HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2018 November 14.

Published in final edited form as:

Cell Rep. 2018 September 25; 24(13): 3413–3422.e4. doi:10.1016/j.celrep.2018.08.081.

Co-dependent Assembly of *Drosophila* piRNA Precursor Complexes and piRNA Cluster Heterochromatin

Gen Zhang[1,3], **Shikui Tu**[2,3,4], **Tianxiong Yu**[2], **Xiao-Ou Zhang**[2], **Swapnil S. Parhad**[1], **Zhiping Weng**[2,*], and **William E. Theurkauf**[1,5,*]

[1]Program in Molecular Medicine, University of Massachusetts Medical School, 373 Plantation Street, Worcester, MA 01605, USA

[2]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, 373 Plantation Street, Worcester, MA 01605, USA

[3]These authors contributed equally

[4]Present address: Center for Cognitive Machines and Computational Health and Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

[5]Lead Contact

Abstract

In *Drosophila*, the piRNAs that guide germline transposon silencing are produced from heterochromatic clusters marked by the HP1 homolog Rhino. We show that Rhino promotes cluster transcript association with UAP56 and the THO complex, forming RNA-protein assemblies that are unique to piRNA precursors. UAP56 and THO are ubiquitous RNAprocessing factors, and null alleles of *uap56* and the THO subunit gene *tho2* are lethal. However, $uap56^{sz15}$ and mutations in the THO subunit genes *thoc5* and *thoc7* are viable but sterile and disrupt piRNA biogenesis. The $uap56^{sz15}$ allele reduces UAP56 binding to THO, and the *thoc5* and *thoc7* mutations disrupt interactions among the remaining THO subunits and UAP56 binding to the core THO subunit Hpr1. These mutations also reduce Rhino binding to clusters and trigger Rhino binding to ectopic sites across the genome. Rhino thus promotes assembly of piRNA precursor complexes, and these complexes restrict Rhino at cluster heterochromatin.

Graphical Abstract

*Correspondence: zhiping.weng@umassmed.edu (Z.W.), william.theurkauf@umassmed.edu (W.E.T.).

AUTHOR CONTRIBUTIONS

G.Z., Z.W., and W.E.T. conceived the project. G.Z. performed all the experiments with help from S.S.P. S.T. performed most bioinformatics analyses with help from T.Y. and X.-O.Z. G.Z. and W.E.T. wrote the paper, with help from all authors.

DECLARATION OF INTERESTS

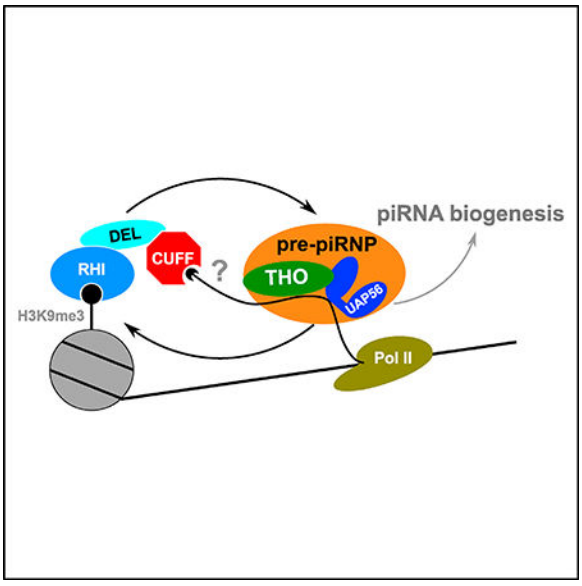
The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and five tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.08.081.

DATA AND SOFTWARE AVAILABILITY

High-throughput sequencing data was deposited in the NCBI Sequence read archive (SRA: SRP151054). The scaffold files for Mass spectrometry analysis were deposited in Mendeley Data (https://doi.org/10.17632/j4yktssk9g.1)



In Brief

Zhang et al. show that the piRNAgenerating loci in Drosophila female germline are co-dependently assembled between unique chromatin factor Rhino (with Cutoff and Deadlock) and general RNA-processing factor TREX complex

INTRODUCTION

Transposable elements are ubiquitous genome constituents with the potential to mobilize and trigger catastrophic genome instability (Belancio et al., 2008; Hedges and Deininger, 2007; McClintock, 1950). The PIWI-interacting RNA (piRNA) pathway is an adaptive genome immune system that silences transposons and maintains genome integrity during germline development (Aravin et al., 2007; Brennecke et al., 2007; Ghildiyal and Zamore, 2009; Khurana et al., 2010; Siomi et al., 2010, 2011). The 23–30 nucleotide long piRNAs, loaded into PIWI clade Argonaut proteins, direct sequence-specific transcriptional and post-transcriptional transposon silencing (Brennecke et al., 2007; Girard et al., 2006; Gunawardane et al., 2007; Lau et al., 2006; Malone et al., 2009; Vagin et al., 2006).

