UC Riverside
UC Riverside Previously Published Works

Title

Splicing-independent loading of TREX on nascent RNA is required for efficient expression of dual-strand piRNA clusters in Drosophila

Permalink

https://escholarship.org/uc/item/7ss4t5d5

Journal

Genes & Development, 30(7)

ISSN

0890-9369

Authors

Hur, Junho K
Luo, Yicheng
Moon, Sungjin
et al.

Publication Date

2016-04-01

DOI

10.1101/gad.276030.115

Copyright Information

Peer reviewed

Splicing-independent loading of TREX on nascent RNA is required for efficient expression of dual-strand piRNA clusters in *Drosophila*

Junho K. Hur,[1,4] Yicheng Luo,[1,4] Sungjin Moon,[2,3,4] Maria Ninova,[1] Georgi K. Marinov,[1] Yun D. Chung,[2] and Alexei A. Aravin[1]

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California 91125, USA; [2]Department of Life Science, University of Seoul, Seoul 130-743, Korea

The conserved THO/TREX (transcription/export) complex is critical for pre-mRNA processing and mRNA nuclear export. In metazoa, TREX is loaded on nascent RNA transcribed by RNA polymerase II in a splicing-dependent fashion; however, how TREX functions is poorly understood. Here we show that Thoc5 and other TREX components are essential for the biogenesis of piRNA, a distinct class of small noncoding RNAs that control expression of transposable elements (TEs) in the *Drosophila* germline. Mutations in TREX lead to defects in piRNA biogenesis, resulting in derepression of multiple TE families, gametogenesis defects, and sterility. TREX components are enriched on piRNA precursors transcribed from dual-strand piRNA clusters and colocalize in distinct nuclear foci that overlap with sites of piRNA transcription. The localization of TREX in nuclear foci and its loading on piRNA precursor transcripts depend on Cutoff, a protein associated with chromatin of piRNA clusters. Finally, we show that TREX is required for accumulation of nascent piRNA precursors. Our study reveals a novel splicing-independent mechanism for TREX loading on nascent RNA and its importance in piRNA biogenesis.

THO is a multiprotein complex, which itself is part of the TREX (transcription/export) complex. The THO complex was first identified in *Saccharomyces cerevisiae* and is conserved throughout metazoan evolution (Zhou et al. 2000; Katahira 2012). In metazoans, THO contains six proteins; three proteins are homologous to yeast proteins (Hpr1/Thoc1, Thoc2, and Thoc3/Tex1), and three are unique (Thoc5/FMIP, Thoc6, and Thoc7) (Aguilera 2005). In addition to the THO subunits, the complete TREX contains two proteins conserved between yeasts and mammals: Yra1, also known as REF/Aly, and Sub2, also known as UAP56. In yeast, the THO/TREX complex participates in transcription, pre-mRNA processing, and nuclear mRNA export of the majority of genes. Yeast TREX associates with transcribing RNA polymerase II (Pol II) and is necessary for transcription elongation (Chavez et al. 2001; Voynov et al. 2006). TREX is loaded onto nascent transcripts during transcription (Lei et al. 2001; Strasser et al. 2002; Abruzzi et al. 2004) and is essential for mRNA export from the nucleus, as the Yra1 subunit of TREX recruits the Mex67p protein, which interacts with nucleoporins and mediates passage of complexes of mRNAs and proteins (mRNPs) through the nuclear pore (Strasser and Hurt 2000; Stutz et al. 2000).

Despite the conservation of the THO/TREX complex in yeast and metazoa, there are important differences in its function in different organisms. As it is in yeast, metazoan TREX is loaded onto nascent transcripts; however, it does not associate with RNA polymerase. Instead, association of TREX with pre-mRNA depends on splicing of pre-mRNA (Masuda et al. 2005) and requires the interaction of several TREX subunits with the nuclear cap-binding proteins CBC20 and CBC80 (Cheng et al. 2006). Most metazoan genes contain introns, leading to effective splicing-dependent TREX loading onto mRNAs. Furthermore,

TREX is recruited to several intronless genes in *Drosophila* and mammals (Nojima et al. 2007; Kopytova et al. 2010). For some of these genes, TREX recruitment is mediated by sequence-specific RNA-binding proteins that recognize distinct motifs on the mRNAs (Chi et al. 2014); in other cases, the mechanism of splicing-independent loading is not understood.

The conserved composition of TREX suggests that it plays the same function in the nuclear export of mRNPs in metazoa as it does in yeast. Indeed, in *Drosophila*, depletion of Sub2/UAP56 causes accumulation of polyadenylated RNAs within the nucleus (Gatfield et al. 2001). In contrast, however, depletion of other TREX components has a surprisingly mild effect on mRNA export and gene expression in *Drosophila* and mammals (Gatfield and Izaurralde 2002; Rehwinkel et al. 2004; Chi et al. 2013). Depletion of the THO complex in *Drosophila* S2 cells results in changes in expression of only a small fraction of genes (Rehwinkel et al. 2004). Similarly, in HeLa cells deficient in Thoc5, no significant accumulation of poly(A) RNAs is observed in the nucleus (Chi et al. 2013). Transcriptome analyses of mouse embryonic fibroblasts in which *thoc5* expression was inhibited showed down-regulation of only 143 genes, and these were efficiently spliced but retained in the nucleus (Guria et al. 2011). In *Drosophila*, *thoc5* mutants are viable but have spermatogenesis defects (Moon et al. 2011).

Here, we show that Thoc5 and other THO subunits are required for female fertility, oocyte patterning, and transposable element (TE) repression in the *Drosophila* germline. We found that Thoc5 is required for biogenesis of piRNAs, a distinct class of small noncoding RNAs (ncRNAs) that are expressed in germ cells and guide transposon silencing. Mature 23- to 28-nucleotide (nt) piRNAs are processed from long, noncoding transcripts (piRNA precursors) generated from distinct genomic regions dubbed piRNA clusters (Brennecke et al. 2007; Siomi et al. 2011). Most piRNA clusters in the *Drosophila* genome are transcribed from both genomic strands and are therefore called dual-strand clusters; the rarer unistrand clusters are transcribed from one strand (Brennecke et al. 2007). A group of proteins composed of the HP1 homolog Rhino (Rhi), the RNA helicase UAP56, and two proteins of unknown function, Cutoff (Cuff) and Deadlock (Del), were shown to be essential for piRNA biogenesis from dual-strand but not unistrand clusters, indicating that piRNA biogenesis from these two types of clusters is quite different (Chen et al. 2007; Klattenhoff et al. 2009; Pane et al. 2011; Zhang et al. 2012; Czech et al. 2013). Subsequent studies revealed that Rhi, Del, and Cuff form the RDC complex that associates with chromatin of dual-strand but not unistrand clusters (Le Thomas et al. 2014; Mohn et al. 2014; Zhang et al. 2014). The chromodomain of Rhi directly binds the H3K9me3 mark, which is enriched in the chromatin of dual-strand clusters (Le Thomas et al. 2014; Mohn et al. 2014; Yu et al. 2015).

We found that Thoc5 localizes in distinct nuclear foci that overlap with sites of piRNA transcription. In addition, we observed that Thoc5 is highly enriched on noncoding piRNA transcripts compared with protein-coding pre-mRNAs. Binding of Thoc5 to unspliced piRNA precursors and localization to nuclear foci require Cuff protein that is present on the chromatin of piRNA clusters. Our data reveal a novel, splicing-independent mechanism of TREX loading on nascent RNA. We also found that Thoc5 is necessary for accumulation of nascent piRNA precursors, indicating that TREX has a role in transcriptional control.

Results

Thoc5 is required for female fertility and dorsal–ventral embryo patterning

Previous studies of *thoc5* mutant flies detected male sterility and spermatogenesis defects (Moon et al. 2011). To test whether gametogenesis defects also occur during oogenesis, we studied fertility of $thoc5^{e00906/1}$ transheterozygous females (for brevity, we refer to this genotype as $thoc5^{e/1}$). The $thoc5^1$ allele has an in-frame deletion and seems to be functionally null, while $thoc5^e$ is a hypomorphic allele caused by the insertion of a TE upstream of the ORF (Moon et al. 2011). We crossed $thoc5^{e/1}$ females with wild-type (w^{1118}) males and counted the numbers of eggs and hatched larvae (Fig. 1A). $thoc5^{e/1}$ females laid numbers of eggs similar to those of $thoc5^{1/+}$ heterozygous and wild-type flies; however, only 1.5% of eggs laid by $thoc5^{e/1}$ females developed into larvae compared with ~70% for heterozygous and wild-type females. A genomic construct that contains the *thoc5* gene that rescued the male sterility in the previous study (Moon et al. 2011) was also able to rescue the egg development defects in embryos laid by $thoc5^{e/1}$ females (data not shown). These results indicate that Thoc5 is critical for female fertility.

The majority of eggs laid by $thoc5^{e/1}$ females have abnormal appendages indicative of a dorsal–ventral patterning defect. We quantitatively assayed the defect in dorsal–ventral patterning by categorizing the morphology of dorsal appendages (Fig. 1B). Almost all eggs laid by wild-type females showed normal appendages. In contrast, significant fractions of eggs laid by mothers with different mutant *thoc5* alleles showed dorsal–ventral patterning defects. Hypomorphic mutations in two other components of the THO complex, Thoc6 and Thoc7, also resulted in patterning defects. Variation in the appendage phenotype caused by the *thoc5*, *thoc6*, and *thoc7* mutations is likely attributable to the nature of the mutations, which, with exception of the $thoc5^1$ allele, are hypomorphic and not null mutations (Kim et al. 2011; Moon et al. 2011). These results suggest that multiple components of the THO complex are nonredundantly required for dorsal–ventral patterning of the early fly embryo.

