. (D) Different PIWI CLIP-seq profiles and corresponding piRNAs profiles of TEs with top PIWI CLIP scores. (Red) Plus strand reads; (blue) minus strand reads. Normalized CLIP-seq reads were deemed significant from our CLIP-seq processing algorithm. (RPM) Reads per million. (E) RNP-IP (RIP) assays validate TE transcript association with PIWI. Error bars correspond to standard deviation from five biological replicates. (F) Heat map showing TE transcript level changes in different compartments after PIWI knockdown in comparison to PIWI CLIP-seq scores. PIWI CLIP-seq scores with only red colors reflect primarily sense-strand patterns (Fig. 1D), only blue colors reflect primarily antisense-strand patterns (Supplemental Fig. S2F), and both strands when both red and blue colors are shown.

cells, and we report unexpected dynamics in the TE landscape across earlier and current OSCs and OSS cell passages. Our data also shed light on how gonadal cell genomes and the Piwi pathway appear to tolerate TE mobilization events instead of absolutely suppressing TEs. Finally, we suggest that the transcriptome differences between OSC and OSS cells may be the result of novel transcription of long noncoding RNAs (lncRNAs) driven by their unique TE landscapes and the Piwi pathway.

Results

The *Drosophila* OSS cell line expresses only primary piRNAs and the single PIWI protein since it is derived from the follicle cells of the *Drosophila* ovary (Niki et al. 2006; Lau et al. 2009; Haase et al. 2010). As such, OSS cells do not express the other Piwi pathway proteins, AUB and AGO3, and lack secondary piRNAs. Thus they are a simpler system to analyze PIWI-dependent gene regulation compared to the nurse cells and oocyte which comprise the *Drosophila* germline. We and other groups have maintained independent lines of these *Drosophila* follicle cell cultures which originated from Niki et al. (2006), and a variant called the OSCs has been utilized in functional studies of the *Drosophila* Piwi pathway (Saito et al. 2009; Sienski et al. 2012). Despite similar morphology and primary piRNA populations, there are notable gene expression profile differences between the OSCs and our OSS cells (Cherbas et al. 2011) as well as some differences in cell culture ploidy (Supplemental Fig. S1A). Since many endogenous cells in *Drosophila* (i.e., follicle, nurse, and salivary gland cells) naturally undergo polyploidization, the different ploidy in OSC and OSS cells may be a natural characteristic.

Many *Drosophila* cell cultures are persistently infected with viruses, such as *Drosophila* S2 cells (Aliyari et al. 2008; Czech et al. 2008; Ghildiyal et al. 2008; Kawamura et al. 2008; Flynt et al. 2009; Goic et al. 2013) as well as OSS cells (Wu et al. 2010). *Drosophila* cells stem this viral overload with RNA interference (RNAi) pathways, including the Piwi pathway; however, we have frequently observed newly thawed and stressed OSS and OSC cultures succumb after a few weeks of growth as cells lose adherence to the plastic substrate and lift off in clumps. We can eventually stabilize stressed OSS cell cultures with a protocol that nurtures the surviving cells back to a state of rapid mitotic growth (Supplemental Fig. S1B).

Recovery of TE transcripts confirms approach to identify PIWI-associated and PIWI-regulated transcripts

Since crosslinking and immunoprecipitation (CLIP) approaches with Argonaute/microRNA (AGO/miRNA) complexes have successfully discovered transcript targets (Chi et al. 2009; Hafner et al. 2010; Zisoulis et al. 2010; Leung et al. 2011), we applied the CLIP-seq technique (also known as HITS-CLIP) to the PIWI protein from OSS cell lysates after UV-light crosslinking (Fig. 1B; Supplemental Fig. S2). Initial tests of the standard CLIP protocols developed for AGO proteins (Chi et al. 2009; Zisoulis et al. 2010; Leung et al. 2011) onto PIWI in OSS cells yielded suboptimal RNA tag recovery, possibly because PIWI targeting and nuclear localization is distinct from AGO/miRNA complexes (Lau et al. 2009; Saito et al. 2009). Therefore, we modified our PIWI CLIP-seq procedure to increase RNA tag recovery while maintaining key reproducibility and specificity controls (see Supplemental Text and Supplemental Fig. S2). Reproducibility was ensured by using two independent PIWI antibodies (one mouse monoclonal from Saito et al. (2006) and

a second rabbit polyclonal, each raised against different epitopes) in altogether four biological replicates. At the same time, specificity was controlled by performing an IP with the rabbit polyclonal antibody pre-incubated with PIWI epitope peptide that effectively blocked PIWI pull-down (Supplemental Fig. S2A). Strong signals of radiolabeled RNAs in PIWI IPs appeared after UV crosslinking, and we were able to construct RNA fragment libraries after optimizing sonication and RNase T1 treatments (Supplemental Fig. S2B–E). To determine the PIWI CLIP tag patterns from deeply sequenced libraries (Supplemental Table S1), we employed a similar mapping and counting strategy of CLIP tags employed by other groups analyzing AGO-protein CLIP tags (Chi et al. 2009; Zisoulis et al. 2010; Leung et al. 2011).

PIWI CLIP-seq identified TEs known to be regulated by the Piwi pathway in *Drosophila* follicle cells, such as *gypsy1*, *ZAM*, and *Idefix* (Fig. 1D; Supplemental Fig. S2F; Pelisson et al. 1994; Tcheressiz et al. 2002; Sarot et al. 2004; Brassat et al. 2006). In fact, >100 TEs exhibited an enrichment PIWI CLIP score greater than at least 1.5-fold over the antigen blocking peptide library signal (Fig. 1F). Most TE-associated piRNAs are antisense to the coding strand of TEs, and consistent with this configuration, TE-associated PIWI CLIP tags for some TEs like *ZAM* and *mdg1* were primarily sense-strand-oriented (Fig. 1D). These tags were broadly distributed across the entire lengths of these TE consensus sequences. Interestingly, other TEs displayed PIWI CLIP tags that were in the same strand polarity as the bulk of TE-associated piRNAs (i.e., *412*, *Idefix*) (Supplemental Fig. S2F). In general, TEs with more antisense piRNAs also exhibited a greater number of antisense PIWI CLIP tags (Supplemental Fig. S3A; Supplemental Table S2). These PIWI CLIP tag patterns may represent putative precursors for TE-piRNAs, which is consistent with the abundant PIWI CLIP tags from the *flamenco* locus (data not shown) and is similar to putative piRNA precursors detected in the CLIP-seq of mouse Piwi proteins MIWI and MILI (Vourekas et al. 2012). We confirmed PIWI association with TE transcripts enriched in ribonucleoprotein-IP (RIP) experiments analyzed by quantitative RT-PCR (Fig. 1E).

To complement our PIWI CLIP-seq approach, we profiled PIWI-regulated endogenous transcripts from OSS cells treated with siRNAs knocking down PIWI compared to cells receiving a negative control siRNA (siGFP). In addition to total mRNA, we measured cytoplasmic, nucleoplasmic, and nascent RNAs to determine the compartments where PIWI regulates target transcripts (Fig. 1C,F). After normalization, we compared expression changes for various TEs with their PIWI CLIP scores. In accordance with strong PIWI CLIP enrichment scores, most TE transcripts that were strongly up-regulated upon PIWI knockdown at the total mRNA level also exhibited up-regulated nascent RNAs (Fig. 1F; Supplemental Fig. S3B,C). Although other regulation patterns in the cytoplasm and nucleoplasm were more complicated, the up-regulated nascent RNA changes suggested a transcriptional gene silencing mechanism, which is discussed in greater detail below as well as tested in a second study (Post et al. 2014).

The level of TE regulation by PIWI largely correlated with an increasing number of piRNAs that are antisense to TEs but not sense piRNAs that are unable to pair with TEs (Supplemental Fig. S3B,C). In fact, the TEs with the greatest change in expression after PIWI knockdown frequently were targeted by at least 2000 reads per million (RPM) of antisense piRNAs (Supplemental Fig. S3B). This trend was consistent with TE-sequence reporters that required a similar bulk of PIWI/piRNA to target the reporter transcript for gene silencing (Post et al. 2014). However, we also observed exceptions, such as *copa2*, which were strongly up-regulated after

PIWI knockdown but had relatively fewer (~100 RPM) mapping piRNAs, and *rover* and *roo*, which had almost 10,000 RPM mapping piRNAs but hardly any change after PIWI knockdown (Supplemental Fig. S3B). Finally, a few TEs appeared to be regulated only in the cytoplasmic fraction when no change was observed at the nascent RNA level (Supplemental Fig. S3D). This may reflect the cytoplasmic reservoir of PIWI, perhaps in organelles like the Yb body (Saito et al. 2010; Olivieri et al. 2011; Qi et al. 2011). In summary, the recovery of abundant TE sequences in the PIWI CLIP-seq and comprehensive RNA-seq analyses validates our approach.

Genic transcripts associated with or regulated by PIWI

Messenger RNA transcripts that were associated with or regulated by PIWI fell into two groups. The first group were mRNAs with high PIWI CLIP scores but were only modestly regulated by PIWI and mainly in the cytoplasm (Supplemental Fig. S4; and Supple-

mental Text). Despite consistently enriched PIWI CLIP-seq scores, cytoplasmic mRNA changes in RNA-seq and qPCR analyses and reproducible RIP assay enrichment (Supplemental Fig. S4B–E), these PIWI-associated transcripts did not exhibit changes in Western blots and luciferase reporter tests (Supplemental Fig. S4F; Post et al. 2014). Furthermore, gel shift assays with recombinant PIWI and RNA sequences from these mRNAs only indicated promiscuous RNA binding activity by PIWI (Supplemental Fig. S4G,H). These transcripts may reflect one tendency of the PIWI protein to bind to a broad range of transcripts, such as diverse piRNA sequences.