In the *Drosophila* female germline, piRNAs are processed from RNA polymerase II (RNA Pol II) transcripts of discrete genomic domains called piRNA clusters, composed of nested transposon fragments, which provide an archive of invading transposon sequences (Andersen et al., 2017; Bergman et al., 2006; Brennecke et al., 2007; Chen et al., 2016; Mohn et al., 2014; Zhang et al., 2012a, 2014). Transposition of an invading mobile element into a cluster is proposed to trigger adaptation, and the system must therefore have ability to process any inserted sequence into piRNAs. Consistent with this flexibility, cluster transcripts do not have well-defined sequence or secondary structure signatures. This is in striking contrast to the precursors for small interfering RNAs (siRNAs) and microRNAs (miRNAs), which form double-stranded structures that are recognized by proprietary processing machines (Brennecke et al., 2007; Ghildiyal and Zamore, 2009; Iwasaki et al.,

2015; Kim et al., 2009; Zhang et al., 2012a, 2014). How transcripts from piRNA clusters are distinguished from bulk RNA Pol II transcripts, including pre-mRNAs and mRNAs, remains an open question.

Germline piRNA clusters in *Drosophila* ovaries are uniquely marked by the heterochromatin protein 1 (HP1) homolog Rhino (Rhi), which associates with Deadlock (Del) and Cutoff (Cuff) to form the RDC complex (Klattenhoff et al., 2009; Mohn et al., 2014; Zhang et al., 2014). These three factors are co-dependent for localization to cluster heterochromatin, drive non-canonical transcription of piRNA clusters from both genomic strands, suppress cluster transcript splicing and polyadenylation, and promote piRNA biogenesis (Andersen et al., 2017; Chen et al., 2016; Klattenhoff et al., 2009; Mohn et al., 2014; Pane et al., 2011; Parhad et al., 2017; Zhang et al., 2014). Like the founding member of the HP1 family, HP1a, Rhi binds to trimethylated lysine 9 of histone 3 (H3K9me3) through its C-terminal chromo domain (Le Thomas et al., 2014; Mohn et al., 2014; Yu et al., 2015). However, H3K9me3 is broadly distributed over heterochromatin, and Rhi localizes specifically to piRNA clusters (Mohn et al., 2014). It is unclear how Rhi distinguishes between H3K9me3 marks on piRNA clusters and bulk heterochromatin.

UAP56 is a ubiquitously expressed DEAD box protein with conserved functions in RNA processing and export, and null uap56 alleles are lethal (Eberl et al., 1997; Gatfield et al., 2001). However, the *uap56*^{sz15} point mutation, when combined with a strong hypomorphic allele (*uap56*^{28}) that produces low levels of wild-type protein, is viable and supports normal gene expression but leads to sterility and disrupts piRNA biogenesis (Meignin and Davis, 2008; Zhang et al., 2012a). We show that this point mutation reduces UAP56 binding to the heteropentameric THO complex (composed of Hpr1, Tho2, Thoc5, Thoc6, and Thoc7), which interacts with UAP56 in the transcription and export (TREX) complex (Chi et al., 2013; Gatfield et al., 2001; Reed and Cheng, 2005; Rehwinkel et al., 2004; Viphakone et al., 2012). Mutations in the *thoc5* gene are sterile and disrupt piRNA production and transposon silencing (Hur et al., 2016). We show that this mutation, and a null allele of *thoc7*, destabilize the THO complex and block UAP56 co-precipitation with Hpr1. We also show that stable binding to both UAP56 and THO is unique to cluster transcripts and that assembly of these pre-piRNA complexes requires Rhi. Significantly, mutations in *thoc7*, *thoc5*, and *uap56* that disrupt transposon silencing also reduce Rhi binding to major piRNA clusters and trigger ectopic Rhi binding to heterochromatic and euchromatic H3K9me3 marks across the genome. Rhi thus promotes assembly of pre-piRNA complexes containing UAP56, THO, and cluster transcripts, and these complexes restrict Rhi at piRNA clusters. We propose that this feedforward system drives efficient and specific piRNA biogenesis.