Abnormal dorsal–ventral patterning in early embryos was previously linked to activation of the Chk2-dependent DNA damage response during oogenesis caused by unrepaired DNA double-stranded breaks (DSB) in germ cells (Klattenhoff et al. 2007). We examined the formation of DSBs in the ovaries of wild-type and $thoc5^{e/1}$ flies by immunofluorescence staining for phosphorylated histone variant H2Av, which marks sites of DNA DSBs. As
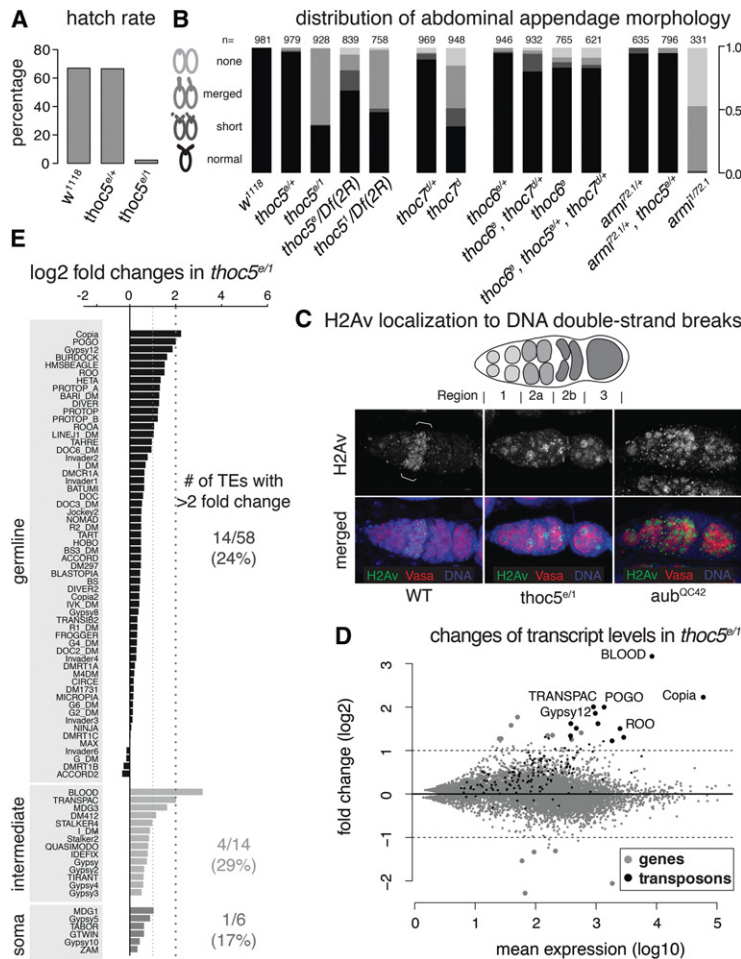
Figure 1. Mutation of *thoc5* leads to oogenesis defects and derepression of multiple transposon families. (*A*) Few eggs laid by Thoc5-deficient females hatch. The hatch rates of the eggs laid by *thoc5^{e/1}*, heterozygous *thoc5^{e/+}*, and wild-type (*w^{1118}*) flies are plotted. (*B*) Dorsal–ventral patterning is impaired in eggs laid by *thoc5^{e/1}* females. Eggs laid by *thoc5*, *thoc7*, *thoc6*, and *armi* mutants and heterozygote controls were categorized by the morphology of their dorsal appendages: normal, shortened, merged, and atrophied. Various allelic combinations of *thoc5* mutants show dorsal–ventral patterning defects. Numbers of eggs analyzed are indicated *above* the bar plot. (*C*) DNA double-stranded breaks (DSBs) accumulate in the ovaries of *thoc5^{e/1}* females. In wild-type (WT) flies, DSBs (visualized by staining with antibody against histone H2Av [green]) occur exclusively in region 2a of the germarium (marked with white brackets). In *thoc5^e/Df(2R)MR43* and *aub^{QC42}* mutant ovaries, DSBs are observed in other regions of the germarium. Germ cells are marked by staining with anti-Vasa antibody (red). DNA (blue) was stained with Hoechst 33258. The *top* panel shows a schematic drawing of a germarium. (*D*) Transposons are derepressed in *thoc5* mutants. Differential expression analyses of gene and transposon levels in *thoc5^{e/1}* and control (*thoc5^{e/+}* heterozygous) ovaries. Each dot indicates a protein-coding gene (gray) or a transposon family (black). The mean expression levels are plotted on the *X*-axis, and the log$_2$ fold differences between the *thoc5^{e/1}* mutant and control are on the *Y*-axis. Genes and transposons that show statistically significant changes (false discovery rate [FDR] < 0.2) are marked with large dots. Data from two independent biological RNA sequencing (RNA-seq) replicates were analyzed using DESeq2. (*E*) The *thoc5* mutation leads to derepression of germline-expressed transposon families. TEs were classified as expressed in the germline, soma, or intermediate (Czech et al. 2013). Shown are the fold changes in expression in the *thoc5^{e/1}* mutant relative to control heterozygous flies from two independent biological RNA-seq replicates calculated using DESeq2.

expected, in wild-type flies, the H2Av signal was present exclusively in region 2a of the germarium, where meiotic recombination leads to formation of DSBs. In contrast, H2Av staining was observed in other regions in the germarium of *thoc5^{e/1}* females, indicating accumulation of unrepaired DSBs (Fig. 1C). A similar phenotype was previously reported in the ovaries of *aubergine* (*aub*) and *armitage* (*armi*) mutants (Klattenhoff et al. 2007). AUB and ARMI are involved in the piRNA pathway that represses expression of TEs in germ cells during oogenesis. Activation of transposons upon impairment of the piRNA pathway correlates with accumulation of DNA DSBs likely caused by transposon activity (Klattenhoff et al. 2007). Taken together, our experiments indicate that, similar to piRNA pathway components, the THO complex is required for proper embryo patterning and suppression of DSB formation during oogenesis.

TEs are derepressed in thoc5 mutants

To determine the cause of the developmental defects and formation of DSBs in the ovaries of *thoc5^{e/1}* mutants, we

analyzed mRNA expression levels in the ovaries of *thoc5^{e/1}* mutant and heterozygous control flies by RNA sequencing (RNA-seq); two independent biological replicas were used for each genotype. In addition to profiling total cellular RNA expression, we prepared RNA-seq libraries of RNA isolated from the chromatin fraction of *thoc5^{e/1}* mutant and control ovaries. Analysis of RNA-seq by DESeq2 revealed that levels of the majority of protein-coding genes did not differ in *thoc5^{e/1}* mutants and control flies in either the total or the chromatin-associated RNA-seq libraries (Fig. 1D; Supplemental Figs. S1A, S2; Supplemental Tables S1, S2). Thirteen and 292 genes (0.15% and 3.6%) have a significant, greater than twofold change in expression between the mutant and control total RNA and chromatin RNA samples, respectively. Genes that showed significant changes were not enriched in particular pathways according to gene ontology analysis. Importantly, mRNA levels of genes required for piRNA-mediated repression were not significantly affected by the *thoc5* deficiency (Supplemental Fig. S1B). In contrast to protein-coding genes, expression of 12 and 51 different families of TEs (8.3% and 34%) was significantly increased in the

ovaries of $thoc5^{e/1}$ mutants compared with control flies in total and chromatin RNA-seq data, respectively (Fig. 1D; Supplemental Figs. S1A, S4; Supplemental Tables S3, S4). These alterations were confirmed by RT-qPCR on RNA isolated from independent biological samples (Supplemental Fig. S1C). The most prominent change was observed for the blood LTR retrotransposon; expression was higher by >100-fold in $thoc5^{e/1}$ mutants compared with control flies. To confirm that the effect was indeed due to the $thoc5$ mutation, we examined the derepression of transposons in $thoc5^{1}/Df(2R)MR43$, where $Df(2R)MR43$ contains a genomic deletion encompassing $thoc5$ (Supplemental Fig. S1D). We found that the up-regulation pattern was consistent with changes seen in the $thoc5^{e/1}$ mutant and similar to the effect observed in the aub mutant.

Previous studies classified TE families according to their expression patterns in the fly ovary: Some transposon families are expressed predominantly in germ or somatic cells, whereas some are expressed in both cell types (Malone et al. 2009; Czech et al. 2013). Separation of TE families according to their expression patterns showed that $thoc5$ deficiency affects primarily transposons expressed in the germline. Fourteen families expressed in germ cells were derepressed more than twofold in $thoc5$ mutants, and an additional four families expressed in both cell types were also up-regulated in the ovaries of $thoc5^{e/1}$ mutant flies relative to levels in control heterozygous ovaries (Fig. 1E; Supplemental Fig. S1E,F). In contrast, only one family classified as somatically expressed was up-regulated more than twofold in the ovaries of $thoc5$-deficient flies. Overall, our results suggest that $thoc5$ deficiency does not have a global effect on the expression of protein-coding genes but instead causes derepression of diverse transposon families primarily in germ cells.

Thoc5 is necessary for piRNA biogenesis but does not affect ping-pong processing

The defects observed in $thoc5$ mutants resemble the phenotypes seen in piRNA pathway mutants. To test whether Thoc5 is involved in piRNA biogenesis and function, we profiled small RNAs in the size range between 19 and 29 nt from ovaries of control and $thoc5^{e/1}$ flies. This size range includes three major classes of small RNAs in $Drosophila$: microRNAs (miRNAs; 21–23 nt), siRNAs (21 nt), and piRNAs (24–28 nt). The size profile and annotation of total small RNAs showed that all three classes of small RNAs are expressed in $thoc5^{e/1}$ ovaries (Supplemental Fig. S2A). The total fraction of miRNAs as well as the levels of individual miRNAs were very similar in ovaries of $thoc5^{e/1}$ mutants and control flies (Fig. 2A,B). Similarly, the levels of endogenous siRNAs derived from hairpin loci (hp-endo-siRNA) were not decreased in ovaries of $thoc5$ mutants (Fig. 2C,D). In contrast, the amount of small RNAs mapping to TEs was reduced independently of whether read numbers were normalized to the total number of reads (RPM [reads per million mapped reads]) or to the level of miRNAs or hp-endo-siRNA (Supplemental Fig. S2B). Interestingly, both siRNAs and piRNAs that map to TE sequences were decreased in $thoc5^{e/1}$ ovaries

relative to levels in control heterozygous flies. These results suggest that expression of piRNAs and repeat-derived siRNAs is reduced but not eliminated in ovaries of $thoc5$ mutants. Indeed, for almost all TE families expressed in germ cells, the amount of antisense piRNAs that can target these transposons was reduced at least twofold, with reductions exceeding fourfold for many transposons (Fig. 2E). No significant reduction was observed in the levels of piRNAs that are antisense to TE families expressed predominantly in the soma. The mapping of piRNAs to consensus sequences of germline-expressed transposons, such as TART, also showed dramatic decreases in piRNAs along the whole body of the element (Fig. 2F). Overall, our data indicate that there is a severe reduction of piRNAs in germ cells of $thoc5$ mutants.

In germ cells, piRNAs are generated through two distinct, although linked, mechanisms: ping-pong processing and primary biogenesis. Ping-pong processing relies on the endonuclease activity of two Piwi proteins, AUB and AGO3, to generate complementary piRNA pairs with 5′ ends that are separated by exactly 10 nt (so called ping-pong pairs) (Brennecke et al. 2007; Gunawardane et al. 2007). To explore whether ping-pong processing is affected in $thoc5$ mutants, we analyzed the fraction of complementary piRNAs that have a 10-nt overlap between their 5′ ends—the signature of ping-pong processing. The prominent signature of ping-pong processing was not significantly reduced in $thoc5^{e/1}$ mutants when piRNA matching TART and other TE families were analyzed (Fig. 2G; Supplemental Fig. S3). To account for reduction in total piRNA levels in $thoc5^{e/1}$ flies, we sampled the same number of piRNA reads from control and mutant libraries for each TE family and determined the fraction of reads in ping-pong pairs (Fig. 2H). This analysis revealed that different TE families have very different numbers of piRNAs generated through ping-pong biogenesis. As expected, TEs expressed exclusively in somatic cells, which lack the ping-pong machinery, showed negligible levels of ping-pong processing. Importantly, $thoc5$ deficiency did not significantly affect the fraction of ping-pong piRNAs for transposons that have prominent levels of ping-pong processing (Fig. 2H; Supplemental Fig. S3). Finally, we analyzed the ping-pong signature using piRNAs, which uniquely mapped to piRNA clusters and overlap in the genome in cis. We found that $thoc5$ mutation does not significantly affect the level of cis ping-pong processing (Supplemental Fig. S4). Taken together, our results indicate that $thoc5$ is not required for piRNA biogenesis through ping-pong. Our data show that Thoc5 is necessary for expression of piRNA in germ cells; however, the ping-pong mechanism of piRNA biogenesis is not inhibited in the absence of Thoc5.