The second group of PIWI-regulated genes were strongly up-regulated upon PIWI knockdown but had low PIWI CLIP scores, and there were few piRNAs mapped to them (Fig. 2; Supplemental Table S2). These transcripts were strongly increased during PIWI knockdown at the nascent transcripts level, resulting in their elevation in the nucleoplasm and cytoplasm. Given the nascent RNA changes at TE transcripts, we inspected the nascent RNAs from these gene loci more closely and discovered detached, antisense

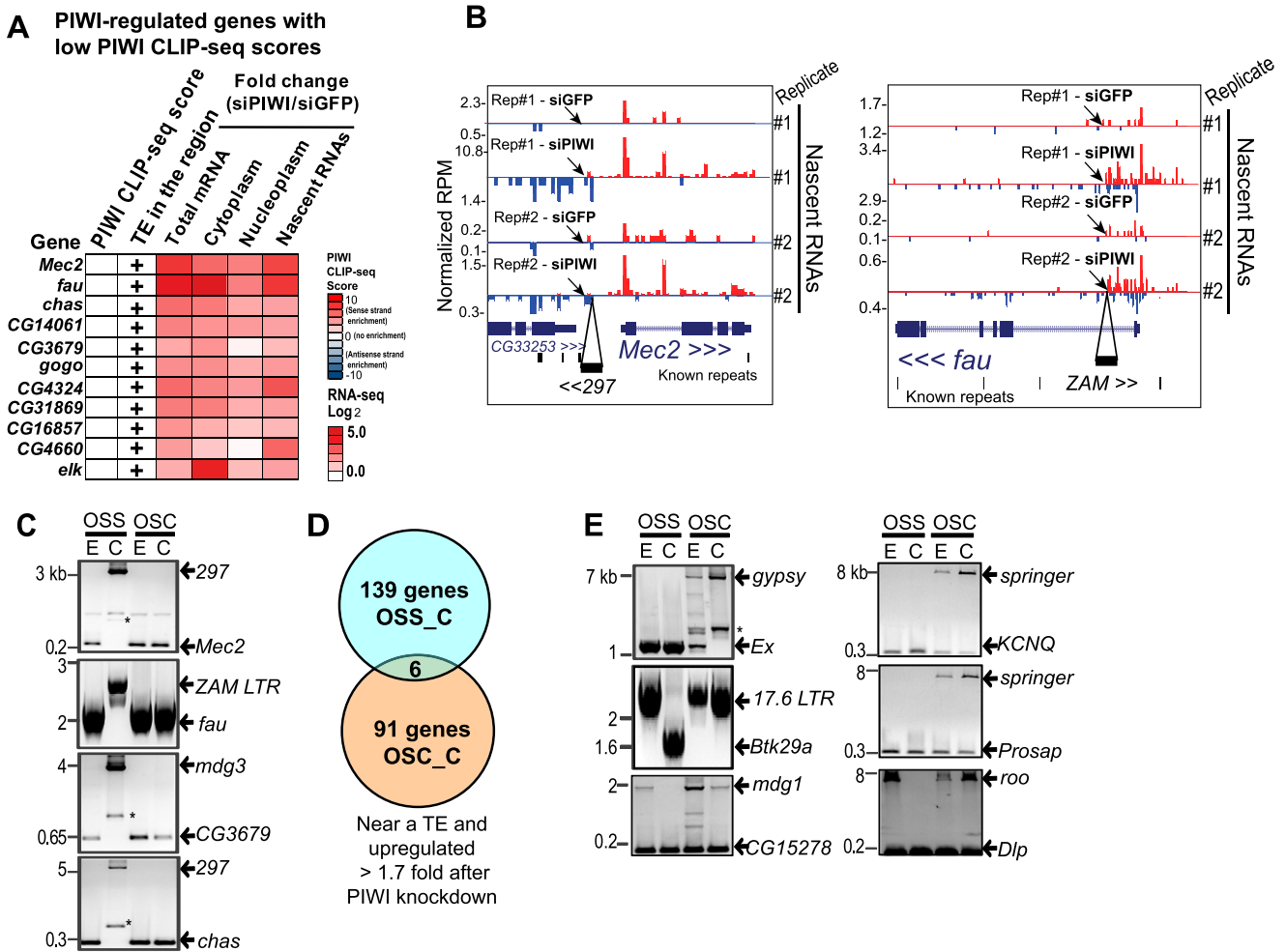


Figure 2. De novo TE insertions near genes impart PIWI-mediated gene silencing. (A) Heat map of top PIWI-regulated genes identified in OSS cells with low CLIP scores and which have a de novo TE insertion nearby. (B) Nascent transcript profiles for two loci that contain de novo TE insertions and are strongly regulated by PIWI silencing. Arrows point to the changing levels of reads for genes and the independent nascent transcripts arising from the TE insertion. (C) Genomic PCR confirms these persistent TE insertions specifically in the OSS_C line. “E” and “C” represent early and current passage of cells (see Fig. 3A). These are persistent TE insertions because no reference genome amplicon (the lowest band) is detected. Asterisks mark side-product amplicons. (D) Venn diagram showing distinct group of PIWI-regulated genes between OSCs and OSS cells. (E) Genomic PCR shows TE insertions specific to the OSC cells. (Left panel) Genes noted in Sienski et al. (2012); (right panel) additional insertions near genes we validated in OSC cell genomes. These are heterogeneous TE insertions because a reference genome amplicon is detected in addition to the TE amplicon. Asterisks mark side-product amplicons.

transcripts near genes, such as *Mec2* and *fau* (Fig. 2B). Upon amplifying by PCR and sequencing the amplicons, we confirmed that the start of nascent RNA reads at these loci corresponded to precise de novo insertion of a TE (Fig. 2C). We also confirmed by PCR that the de novo TE insertion loci, such as *CG3679* and *chas*, were very persistent in the cultures by the lack of a reference genome amplicon. Subsequent genome sequencing identified de novo TE insertions for the remaining loci highlighted in Figure 2A.

Our OSS cell transcriptome analysis indicated that de novo TEs were independently generating nascent transcripts. In a separate study using reporter genes in OSS cells, we showed that PIWI-mediated silencing requires piRNAs pairing with a nascent transcript (Post et al. 2014); therefore we interpret that the nascent transcripts arising from TEs may serve as a platform for recruiting PIWI/piRNA transcriptional silencing that spreads to the adjacent gene. This mechanism is consistent with the similar observations reported by Sienski et al. (2012) that utilized OSCs, as well as in fly ovaries reported by Huang et al. (2013), Le Thomas et al. (2013), Ohtani et al. (2013), and Rozhkov et al. (2013).

However, the vast majority of the PIWI-regulated genes in OSS cells were nonoverlapping with PIWI-regulated genes in OSCs (only ~6% of PIWI-regulated genes shared between OSCs and OSS cells) (Fig. 2D; Supplemental Table S3). Indeed, OSS cells and OSCs exhibit distinct gene expression profiles despite sharing a common origin from the Niki laboratory (Cherbas et al. 2011). Additionally, our genomic PCR analysis confirmed differences in de novo TE insertion patterns between OSCs and OSS cells as well as in different passages of the cells from previously cryopreserved lines that were thawed and revived for genomic DNA extraction (Fig. 2E). These data hinted at unique TE landscapes between OSCs and OSS cells.

A new pipeline to assess TE landscapes in OSC and OSS cell genomes

To comprehensively determine the genomic differences between follicle cell lines, we deeply sequenced the genomic DNA of our current OSS cells ("OSS_C") as well as two earlier passages of OSS ("OSS_E") and OSC ("OSC_E") cells that we had revived from our cryopreserved stocks (Fig. 3A). We then analyzed the deposited genomic sequence from Sienski et al. (2012) as current OSC ("OSC_C") cells, as well as sampled a vial of OSC_C from the Brennecke lab. All genomes were sequenced with either 100 or 150 bp reads on the Illumina platform to at least a depth of ~25 million reads and an average 10-fold genome coverage (Supplemental Table S1).

Inspired by previous methodologies for detecting de novo TE insertions in *Drosophila* (Khurana et al. 2011; Kofler et al. 2012; Linheiro and Bergman 2012; Sienski et al. 2012), we developed our own custom pipeline that combined the efficiency of Bowtie mapping of split reads (Langmead et al. 2009) with BLAT's ability to score the mappings according to degrees of matches (Kent 2002), enabling us to achieve greater sensitivity and specificity for detecting de novo TE insertions from single-end (SE) reads (Supplemental Fig. S5; Supplemental Text). We pinpointed reads with one end matching a specific TE while the other end matched the euchromatic *D. melanogaster* Release 5/dm3 reference genome. Reads were further clustered and then validated by BLAT so that de novo TEs were frequently represented by reads spanning both sides of the insertion. Our pipeline faithfully detected key TE insertions described in Sienski et al. (2012) and all the TE insertions we had determined by genomic PCR and resulted in 1196, 2847, 1143, and 1152 insertions from the OSS_E, OSS_C, OSC_E, and OSC_C genomes, respectively

(Supplemental Table S4). We empirically determined that our algorithm's false discovery rate (FDR) was on average <12.1% for all four libraries, and this was ~4-fold lower than other algorithms that only used a split-read mapping approach with Bowtie (see Supplemental Text). Importantly, the TE insertion reads pinpointed by our algorithm displayed signatures of target site duplications (TSD), which are short duplicated sequences flanking the TE insertion (i.e., "CG-[TE insertion]-CG") (Supplemental Fig. S5C; Fig. 3F). This molecular signature results from the repair and duplication of a staggered DNA cut during the TE mobilization process (Bowen and McDonald 2001; Linheiro and Bergman 2012).

Despite similar genomic categories for TE insertions, the breakdown of the most prevalent de novo TEs was strikingly different between the cell passages (Fig. 3B). *gypsy* and *roo* were the dominant de novo TEs in OSS_E cells, and a specific variant of *gypsy* called *springer* expanded significantly in the OSC_E and OSC_C line. However, the OSS_C genome truly stood out with more than twice as many TE insertions compared to the other lines, and *ZAM* was the main dominant TE, accounting for a surprising 41% of the de novo insertions. Indeed, the large presence of de novo *ZAM* insertions in OSS_C cell genomes is consistent with high PIWI CLIP-seq scores corresponding to *ZAM* (Fig. 1D), *ZAM* being the second most abundantly expressed TE (after *copia*) at steady state in siGFP-treated OSS cells (data not shown).

From genomic PCR analyses, TEs near the *fau* and *Mec2* genes were fully persistent in the culture, whereas other loci, such as in *CG15278*, *KCNQ*, *Prosap*, and *Dlp*, exhibited both the "reference genome allele" and the "TE allele" (Fig. 2E). Therefore, we computed a coverage ratio (CR) for each TE insertion by dividing the number of TE insertion reads by the number of reference genome-mapping reads falling within the same coordinates covered by the TE insertion reads (see Supplemental Text). The karyotypes for both OSCs and OSS cells are a mixture of diploid and putatively polyploid cells (Supplemental Fig. S1A), and OSS cells tended to be more polyploid. Because our genome sequencing covers the entire spectrum of insertion types within the culture, we designate "Persistent" TE insertions having a CR > 4.0, while other TE insertions with CR ≤ 4.0 were considered "Heterogeneous."