RESULTS

The *uap56*^{sz15} Allele Reduces Binding to the THO Complex

UAP56 is a conserved DEAD box protein implicated in RNA processing and export, and null mutations in *Drosophila uap56* are lethal (Eberl et al., 1997; Gatfield et al., 2001). However, the *uap56*^{sz15} point mutation, combined with a strong hypomorphic allele (*uap56*^{28}) that produces low levels of wild-type protein, is viable but sterile and disrupts germline piRNA biogenesis and transposon silencing (Eberl et al., 1997; Zhang et al.,

2012a). We refer to this allelic combination as uap56 mutant in the balance of the text. UAP56 co-localizes with Rhi and binds to cluster transcripts, while the UAP56^{sz15} protein does not co-localize with Rhi and shows reduced binding to germline cluster transcripts (Zhang et al., 2012a). These findings suggest that *uap56^{sz15}* disrupts a protein-protein interaction that is essential to piRNA biogenesis (Zhang et al., 2012a).

To identify proteins that show altered binding to the *uap56^{sz15}* gene product, we affinity-purified Venus-tagged UAP56 and UAP56^{sz15} proteins from wild-type ovaries (Zhang et al., 2012a) and assayed bound proteins by mass spectrometry. The relative abundance of co-precipitating proteins was estimated using iBAQ values (Schwanhäusser et al., 2011), normalized to the Venus tag. We then calculated the average fold difference in protein binding to UAP56^{sz15}venus relative to UAP56venus, from three biological replicates. Figure 1A shows ranked fold differences in protein abundance, with highest fold reduction on the left. All five subunits of the THO complex (labeled red in Figure 1A, inset) rank among the top proteins showing reduced binding to UAP56^{sz15}venus. We independently verified this observation by quantitative western blotting for the THO subunit Hpr1 in UAP56venus and UAP56^{sz15}venus immunoprecipitates (Figures 1B and 1C). For quantification, Hpr1 signal was compared with signal for the Venus tag on the UAP56 fusions. To control for non-linearity in the assay, blots were performed on a series of dilutions of the precipitates. Figure 1C summarizes quantification of four independent biological replicates, each assayed at three different dilutions. These studies revealed a 3-fold reduction in Hpr1 binding to UAP56^{sz15} relative to wild-type, consistent with our proteomic data. The point mutation in *uap56^{sz15}* thus reduces UAP56 binding to the THO complex (Figure 1D).

THO interacts with UAP56 within the evolutionarily conserved TREX complex and has been previously implicated in germline piRNA biogenesis (Hur et al., 2016; Reed and Cheng, 2005). UAP56 and THO also localize to Rhi nuclear foci (Hur et al., 2016; Zhang et al., 2012a). To determine if THO localization depends on UAP56, we labeled Tho2 in wild-type and *uap56* mutant ovaries. In wild-type, Tho2 co-localizes with Rhi nuclear foci throughout oogenesis (Figure S1A, top). In *uap56* mutant ovaries, Tho2 shows reduced localization to Rhi foci in early stage egg chambers, and the signal breaks down in later stages egg chambers (Figure S1A, middle). Wild-type interactions between UAP56 and THO are therefore required to maintain THO localization to Rhi nuclear foci.

Intact THO Is Dispensable for Development but Required for piRNA Production

The *thoc5^{e}/thoc5^{1}* allelic combination (referred to as *thoc5* mutant), which disrupts piRNA biogenesis (Hur et al., 2016), appears to be hypomorphic. Low levels of Thoc5 could therefore support zygotic development. By contrast, *thoc7^{d05792}* blocks *thoc7* splicing and produces no detectable Thoc7 protein, as assayed by western blotting (Figure S1C) (Kim et al., 2011). Furthermore, homozygous *thoc7^{d05792}* and hemizygous *thoc7^{d05792}*/Df(3L)BSC128 combinations are phenotypically identical (Figure S1C and data not shown). The *thoc7$^{d/Df}$* combination thus appears to be null (referred to as *thoc7* mutant below). This allelic combination is viable but sterile, disrupts transposon silencing, and reduces germline piRNA production but does not lead to global changes in gene expressions (Figures S1D–

S1F). The intact five-member THO complex thus appears to be dispensable for *Drosophila* development but essential to piRNA production.

To define the role of Thoc7 in assembly of the remaining THO complex components, we used IP-mass spectrometry to characterize proteins that co-precipitate with Hpr1 and Thoc5GFP in wild-type and *thoc7* mutants (Tables S1 and S2, respectively). In wild-type ovary extracts, all five THO subunits co-precipitate with both Hpr1 and Thoc5GFP (Tables S1 and S2). In the *thoc7* mutant ovaries, in contrast, Tho2 co-precipitates with Hpr1, but UAP56 and other THO subunits are undetectable (*thoc7*$^{d/Df}$ in Table S1). Similarly, Tho2 was the only subunit that co-precipitated with Hpr1 from *thoc5* mutant ovaries (*thoc5*$^{e/1}$ in Table S1). In addition, Thoc5 was the only subunit detected when Thoc5GFP was precipitated from thoc7 mutantovaries(*thoc7*$^{d/Df}$inTableS2). Finally, no THO subunits co-precipitate with UAP56venus in thoc7 mutant ovaries (*thoc7*$^{d/Df}$ in Table S3). Consistent with these biochemical observations, Tho2 and UAP56 do not co-localize with Rhi in thoc7 mutant ovaries (Figures S1A and S1B). The results of these proteomic and localization studies are summarized diagrammaticallyinFigure1DandindicatethatlossofThoc7leads to THO complex dissociation into single Thoc5 and Thoc6 subunits and a sub-complex containing Hpr1 and Tho2, which does not stably interact with UAP56.