Thoc5 is required for piRNA biogenesis from dual-strand piRNA clusters

In $Drosophila$, many piRNAs are generated from piRNA clusters. Transcripts from these genomic loci serve as substrates for further processing into mature piRNAs (Brennecke et al. 2007). Two types of piRNA clusters,
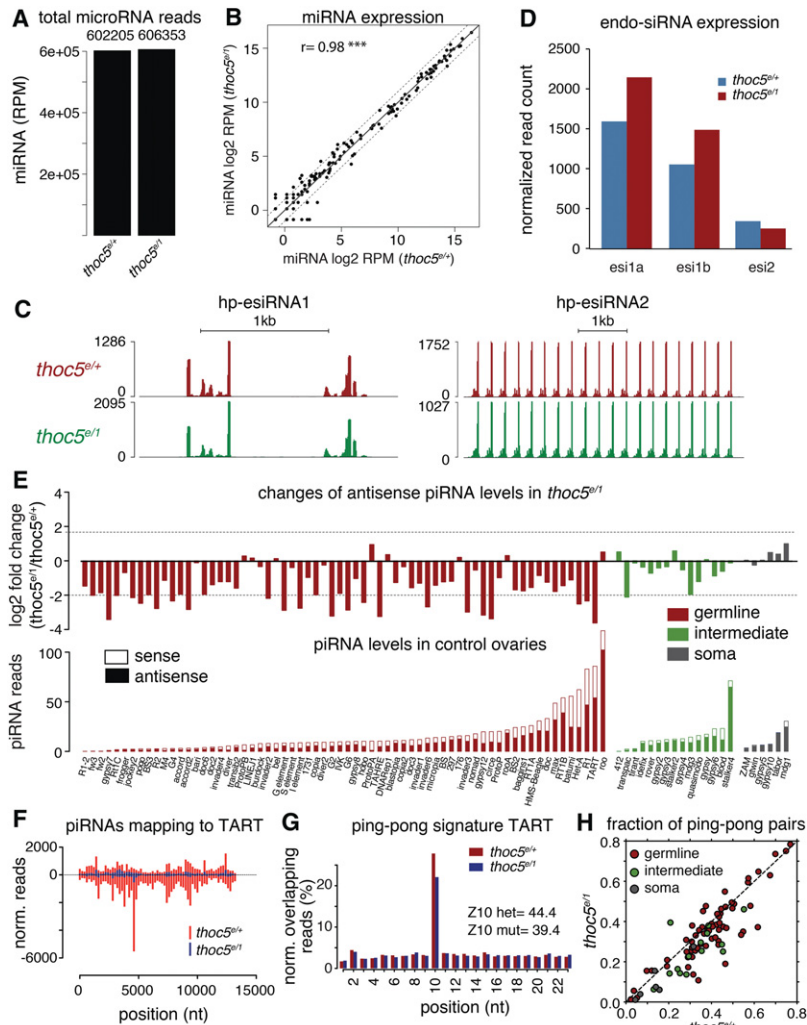
Figure 2. piRNAs targeting germline transposons are depleted in *thoc5* mutants. (*A*) miRNA levels are unaffected by *thoc5* mutation. miRNA expression levels in small RNA libraries from *thoc5^{e/1}* and control (*thoc5^{e/+}* heterozygous) ovaries. Shown are total reads mapping to *Drosophila melanogaster* miRNAs (miRBase version 21) per million mapped reads (RPM). (*B*) Expression of individual miRNAs is unaltered upon *thoc5* mutation. Correlation of miRNA expression between *thoc5^{e/1}* and control ovaries. Each dot represents $\log_2$ transformed RPM values for individual miRNAs. The solid line represents perfect correlation, and the dotted lines represent twofold change. The Pearson's correlation value (r) is shown. (***) $P <$ 0.001. (*C*) Endo-siRNA profiles are unaffected by *thoc5* mutation. The plots show the small RNA read distribution across two loci encoding hp-endo-siRNAs—hp-esiRNA1 (chromosome 2L, 9,788,500–9,790,500) and hp-esiRNA2 (chromosome X, 4,815,000–4,822,000)—in ovaries of *thoc5^{e/1}* and control flies. Signal was normalized to total levels of miRNAs. For the repetitive hp-endo-siRNA2 locus, read numbers were also normalized to the numbers of mapping positions. (*D*) Endo-siRNA expression is unaffected by *thoc5* mutation. RPM normalized read count of hp-endo-siRNAs from three genomic loci in the ovaries of *thoc5^{e/1}* and control flies. (*E*) Expression of piRNAs targeting multiple transposon families is decreased in *thoc5^{e/1}* flies. The *top* graph shows changes in the levels of antisense piRNAs mapping to each transposon family in *thoc5^{e/1}* mutants relative to levels in heterozygous controls. TEs are classified as germline, soma, or intermediate based on expression pattern. The *bottom* graph shows expression levels of sense and antisense piRNAs mapping to each transposon family in ovaries of *thoc5^{e/+}* heterozygous control. (*F*) The level of piRNAs targeting the TART retrotranspo-

son is decreased in the *thoc5^{e/1}* mutant along the whole TE sequence. Shown are the normalized piRNA read numbers mapping along the TART consensus sequence in 200-nt bins from *thoc5^{e/1}* (blue) and heterozygous control (red) libraries. (*G*) Ping-pong is not affected by *thoc5* deficiency. Analysis of ping-pong piRNA biogenesis in *thoc5^{e/1}* and control flies. The *Y*-axis shows the relative proportion of reads having 5′–5′ sense–antisense overlap at a given position compared with all reads mapping to the TART retrotransposon consensus sequence. The graphs for other transposons are depicted in Supplemental Figure S3. The peak at 10 nt corresponds to the signature of ping-pong piRNA biogenesis. *Z*-scores for position 10 are shown. (*H*) The fraction of ping-pong piRNA pairs (complementary piRNA pairs in which the 5′ ends are separated by 10 nt) for each TE family are similar in *thoc5^{e/1}* and control flies. For each TE family, the same number of piRNAs was sampled 1000 times from *thoc5^{e/1}* and heterozygous control libraries, and the mean number is shown.

unistrand and dual-strand, are expressed in the *Drosophila* ovary. Mapping of piRNAs from *thoc5^{e/1}* mutants to major clusters showed that piRNAs from dual-strand clusters (*42AB*, *38C*, *80EF*, and *X-TAS*) were strongly depleted, whereas expression of unistrand clusters (*20A* and *flamenco* [*flam*]) was not affected or was increased (Fig. 3A,B; Supplemental Fig. S5). The specific effect of *thoc5* deficiency on dual-strand and not unistrand clusters was similar to the effects of deletion of genes encoding the HP1 paralog Rhi, the RNA helicase UAP56, and a protein of unknown function called Cuff (Chen et al. 2007; Klattenhoff et al. 2009; Pane et al. 2011; Zhang et al. 2012, 2014; Le Thomas et al. 2014; Mohn et al. 2014). Analysis of small RNA libraries from *thoc5* and *cuff* mutants in parallel with previously reported *uap56* (Zhang et al. 2012)

and *rhi* (Mohn et al. 2014) libraries showed some variability of piRNA expression in control flies, which is probably attributable to differences in the starting material and protocols used. Nevertheless, it is clear that *thoc5* mutation along with *rhi*, *cuff*, and *uap56* mutations leads to a decrease in the amount of piRNAs from dual-strand clusters, while the levels of piRNAs from unistrand clusters are not affected or increase in abundance (Fig. 3A,B). Furthermore, similar to *rhi* and *cuff* mutants, the levels of piRNA generated from mRNAs of hundreds of protein-coding genes are also increased in the *thoc5* mutant (Fig. 3C).

Transcripts were generated from both genomic strands in dual-strand piRNA clusters, resulting in the formation of dsRNA that was processed into 21-nt endo-siRNA. Interestingly, the levels of endo-siRNA from dual-strand
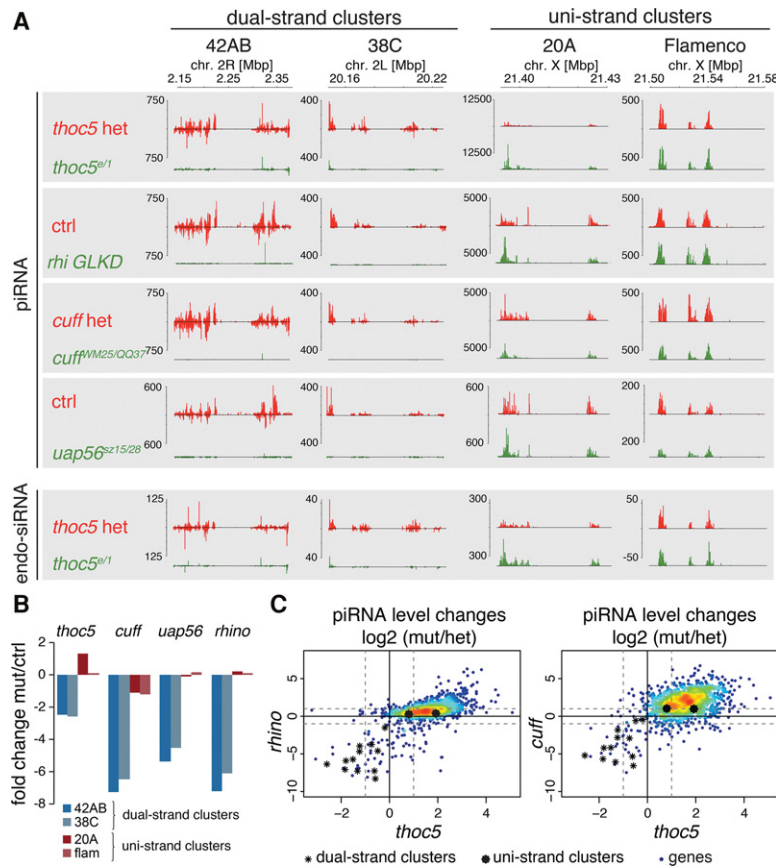
Figure 3. Thoc5 is required for biogenesis of piRNAs and endo-siRNAs from dual-strand clusters. (*A*) The levels of piRNAs and endo-siRNAs generated from dual-strand piRNA clusters are decreased in the ovaries of $thoc5^{e/1}$ mutants compared with heterozygous controls. The plots show the distribution of uniquely mapping piRNAs and endo-siRNAs over dual-strand (*42AB* and *38C*) and unistrand (*20A* and *flam*) clusters in the ovaries of $thoc5^{e/1}$, $cuff^{WM25/QQ37}$ (Le Thomas et al. 2014), *rhi* germline knockdown (GLKD) (Mohn et al. 2014; Zhang et al. 2014), $uap56s^{z15/28}$ (Zhang et al. 2012), and corresponding control (het) flies. The piRNA and endo-siRNA levels were normalized to miRNA reads to adjust for sequencing depth. (*B*) Thoc5, Cuff, UAP56, and Rhi are required for generation of piRNAs exclusively from dual-strand clusters. Fold changes (log₂) in piRNA expression from two dual-stranded (*42AB* and *38C*) and two unistranded (*20A* and *flam*) clusters in the ovaries of different mutant flies with respect to the corresponding controls. piRNA reads (23–28 nt) uniquely mapping to clusters were normalized to miRNA reads to adjust for sequencing depth. (*C*) The effect of the loss of Thoc5 on piRNA generation from major piRNA clusters and genic loci are similar to those of loss of Rhi and Cuff. Shown are scatter plots of changes (log₂ of mutant/control) in the normalized levels of piRNAs that uniquely map to each piRNA cluster or a gene in the $thoc5^{e/1}$, $cuff^{WM25/QQ37}$, *rhi* germline knockdown, and corresponding control flies. The levels of piRNAs produced from all dual-strand clusters are decreased in all three mutants, whereas piRNA levels from unistrand clusters and genes are increased in mutants relative to heterozygous controls.

clusters were also decreased in $thoc5^{e/1}$ flies (Fig. 3A). In contrast, endo-siRNA processed from hairpin loci were not reduced in $thoc5^{e/1}$ flies, indicating that Thoc5 is not generally required for endo-siRNA biogenesis (Fig. 2C,D). Endo-siRNAs are processed by the nuclease Dicer in a process that is distinct from piRNA biogenesis (Vagin et al. 2006; Czech et al. 2008). Therefore, Thoc5 is necessary for the generation of precursors for endo-siRNA and piRNA generated from dual-strand piRNA clusters, and its function is not restricted to piRNA biogenesis. The effect of Thoc5 on dual-stand piRNA clusters, but not unistrand clusters, suggests that it might cooperate with Rhi, Cuff, and UAP56 to generate piRNA and endo-siRNA from these loci.

TREX localizes in distinct foci at nurse cell nuclei

To get insight into the function of Thoc5 in piRNA biogenesis, we used immunofluorescence to determine the subcellular localization of Thoc5 in fly ovaries. GFP-tagged Thoc5 was expressed in transgenic flies under the control of the native genomic regulatory elements of the *thoc5* gene to ensure physiological expression level (Moon et al. 2011). Importantly, expression of GFP-Thoc5 suppressed the overexpression of transposons observed in $thoc5^{e/1}$ mutants, indicating that the tagged protein is functionally intact and can participate in piRNA-

mediated TE repression (Fig. 4A). Tagged Thoc5 localized to several prominent foci in nurse cell nuclei. Two other THO subunits, Thoc2 and Thoc7, which we detected using immunofluorescence with antibodies against the native proteins, also localized to nuclear foci in nurse cells (Fig. 4G). Similar localization patterns were previously described for Rhi, Cuff, and UAP56, the proteins involved in piRNA biogenesis from dual-strand clusters (Mohn et al. 2014; Zhang et al. 2014). Indeed, double labeling revealed that GFP-Thoc5 colocalizes with tagged UAP56 and Rhi in nuclear foci (Pearson's correlation coefficients of 0.69 and 0.68) (Fig. 4B–D). In addition, tagged UAP56 and Rhi colocalized with native Thoc7 in nuclear foci (Supplemental Fig. S6). These data show that several subunits of the THO complex colocalize in distinct nuclear foci together with UAP56 and Rhi. Coimmunoprecipitation from S2 cells showed that Thoc5 binds to UAP56 in an RNA-independent fashion, indicating that the two proteins are present in the same TREX complex (Fig. 4E).