When these classes of TE insertions were plotted (Supplemental Fig. S6; Fig. 3C), the patterns were highly dispersed across the lengths of the major euchromatic chromosomes 2, 3, and X. There was a range of proportions between persistent TE insertions versus heterogeneous TE insertions among the cell passages and within the different major chromosomes. The right arms of chromosome 2 from OSC_E, OSC_C, and OSS_E were depleted in heterogeneous TE insertions, while the entire genome of OSS_C had a greater proportion of heterogeneous TE insertions (Fig. 3D). As expected, de novo TE insertions tended to avoid exons, whose disruption might nullify protein expression (Fig. 3E). However, in all cell lines there were statistically significant preferences for TEs to insert in intronic and intergenic regions within 1 kb of a gene's annotation when compared to chance. This could be attributed to greater chromatin accessibility (Fontanillas et al. 2007; Spradling et al. 2011) and, perhaps consequently, more frequent impact on gonadal cell transcriptomes.

De novo TEs did not accumulate in "hot spots" like existing TEs, which concentrate in the piRNA-generating clusters and the assembled pericentromeric heterochromatin in the *D. melanogaster* Release 5 reference genome. Our algorithm, which requires one

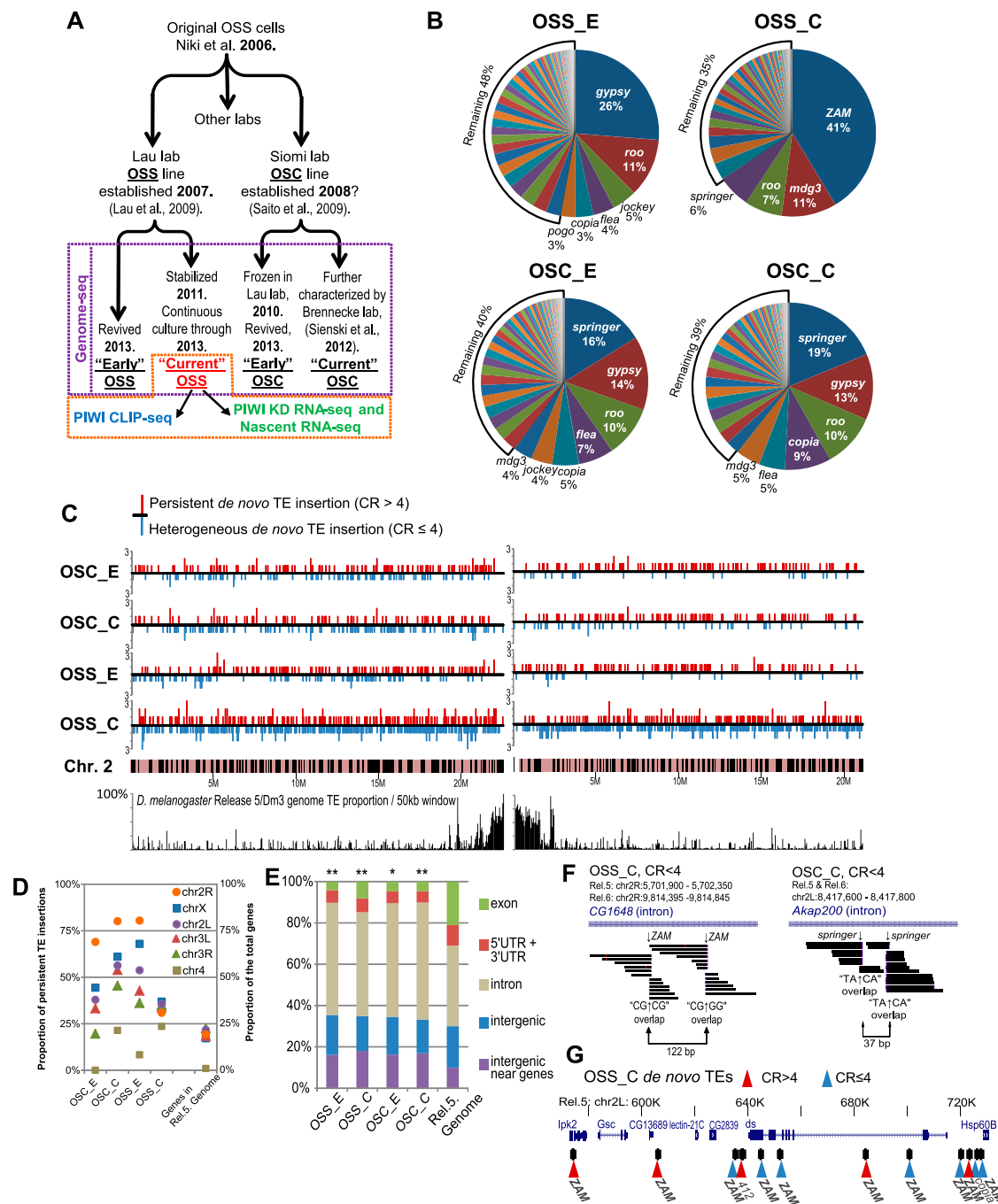


Figure 3. Genome sequencing of follicle cell lines reveals *de novo* TE landscape diversity. (A) Diagram of the history of the OSS and OSC cell lines used in this study and which genomics approaches were applied to specific cell lines. (B) Proportions of the classes of TEs comprising all *de novo* TE insertions. (C) Distribution of persistent and heterogeneous TE insertions on the *Drosophila* chromosome 2, overlaid on the *D. melanogaster* Release 5/dm3 genome proportions of TEs comprising the left and right arms. Complete genome-TE landscape maps are in Supplemental Figure S6. (D) Dot graph of the ratios of persistent TE insertions among the different chromosomes in the four cell lines. Whereas all chromosomes in OSS_C cells have a similar lower ratio of persistent TE insertions, chromosome 2R has a notably higher proportion of persistent TE insertions in OSC cells and OSS_E cells. (E) The frequencies of TE insertions in different genome functional regions are significantly distinct from the expected proportions of functional regions in the Release 5/dm3 reference genome. (**) *P*-value < 0.001, (*) *P*-value < 0.05; Fisher’s exact test. (F) Examples of tandem *de novo* TE insertions in OSS_C and OSC_C cells. (G) An emerging “hot spot” for multiple ZAM insertions in the OSS_C genomes.

uniquely mapping reference genome anchor, is limited in resolving *de novo* TEs in heterochromatin. However, we noted several instances in both OSS and OSC genomes of tandem insertions of the same TE class as close as <50 bp apart (Fig. 3F). The

major expansion of ZAM in OSS_C cells also created a 30-kb region containing 10 ZAM insertions with two other TEs, suggesting this could be an emerging “hot spot” for *de novo* TEs to land (Fig. 3G).

TE landscapes reveal the relatedness between early and current OSC and OSS cell passages

To examine the dynamics of TE landscapes between cell passages, we generated similarity and difference maps of TE insertions between two cell genomes (Fig. 4A; Supplemental Fig. S7). We first compared the genomes of early passages of OSCs and OSS cells to current passages and partitioned persistent TE insertions ($CR > 4$) from heterogeneous insertions to illustrate dynamic TE insertion emergence. This is highlighted in the heterogeneous TE insertions that comprised the bulk doubling of TE insertions in the OSS_C line compared to the “ancestral” OSS_E genome (Fig. 4B). In contrast, the total number of de novo TEs did not substantially increase between OSC_C and OSC_E passages, because the gain of heterogeneous TE insertions was offset by the loss of persistent insertions between OSC cultures (Fig. 4A,B).

These de novo TE difference maps enabled us to examine cell-passage relatedness, as visualized in a Euler plot (Fig. 4C). The majority of de novo TE insertions (>76%) were shared in OSC_E and OSC_C cells (see also Fig. 3B), confirming their close relatedness in originating from the Siomi laboratory and being passaged for a shorter time compared to OSS_C cells. In contrast, only a minor proportion (12%) of OSS_C TE insertions were shared with the OSS_E cell passage, with OSS_C insertions largely being distinct from OSC_E, OSC_C, and OSC_E genomes. Rather, OSS_E shared a greater proportion of its insertions (~47%) with the

OSC_E and OSC_C lines, which is consistent with the history of closer time frames from when these cells were first obtained from the Niki laboratory source (Fig. 4D). Therefore, the duration of culturing and other factors may have greatly distorted the TE landscape of OSS_C cells so that they have become a distinct follicle cell population, unique from OSCs and earlier passages of OSS cells. Despite a modest increase in overall de novo TE insertions in current versus early OSCs, these data also suggest OSCs generally exhibit more stable TE landscapes compared to OSS cells.

De novo TE insertions stimulate novel candidate long-noncoding RNAs (lncRNAs) that associate with PIWI and overlap with PIWI-regulated genes

The distinct TE landscapes between OSCs and OSS cells prompted us to examine how extensively TE insertions could alter transcriptomes beyond TEs and mRNAs. By searching nascent RNA reads around the vicinities of de novo TE insertions, we also discovered 204 and 289 candidate lncRNAs in current OSCs and OSS cells, respectively (Fig. 5; Supplemental Table S5). Our conservative criteria for determining unambiguous candidate lncRNAs near TEs is defined by at least 10 RPM of nascent RNA tags, being at least 1-kb long, and were within 1 kb or overlapping a de novo TE insertion. These candidate lncRNAs were transcribed completely antisense to coding transcripts (~71%), or from unannotated genomic regions (~23%), or beginning within a large intron of