Hpr1 and Tho2 are the only THO subunits conserved from yeast to humans (Reed and Cheng, 2005), and *Drosophila tho2*-null alleles are lethal (Jagut et al., 2013). These findings, with the data presented here, suggest that an Hpr1Tho2 sub-complex is sufficient for viability, but intact THO and wild-type THO binding to UAP56 (forming TREX) are essential for piRNA biogenesis. However, weak interactions between UAP56 and THO subunits are unlikely to be preserved during immuno-precipitation and could support zygotic development.

StableBindingtoUAP56andTHOIsSpecifictoGermline piRNA Cluster Transcripts

Previous studies indicate that germline cluster transcripts coprecipitate with UAP56 and Thoc5, while mature mRNAs fail to associate with either protein (Hur et al., 2016; Zhang et al., 2012a). We have confirmed these observations by RNA immuno-precipitation (RIP) with deep sequencing and found that cluster transcripts also co-precipitate with endogenous Hpr1 (Figure 2). To control specificity in these experiments, we expressed and immuno-precipitated GFP alone, performed immuno-precipitation with a non-specific IgG, and quantified enrichment in precipitates relative to input or control RIP. Nearly identical results were obtained in both cases. To simplify presentation, abundance in RIP relative to the corresponding input is shown (Figure 2).

Consistent with previous studies, *42AB* piRNA cluster transcripts, but not mature mRNAs, bind to Hpr1, Thoc5GFP, and UAP56venus (Figure 2A, *42AB* and *CG7747*) (Hur et al., 2016; Zhang et al., 2012a). However, direct visual inspection of our data also revealed significant binding of intronic transcripts to Hpr1 and Thoc5GFP but not with UAP56venus (Figure 2A, *Doa*). The scatterplots in Figures 2B–2D confirm that this pattern extends across the transcriptome, with cluster transcripts uniformly enriched with Hpr1, Thoc5GFP, and UAP56, intron transcripts enriched with Hpr1 and Thoc5GFP, but not UAP56, and mature mRNAs showing only background binding to these three proteins.

To quantify intron co-precipitation, we computationally defined introns used in ovaries, as previously described (Zhang et al., 2014), and analyzed introns from protein-coding genes with a minimum of 1 read per kilobase of transcript per million mapped reads (RPKM) in the input. In Thoc5GFP and Hpr1 RIP, 18% and 29% of all introns were enriched by more than 2-fold, respectively, and 14% of all introns were enriched by more than 2-fold with both Thoc5GFP and Hpr1. For this subset of introns, the average fold enrichment (over input) with Hpr1 and Thoc5GFP was 7.5 ± 6.8 and 4.5 ± 3.5, respectively. By contrast, the same set of introns showed an average fold enrichment of only 1.7 ± 1.5 with UAP56venus.

The intron mapping sequences bound by THO could represent unspliced pre-mRNAs or free lariats produced by splicing. To distinguish between these alternatives, we quantified enrichment of RNA sequencing (RNA-seq) reads mapping across 5' splice sites, 3' splice sites, and spliced junctions. Consistent with binding to unspliced introns, reads mapping across both 5' and 3' splice sites were highly enriched with Thoc5GFP and Hpr1, but not UAP56venus, and splice junction mapping reads were not significantly enriched with UAP56, Hpr1, or Thoc5GFP (Figure S2). The THO complex, but not TREX, thus stably interacts with a significant fraction of unspliced pre-mRNAs.

Our studies define three classes of ovarian RNA Pol II transcripts (Figure 2A). Class I is composed of germline cluster transcripts, which are stably bound by THO and UAP56. Class II includes unspliced pre-mRNAs, which are stably bound by THO (defined by Thoc5GFP and Hpr1) but not UAP56. Class III transcripts include mature mRNAs and uni-strand cluster transcripts, which are not stably bound by THO or UAP56. Signifitercantly, our biochemical studies indicate that piRNA biogenesis, but not mRNA expression, requires wild-type interactions between UAP56 and THO. Stable binding by UAP56 and THO is therefore specific to piRNA precursors, and wildtype interaction between these factors is required for piRNA biogenesis.