Next, we tested whether the interaction between Thoc5 and UAP56 is required for their localization to nuclear foci. In $thoc5^{e/1}$ cells, UAP56 was delocalized from nuclear foci and instead dispersed in the nucleoplasm, although the amount of the protein was not reduced (Fig. 4F). Contrary to UAP56, the amounts of other THO subunits, specifically Thoc2 and Thoc7, were severely reduced in $thoc5^{e/1}$ cells (Fig. 4G). Tagged GFP-Thoc5 as well as the
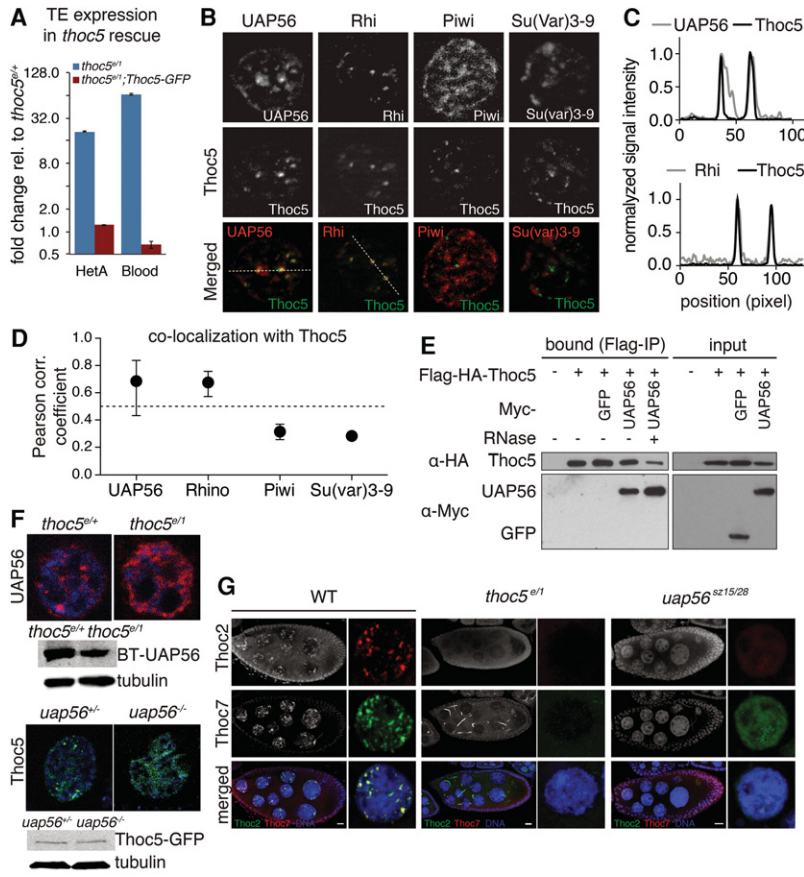
Figure 4. Thoc5 colocalizes with Rhi and UAP56 in nuclear foci and interacts with UAP56. (*A*) The GFP-Thoc5 genomic construct suppresses overexpression of transposons in the $thoc5^{e/1}$ mutant. Expression of HetA and blood transposons was measured by RT-qPCR in the ovaries of the $thoc5^{e/1}$ mutant, the $thoc5^{e/1}$ mutant that expresses the GFP-Thoc5 construct, and control (heterozygous) flies. The *Y*-axis shows change in expression ($\log_2$ scale) relative to the heterozygous control. Error bars indicate the standard deviation of two biological replicates. (*B*) Thoc5 colocalizes with UAP56 and Rhi in distinct nuclear foci. Thoc5-GFP, UAP56, Rhi, Piwi, and Su(Var)3-9 were visualized by immunofluorescence microscopy in the nuclei of nurse cells of *Drosophila* ovaries. (*C*) Signal intensities for Thoc5-GFP, UAP56, and Rhi across the nucleus along the lines in *B* are plotted. (*D*) Quantification of colocalization of Thoc5, UAP56, and Rhi. Shown are the average Pearson's correlation coefficients for colocalization of signal intensities between Thoc5 and the indicated proteins. Three nuclei were analyzed for each protein pair. Error bars indicate 95% confidence intervals. Student's *t*-test showed that there are significant differences (*P*-values < 0.025) between colocalization of Thoc5 with UAP56 and Rhi compared with control nuclear proteins Piwi and Su(Var)3-9. (*E*) Thoc5 interacts with UAP56 in an RNA-independent fashion. Flag-HA-tagged Thoc5 was coexpressed in S2 cells with Myc-tagged GFP or Myc-tagged UAP56. Flag-HA-Thoc5 was immunoprecipitated using anti-Flag beads with or without RNase A treatment, and copurified proteins were analyzed by Western blot. (*F*) The localization of Thoc5 and UAP56 in distinct nuclear foci is interdependent. In $thoc5^{e/1}$ mutants, UAP56 appears dispersed in the nuclei. Reciprocally, Thoc5 is delocalized from nuclear foci in $uap56^{28/SZ15}$ flies. The Western blots *below* show that the amount of UAP56 and Thoc5 protein is not decreased in the reciprocal mutants. Tubulin was used as a loading control. (BT) Biotin–protein A tag. (*G*) Localization of THO subunits into nuclear foci depends on Thoc5 and UAP56 expression. Shown is the subcellular localization of two THO subunits, Thoc2 and Thoc7, in wild-type (WT) ovaries and the ovaries of $thoc5^{e/1}$ and $uap56^{28/SZ15}$ flies. Both subunits are colocalized in discrete nuclear foci in the nuclei of nurse cells in wild-type ovaries. Thoc2 and Thoc7 are destabilized and cannot be detected in *thoc5* mutants. On the contrary, Thoc2 and Thoc7 are expressed in the egg chambers of *uap56* mutants but are delocalized from the discrete nuclear foci. DNA (blue) was visualized by Hoechst 33258 staining. Bar, 10 μm.

endogenous Thoc2 and Thoc7 proteins were dispersed in the nucleoplasm in the *uap56* mutant instead of localizing to nuclear foci (Fig. 4F,G). Importantly, the *uap56* mutation did not destabilize THO subunits. These experiments revealed important interactions between components of the TREX complex. Within the THO subcomplex, Thoc5 is required for proper expression of other THO subunits (Thoc2 and Thoc7), which are greatly reduced in the ovaries of *thoc5* mutants. The same result was previously observed in the testis (Moon et al. 2011) and the whole body (Kim et al. 2011). In contrast, depletion of Thoc5 has no effect on the stability of UAP56, and, reciprocally, UAP56 depletion does not affect Thoc5 expression. However, both proteins are required for proper localization of the other factor to nuclear foci. Taken together, our data indicate that different subunits of the TREX complex, components of the THO subcomplex and UAP56, colocalize in distinct nuclear foci and require each other for this localization pattern.

Cuff recruits the TREX complex to piRNA precursors

Rhi binds the H3K9me3 histone mark and is enriched on chromatin of dual-strand piRNA clusters (Klattenhoff et al. 2009; Le Thomas et al. 2014; Mohn et al. 2014; Zhang et al. 2014). Colocalization of Thoc5 with Rhi suggests that TREX accumulates at the sites where nascent pre-piRNAs are transcribed. To determine whether Thoc5 binds piRNA precursors, we immunoprecipitated two different subunits of the THO complex—Thoc5 and Thoc7—from ovary extracts and analyzed the associated RNAs by RT-qPCR. RNA copurified with Thoc5 showed ∼50-fold enrichment in transcripts from dual-strand piRNA clusters (Fig. 5A). Similarly, ncRNAs from dual-strand piRNA clusters were enriched in Thoc7 immunoprecipitates. In contrast, transcripts from unistrand clusters and pre-mRNAs were not enriched in Thoc5 and Thoc7 immunoprecipitates. Next, we profiled RNAs associated with the THO complex in fly ovaries using GFP-
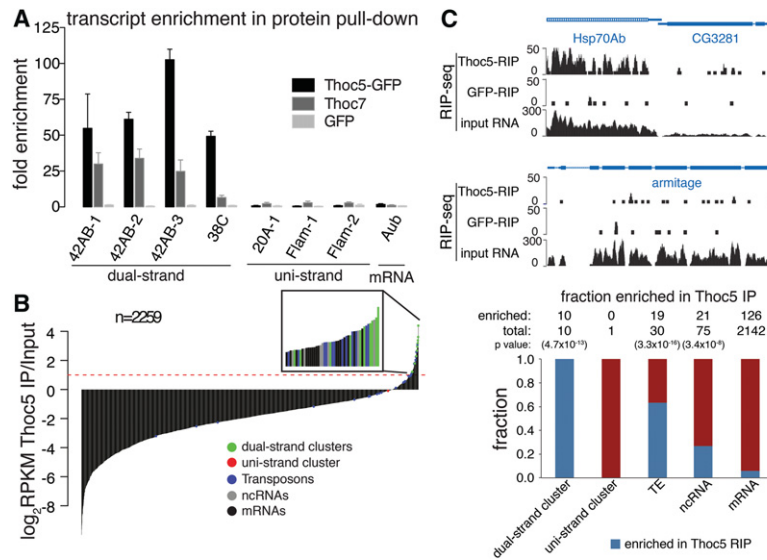
Figure 5. TREX binds piRNA transcripts from dual-strand clusters. (*A*) TREX subunits Thoc5 and Thoc7 are enriched on long RNAs from dual-strand piRNA clusters but not on transcripts from unistrand clusters or *aub* pre-mRNA. Thoc7 was immunoprecipitated using an antibody against the native protein, whereas GFP-Thoc5 and control GFP were immunoprecipitated using an anti-GFP antibody from the ovaries of transgenic flies expressing the corresponding proteins in the germline. Copurified RNA was analyzed by RT-qPCR to measure the amount of RNA in immunoprecipitates relative to input. Levels of RNAs were normalized to levels of *rp49* mRNA. Error bars indicate the standard error of the mean of two independent experiments. (*B*) Thoc5 preferentially associates with dual-strand piRNA cluster transcripts, transposons, and some ncRNAs. The bar plot shows ordered transcript enrichment in GFP-Thoc5 RIP-seq data. Thoc5 RIP-seq data were normalized by subtracting the signal from negative control (GFP IP) followed by plotting the enrichment in Thoc5 RIP, measured as the log$_2$ ratio of RPKM (reads per kilobase per million mapped reads) in RIP relative to input. Dots indicate transcripts from dual-strand piRNA clusters (green), unistrand piRNA clusters (red), transposons (blue), ncRNAs (gray), and mRNAs (black). The *inset* shows transcripts that are enriched in Thoc5 RIP more than twofold. The graph at the *right* shows the numbers of mRNAs, ncRNAs, transposons, and piRNA cluster transcripts that are enriched in Thoc5 RIP (log$_2$ RIP/input >0). Dual-stranded clusters, transposons, and ncRNAs are significantly overrepresented among the Thoc5-enriched transcript group, as estimated by a hypergeometric distribution test. (*C*) Thoc5 binds to the *Hsp70* transcript, as measured by RIP-seq. No enrichment was seen for other transcripts (*armi* mRNA is shown as an example). GFP-Thoc5 and control GFP were immunoprecipitated from the ovaries using GFP antibodies, and associated RNA was cloned and sequenced.