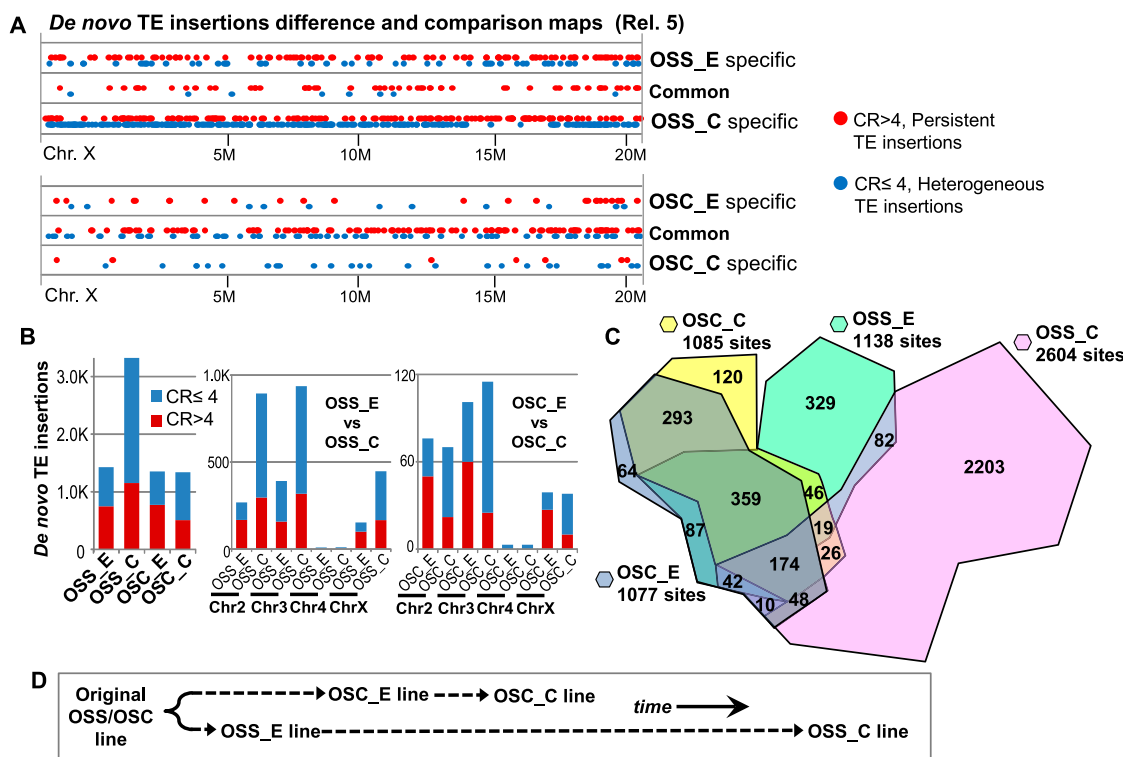


Figure 4. Comparison of TE landscapes indicates dynamics and relatedness of gonadal cell lines. (A) Difference maps of *Drosophila* chromosome X comparing OSS early and current and OSC early in current cell lines. Each dot represents a TE insertion locus that is either specific to one cell line or commonly shared by both compared lines. Dots for persistent and heterogeneous insertions are offset to aid visualization. Complete *D. melanogaster* Release 5 reference comparison maps are in Supplemental Figure S7. (B) Tally of total de novo TE insertions (left) and chromosomal breakdown of cell line-specific TE insertions (middle and right). (C) Euler plot showing the greatest overlap in de novo TE insertions between OSC_E and OSC_C cells, greater overlap between OSC cells and OSS_E cells, and unexpectedly low overlap between OSS_C cells and other lines. The overlapping regions are the number of individual TE insertions that are located at the same genomic position between the cell lines. (D) Model for cell line relatedness based on TE landscape similarities and expansion of TEs during prolonged culture of OSS cells.

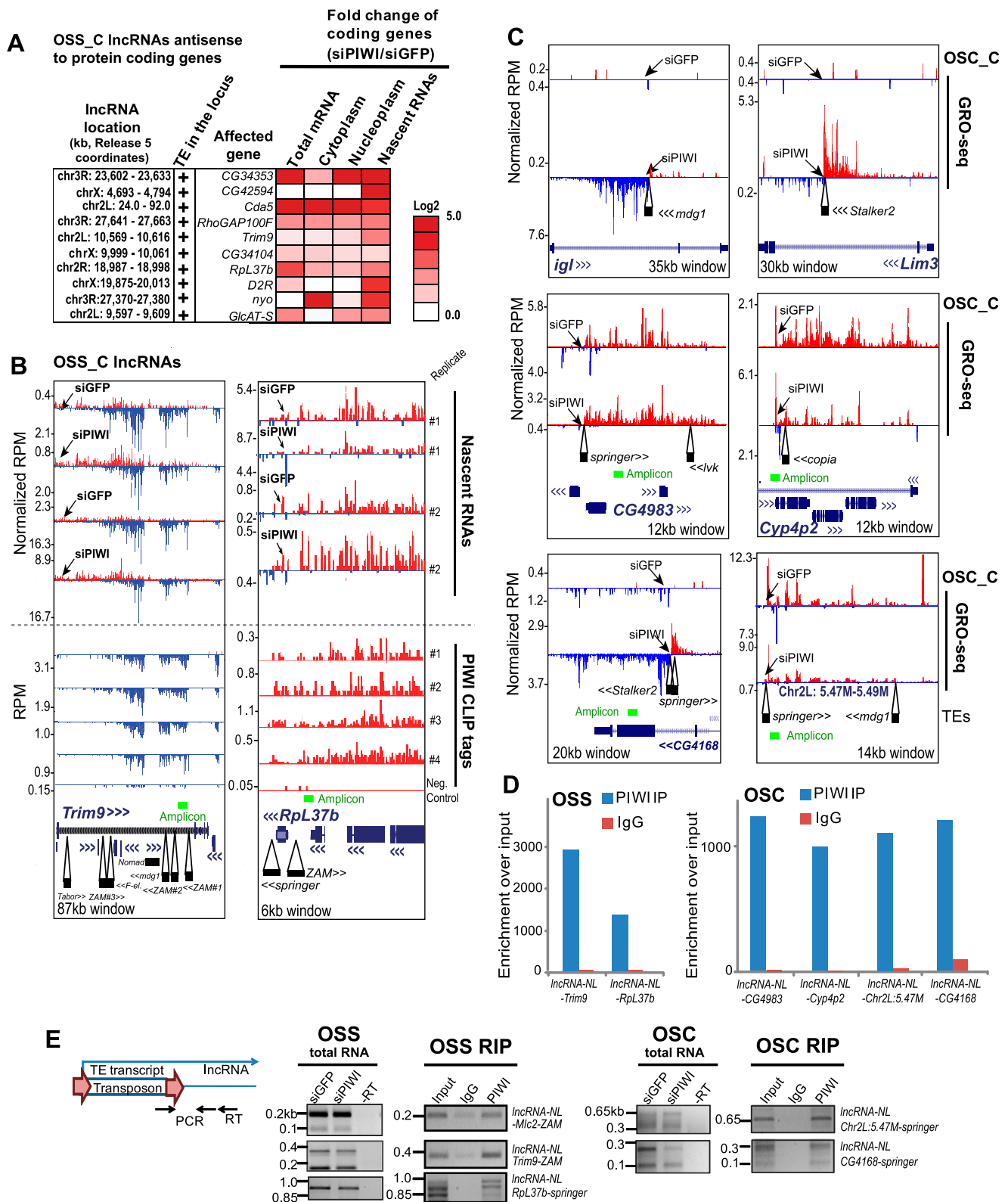


Figure 5. Novel long noncoding RNAs (lncRNAs) are stimulated by de novo TE insertions. (A) Heat map of representative OSS cell genes overlapped by lncRNAs and up-regulated during PIWI knockdown. *D. melanogaster* Release 5/dm3 reference genome coordinates are shown here; updated Release 6 coordinates are in Supplemental Figure S8E. (B) Representative lncRNA loci in OSS cells, with nascent RNAs in upper tracks and PIWI CLIP tag profiles in lower tracks. The *lncRNA-NL-Trim9* is on the minus strand (blue reads), while the *lncRNA-NL-RpL37b* is on the plus strand (red reads). Arrows point to coding transcripts for *Trim9* (red reads) and *RpL37b* (blue reads) that increase when PIWI is knocked down. Location of de novo TE insertions are at the bottom of the diagrams, whereas green bars mark amplicons for evaluating lncRNA enrichment in a PIWI RIP experiment and RT-qPCR. (C) Representative TE-associated lncRNAs in OSC cells identified from GRO-seq reads. The lncRNAs at *igl* and *Lim3* are unambiguously antisense to the coding transcripts, whereas the transcripts overlapping the same sense of *CG4983* and *Cyp4p2* may not adhere to the strict definition of lncRNAs, but their long extensions in noncoding regions are reminiscent of defined lncRNAs. Arrows point to the putative start of the lncRNA. (D) RIP assay validates lncRNA association with PIWI. The lncRNAs at *igl* and *Lim3* were not analyzed by PIWI RIP, since they are expressed only upon PIWI knockdown. (E) RT-PCR of amplicons that span the inserted TE sequence and the lncRNA. These data are consistent with lncRNA nascent reads not affected by PIWI knockdown and enrichment of lncRNA in PIWI RIP.

a gene and extended at least 1 kb beyond the stop codon (~6%). The mean and median lengths of these transcripts were ~22 kb and ~12 kb long in OSS cells, respectively, and ~17 kb for both mean and median lengths in OSCs. These differences in average candidate lncRNA lengths may be due to OSS cells being profiled by Nascent-seq, while OSCs were profiled by GRO-seq. Although both techniques effectively measure nascent RNAs, slightly different nascent read profiles have been observed (Ferrari et al. 2013). Despite the fact that our list of candidate lncRNAs in OSCs and OSS cells was distinct from another list of *Drosophila* lncRNAs derived from modENCODE data sets (Brown et al. 2014), the bulk of these lncRNA transcripts were predicted by the Coding-Potential Assessment Tool (CPAT) (Wang et al. 2013) to have overall protein-coding probabilities below 2%, well under the coding potential of equivalent protein-coding transcripts and supporting their characterization as lncRNAs (Supplemental Fig. S8A).

Although most of these TE-associated lncRNAs were transcribed antisense to a coding gene, only ~5% and ~15% of OSS and OSC TE-linked lncRNAs, respectively, overlapped a coding gene that was up-regulated upon PIWI knockdown (Fig. 5A; Supplemental Table S5). For example, *RpL37b* and *Trim9* transcripts were up-regulated ~8- and ~3-fold upon PIWI knockdown in OSS cells even though the lncRNA transcripts that overlapped them were unaffected. The PIWI CLIP-seq scores at these two transcripts were complicated by the fact that the majority of the CLIP tags actually corresponded to the TE-linked lncRNAs (Fig. 5B). We conducted RIP assays to confirm that these lncRNAs were associated with PIWI (Fig. 5D), yet few piRNAs mapped to these lncRNA loci. In RT-PCR experiments with one primer within the de novo TE insertion and a second primer for the lncRNA, we found that TE sequences are part of the same transcript as the lncRNAs and were enriched in PIWI IPs (Fig. 5E). PIWI recruitment to lncRNAs could begin with piRNAs base-pairing to the TE sequences; however, the much broader PIWI CLIP-seq tags throughout the rest of the lncRNAs could also be attributed to promiscuous RNA binding by PIWI, which was observed in our in vitro experiments (Supplemental Fig. S4H), as well as in previously reported studies with mouse Piwi proteins (Vourekas et al. 2012). Further study will be needed to understand why some lncRNAs are unaffected by PIWI while the overlapping gene nearest the TE is silenced. We propose the hypothesis that lncRNAs could act as a “decoy” for PIWI binding without a piRNA, and thus not yet activating PIWI for silencing.