Rhi Promotes Stable Association of UAP56 with Cluster Transcripts

The *Drosophila* HP1 homolog Rhi anchors a chromatin complex that promotes cluster transcription (Andersen et al., 2017;Mohn et al., 2014), suppresses clustertranscript splicing (Zhang et al., 2014) and cluster transcript polyadenylation (Chen et al., 2016), and is essential to germline piRNA biogenesis (Klattenhoff et al., 2009; Mohn et al., 2014; Zhang et al., 2014). To determine if Rhi is required for cluster transcript binding to UAP56 and THO, we performed RIP sequencing (RIP-seq) with Thoc5GFP and UAP56venus in $rhi^{2/KG}$ mutant ovaries. This null combination is referred to as *rhi* mutant in the balance of the text. This combination reduces overall cluster transcription, but clusters continue to be transcribed (Mohn et al., 2014; Zhang et al., 2014). For example, low-level production of unspliced transcripts continues over much of the left side of the *42AB* cluster, and high-level transcription of spliced transcripts from an adjacent partial *gypsy12* transposon is induced (Figures 3A and 3B, input) (Zhang et al., 2014). In wild-type controls, transcripts over this entire region bind to UAP56venus and Thoc5GFP (Figure 3A). In *rhi* mutants, the unspliced transcripts stably associate with Thoc5GFP, but not with UAP56venus (Figures 3B and S3A), and spliced *gyspy12* transcripts do not precipitate with either Thoc5GFP or UAP56venus (Figure 3B, highlighted on the right). In the absence of Rhi, unspliced cluster

transcripts thus mirror unspliced pre-mRNAs, spliced *gypsy12* transcripts mimic mature mRNAs, and neither RNA is processed into piRNAs. As shown in the scatterplots in Figure 3C, the *rhi* mutation essentially eliminates UAP56venus binding to all cluster transcripts but has a relatively modest effect on cluster transcript binding to Thoc5GFP (Figure 3C, $rhi^{2/KG}$). In contrast, the *rhi* mutation does not affect Thoc5GFP binding to pre-mRNAs (Figures 3D and S3B, $rhi^{2/KG}$). Rhi is therefore required for assembly of piRNA precursor complexes containing cluster transcripts, UAP56, and THO.

To determine if UAP56 and THO subunits are interdependent for binding to cluster transcripts, we performed Thoc5GFP RIP-seq from *uap56* and *thoc7* mutant ovaries and UAP56venus RIP-seq from *thoc7* mutant ovaries (Figure 3C). The *thoc7* mutation completely abrogates cluster transcript binding to Thoc5GFP and significantlyreduces cluster transcript binding to UAP56venus (Figures 3C and S3A, $thoc7^{d/Df}$). The *thoc7* mutation also blocks Tho5GFP binding to unspliced introns (Figures 3D and S3B, $thoc7^{d/Df}$). In *uap56* mutant ovaries, in contrast, cluster transcript and intron binding to Thoc5GFP show only modest reductions (Figures 3C, 3D, and S3, $uap56^{28/sz15}$). Stable THO subunit binding to RNA thus requires an intact THO complex but does not require wild-type interactions with UAP56.

UAP56andTHORestrictRhiatpiRNAClusterChromatin

The studies presented here, with extensive works from a number of laboratories, indicate that Rhi has a central role in producing piRNA precursors (Andersen et al., 2017; Hur et al., 2016; Klattenhoff et al., 2009; Mohn et al., 2014; Zhang et al., 2012a, 2014). The chromo domain of Rhi binds to H3K9me3, and this mark is present at clusters but is also broadly distributed over heterochromatin. Rhi localization to H3K9me3 marks on clusters is therefore critical to the specificity of piRNA biogenesis, but how this is achieved is not understood.

Rhi forms distinct nuclear foci in *thoc5, thoc7*, and *uap56* mutant ovaries (Figure S1) (Hur et al., 2016; Zhang et al., 2012a), but cytological localization is not sufficient to determine if these Rhi foci correspond to clusters. We therefore used chromatin immuno-precipitation sequencing (ChIP-seq) to directly determine Rhi localization in *thoc7, thoc5*, and *uap56* mutants and in w^1 controls. Because Rhi binds to H3K9me3, we also profiled this histone modification in all four genotypes. These studies revealed a striking genome-wide redistribution of Rhi in all three mutants (Figure 4). Figure 4A shows Rhi ChIPseq signal over the right arm of *Drosophila* chromosome 2. In wild-type ovaries, Rhi is highly enriched at the major germline piRNA cluster *42AB* (Figure 4A, gray dashed box). Note that H3K9me3 is present at this cluster but is also widely distributed over pericentromeric regions that do not have Rhi signal (Figure 4A, compare w^1 red track with blue track). In the three TREX component mutants, in contrast, Rhi shows reduced binding to *42AB* and ectopic localization to both pericentromeric heterochromatin and euchromatin (Figure 4A, compare the red tracks in the mutants with the w^1 control). Figure 4B shows an enlarged view of the ChIP-seq profiles at *42AB*. Rhi is shown for *w1* and all three mutants, with H3K9me3 distribution in wild-type. As shown in Figure S4B, the TREX mutants do not significantly alter H3K9me3 at this cluster. We independently confirmed the reduction in

Rhi binding to *42AB* by ChIP-qPCR (Figure 4D). THO and UAP56 thus restrict Rhi to germline piRNA clusters.