Thoc5 RIP-seq (RNA immunoprecipitation [RIP] combined with deep sequencing) with GFP RIP-seq as the negative control. Transcriptome-wide analysis revealed that transcripts from all 10 dual-strand piRNA clusters that passed an expression-level threshold were enriched in the Thoc5 immunoprecipitate between twofold and 16-fold (Fig. 5B). In contrast, the transcripts from the unistrand cluster *flam* were not enriched. Many transposon transcripts and ncRNAs were also present in the Thoc5 RIP. Indeed, the hypergeometric distribution test showed significant enrichment of transposons and ncRNA in the Thoc5 RIP (Fig. 5B). In contrast, <5% of protein-coding genes display greater than twofold enrichment in the Thoc5 RIP. We found no overlap between genes that are enriched in the Thoc5 RIP and genes that change their expression in the *thoc5* mutant (data not shown). One example of mRNA that copurified with Thoc5 is the *Hsp70* transcript (Fig. 5C) that previously was shown to generate piRNA in a Rhi-dependent manner (Mohn et al. 2014). Overall, our data indicate that the TREX complex preferentially binds a distinct set of nuclear RNAs that includes ncRNA transcripts from dual-strand piRNA clusters, transposon transcripts, and a few mRNAs.

Rhi and Cuff form a complex that is present on chromatin of dual-strand piRNA clusters (Le Thomas et al. 2014; Mohn et al. 2014; Zhang et al. 2014). To determine whether Thoc5 interacts with Rhi and Cuff, we coimmunoprecipitated tagged proteins from S2 cell extracts. These experiments showed that Thoc5 forms a complex with Cuff but not with Rhi (Fig. 6A). Importantly, while RNase treatment of the extract reduced the fraction of Cuff that

coimmunoprecipitated with Thoc5, a significant amount of Cuff still coprecipitated with Thoc5 under these conditions, indicating that at least some fraction forms a complex in an RNA-independent fashion. It is important to note that the interaction between Cuff and Thoc5 detected using coimmunoprecipitation can be indirect and mediated by other proteins; for example, other subunits of the TREX complex, which are expressed in S2 cells. To further dissect the Thoc5–Cuff interaction, we performed coimmunoprecipitations using truncated proteins. Splitting of Cuff into two parts prevented interaction of either part with Thoc5 (Supplemental Fig. S7). In contrast, both the N-terminal and C-terminal portions of Thoc5 formed a complex with full-length Cuff (Supplemental Fig. S7). Finally, we tested whether Thoc5 interacts with a close paralog of Cuff, CG9125 (dRai1). We found that CG9125 did not form the complex with Thoc5, indicating that the interaction is specific for Cuff (Fig. 6B).

To test whether the interaction between Cuff and the TREX complex plays a role in TREX recruitment to nuclear foci, we analyzed localization of Cuff, Thoc5, and UAP56 in the reciprocal mutants. Lack of Cuff completely delocalized the two TREX subunits Thoc5 and UAP56 from nuclear foci (Fig. 6C). Similarly, THO subunits were delocalized from nuclear foci upon knockdown of *rhi* (Fig. 6D). In contrast, the lack of Thoc5 did not impact Cuff localization, as Cuff remained localized in nuclear foci in the *thoc5*$^{e/1}$ mutant cells (Fig. 6E). These data suggest that Cuff is required for recruitment of the TREX complex to nascent transcripts of dual-strand piRNA loci. To further test this hypothesis, we recruited Cuff to

an artificial mRNA reporter expressed in the *Drosophila* germline using the λ-RNA-binding domain and then measured the association between TREX and the reporter mRNA (Fig. 6F). It is important to note that the reporter mRNA lacks introns and therefore should not associate with TREX in a splicing-dependent manner, which is the established mechanism of TREX loading on nascent transcripts in metazoa (Masuda et al. 2005). Tethering of Cuff to the reporter lead to more than twofold enrichment (*P* = 0.019) of TREX on the reporter mRNA, as measured by RIP of the native Thoc7 protein (Fig. 6G). This result indicates that the presence of Cuff in the vicinity of an intronless nascent transcript enhances association of the RNA with TREX. The most plausible explanation that takes into account the physical interaction between Cuff and TREX (Fig. 6A,B) suggests that Cuff recruits TREX to nascent transcripts through formation of a Cuff/TREX complex. Together, our data can be combined into a model that indicates that Cuff interacts with TREX to promote TREX loading onto nascent transcripts.

Thoc5 is required for efficient transcription of dual-strand piRNA clusters

To understand the function of TREX association with piRNA precursors, we tested whether *thoc5* deficiency changes the levels of long ncRNAs generated from clusters. Both RNA-seq and RT-qPCR analyses showed that transcript levels from dual-strand, but not unistrand, piRNA clusters decreased in the ovaries of $thoc5^{e/1}$ mu-

tants relative to levels in the ovaries from controls (Fig. 7A). We also observed a correlation between changes in long piRNA precursor levels and changes in mature piRNA levels from individual piRNA clusters in $thoc5^{e/1}$ mutants (Fig. 7B).

To understand at which step of piRNA precursor expression TREX plays a role, we analyzed the abundance of nascent chromatin-bound transcripts in $thoc5^{e/1}$ and control ovaries. The analysis of chromatin-associated RNA by RT-qPCR and RNA-seq showed depletion of rRNA and enrichment of reads derived from introns of protein-coding genes, indicating that nascent transcripts were purified (Supplemental Fig. S8A,B). Chromatin-associated RNA from dual-strand clusters was depleted in $thoc5^{e/1}$ mutants, indicating that Thoc5 is required for accumulation of nascent piRNA precursors (Fig. 7C,D). Similarly, the levels of readthrough transcripts downstream from the *Hsp70Aa* gene that give rise to piRNAs were decreased in chromatin-associated RNA from $thoc5^{e/1}$ and *cuff* mutants (Fig. 7E). These data indicate that Thoc5 works to ensure proper nascent transcript levels from dual-strand clusters.

As an alternative approach to detect nascent piRNA transcripts, we performed in situ RNA hybridization on ovarian cell nuclei using probes for dual-strand (*42AB*)
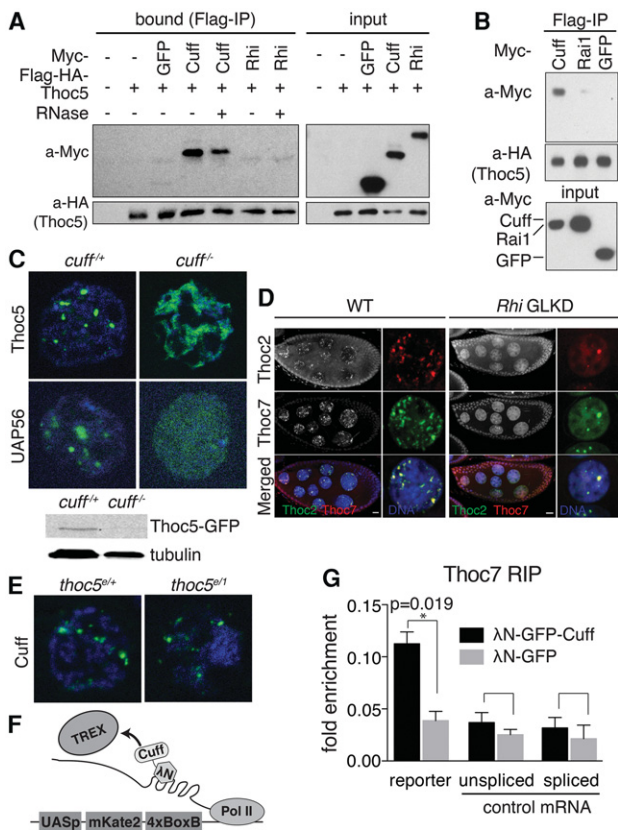

Figure 6. Cuff recruits TREX to transcripts from dual-strand piRNA clusters. (*A*) Cuff, but not Rhi, copurifies with Thoc5, and this interaction is partially resistant to RNase A treatment. Flag-HA-tagged Thoc5 was coexpressed with Myc-tagged GFP, Myc-tagged Cuff, or Myc-tagged Rhi in S2 cells. Flag-HA-Thoc5 was immunoprecipitated with anti-Flag beads in the presence or absence of RNase A, and copurified proteins were analyzed by Western blot (anti-Myc and anti-HA). (*B*) Thoc5 interacts with Cuff but not with its paralog, Rai1 (CG9125). Flag-HA-Thoc5 was coexpressed with Myc-tagged Cuff or Myc-tagged Rai1 in S2 cells. Thoc5 was immunoprecipitated, and copurified proteins were assayed by Western blot. (*C*) Thoc5 and UAP56 (green) are delocalized from nuclear foci in *cuff* mutant ovaries. The Western blots *below* show reduction of Thoc5 expression in *cuff* mutant ovaries; however, this is likely caused by the severe morphological defects (reduction of germ cells) observed in cuff mutants, as Thoc5 was detected in individual germ cells examined by immunofluorescence. DNA (blue) was stained with DAPI. (*D*) Subcellular localization of two THO subunits, Thoc2 and Thoc7, in *rhi* knockdown (*rhi* GLKD) ovaries. Thoc2 and Thoc7 are expressed in egg chambers of *rhi* knockdown but are delocalized from discrete nuclear foci observed in wild-type (WT) nuclei. DNA (blue) was visualized by Hoechst 33258 staining. Bar, 10 μm. (*E*) Cuff (green) localization in nuclear foci is not perturbed in $thoc5^{e/1}$ mutants. DNA (blue) was stained with DAPI. (*F*) Schematic diagram of the reporter used to study the effect of Cuff recruitment to mRNA. The reporter has four BoxB sequences in the 3′ untranslated region (UTR) that recruit the Cuff protein fused to the λN peptide. Both the mKate2-4×BoxB reporter and λN-GFP-Cuff were coexpressed in the ovaries of transgenic flies using a germline driver. (*G*) Cuff recruitment to reporter mRNA promotes reporter association with TREX. λN-GFP-Cuff or control (λN-GFP) protein was recruited to a reporter mRNA, and the association of TREX was analyzed by Thoc7 immunoprecipitation followed by RT-qPCR. Error bars indicate the standard error of the mean of two biological replicas.
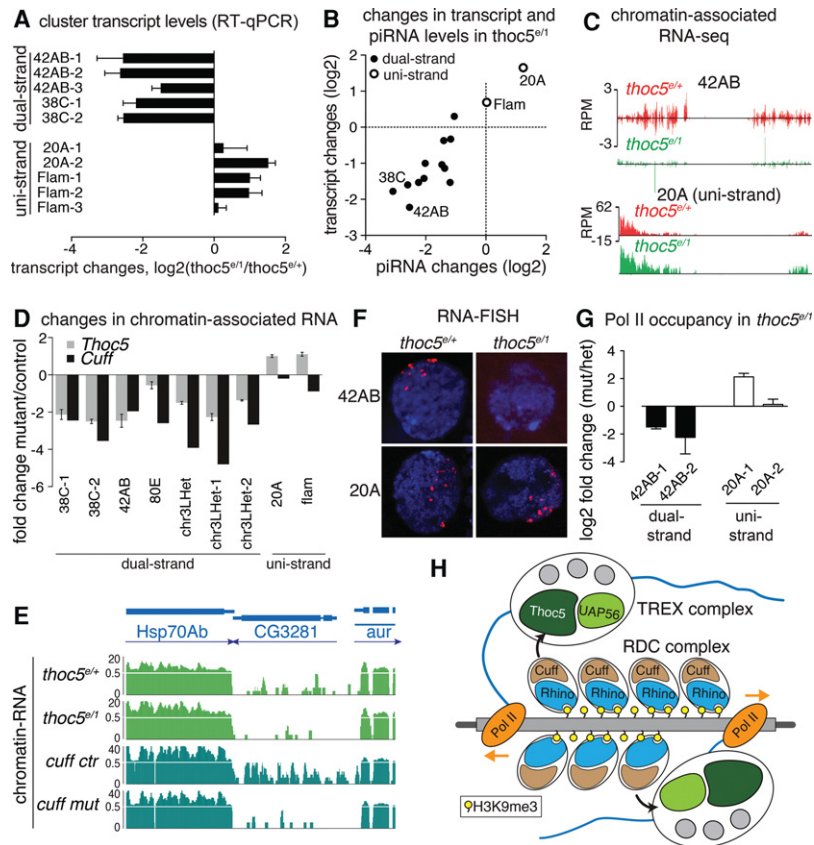
Figure 7. TREX is required for transcription of piRNA precursors from dual-strand clusters. (*A*) The levels of piRNA precursors from dual-strand, but not unistrand, clusters are lower in $thoc5^{e/1}$ mutants than in controls. Levels of long RNAs in the ovaries of $thoc5^{e/1}$ and control flies were measured by RT-qPCR (average of two biological replicates). Plotted are the transcript levels in mutant ovaries normalized to transcript levels in the heterozygous control. Error bars indicate the standard error of the mean of three biological replicates. (*B*) Changes in precursor levels in $thoc5^{e/1}$ mutants correlate with changes in mature piRNA levels. A scatter plot of changes of piRNA precursor levels (determined by RNA-seq) and mature piRN levelsA (determined by small RNA-seq) in $thoc5$ mutants is shown. Each point indicates fold changes in the levels of a unistrand or dual-strand piRNA cluster relative to the levels in the control heterozygous ovaries. (*C*) $thoc5$ deficiency diminishes the levels of nascent precursor piRNA transcripts from dual-strand clusters but not from unistrand clusters. The levels of nascent RNA in dual-strand (*42AB*) and unistrand (*20A*) clusters assayed by RNA-seq of chromatin-associated ovarian RNA are plotted in RPM for transcripts from $thoc5^{e/1}$ and control ovaries. (*D*) Thoc5 is necessary for accumulation of nascent piRNA precursors from dual-strand, but not unistrand, clusters. Shown are log_2 fold changes of chromatin-associated RNA levels from dual-strand and unistrand piRNA clusters in $thoc5$ and $cuff$ mutant ovaries relative to the respective heterozygous controls, as measured by RNA-seq. For each cluster, the number of reads was normalized (RPM), and clusters with an expression level >10 RPM were considered for analyses. Error bars show the standard deviation of two biological replicates. (*E*) Thoc5 and Cuff are required for accumulation of nascent transcripts downstream from the *Hsp70* gene. Shown are the reads per million (RPM) tracks of strand-specific chromatin-associated RNA-seq data. To visualize the signal over the repetitive *Hsp70* gene, reads mapping to multiple positions in the genome were allowed. Note that although the transcripts are mapped to the *CG3281* gene, they are in the opposite orientation to this gene and were generated by readthrough of the upstream *hsp70* gene. (*F*) Mutation of *thoc5* disrupts transcription of dual-strand (*42AB*), but not unistrand (*20A*), clusters. RNA fluorescent in situ hybridization (FISH) was performed in nurse cell nuclei. In control heterozygous flies, both *42AB* and *20A* signals appear as multiple nuclear foci (red) in the polyploid nurse cells (stained with DAPI [blue]). In the $thoc5^{e/1}$ mutant ovary, the signal from *42AB* is reduced below background, whereas the *20A* signal is not affected. (*G*) Thoc5 is required for efficient transcription of dual-strand clusters. The amount of Pol II on chromatin of dual-strand (*42AB*) and unistrand (*20A*) clusters in ovaries of $thoc5^{e/1}$ mutants and control flies was assayed by chromatin immunoprecipitation (ChIP)-qPCR. Pol II levels were decreased approximately fourfold at the dual-strand (*42AB*) cluster locus but not at the unistrand (*20A*) locus. Error bars indicate the standard error of the mean of two biological replicates. (*H*) A model of TREX recruitment and loading on piRNA cluster transcripts. The loading of TREX, composed of THO subunits and UAP56, is mediated by Cuff. Cuff is recruited to dual-strand clusters through its interaction with the chromodomain-containing Rhi protein, which binds the H3K9me3 mark. TREX is required for efficient transcription of piRNA precursors.