In fact, most lncRNAs (77%) in OSS cells like the *lncRNA-NL-Trim9* and *lncRNA-NL-RpL37b* were not significantly affected after PIWI knockdown, but *lncRNA-NL-CG15483* and *lncRNA-NL-EcR* represent two clear examples of OSS lncRNAs clearly affected by PIWI silencing (Supplemental Fig. S8B). In contrast, the expression levels for the majority (62%) of lncRNAs in OSCs were up-regulated upon PIWI knockdown, such as *lncRNA-NL-igl* and *lncRNA-NL-Lim3* (Fig. 5C). We also do not fully understand the differences in lncRNA regulation between OSCs and OSS cells, which could be due to other epigenetic mechanisms operating differently between these two cell lines.

To show that PIWI also associated with lncRNAs in OSCs, we need to examine the stable lncRNAs in OSCs in untreated cells, but most OSC lncRNAs were lowly expressed. However, we found two putative TE-linked lncRNAs that were highly expressed in untreated OSCs, *lncRNA-NL-CG4983* and *lncRNA-NL-Cyp4p2* (Fig. 5C), and were enriched in the PIWI RIP assay (Fig. 5D). These lncRNAs were not automatically classified in our algorithm because they were transcribed from the same sense strand of coding genes, yet extended many kilobases beyond the shorter annotated

coding transcripts. These examples suggest that we are underestimating the total lncRNA diversity in *Drosophila* and highlight the challenge imposed by the compactness of the *Drosophila* genome on lncRNA determination.

Less than 13% of the unambiguous lncRNAs had overlapping coordinates between OSCs and OSS cells (Fig. 6A), with two examples in *lncRNAs-NL-fau* and *lncRNA-NL-Sp7* (Fig. 6B). Despite the same genomic coordinate window on these loci, the lncRNAs between the cell lines were actually different in their start sites and profiles as well as distinct de novo TE insertions. Therefore, we hypothesized that the de novo TE diversity between cell lines could actually be the major driver of differences in lncRNA expression. When we counted only the de novo TE classes proximal to lncRNAs (24% and 18% of the total de novo TEs in OSCs and OSS cells, respectively), *ZAM* was still the dominant TE in OSS cells, while *springer* and *gypsy* were the dominant TEs in OSCs (Fig. 6C); thus, lncRNA-associated TEs reflected the overall TE expansion in the cells' genomes. We also calculated that the majority of the lncRNA-associated TEs were inserted at the 5' end of the lncRNA (~77% and ~62% in OSCs and OSS cells, respectively), suggesting a bias in the insertion position for these TEs to favor stimulating lncRNA expression.

To examine this hypothesis experimentally, we scrutinized our deep-sequencing data sets for lncRNAs that were present in current OSCs but not in current OSS cells and vice versa (Fig. 6B). We performed RT-qPCR on total RNA from OSCs and OSS cells on two OSS cell-specific lncRNAs (*lncRNA-NL-Trim9* and *lncRNA-NL-RpL37b*) and four lncRNAs specific to OSCs (*lncRNA-NL-Chr2L:5.47M*, *lncRNA-NL-Cyp4p2*, *lncRNA-NL-CG4983*, and *lncRNA-NL-CG4168*) to confirm that lncRNA expression was only occurring in one but not the other cell line (Fig. 6D). We then confirmed by genomic PCR that only in the cell line where a lncRNA was expressed was there a de novo TE insertion present; whereas genomic PCR and RT-qPCR also confirmed that the reference genome amplicon was present in those cell lines which then lacked lncRNA expression (Fig. 6E). Altogether, these data imply that the de novo TEs are responsible for stimulating lncRNA expression in both OSCs and OSS cells and contributing to transcriptome diversity.

Discussion

In this study, we determined the spectrum of transcripts that are regulated by the Piwi pathway in OSS cells, which serve as a simplified model system compared to the fly ovary because they express only a single PIWI protein and primary piRNAs (Lau et al. 2009; Saito et al. 2009). Our approaches confirmed TE transcripts are the main targets of PIWI-mediated repression, whereas a set of genes with de novo inserted TEs were also suppressed by PIWI at the nascent RNA level indirectly through silencing of the adjacent TE. These data were further supported by our complementary reporter assay study that indicates PIWI-mediated gene silencing requires a bulk of antisense piRNAs to pair with a TE target, and this mechanism can rapidly repress transcription prior to RNA splicing on transgenes (Post et al. 2014). Our conclusions are consistent with other studies examining PIWI gene silencing (Li et al. 2009; Haase et al. 2010; Saito et al. 2010; Sienski et al. 2012; Donertas et al. 2013; Huang et al. 2013; Le Thomas et al. 2013; Ohtani et al. 2013). However, our study distinctly shows that despite many coding gene transcripts predicted to base-pair with piRNAs (Fig. 1A), the association of PIWI/piRNA complexes with coding gene transcripts is much more limited, whereas PIWI-mediated silencing is triggered through a bulk interaction of PIWI/piRNAs with targets like TEs (Post et al. 2014).

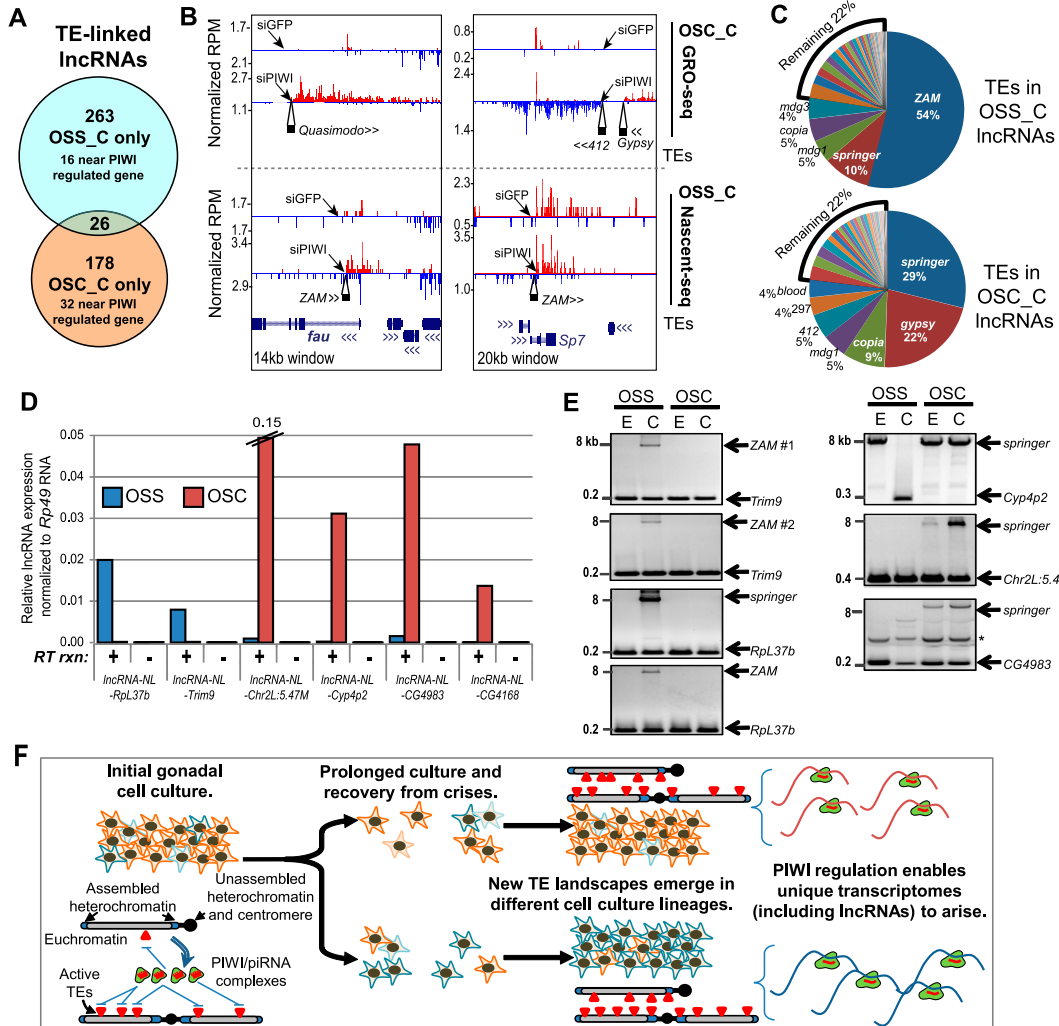


Figure 6. TE landscape differences correlate with lncRNA diversity. (A) Venn diagram of unambiguous lncRNAs that are mostly distinct populations between OSCs and OSS cells. The overlap represents lncRNAs that overlap by at least 1 kb in genomic coordinates from the *D. melanogaster* Release 5/dm3 genome. (B) Representative loci where both OSC and OSS cell genomes contain TE insertions, but the differences in TE composition are linked to differences in lncRNA configurations. (C) Composition of the major TE classes falling within lncRNA loci. (D) Quantitative RT-PCR of lncRNA expression, normalized to *Rp49* levels, for loci where TE insertions are present in just one follicle cell line (lncRNA-*NL-Trim9* and -*NL-RpL37b* in OSS cells; -*NL-Chr2L:5.47M*, -*NL-Cyp4p2*, -*NL-CG4983*, and -*NL-CG4168* in OSCs). (E) Genomic PCR confirmation of de novo TE insertions that are either only in OSCs or only in OSS cells. Asterisk marks a nonspecific amplicon, and sequence-specific obstacles in primer design prevented genomic PCR validation of TE insertions in the *CG4168* locus. (F) Model for TE dynamics in follicle cell cultures that result in distinct TE landscapes and unique transcriptome profiles, including the production of novel lncRNAs.