In wild-type ovaries, Rhi overlaps with H3K9me3 marks at piRNA clusters (Figures 4A and 4B, Figure S4B for H3K9me3 signal in the mutants) (Klattenhoff et al., 2009; Mohn et al., 2014; Parhad et al., 2017). In *thoc7, thoc5*, and *uap56* mutants, ectopic Rhi peaks also correspond to H3K9me3 domains. In wild-type, these domains are frequently marked by H3K9me3 but do not bind Rhi. Figure 4C shows an example of this class of ectopic locus, which is marked by H3K9me3 in all genotypes (Figure S4B for H3K9me3 signal in the mutants). We independently confirmed Rhi binding at this site by ChIP-qPCR (Figure 4D). To extend this analysis genome wide, we used MACS2 (Zhang et al., 2008) to computationally define high-confidence Rhi domains from two independent biological replicates from each genotype (STAR Methods). The genomic positions of the high-confidence Rhi domains are shown in the “Rhino domain” track in the genome browser views in Figures 4 and S4A. In wild-type ovaries, approximately two-thirds of Rhi peaks map to piRNA clusters (Figure 4E, orange pie chart). Each of the mutants leads to an increase in Rhi peaks mapping outside piRNA clusters (Figure 4E, orange pie charts). Intriguingly, these mutants also increased the number of Rhi peaks mapping to clusters, which reflects increased Rhi binding to a number of minor clusters that normally show only weak Rhi accumulation (Figure S4A), coupled with a general decline in peak intensity at major clusters.

We then used MACS2 (Zhang et al., 2008) to call H3K9me3 domains in each genotype and determined the overlap between Rhi and H3K9me3 domains (Figure 4E, blue pie charts). Strikingly, 92%–94% of Rhi domains overlap with H3K9me3 marks in wild-type and the TREX component mutants. The *thoc7, thoc5*, and *uap56* mutations thus trigger Rhi binding to H3K9me3 domains outside piRNA clusters. In the example shown in Figure 4C, these mutations lead to Rhi binding to a site that is marked by H3K9me3 in all genotypes (H3K9me3 profiles in mutants shown in Figure S4B). However, each of the TREX mutations produces a significant number of new H3K9me3 domains, and Rhi binds to some of these genotypespecific H3K9me3 sites (Table S4). However, 68%–80% of Rhi peaks in the mutants correspond to H3K9me3 domains that are present in control ovaries. Stable TREX is therefore required to restrict Rhi to cluster chromatin and suppresses H3K9me3 modification of other chromatin domains.

Rhi promotes cluster transcription and piRNA biogenesis (Andersen et al., 2017; Klattenhoff et al., 2009; Mohn et al., 2014; Zhang et al., 2014), raising the possibility that ectopic binding may enhance transcription and trigger piRNA production from ectopic loci. We observed a modest increase in steady-state transcript accumulation at ectopic Rhi loci in the TREX mutants (Figure S4C). However, this is not linked to enhanced piRNA production (Figure S4D). These findings suggest that Rhi, presumably acting through the RDC, triggers transcription at ectopic sites (Andersen et al., 2017). However, in the absence of the TREX, Rhi binding cannot induce piRNA production.

DISCUSSION

In *Drosophila* ovaries, primary piRNAs are derived from heterochromatic clusters composed of nested transposon fragments (Brennecke et al., 2007). These loci serve as an archive of invading transposon sequences, and transposition of an invading element into a cluster is proposed to trigger adaptation, as the inserted sequences are incorporated into cluster transcripts and processed into mature piRNAs (Bergman et al., 2006; Brennecke et al., 2007; Muerdter et al., 2012). Clusters are therefore proposed to determine piRNA pathway specificity and adaptability. However, the mechanisms that specify cluster location and differentiate cluster transcripts from gene transcripts are not understood. We present evidence that cluster chromatin promotes assembly of pre-piRNA complexes defined by stable UAP56 and THO binding and that assembly of these complexes restricts Rhi to clusters chromatin.