and single-strand (*20A*) piRNA clusters (Fig. 7F; Supplemental Fig. S8C). For each locus, we detected several spots per nuclei, which correspond to sites of nascent transcript accumulation in these polytene cells, similar to previously described results (Mohn et al. 2014). The signal from the dual-strand cluster *42AB* completely disappeared in the nuclei of $thoc5^{e/1}$ mutants; there was no difference in the signal from unistrand cluster *20A* in mutant and control cells, indicating a specific effect of Thoc5 on accumulation of nascent transcripts from dual-strand piRNA clusters. Together, the analysis of chromatin-associated RNA and in situ hybridization show that Thoc5 is necessary to ensure proper nascent transcript levels from dual-strand clusters either through direct effect on transcription or by stabilizing nascent transcripts. To determine whether Thoc5 has a direct effect on transcription, we determined the RNA Pol II signal on dual-strand clusters by chromatin immunoprecipitation (ChIP)-qPCR. In $thoc5^{e/1}$ nuclei, the Pol II signal was twofold to fourfold lower than in control cells (Fig. 7G), suggesting that Thoc5 leads to accumulation of piRNA precursors by enabling efficient transcription of dual-strand cluster loci.

Discussion

THO is an evolutionarily conserved multisubunit protein complex that associates with nascent RNA in a sequence-

independent fashion. Despite its conservation and general role in gene expression, the functions of THO and the mechanism of its loading on RNA remain poorly understood. In this study, we found that Thoc5 and other components of the THO complex are required for *Drosophila* ovarian germ cell development and fertility (Fig. 1A,B). Further experiments revealed that THO is necessary for repression of multiple TE families (Fig. 1D,E). In agreement with our results, several components of the THO complex were identified in a genome-wide RNAi screen for factors involved in TE repression in germ cells (Czech et al. 2013).

We found that TE derepression in Thoc5-deficient ovaries is caused by a decrease in the abundance of piRNAs, ncRNAs that guide transcriptional and post-transcriptional repression of TEs in germ cells (Figs. 2, 3). Multiple lines of evidence suggest that THO is directly involved in piRNA biogenesis. First, several THO subunits (Thoc5, Thoc2, and Thoc7) localize at distinct nuclear foci that coincide with Rhi (Fig. 4B,C,G), a factor that is enriched on piRNA clusters and required for piRNA biogenesis (Mohn et al. 2014). Second, we found that two THO components (Thoc5 and Thoc7) are enriched on long piRNA precursor transcripts and are necessary for their expression (Figs. 5, 7A–D).

In yeast and mammals, THO, the RNA helicase UAP56 (called Sub2 in yeast), and Aly/Ref together form the TREX complex, which is loaded on nascent RNA. UAP56 is required for piRNA biogenesis and binds to piRNA precursors (Zhang et al. 2012). We found that Thoc5 forms a complex with UAP56 in an RNA-independent fashion (Fig. 4E). Furthermore, Thoc5 and UAP56 colocalize in nuclear foci, and their localization, but not protein stability, is interdependent (Fig. 4B–F). These data suggest that a preformed TREX complex is loaded on piRNA precursors. Although TREX is considered to be a general factor, which is loaded on nascent RNA transcribed by RNA Pol II (Strasser et al. 2002), our RIP-seq data showed that association of TREX with different nascent transcripts is quantitatively different (Fig. 5). Indeed, while <5% of mRNAs display greater than twofold enrichment in Thoc5 RIP, transcripts from all dual-strand piRNA clusters are enriched. Enrichment of TREX on piRNA precursors raises the question about the mechanism for its specific loading to these transcripts. This question is particularly intriguing, since association of TREX with nascent RNAs in metazoa was previously shown to be dependent on splicing (Masuda et al. 2005), but piRNA precursors are not spliced (Zhang et al. 2014). Furthermore, splicing of piRNA precursors induced by depletion of Rhi was shown to have a negative effect on piRNA biogenesis (Zhang et al. 2014). The clue to this puzzle was provided by the fact that both Thoc5 and UAP56 are required for piRNA biogenesis from only a subset of all genomic regions that generate piRNA; namely, dual-strand piRNA clusters. The same exclusive effect on dual-strand clusters was previously shown for Rhi and Cuff, two proteins that form a complex on chromatin of dual-strand clusters due to direct binding of the H3K9me3 histone mark by the chromodomain of Rhi

(Chen et al. 2007; Klattenhoff et al. 2009; Pane et al. 2011; Zhang et al. 2012, 2014; Le Thomas et al. 2014; Mohn et al. 2014). We found that TREX and Cuff physically and genetically interact. First, Cuff is required for localization of both Thoc5 and UAP56 to nuclear foci, the sites of nascent piRNA precursor transcription (Fig. 6C). Second, Cuff copurifies with Thoc5 from cellular extract, and this interaction is at least partially RNA-independent, indicating that Cuff can bind the TREX complex either directly or through other proteins (Fig. 6A,B). Importantly, a close paralog of Cuff, CG9125, does not form a complex with TREX, supporting the specificity of the interaction (Fig. 6B). Finally, tethering of Cuff to an artificial intronless reporter enhances binding of TREX to the reporter transcript (Fig. 6F,G). This experiment provides direct evidence that the presence of Cuff in the vicinity of a nascent transcript stimulates the association of TREX with the RNA. Together, our results suggest that Cuff and TREX form a complex on dual-strand piRNA loci, and Cuff enhances loading of TREX on nascent RNA.

Together with previously reported results, our data can be integrated into a model that suggests that interaction with chromatin-bound Cuff helps to recruit TREX to dual-strand piRNA cluster loci in the nucleus to ensure its binding with nonspliced nascent piRNA transcripts (Fig. 7H). Previous studies showed that Cuff is enriched on chromatin of dual-strand piRNA clusters, likely through formation of the complex with Rhi that directly binds the histone H3K9me3 mark (Le Thomas et al. 2014; Mohn et al. 2014; Zhang et al. 2014). We propose that, similar to the result that we obtained using an artificial reporter, Cuff enhances loading of TREX on nascent piRNA precursors. Unfortunately, it is impossible to test the association of TREX with piRNA precursors in *cuff* mutants, as *cuff* deficiency leads to almost complete elimination of these RNA (Le Thomas et al. 2014; Mohn et al. 2014). Our model proposes a novel splicing-independent mechanism of TREX loading in which a chromatin-binding protein (Rhi) and an adapter protein (Cuff) result in locus-specific, but sequence-independent, loading of an RNA-binding protein complex on nascent transcripts (Fig. 7H). It will be interesting to explore whether other proteins that associate with specific chromatin marks mediate loading of TREX on other intronless transcripts. We hypothesize that chromatin structure and associated chromatin proteins might play crucial roles in guiding loading of other non-sequence-specific RNA-binding proteins on nascent RNA. Several RNA-binding proteins are cotranscriptionally loaded on nascent RNA and may even shuttle with processed mRNA to the cytoplasm (Dreyfuss et al. 2002). Therefore, the origin of an RNA in a particular chromatin environment might have long-lasting and far-reaching effects on its stability, subcellular localization, and translation efficiency. Quantitative profiling of the association between RNA-binding proteins and RNA originating from different chromatin environments will provide the ultimate test of this hypothesis.

TREX was proposed to play roles in post-transcriptional pre-mRNA processing and nuclear export of mRNA. However, our results showed that only a small number

of genes change their expression in TREX mutants. Similar results were previously reported using knockdown of TREX components in fly and mammalian cells, suggesting that the function of TREX in mRNA biogenesis in metazoa might be redundant with functions of other factors (Rehwinkel et al. 2004; Katahira et al. 2009; Viphakone et al. 2012; Chi et al. 2013). Using two independent approaches, analysis of chromatin-associated nascent transcripts (Fig. 7C,D) and in situ RNA hybridization (Fig. 7F), we showed that TREX is required for the accumulation of nascent piRNA precursors at sites of transcription. These experiments cannot formally rule out the possibility that TREX stabilizes nascent piRNA transcripts and prevents their cotranscriptional degradation. However, the decrease in Pol II occupancy on piRNA clusters in the $thoc5^{e/1}$ mutant (Fig. 7G) suggests that TREX might be required for efficient transcription of piRNA precursors. TREX has previously been shown to prevent formation of RNA–DNA hybrids (R loops) between the nascent transcript and the DNA template in mammalian cells (Dominguez-Sanchez et al. 2011), establishing a possible mechanism by which TREX could facilitate transcription. The impact of TREX on transcription does not eliminate the possibility that it also functions on later steps of piRNA biogenesis. Particularly, it will be important to determine whether TREX plays a role in the export of piRNA precursors from the nucleus to the cytoplasm, where final processing into mature piRNAs takes place.