By comparing PIWI-regulated transcriptomes between OSCs and OSS cells, we discovered an unexpected diversity of de novo TE landscapes. Furthermore, after analyzing earlier cryopreserved passages of OSCs and OSS cells, we propose a model for TE dynamics under the influence of the Piwi pathway (Figs. 4D, 6F). The similarities in TE landscapes between OSC_E, OSC_C, and OSS_E lines may suggest that they are the earliest and most stable descendants of an initial gonadal cell culture. However, standard culture conditions can still generate variation among individual cell genomes within a culture, as groups of cells fluctuate in their growth proportions and undergo “bottlenecks” during crises. This can then create heterogeneity of TE persistence within these group-cell analyses and a general increase in the number of de novo TE insertions. Alternatively, the dynamic TE landscapes might also reflect differential selection of pre-existing heterogeneity in the

aggregate cell population or reflect discrepancies between piRNA abundance and TE de-repression (Khurana et al. 2011).

With prolonged continuous culture of the OSS_C line, a more extreme scenario emerged where TEs like *ZAM* have greatly expanded in the genome and escaped suppression by the Piwi pathway. *ZAM* is naturally expressed in *Drosophila* follicle cells (Meignin et al. 2004), where it can generate virus-like particles from follicle cells that can be transmitted to the oocyte (Brasset et al. 2006). However, the two piRNA clusters *flamenco* and *COM* typically keep this TE in check by providing *ZAM*-directed piRNAs that will direct PIWI-mediated gene silencing to the *ZAM* loci (Desset et al. 2003, 2008; Saito et al. 2006; Vagin et al. 2006; Brennecke et al. 2007; Mevel-Ninio et al. 2007). *ZAM* mRNAs are prevalent in OSCs and OSS cells despite the presence of *ZAM*-directed piRNAs (Lau et al. 2009; Saito et al. 2009; Sienski et al. 2012),

yet OSCs have few *ZAM* de novo insertions and instead have allowed *springer* to gain a foothold. Determining how the *ZAM* TE specifically accumulated in OSS cells and evaded suppression by the PIWI/piRNA complex will be an important future direction.

Our analysis reveals that TE dynamics and landscape diversity can profoundly affect transcriptomes in follicle cells, rendering different genes to become targets of PIWI regulation while also promoting the expression of novel repertoires of lncRNAs. Some de novo TEs at lncRNA loci may exhibit dually opposing functions: on the one hand imparting PIWI-mediated gene silencing to nearby coding genes and on the other hand stimulating lncRNA expression. Furthermore, our data suggest PIWI may associate with lncRNAs independently of piRNA base-pairing, similar to a model for MIWI interaction with certain transcripts (Vourekas et al. 2012). Additionally, some lncRNAs were transcribed antisense to other genes that appeared to decrease in expression after PIWI knockdown, most significantly in the cytoplasm (Supplemental Fig. S8C,D). Because the multi-kilobase lengths of lncRNAs pose a challenge to test their function, (e.g., in our reporter assay for PIWI-silencing) (Post et al. 2014), the down-regulation mechanism of these lncRNA-associated coding genes after PIWI knockdown remains unclear. However, this is reminiscent of PIWI-linked gene activation at a telomeric gene locus in *Drosophila* (Yin and Lin 2007). Future approaches will be needed to evaluate the function of these novel PIWI-associated lncRNAs.

The regulation of TE-associated lncRNAs is highly complex, perhaps integrating additional levels of epigenetic regulation beyond the Piwi pathway. For instance, we observed hints at a few lncRNA loci that both the silencing mark of histone H3-lysine-9-trimethylation (H3K9me3) as well as the gene expression marks of RNA polymerase II (RNA Pol II) and H3K36me3 were present (Supplemental Fig. S9A,B). We cannot yet determine whether these lncRNAs are functionally required for PIWI-mediated gene silencing or if they are perhaps acting as a “decoy” for PIWI binding and contributing to the de novo TE’s expression. However, it is clear that lncRNA expression differences between OSCs and OSS cells correlates with particular de novo TEs inserted in one line and not the other. Additionally, this mechanism to generate novel lncRNAs via new TE insertions that could be subjected to Piwi regulation might be a route for animal genomes to sample newly emerging lncRNAs, modulating their expression via the Piwi pathway until a function for the lncRNA is selected. Recently, a number of vertebrate lncRNAs have also been suggested to be stimulated by TEs (Kelley and Rinn 2012; Kapusta et al. 2013); thus it is tempting to speculate that Piwi pathways in vertebrates may also regulate TE-associated lncRNAs.

Deep-sequencing of inbred fly lines from the *Drosophila* Genetic Reference Panel has also revealed extensive diversity in de novo TE insertions, with multiple insertions detected per fly line and in all the inbred lines (Linheiro and Bergman 2012; Mackay et al. 2012). In contrast to a whole animal, OSS cells are unencumbered by the sensitivity of gonad development, and a cell culture selection experiment can sample millions of genomes in a culture dish simultaneously. Therefore, OSS cells suitably complement fly genetic studies and are a rapid system to explore the effect of the Piwi pathway on the emergence of new de novo TE landscapes. Furthermore, the biochemical tractability and RNA-seq approaches we have described here will enable future efforts to show how each distinct TE landscape can impact gonadal cell transcriptomes including the generation of novel lncRNAs. It will be interesting to determine what may be the upper limit for animal cell genomes to accommodate additional de novo TE insertions while under PIWI regulation and during changes in cell ploidy (Supplemental Fig. S1; Ng et al. 2012; Arkhipova and Rodriguez

2013). The natural dynamism of TE landscapes in *Drosophila* follicle cell cultures will be integral for further studies on how the Piwi pathway and TEs influence gene and genome regulation.

Methods

OSS cell cultivation and siRNA treatments

OSS and OSC cell cultures and siRNA knockdowns were performed according to Niki et al. (2006), Lau et al. (2009), and Saito et al. (2009), with additional culture modifications only during the stabilization process for culturing our current OSS cells. For PIWI depletions, 500 pmol siRNAs were electroporated into a 10-cm plate of OSS cells using Amaxa kit V (program D013), and cells were cultured for 6 d before harvesting. Successful depletion was assessed by Western blotting. All oligonucleotide sequences utilized for knockdowns, genomic PCR, qRT-PCR, and library constructions are listed in Supplemental Table S6.

Western blots, antibodies, recombinant GST-PIWI purification, and EMSA

Western blotting was performed according to standard protocols. The mouse anti-PIWI antibody was a gift from the Siomi laboratory. Rabbit polyclonal antibody to PIWI was raised to the synthetic peptide that corresponds to the N terminus (CADDQGRGRRRPLNEDD). Mouse anti-alpha Tubulin (E7A) antibody was purchased from the Developmental Studies Hybridoma Bank. GST-PIWI was expressed from pGEX 6P-1-piwi in BL21(DE3)Rosetta cells induced overnight at 16°C. Cells were lysed in a lysis buffer (50 mM KH₂PO₄, pH 8; 1 M NaCl, 5 mM β-mercaptoethanol, 3% glycerol, 1 mM PMSF) with sonication and purified on a GST resin column (GE Healthcare) according to manufacturer instructions. A mobility shift assay was performed essentially as described (Cho et al. 2012).

PIWI CLIP-seq procedure

PIWI CLIP-seq procedures were adapted from Chi et al. (2009) and Leung et al. (2011) (see Supplemental Methods). In brief, two 10-cm dishes of OSS cells were irradiated with UV light and then lysed in a dounce homogenizer, followed by sonication in Q buffer (20 mM HEPES-KOH, pH 7.9; 10% glycerol; 0.1 M KOAc; 0.2 mM EDTA; 1.5 mM MgCl₂; 0.5 mM DTT; 1× Roche Complete EDTA-free Protease Inhibitor Cocktail; 0.5% NP40). Protein A/G magnetic beads (Pierce) were coated with anti-PIWI antibodies at 60 μg per 60 μl of beads for 2 h at 23°C, and washed with PBS. For the antigen-blocking peptide block negative control experiment, 0.1 mg/mL of the peptide was pre-incubated with antibody-coated beads for 1 h prior to adding OSS cell lysate. Extracted RNAs from PIWI complexes were labeled with PNK and ³²P-γATP.

To generate CLIP-seq libraries, a pre-adenylated 3’ adaptor was ligated to PIWI-bound RNAs attached to beads. T4 PNK and ATP were then added to the beads for 5’-end phosphorylation. Beads were then washed four times with PNK buffer and RNA, and proteins were eluted in Laemmli SDS-PAGE loading buffer and resolved on a Bis-Tris Nu-PAGE gel, 4%–12% (Invitrogen). The region around ~80–200 kDa was cut out, and RNA was extracted from gel slices by passive elution and cleaned up with the RNA Clean & Concentrator kit (Zymo Research). 5’ adaptors were ligated overnight and reverse transcription was performed with SuperScript III RT. DNA was amplified with Phusion polymerase, and 100- to 200-bp products were sequenced on the Illumina HiSeq 2000 platform (50-bp run).

RNP immunoprecipitation (RIP) assay

Cell lysates were subjected to IP with 25 μ l of protein A/G magnetic beads (Pierce) and 20 μ g of antibody in the presence of 1 unit/ μ l of Ribolock RNase inhibitor. Beads were washed five times with Q buffer; RNA was extracted with TRI reagent from 80% of the beads. RT-qPCR was done using M-MLV reverse transcriptase and GoTaq SYBR Green Master Mix (Promega). Twenty percent of the beads were subjected to Western immunoblot. The amount of immunoprecipitated RNA was normalized to the amount in 10% of input and rabbit IgG was used as a negative control.

Cell fractionation, messenger RNA, nascent RNA, and qRT-PCR analysis

Cell fractionation and isolation of native nascent transcripts (NUN fraction) was done according to Khodor et al. (2011). RNA was extracted from OSS cells and fractions by TRI reagent RT (MRC Inc). The remaining DNA was digested by 6 units of DNase I (NEB). Total RNA samples as well as RNA from the cytoplasm and nucleoplasm were subjected to two rounds of polyA enrichment using biotinylated (dT)₁₈ oligo and a PolyAtract kit (Promega). RNA from the NUN fraction was depleted from ribosomal RNA using a biotinylated oligo set (Pennington et al. 2013). RNA-seq libraries were obtained from four independent PIWI knockdown experiments using four different library construction protocols: (1) ScriptSeq V1; (2) ScriptSeq V2 (Epicenter; performed according to manufacturer's instructions); (3) random primer-based library construction protocol (Pennington et al. 2013); and (4) mRNA fragmentation followed by the small RNA library construction protocol (sRNA-seq) (Matts et al. 2014). Sequencing was performed on an Illumina HiSeq 2000, and 50-bp-long reads were processed and split according to their index primer barcodes.