UAP56-THO Interactions Are Critical to piRNA Biogenesis

UAP56 and the THO complex are conserved RNA splicing and export factors (Reed and Cheng, 2005), and null alleles of *uap56* and the core THO subunit *tho2* are lethal (Eberl et al., 1997; Gatfield et al., 2001; Jagut et al., 2013). However, the $uap56^{sz15}$ allele and mutations in *thoc5* and *thoc7* are viable but sterile, and disrupt piRNA biogenesis (Hur et al., 2016; Zhang et al., 2012a). We show that the $uap56^{sz15}$ mutation reduces UAP56 binding to THO and that the *thoc5* and *thoc7* mutations lead to dissociation of the remaining THO subunits and block interactions between UAP56 and the remaining subunits. High-affinity interactions between UAP56 and the THO complex are therefore required for piRNA biogenesis but dispensable for viability. Intriguingly, Hpr1 binding to Tho2 is retained in viable *thoc5* and *thoc7* mutants, and Hpr1 and Tho2 are the only THO subunits conserved from yeast to humans (Reed and Cheng, 2005). These observations suggest that an Hpr1-Tho2 heterodimer is sufficient to support basic cellular functions and zygotic development, but this remains to be rigorously tested.

Previous studies indicated that germline piRNA cluster transcripts co-precipitate with UAP56 and the THO subunit Thoc5 (Hur et al., 2016; Zhang et al., 2012a). Here we show that germline cluster transcripts are stably bound by UAP56 and THO, a significant fraction of unspliced pre-mRNAs are stably bound by THO but show only weak binding to UAP56, and mature mRNAs and somatic piRNA cluster transcripts are not enriched with UAP56venus, Hpr1, or Thoc5GFP. Significantly, *rhi* mutations block stable UAP56 binding to germline cluster transcripts but do not prevent THO binding to cluster transcripts or premRNAs. Rhi is therefore required for stable binding of UAP56 to cluster transcripts, generating complexes that are specific to germline piRNA precursors.

How does Rhi promote assembly of these RNA-protein complexes? Rhi interacts with the linker protein Del and the DXO homolog Cuff, forming the RDC complex, which promotes cluster transcription and suppresses cluster transcript splicing and polyadenylation (Chen et al., 2016; Mohn et al., 2014; Parhad et al., 2017; Zhang et al., 2014). A number of observations suggest that Rhi, functioning through Cuff, may indirectly promote UAP56 binding to cluster transcripts. Recent studies indicate that Rhi recruits transcription initiation factors that drive RNA Pol II transcription from both genomic strands (Andersen et al.,

2017). The resulting “non-canonical” transcripts are capped, and cap binding by the nuclear cap binding complex (CBC) promotes efficient splicing and polyadenylation (Proudfoot et al., 2002). However, cluster transcripts are not spliced or polyadenylated (Chen et al., 2016; Zhang et al., 2014). Cuff is a homolog of the decapping enzyme DXO, but the catalytic pocket is not conserved, and residues that interact with the RNA backbone are retained (Zhang et al., 2014). These observations suggest that Rhi localizes Cuff to clusters, where it binds nascent capped transcripts, blocking access to the CBC. This may indirectly suppress splicing and polyadenylation, generating complexes with stably bound THO and UAP56. By contrast, UAP56 appears to only transiently associate with pre-mRNAs, reflected in the weak enrichment observed in our native RIP-seq assays. Supporting this model, *cuff* mutations that block piRNA production also lead to a pronounced increase in cluster transcript splicing (Zhang et al., 2014).

Rhi binds to H3K9me3 *in vitro* and co-localizes with H3K9me3 at piRNA clusters (Le Thomas et al., 2014; Mohn et al., 2014; Yu et al., 2015). However, clusters represent only a small fraction of the chromatin marked by H3K9me3 (Mohn et al., 2014), and it is unclear how Rhi is restricted to these specialized domains. Here we show that mutations that disrupt the TREX significantly reduce Rhi association with major piRNA clusters and trigger ectopic Rhi localization to euchromatic and heterochromatic sites that are marked by H3K9me3. Stable interactions between UAP56 and an intact THO are therefore required to restrict Rhi to H3K9me3 marks on piRNA clusters.