Materials and methods

Drosophila stocks

$thoc5^1$ and Thoc5-GFP transgenic flies were previously described (Moon et al. 2011). To tag Thoc5, GFP sequence was inserted into an ~8-kB genomic rescue construct that contains the $thoc5$ gene with three additional genes surrounding it, providing expression levels comparable with that of the endogenous protein. $thoc5^{e00906}$, $thoc6^{e00298}$, and $thoc7^{d05792}$ alleles were obtained from the Bloomington Stock Center (Indiana University). $cuff^{WM25}$, $cuff^{QQ37}$ mutants, and eGFP-Cuff transgenic flies were kindly provided by Dr. T. Schupbach. $uap56^{28}$, $uap56^{SZ15}$, aub^{QC42}, $armi^{72.1}$, and $armi^1$ were kindly provided by Dr. W. Theurkauf.

Generation of transgenic fly lines

Plasmids for expression of BioTAP-tagged UAP56, Rhi, and Su(Var)3-9 were generated by insertion of cDNA sequences obtained by RT–PCR into pENTR-D-TOPO followed by transfer into the pUASP-Gateway-PhiC31 vector with BioTAP tag (Alekseyenko et al. 2015). UAP56 was tagged at the N terminus, while Su(Var)3-9 and Rhi were tagged at the C terminus. Transgenic flies were generated by φC31 integrase-mediated insertion into the landing pad on the third chromosome, PBac{y⁺-attP-3B} VK00033. Expression of the transgenes with the yeast upstream activation sequence promoter (UASp) was driven by a Nanos-GAL4 driver. The construct for BioTAP-tagged Piwi was generated by replacing GFP with the protein A sequence in the previously described GFP-Piwi plasmid (Le Thomas et al. 2013). Transgenic flies were generated by P-element-mediated transformation (BestGene).

RNA-seq library preparation and high-throughput data analyses

RNA-seq library preparation and analyses were performed as described (Mortazavi et al. 2008; Le Thomas et al. 2013, 2014). Briefly, total RNA from dissected ovaries was purified using Trizol, and ribosomal RNA was depleted using the Ribo-Zero rRNA removal kit (human/mouse/rat) (Epicenter) with additional probes. Libraries were generated using the NEBNext library preparation kit (New England Biolabs) for Illumina HiSeq 2500 (50-base-pair [bp] reads) platform. The RNA-seq reads were first mapped against *Drosophila* rRNA using Bowtie version 0.12.7 (Langmead et al. 2009) with the following settings: "-v3 -a." The remaining reads were aligned against the *Drosophila melanogaster* genome (dm3) with settings "-v0 -m1," and reads corresponding to RefSeq-annotated genes and piRNA clusters were calculated using eXpress or custom Python scripts. For analyses including repetitive regions (*Hsp70Ab*), data were mapped allowing up to 10,000 mapping positions, and read counts were corrected based on the number of mapped positions. Reads were also aligned against the Repbase repeat database (Jurka et al. 2005) with the following settings: "-v3 -a." Read counts for genes and transposon families were combined, and differential expression was analyzed using the DESeq2 package (Anders and Huber 2010; Love et al. 2014) using default parameters. Statistically up-regulated or down-regulated genes and transposons were defined using a Benjamini-Hochberg adjusted P-value (false discovery rate [FDR]) threshold of 0.2 and fold change >2.

Thoc5 and Thoc7 RIP

RNA associated with Thoc5 was analyzed by RIP from ovarian lysates of flies that express Thoc5-GFP using antibodies against GFP (Development Studies Hybridoma Bank [DSHB], 12A6). Two control samples were used to determine the specificity of RIP: the lysates from flies that express GFP alone and flies that do not express GFP-tagged proteins. Thoc7 was immunoprecipitated using specific antibodies (Moon et al. 2011), and a "no antibody" sample was used as a negative control. One-hundred pairs of ovaries dissected from yeast-fed flies were homogenized in lysis buffer (50 mM Tris-Cl at pH 7.4, 150 mM NaCl, 0.2% Triton X-100, 0.2% NP40, 2 mM DTT, RNase inhibitor, EDTA-free Complete protease inhibitor [Roche], 5% glycerol). The lysate was cleared by centrifugation and incubated with protein G Dynabeads (Invitrogen) that were conjugated with anti-GFP or anti-Thoc7 antibody (Moon et al. 2011) for 2 h at 4°C. Beads were washed five times in PBS with 0.1% Triton X-100 followed by proteinase K treatment for 30 min at 37°C and RNA extraction using Trizol (Invitrogen) following the manufacturer's suggestions. RNA was reverse-transcribed with SuperScript III (Invitrogen) and random hexamer primers and was used for qPCR. Primers used in the RT-qPCR are listed in Supplemental Table S3. For analysis of RNA associated with Thoc5, immunoprecipitation/input ratios were calculated and normalized to $rp49$ mRNA. Next, fold changes relative to control samples from flies that did not express GFP-tagged protein were calculated. For transcriptome-wide analyses of Thoc5-GPF RIP, the signal from the negative GFP-only control was subtracted, and data were normalized to input. Genes with low expression (<10 RPKM [reads per kilobase per million mapped reads]) in input were discarded.

Small RNA library preparation and sequencing

Small RNA libraries were prepared from total RNA of dissected ovaries ($thoc5^{e/1}$ mutant and heterozygous) as described in previous studies (Brennecke et al. 2007; Aravin et al. 2008). Briefly, small RNAs in the range of 19–29 nt were isolated from 12%

polyacrylamide gels. Next, 3′ and 5′ linkers were ligated, and cDNA was synthesized by reverse transcription using Super-Script III (Invitrogen). The cDNA was amplified by PCR for library preparation and Illumina sequencing at BGI. Additional piRNA data sets were extracted from available Gene Expression Omnibus (GEO) data for *cuff* (GSE59610) and *rhino* (GSE55824) germline knockdown. Adapter sequences were removed using the cutadapt tool (https://code.google.com/p/cutadapt), and reads >15 nt were retained. Reads were mapped to rRNA sequences using Bowtie version 0.12.7 (Langmead et al. 2009), allowing three mismatches. The remaining reads were then mapped to the *D. melanogaster* genome (dm3) using Bowtie version 0.12.7, allowing zero mismatches and retaining only uniquely mapping reads. Numbers of reads corresponding to *D. melanogaster* miRNA were calculated based on the miRBase version 21 annotations using a custom script, and total levels of miRNA were used as a reference for data normalization. piRNA cluster annotations were retrieved from Le Thomas et al. (2014), and normalized expression was calculated using a custom script. For visualization of piRNA coverage as genome tracks, reads mapping uniquely to one position were retained. In the analyses of the repetitive hp-endo-siRNA locus, we considered all reads, and read counts were corrected by the number of mapped positions. Genome coverage tracks were generated from mapped reads using BedTools (Quinlan and Hall 2010) and BigWig tools (Kent et al. 2010).

Analysis of ping-pong processing

For evaluating the ping-pong signature of transposons, 23- to 28-nt piRNA reads that mapped to the consensus sequences of each transposon family (as defined in Repbase) allowing up to three mismatches were selected. For the ping-pong signature at piRNA clusters (*cis* ping-pong), we selected 23- to 28-nt-long reads mapping uniquely and perfectly to the piRNA clusters annotated in the fly genome (dm3). The number of reads that form 5′–5′ sense–antisense pairs with different overlaps for each element or cluster were calculated using a custom script. Z-scores for 5′–5′ sense–antisense overlap at position 10 were calculated as described in Zhang et al. (2011), considering overlaps at distances between 1–9 and 11–23 as background. To further determine the fraction of piRNAs participating in ping-pong pairs, piRNAs were grouped according to the transposon family. For each piRNA group, the fraction of piRNAs that had at least one ping-pong pair (10-nt differences between the 5′ ends) was calculated. To control for the variable sizes of the piRNA groups in different libraries, piRNAs were subsampled to the size of the smallest group, the sampling was repeated 1000 times, and the average value was used for subsequent analysis.

ChIP-qPCR

ChIP with antibodies against RNA Pol II (Abcam, ab5408) was performed as described previously (Le Thomas et al. 2014). Dissected fly ovaries were fixed with 1% paraformaldehyde (PFA) in PBS for 10 min at room temperature. After quenching with glycine (25 mM final) for 5 min at room temperature, the samples were dounced and sonicated in RIPA buffer (20 mM Tris-Cl at pH 7.4, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, EDTA-free protease inhibitor [Roche], 10 mM NaF, 0.2 mM Na_3VO_4). Sonication was performed with Bioruptor (Diagenode) set to medium power for 20 cycles (30 sec on and 30 sec off). Samples were centrifuged, and the supernatants were precleared with protein G Dynabeads (Invitrogen) for 2 h at 4°C. Inputs were taken after preclearing. The rest of the samples were incubated with antibody-conjugated protein G beads for 2 h at 4°C. After incubation, the beads were washed five times with LiCl immunoprecipitation buffer (10 mM Tris-Cl at pH 7.5, 500 mM LiCl, 1% NP-40, 1% sodium deoxycholate). The beads were rinsed with TE, and purified DNA was eluted with 100 µg of proteinase K in proteinase K buffer (200 mM Tris-Cl at pH 7.4, 25 mM EDTA, 300 mM NaCl, 2% SDS) for 3 h at 55°C and then overnight at 65°C. The eluted DNA was purified using phenol-chloroform extraction followed by ethanol precipitation. qPCR values were normalized to input and to a region in Flamenco. Primers used for qPCR are listed in Supplemental Table S6.

Coimmunoprecipitation of proteins from fly ovaries and S2 cells

Fly ovaries that expressed tagged transgenes were dissected in PBS. Schneider S2 cells were cultured in Schneider medium with 10% fetal bovine serum and 100 U of penicillin–streptomycin (Life Technologies). Expression plasmids were generated by Gateway cloning (Life Technologies) using destination vectors of the *Drosophila* Gateway Vector Collection. Expression of proteins fused to Flag, Flag-HA, or Myc tags was driven by the Actin5C promoter. Transfection was conducted with TransIT-LT1 (Mirus Biosciences). Ovaries were homogenized in lysis buffer (20 mM Tris-Cl at pH 7.0, 100 mM NaCl, 0.2% NP-40, 0.2% Triton X-100) with EDTA-free proteinase inhibitor (Roche). S2 cells were lysed 2 d after transfection in the same buffer. The lysate was incubated for 10 min on ice and was cleared by centrifuging at 3000g for 10 min at 4°C. Inputs were taken from the supernatant. For RNase-treated samples, RNase A was added to a final concentration of 100 µg/mL. Anti-Flag M2 beads (Sigma-Aldrich) were blocked with 5 mg/mL BSA and then added to the supernatant. After incubation for 90 min at 4°C, beads were washed in PBST (PBS and 0.1% Tween-20), and the bound proteins were eluted by adding SDS loading buffer and heating at 95°C. Input and bound samples were separated on SDS-PAGE and were assayed by Western blot. The antibodies used were horseradish peroxidase (HRP)-conjugated anti-HA antibody (Sigma-Aldrich) at 1:5000, HRP-conjugated anti-Myc antibody (Invitrogen) at 1:5000, and peroxidase–anti-peroxidase (PAP) antibody (Sigma-Aldrich) at 1:5000. For coimmunoprecipitation of the truncated Thoc5 and Cuff proteins, the truncated sequences were cloned into pActin-Flag-HA-Gateway vector, and full-length Thoc5 and Cuff were cloned into pActin-eGFP-Gateway vector from the *Drosophila* Gateway Vector Collection. S2 cells were cotransfected with the plasmids encoding GFP-fused and Flag-HA-fused proteins. The cells were lysed in lysis buffer (20 mM Tris-Cl at pH 7.0, 100 mM NaCl, 0.2% NP-40, 0.2% Triton X-100). To immunoprecipitate tagged proteins, the lysate was incubated with anti-GFP beads (Allele) followed by five washes with wash buffer (0.1% NP40, 0.1% Triton X-100, 400 mM NaCl, 20 mM Tris-Cl at pH 7.4). Flag-tagged proteins were detected by Western blot using anti-Flag antibody (Sigma-Aldrich) at 1:10,000 dilution.