To experimentally validate RNA expression changes, reverse transcription was performed using 0.2 μ g of RNA and M-MLV reverse transcriptase (Promega). qPCR was performed using GoTaq SYBR Green Master Mix (Promega) on a Bio-Rad C1000 quantitative PCR machine.

Bioinformatics analysis of CLIP-seq, RNA-seq, and nascent-seq

We developed our algorithmic procedures modeled after Chi et al. (2009), Khodor et al. (2011), Leung et al. (2011), and Pennington et al. (2013) and were written as custom C and shell scripts. FASTQ file reads were quality checked, adaptor sequences were trimmed, and reads were mapped to the RefSeq transcripts and genomic sequence from the *D. melanogaster* Release 5/dm3 genome by using Bowtie (allowing maximum two mismatches) (Langmead et al. 2009). Structural RNAs were determined by cross-mapping to a custom database and removed from subsequent analyses. TE reads mapping was performed against a list of *Drosophila* consensus TE sequences obtained from the Repbase database, Release 19 (Kapitonov and Jurka 2008) and from FlyBase, Release 5 (Kaminker et al. 2002), while virus sequences were obtained from GenBank. Within each data set of mapped reads, signal merge counts for unique reads were obtained (read frequency information was dismissed to avoid the jackpot effect). The basic processing pipeline is written in the shell script (process-quick.sh).

To subtract the background signal based on transcript expression levels in the PIWI CLIP-seq, we implemented the noise filters previously described in Chi et al. (2009) and Leung et al. (2011). For each gene, the combined expression data from four RNA-seq experiments performed in this study were subjected to log₁₀ transformation. To model the random association of RNA fragments during IP, the in silico CLIP algorithm broke each mRNA sequence into 50-nt windows and then calculated random CLIP as-

sociation probabilities based on the log₁₀ of transcript expression level. The calculated value was subtracted from experimentally merged CLIP counts, and final signals were quantified by RPM. The C code for the simulation is called `ngs_remove_clip_noise_by_mma.c`.

The top 50 genes with the highest PIWI CLIP-seq scores were picked for motif analysis. Up to the three highest peaks in the exon regions of each gene were selected, and the sequences of 150-bp length around the peaks were retrieved for motif analysis. A set of 202 sequences of 150 bp each from genes without CLIP signal were used as the negative sample. Motif analysis was performed using MEME (Bailey et al. 2009), GLAM2 (Frith et al. 2008), and Weeder (Pavesi and Pesole 2006).

Libraries obtained by RNA-seq were sorted according to barcodes in 5'-linker sequence, followed by trimming of linker sequences. RNA-seq reads were mapped with Bowtie to the exons of a RefSeq transcript gene list from the *D. melanogaster* Release 5/dm3 genome. Repbase was used for mapping reads to repetitive elements (Kapitonov and Jurka 2008). Gene expression was quantified by RPKMs (reads per kilobase per million) and each gene's RPKM value was further normalized to the RPKM value of *Rp49* (also known as *RpL32*). In addition, we conducted our differential expression analysis using the Bioconductor packages EdgeR (Robinson et al. 2010) and DESeq (Anders and Huber 2010) and followed the instructions provided in their respective vignettes. For EdgeR, we used an exact negative binomial testing procedure for differential expression.

Nascent RNA reads were merged and mapped to the *D. melanogaster* Release 5/dm3 genome, by using Bowtie (allowing a maximum of two mismatches). WIG files with a step size of 50 bp were generated and viewed by the UCSC Genome Browser. Read counts for each gene were calculated by overlapping the RefSeq track of the Genome Browser on the *D. melanogaster* Release 5/dm3 genome, and nascent RNA gene counts included intron and exon reads within a transcript. The transcript isoform with the highest count was selected as representative for a gene. The C code for counting the nascent RNA reads for each gene interval is called `ngs_genecentric.c`.

To identify candidate lncRNAs from OSCs and OSS cell nascent RNA reads, we took a hybrid approach of combining an automated search with manual curation. Using the defined coordinates for de novo TE insertions, we measured the strand-specific nascent RNA counts in 5-kb windows as well as tracked the annotations of genes orientations in these windows. We then sorted for windows where there were at least 10 RPM of nascent reads not in the same strand as an annotated gene. Using the TE insertion as an anchoring coordinate, we then manually inspected each lncRNA candidate in the UCSC Genome Browser on the *D. melanogaster* Release 5/dm3 genome, noting the following conservative criteria for unambiguous lncRNA transcripts as defined by at least 10 RPM of nascent RNA tags, being at least 1 kb long, and within 1 kb or overlapping a de novo TE insertion. Updated *D. melanogaster* Release 6 genome coordinates for lncRNAs were determined with webpage tools from FlyBase (http://flybase.org/static_pages/downloads/COORD.html) and the UCSC Genome Browser liftOver tool (<http://genome-preview.ucsc.edu/cgi-bin/hgLiftOver>). This update is reflected in Supplemental Figure S8 and Supplemental Table S5.

Genomic DNA library construction, sequencing, and TE insertion analysis

Genomic DNA extracted from 5×10^6 cells was fragmented with a Bioruptor sonicator (Diagenode): 8 cycles (20-sec pulse and 90-sec pause) with power set to "High." Fragmented DNA was used for library construction performed essentially as described (Ensminger et al. 2012). Single-end 150-nt sequencing runs with a v3 SBS kit were performed on a MiSeq.

Genome analyses for de novo TE insertions were initially inspired by the algorithmic procedures described in Khurana et al. (2011), Kofler et al. (2012), Linheiro and Bergman (2012), and Sienski et al. (2012). Our own custom pipeline for TE insertions was modified to increase sensitivity and specificity, and provide rich functional annotations of the regions around each de novo TE insertion. A custom Perl script was written to parse the genome sequencing reads as they were being processed via SAMtools (Li et al. 2009), Bowtie, and BLAT (Kent 2002) commands. We further processed this file by comparing each coordinate for a TE insertion to genome annotation files downloaded from the *D. melanogaster* Release 5/dm3 genome on the UCSC Genome Browser to designate if TE insertion was in an exon, intron, UTR, or intergenic region. Counts of TE insertion reads were then determined for 5-kb windows in the genome to smooth out inconsequential base differences for what were essentially identical de novo TE insertions between cell lines. Counting of TE classes, calculation of the coverage ratios, and building of density and difference maps were accomplished on Microsoft Excel spreadsheets and SQL analysis with Microsoft Access. Updated *D. melanogaster* Release 6 genome coordinates for TE insertions were determined with webpage tools from FlyBase (http://flybase.org/static_pages/downloads/COORD.html) and the UCSC Genome Browser liftOver tool (<http://genome-preview.ucsc.edu/cgi-bin/hgLiftOver>). This update is reflected in Figure 3, Supplemental Fig. S5, and Supplemental Table S4.

Data access

RNA-seq data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE45112. DNA-seq data have been submitted to the NCBI Sequence Read Archive (SRA; <http://www.ncbi.nlm.nih.gov/sra>) under accession number SRP039565. Computational scripts and additional BED files for this study can be found in the Supplemental Material and at http://www.bio.brandeis.edu/lalab/pubs_protocols/Sytnikova_et_al_GenomeRes2014_Additional_Items.html.

Acknowledgments

We thank Haruhiko and Mikiko Siomi for providing early OSCs, GST-PIWI constructs, and the monoclonal antibody to PIWI. We thank Julius Brennecke for his batch of current OSCs. We also thank Nahum Sonenberg (anti-dPABP1), David Glover (anti-larp), and Dorothea Godt (anti-tj) for the additional gifts of antibodies. We thank Dianne Schwarz and Michael Blower for assistance in facilitating genome sequencing on the Illumina MiSeq and for manuscript comments. We thank Michael Marr and Michael Rosbash for additional advice and comments on this manuscript and access to the Illumina HiSeq 2000. We are also grateful to Jack Bateman and Ed Dougherty for karyotyping advice. This work was supported by the National Institutes of Health (Core Facilities Grant P30 NS045713 to the Brandeis Biology Department, T32GM007122 to J.P.C., and R00HD057298 to N.C.L.). N.C.L. is a Searle Scholar.

References

Aliyari R, Wu Q, Li HW, Wang XH, Li F, Green LD, Han CS, Li WX, Ding SW. 2008. Mechanism of induction and suppression of antiviral immunity directed by virus-derived small RNAs in *Drosophila*. *Cell Host Microbe* **4**: 387–397.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.

Arkhipova IR, Rodriguez F. 2013. Genetic and epigenetic changes involving (retro)transposons in animal hybrids and polyploids. *Cytogenet Genome Res* **140**: 295–311.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202–W208.

Bowen NJ, McDonald JF. 2001. *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Res* **11**: 1527–1540.

Brasset E, Taddei AR, Arnaud F, Faye B, Fausto AM, Mazzini M, Giorgi F, Vaury C. 2006. Viral particles of the endogenous retrovirus ZAM from *Drosophila melanogaster* use a pre-existing endosome/exosome pathway for transfer to the oocyte. *Retrovirology* **3**: 25.

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

Brown JB, Boley N, Eisman R, May GE, Stoiber MH, Duff MO, Booth BW, Wen J, Park S, Suzuki AM, et al. 2014. Diversity and dynamics of the *Drosophila* transcriptome. *Nature* **512**: 393–399.

Cherbas L, Willingham A, Zhang D, Yang L, Zou Y, Eads BD, Carlson JW, Landolin JM, Kapranov P, Dumais J, et al. 2011. The transcriptional diversity of 25 *Drosophila* cell lines. *Genome Res* **21**: 301–314.

Chi SW, Zang JB, Mele A, Darnell RB. 2009. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460**: 479–486.

Cho J, Chang H, Kwon SC, Kim B, Kim Y, Choe J, Ha M, Kim YK, Kim VN. 2012. LIN28A is a suppressor of ER-associated translation in embryonic stem cells. *Cell* **151**: 765–777.

Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, Morell M, O'Shea KS, Moran JV, Gage FH. 2009. L1 retrotransposition in human neural progenitor cells. *Nature* **460**: 1127–1131.

Cowley M, Oakey RJ. 2013. Transposable elements re-wire and fine-tune the transcriptome. *PLoS Genet* **9**: e1003234.

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

Desset S, Meignin C, Dastugue B, Vaury C. 2003. COM, a heterochromatic locus governing the control of independent endogenous retroviruses from *Drosophila melanogaster*. *Genetics* **164**: 501–509.

Desset S, Buchon N, Meignin C, Coiffet M, Vaury C. 2008. In *Drosophila melanogaster* the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and -independent pathways. *PLoS One* **3**: e1526.

Donertas D, Sienski G, Brennecke J. 2013. *Drosophila* Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. *Genes Dev* **27**: 1693–1705.

Ensminger AW, Yassin Y, Miron A, Isberg RR. 2012. Experimental evolution of *Legionella pneumophila* in mouse macrophages leads to strains with altered determinants of environmental survival. *PLoS Pathog* **8**: e1002731.

Ferrari F, Plachetka A, Alekseyenko AA, Jung YL, Ozsolak F, Kharchenko PV, Park PJ, Kuroda MI. 2013. “Jump start and gain” model for dosage compensation in *Drosophila* based on direct sequencing of nascent transcripts. *Cell Rep* **5**: 629–636.

Flynt A, Liu N, Martin R, Lai EC. 2009. Dicing of viral replication intermediates during silencing of latent *Drosophila* viruses. *Proc Natl Acad Sci* **106**: 5270–5275.

Fontanillas P, Hartl DL, Reuter M. 2007. Genome organization and gene expression shape the transposable element distribution in the *Drosophila melanogaster* euchromatin. *PLoS Genet* **3**: e210.

Frith MC, Saunders NF, Kobe B, Bailey TL. 2008. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol* **4**: e1000071.

Garcia-Perez JL, Morell M, Scheys JO, Kulpa DA, Morell S, Carter CC, Hammer GD, Collins KL, O'Shea KS, Menendez P, et al. 2010. Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature* **466**: 769–773.

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

Goic B, Vodovar N, Mondotte JA, Monot C, Frangeul L, Blanc H, Gausson V, Vera-Otarola J, Cristofari G, Saleh MC. 2013. RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model *Drosophila*. *Nat Immunol* **14**: 396–403.

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

Hafner M, Landthaler M, Burger L, Khorshid M, Haussler J, Berminger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**: 129–141.

Han JS, Boeke JD. 2004. A highly active synthetic mammalian retrotransposon. *Nature* **429**: 314–318.

- Huang XA, Yin H, Sweeney S, Raha D, Snyder M, Lin H. 2013. A major epigenetic programming mechanism guided by piRNAs. *Dev Cell* **24**: 502–516.
- Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, Frise E, Wheeler DA, Lewis SE, Rubin GM et al. 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol* **3**: research0084.1–0084.20.
- Kapitonov VV, Jurka J. 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat Rev Genet* **9**: 411–412.
- Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, Yandell M, Feschotte C. 2013. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet* **9**: e1003470.
- Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H. 2008. *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**: 793–797.
- Kelley D, Rinn J. 2012. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol* **13**: R107.
- Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Res* **12**: 656–664.
- Khodor YL, Rodriguez J, Abruzzi KC, Tang CH, Marr MT 2nd, Rosbash M. 2011. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev* **25**: 2502–2512.
- Khurana JS, Wang J, Xu J, Koppetsch BS, Thomson TC, Nowosielska A, Li C, Zamore PD, Weng Z, Theurkauf WE. 2011. Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell* **147**: 1551–1563.
- Kofler R, Betancourt AJ, Schlotterer C. 2012. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLoS Genet* **8**: e1002487.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- Lau NC. 2010. Small RNAs in the animal gonad: guarding genomes and guiding development. *Int J Biochem Cell Biol* **42**: 1334–1347.
- Lau NC, Robine N, Martin R, Chung WJ, Niki Y, Berezikov E, Lai EC. 2009. Abundant primary piRNAs, endo-siRNAs, and microRNAs in a *Drosophila* ovary cell line. *Genome Res* **19**: 1776–1785.
- Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, Toth KF. 2013. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev* **27**: 390–399.
- Leung AK, Young AG, Bhutkar A, Zheng GX, Bosson AD, Nielsen CB, Sharp PA. 2011. Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat Struct Mol Biol* **18**: 237–244.
- Levin HL, Moran JV. 2011. Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet* **12**: 615–627.
- Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, Seitz H, Horwich MD, Syrzycka M, Honda BM, et al. 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137**: 509–521.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map (SAM) format and SAMtools. *Bioinformatics* **25**: 2078–2079.
- Linheiro RS, Bergman CM. 2012. Whole genome resequencing reveals natural target site preferences of transposable elements in *Drosophila melanogaster*. *PLoS ONE* **7**: e30008.
- Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, et al. 2012. The *Drosophila melanogaster* Genetic Reference Panel. *Nature* **482**: 173–178.
- Matts JA, Sytņnikova Y, Chirn GW, Igloi GL, Lau NC. 2014. Small RNA library construction from minute biological samples. *Methods Mol Biol* **1093**: 123–136.
- Meignin C, Dastugue B, Vaury C. 2004. Intercellular communication between germ line and somatic line is utilized to control the transcription of ZAM, an endogenous retrovirus from *Drosophila melanogaster*. *Nucleic Acids Res* **32**: 3799–3806.
- Mevl-Ninio M, Pelisson A, Kinder J, Campos AR, Bucheton A. 2007. The flamenco locus controls the gypsy and ZAM retroviruses and is required for *Drosophila* oogenesis. *Genetics* **175**: 1615–1624.
- Ng DW, Lu J, Chen ZJ. 2012. Big roles for small RNAs in polyploidy, hybrid vigor, and hybrid incompatibility. *Curr Opin Plant Biol* **15**: 154–161.
- Niki Y, Yamaguchi T, Mahowald AP. 2006. Establishment of stable cell lines of *Drosophila* germ-line stem cells. *Proc Natl Acad Sci* **103**: 16325–16330.
- Ohtani H, Iwasaki YW, Shibuya A, Siomi H, Siomi MC, Saito K. 2013. DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the *Drosophila* ovary. *Genes Dev* **27**: 1656–1661.
- Olivieri D, Sykora MM, Sachidanandam R, Mechtler K, Brennecke J. 2011. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *EMBO J* **29**: 3301–3317.
- Pavesi G, Pesole G. 2006. Using Weeder for the discovery of conserved transcription factor binding sites. *Curr Protoc Bioinformatics* **15**: 2.11.1–2.11.19.
- Pelisson A, Song SU, Prud'homme N, Smith PA, Bucheton A, Corces VG. 1994. Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila* flamenco gene. *EMBO J* **13**: 4401–4411.
- Pennington KL, Marr SK, Chirn GW, Marr MT II. 2013. Holo-TFIID controls the magnitude of a transcription burst and fine-tuning of transcription. *Proc Natl Acad Sci* **110**: 7678–7683.
- Post C, Clark JP, Sytņnikova YA, Chirn GW, Lau NC. 2014. The capacity of target silencing by *Drosophila* PIWI and piRNAs. *RNA* doi: 10.1261/ma.046300.114.
- Qi H, Watanabe T, Ku HY, Liu N, Zhong M, Lin H. 2011. The Yb body, a major site for Piwi-associated RNA biogenesis and a gateway for Piwi expression and transport to the nucleus in somatic cells. *J Biol Chem* **286**: 3789–3797.
- Quinlan AR, Boland MJ, Leibowitz ML, Shumilina S, Pehrson SM, Baldwin KK, Hall IM. 2011. Genome sequencing of mouse induced pluripotent stem cells reveals retroelement stability and infrequent DNA rearrangement during reprogramming. *Cell Stem Cell* **9**: 366–373.
- Robine N, Lau NC, Balla S, Jin Z, Okamura K, Kuramochi-Miyagawa S, Blower MD, Lai EC. 2009. A broadly conserved pathway generates 3' UTR-directed primary piRNAs. *Curr Biol* **19**: 2066–2076.
- Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
- Rozhkov NV, Hammell M, Hannon GJ. 2013. Multiple roles for Piwi in silencing *Drosophila* transposons. *Genes Dev* **27**: 400–412.
- Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC. 2006. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev* **20**: 2214–2222.
- Saito K, Inagaki S, Mituyama T, Kawamura Y, Ono Y, Sakota E, Kotani H, Asai K, Siomi H, Siomi MC. 2009. A regulatory circuit for piwi by the large Maf gene *traffic jam* in *Drosophila*. *Nature* **461**: 1296–1299.
- Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, Siomi H, Siomi MC. 2010. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes Dev* **24**: 2493–2498.
- Sarot E, Payen-Groschene G, Bucheton A, Pelisson A. 2004. Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the *Drosophila melanogaster* flamenco gene. *Genetics* **166**: 1313–1321.
- Sienski G, Donertas D, Brennecke J. 2012. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151**: 964–980.
- Siomi MC, Sato K, Pezic D, Aravin AA. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* **12**: 246–258.
- Spradling AC, Bellen HJ, Hoskins RA. 2011. *Drosophila* P elements preferentially transpose to replication origins. *Proc Natl Acad Sci* **108**: 15948–15953.
- Tcheressiz S, Calco V, Arnaud F, Arthaud L, Dastugue B, Vaury C. 2002. Expression of the *Idefix* retrotransposon in early follicle cells in the germline of *Drosophila melanogaster* is determined by its LTR sequences and a specific genomic context. *Mol Genet Genomics* **267**: 133–141.
- Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**: 320–324.
- Vourekas A, Zheng Q, Alexiou P, Maragkakis M, Kirino Y, Gregory BD, Mourelatos Z. 2012. Mili and Miwi target RNA repertoire reveals piRNA biogenesis and function of Miwi in spermiogenesis. *Nat Struct Mol Biol* **19**: 773–781.
- Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. 2013. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res* **41**: e74.
- Wu Q, Luo Y, Lu R, Lau N, Lai EC, Li WX, Ding SW. 2010. Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proc Natl Acad Sci* **107**: 1606–1611.
- Yin H, Lin H. 2007. An epigenetic activation role of Piwi and a Piwi-associated piRNA in *Drosophila melanogaster*. *Nature* **450**: 304–308.
- Zisoulis DG, Lovci MT, Wilbert ML, Hutt KR, Liang TY, Pasquinelli AE, Yeo GW. 2010. Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. *Nat Struct Mol Biol* **17**: 173–179.

Received May 8, 2014; accepted in revised form September 12, 2014.