Immunofluorescence microscopy

Immunofluorescence microscopy was performed as previously described (Rozhkov et al. 2013). Ovaries were fixed in 4% PFA in PBS supplemented with 0.1% Triton X-100 (PBST) for 20 min. Samples were blocked in PBST with 1% BSA for 1 h at room temperature and incubated with primary antibody overnight at 4°C. Primary antibodies used were PAP (Sigma-Aldrich), anti-H2Av (Millipore), and anti-Vasa (DSHB), anti-Thoc7 (Kim et al. 2011), and anti-Thoc2 (Rehwinkel et al. 2004). Samples

were washed and incubated with a secondary antibody (Alexa fluor 488, 546, or 568; Life Technologies) at 1:1000 dilution for 1 h. The samples were washed and stained with 1 μM DAPI (Sigma-Aldrich). Next, the ovaries were washed again in PBS and mounted in VectaShield medium (Vector Laboratories). Images were acquired using an Axio Imager microscope with an Apotome structured illumination system for optical sectioning (Carl Zeiss).

Colocalization analysis

Images of ovary samples coexpressing Thoc5-GFP and other proteins (probed with Alexa fluor 568) were analyzed by Fiji software using the "coloc2" plug-in. Pearson's correlation coefficient (r) was calculated for three independent nuclei. Individual r values were converted to Fisher's z values. The mean and 95% confidence interval were calculated for the z values, and the mean and range were converted back to r space.

Isolation of chromatin-bound nascent RNA transcripts

Chromatin-bound RNA was isolated as described (Khodor et al. 2011) with modifications. For each experiment, 200–300 female flies were fed on yeast for 2–3 d prior to dissection. The ovaries were dissected into PBS and washed in ice-cold PBS once. To isolate the nuclei, the ovaries were transferred into 1 mL of buffer AT [15 mM HEPES-KOH at pH 7.6, 10 mM KCl, 5 mM Mg($C_2H_3O_2$)$_2$, 3 mM CaCl$_2$, 300 mM sucrose, 0.1% Triton X-100, 1 mM DTT, complete protease inhibitor, EDTA-free (Roche)] and dounced 30 times with a tight pestle. To clear the lysate of debris, it was centrifuged through a miracloth strainer. Approximately one-third of a 0.6-mL Eppendorf tube was cut off from the tip, covered with a piece of miracloth filter, and put inside a 1.5-mL Eppendorf tube. The lysate was pipetted into the inner strainer tube and briefly centrifuged through the miracloth in a minifuge. Aliquots of 0.5 mL of the filtered lysate were layered on top of 1 mL of buffer B [15 mM HEPES-KOH at pH 7.6, 10 mM KCl, 5 mM Mg ($C_2H_3O_2$)$_2$, 3 mM CaCl$_2$, 1 M sucrose, 0.1% Triton X-100, 1 mM DTT, Complete protease inhibitor, EDTA free (Roche)] and centrifuged at 5900g for 15 min at 4°C. The volume of the pellet was estimated, and the pellet was resuspended in 5 vol of nuclear lysis buffer (10 mM HEPES-KOH at pH 7.6, 100 mM KCl, 0,1 mM EDTA, 10% glycerol, 0.1 mM ZnCl$_2$, 0.15 mM spermine, 0.5 mM spermidine, 10 mM NaF, 0.2 mM Na$_3$VO$_4$, 1 mM DTT, Complete protease inhibitor, EDTA-free [Roche], 1 U/μL RNasin plus) and dounced three times with a loose pestle and twice with a tight pestle. The lysate was transferred into a tube large enough to encompass double the lysate volume. An equal volume of 2xNUN buffer (50 mM HEPES-KOH at pH 7.6, 600 mM NaCl, 2 M urea, 2% Ipegal, Complete protease inhibitor, EDTA-free [Roche]) was added in small portions. After each addition of 2xNUN buffer, the sample was briefly vortexed and put on ice. The sample was spun at 16,100g for 30 min at 4°C and washed in a 1:1 mix of nuclear lysis buffer and 2xNUN buffer twice. The chromatin pellet was resuspended in 1 mL of Ribozol and passed through a 21-guage needle until it was completely dissolved. RNA was extracted following the procedure in the Ribozol manual.

RNA fluorescence in situ hybridization (FISH)

RNA-FISH protocol and probe sequences were adapted from Mohn et al. (2014). Oligonucleotide probes were chemically synthesized with 5′ amine modification (IDT) and then coupled with fluorescent NHS-ester dye (Alexa-594). Fly ovaries were dissected into cold PBS and fixed with 4% formaldehyde and 0.15% Triton X-100 in PBS at room temperature. The ovaries were washed three times with PBS containing 0.3% Triton X-100 at room temperature. The ovaries were then shaken overnight in 70% ethanol at 4°C. After rehydration for 5 min in wash buffer (2× SSC with 10% formamide), the ovaries were incubated with FISH probes (~10 pmol of probes) in 50 μL of hybridization buffer (2× SSC, 10 mg/mL dextran sulfate, high-molecular-weight >500,000 g/mol, 10% formamide) overnight at 37°C. The samples were then rinsed twice and washed twice for 30 min each with wash buffer. The samples were counterstained with 1 μg/mL DAPI in 2× SSC for 10 min at room temperature and then washed twice with 2× SSC. The samples were mounted on slide glass using VectaShield mounting medium. The samples were incubated for 24 h at room temperature and imaged using confocal microscopy.

Accession codes

High-throughput sequencing data for RNA-seq and piRNA-seq experiments are available through Gene Expression Omnibus (accession no. GSE79325).

References

Abruzzi KC, Lacadie S, Rosbash M. 2004. Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. *EMBO J* **23:** 2620–2631.

Aguilera A. 2005. Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. *Curr Opin Cell Biol* **17:** 242–250.

Alekseyenko AA, McElroy KA, Kang H, Zee BM, Kharchenko PV, Kuroda MI. 2015. BioTAP-XL: cross-linking/tandem affinity purification to study DNA targets, RNA, and protein components of chromatin-associated complexes. *Curr Protoc Mol Biol* **109:** 21 30 21–21 30 32.

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

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

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.

Chavez S, Garcia-Rubio M, Prado F, Aguilera A. 2001. Hpr1 is preferentially required for transcription of either long or

G+C-rich DNA sequences in *Saccharomyces cerevisiae*. *Mol Cell Biol* **21:** 7054–7064.

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

Cheng H, Dufu K, Lee CS, Hsu JL, Dias A, Reed R. 2006. Human mRNA export machinery recruited to the 5′ end of mRNA. *Cell* **127:** 1389–1400.

Chi B, Wang Q, Wu G, Tan M, Wang L, Shi M, Chang X, Cheng H. 2013. Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA. *Nucleic Acids Res* **41:** 1294–1306.

Chi B, Wang K, Du Y, Gui B, Chang X, Wang L, Fan J, Chen S, Wu X, Li G, et al. 2014. A Sub-Element in PRE enhances nuclear export of intronless mRNAs by recruiting the TREX complex via ZC3H18. *Nucleic Acids Res* **42:** 7305–7318.

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

Czech B, Preall JB, McGinn J, Hannon GJ. 2013. A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway. *Mol Cell* **50:** 749–761.

Dominguez-Sanchez MS, Barroso S, Gomez-Gonzalez B, Luna R, Aguilera A. 2011. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet* **7:** e1002386.

Dreyfuss G, Kim VN, Kataoka N. 2002. Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* **3:** 195–205.

Gatfield D, Izaurralde E. 2002. REF1/Aly and the additional exon junction complex proteins are dispensable for nuclear mRNA export. *J Cell Biol* **159:** 579–588.

Gatfield D, Le Hir H, Schmitt C, Braun IC, Kocher T, Wilm M, Izaurralde E. 2001. The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in *Drosophila*. *Curr Biol* **11:** 1716–1721.

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

Guria A, Tran DD, Ramachandran S, Koch A, El Bounkari O, Dutta P, Hauser H, Tamura T. 2011. Identification of mRNAs that are spliced but not exported to the cytoplasm in the absence of THOC5 in mouse embryo fibroblasts. *RNA* **17:** 1048–1056.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110:** 462–467.

Katahira J. 2012. mRNA export and the TREX complex. *Biochim Biophys Acta* **1819:** 507–513.

Katahira J, Inoue H, Hurt E, Yoneda Y. 2009. Adaptor Aly and co-adaptor Thoc5 function in the Tap-p15-mediated nuclear export of HSP70 mRNA. *EMBO J* **28:** 556–567.

Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. 2010. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26:** 2204–2207.

Khodor YL, Rodriguez J, Abruzzi KC, Tang CH, Marr MT II, Rosbash M. 2011. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev* **25:** 2502–2512.

Kim H, Cho B, Moon S, Chung YD. 2011. The THO complex is required for stress tolerance and longevity in *Drosophila*. *Genes Genomics* **33:** 291–297.

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

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

Kopytova DV, Orlova AV, Krasnov AN, Gurskiy DY, Nikolenko JV, Nabirochkina EN, Shidlovskii YV, Georgieva SG. 2010. Multifunctional factor ENY2 is associated with the THO complex and promotes its recruitment onto nascent mRNA. *Genes Dev* **24:** 86–96.

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.

Lei EP, Krebber H, Silver PA. 2001. Messenger RNAs are recruited for nuclear export during transcription. *Genes Dev* **15:** 1771–1782.

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.

Le Thomas A, Stuwe E, Li S, Du J, Marinov G, Rozhkov N, Chen YC, Luo Y, Sachidanandam R, Toth KF, et al. 2014. Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev* **28:** 1667–1680.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550.

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

Masuda S, Das R, Cheng H, Hurt E, Dorman N, Reed R. 2005. Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev* **19:** 1512–1517.

Mohn F, Sienski G, Handler D, Brennecke J. 2014. The Rhino–Deadlock–Cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell* **157:** 1364–1379.

Moon S, Cho B, Min SH, Lee D, Chung YD. 2011. The THO complex is required for nucleolar integrity in *Drosophila* spermatocytes. *Development* **138:** 3835–3845.

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

Nojima T, Hirose T, Kimura H, Hagiwara M. 2007. The interaction between cap-binding complex and RNA export factor is required for intronless mRNA export. *J Biol Chem* **282:** 15645–15651.

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

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841–842.

Rehwinkel J, Herold A, Gari K, Kocher T, Rode M, Ciccarelli FL, Wilm M, Izaurralde E. 2004. Genome-wide analysis of mRNAs regulated by the THO complex in *Drosophila melanogaster*. *Nat Struct Mol Biol* **11:** 558–566.

Rozhkov NV, Hammell M, Hannon GJ. 2013. Multiple roles for Piwi in silencing *Drosophila* transposons. *Genes Dev* **27:** 400–412.

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.

Strasser K, Hurt E. 2000. Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. *EMBO J* **19:** 410–420.

Strasser K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, Rondon AG, Aguilera A, Struhl K, Reed R, et al. 2002. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417:** 304–308.

Stutz F, Bachi A, Doerks T, Braun IC, Seraphin B, Wilm M, Bork P, Izaurralde E. 2000. REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA* **6:** 638–650.

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.

Viphakone N, Hautbergue GM, Walsh M, Chang CT, Holland A, Folco EG, Reed R, Wilson SA. 2012. TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. *Nat Commun* **3:** 1006.

Voynov V, Verstrepen KJ, Jansen A, Runner VM, Buratowski S, Fink GR. 2006. Genes with internal repeats require the THO complex for transcription. *Proc Natl Acad Sci* **103:** 14423–14428.

Yu B, Cassani M, Wang M, Liu M, Ma J, Li G, Zhang Z, Huang Y. 2015. Structural insights into Rhino-mediated germline piRNA cluster formation. *Cell Res* **25:** 525–528.

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

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

Zhang Z, Wang J, Schultz N, Zhang F, Parhad SS, Tu S, Vreven T, Zamore PD, Weng Z, Theurkauf WE. 2014. The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell* **157:** 1353–1363.

Zhou Z, Luo MJ, Straesser K, Katahira J, Hurt E, Reed R. 2000. The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407:** 401–405.