A Feedforward Model for Cluster Chromatin Assembly

Knock down of *cuff* or *del essentially* eliminates nuclear Rhi foci, and ChIP-seq indicates that Rhi does not bind to clusters or ectopic sites (Mohn et al., 2014). In contrast, Cuff and Rhi still form foci in TREX mutants (Hur et al., 2016; Zhang et al., 2012a), and our Rhi ChIP-seq studies indicate that many of these foci do not correspond to clusters. On the basis of these observations, and the data presented here, we propose that Rhi, within the RDC, samples H3K9me3 marks throughout the genome. However, most of these domains are transcriptionally silent, and RDC binding is unstable (Figure 4F, step 1). piRNA clusters, in contrast, are transcribed and marked by H3K9me3. At these sites, the RDC binds to H3K9me3 through Rhi, and we speculate that capped nascent transcripts are bound through Cuff, which blocks CBC binding, suppressing splicing and polyadenylation (Chen et al., 2016; Zhang et al., 2014) and triggering stable UAP56 binding (Figure 4F, steps 2 and 3). Assembly into these higher order pre-piRNA complexes prevents RDC exchange with the soluble pool (Figure 4F, step 4). Through Del and Moonshiner, the RDC also triggers non-canonical transcription from both genomic strands (Figure 4F, step 5) (Andersen et al., 2017), enhancing RDC binding (Figure 4F, step 6). We note that some cluster transcription persists in rhi mutants, often from one genomic strand, and several major piRNA clusters have flanking canonical RNA Pol II promoters. We speculate that RDC binding to these transcripts initiates this feedforward system, which drives Rhi binding to cluster heterochromatin and promotes piRNA precursor complex assembly.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rat anti-Tho2	Rehwinkel et al., 2004	N/A
Rat anti-Hpr1	Rehwinkel et al., 2004	N/A
Rabbit anti-Thoc7	Kim et al., 2011	N/A
Rabbit anti-UAP56	Eberl et al., 1997	RRID: AB_2567529
Guinea pig anti-Rhi	Klattenhoff et al., 2009	RRID: AB_2568331
Rabbit anti H3K9me3	Abcam	Cat# ab8898
Monoclonal Mouse Anti-α-Tubulin	Sigma Aldrich	Cat# T5168
Rabbit anti-GFP	ThermoFisher Scientific	Cat# A11122
Alexa Fluor 563 Goat anti Guinea pig IgG	ThermoFisher Scientific	Cat# A11075
Alexa Fluor 633 Goat anti Guinea pig IgG	ThermoFisher Scientific	Cat# A21105
Alexa Fluor 488 Goat anti Rat IgG	ThermoFisher Scientific	Cat# A11006
Alexa Fluor 488 Goat anti Rabbit IgG	ThermoFisher Scientific	Cat# R37116
IRDye 800CW Goat anti-Rabbit IgG	LI-COR	Cat# 926–32211
IRDye 680 Goat anti-Rabbit IgG	LI-COR	Cat# 926–32221
IRDye 800CW Goat anti-Rat IgG	LI-COR	Cat# 926–32219
IRDye 800CW Goat anti-Mouse IgG	LI-COR	Cat# 926–32210
Chemicals, Peptides, and Recombinant Proteins		
Superscript III	ThermoFisher Scientific	Cat# 18080–085
RNase OUT	ThermoFisher Scientific	Cat# 10777–019
TURBO DNase	ThermoFisher Scientific	Cat# AM2238
RNaseH	ThermoFisher Scientific	Cat# 18021–071
T4 RNA Ligase	ThermoFisher Scientific	AM2141
AccuPrime Pfx DNA Polymerase	ThermoFisher Scientific	Cat# 12344024
UltraPure Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v)	ThermoFisher Scientific	Cat# 15593031
dNTP Set (100 mM)	ThermoFisher Scientific	Cat# 10297018
dUTP Solution (100 mM)	ThermoFisher Scientific	Cat# R0133
Hybridase	Lucigen	Cat# H39500
dNTP mix	NEB	Cat# N0447L
DNA polymerase I	NEB	Cat# M0209S
T4 DNA polymerase	NEB	Cat# M0203L
Klenow DNA polymerase	NEB	Cat# M0210S
T4 PNK	NEB	Cat# M0201L

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Klenow 3' to 5' exo	NEB	Cat# M0212L
UDG	NEB	Cat# M0280S
Phusion Polymerase	NEB	Cat# M0530S
T4 RNA Ligase 2, truncated	NEB	Cat# M0242L
50% PEG8000	NEB	Cat# B1004S
T4 DNA ligase	Enzymatics	Cat# L6030-HC-L
16% formaldehyde	Ted Pella	Cat# 18505
Critical Commercial Assays		
mirVANA miRNA isolation kit	ThermoFisher Scientific	Cat# AM1560
RNeasy Mini Kit	QIAGEN	Cat# 74104
QuantiTect SYBR Green PCR Kits	QIAGEN	Cat# 204145
Dynabeads Protein G	ThermoFisher Scientific	Cat# 10004D
Dynabeads Protein A	ThermoFisher Scientific	Cat# 10001D
GFP-Trap_A beads	Chromotek	Cat# gta-100
ANTI-FLAG M2 Agarose Affinity Gel	Sigma Aldrich	Cat# A2220
RNA Clean & Concentrator-5	Zymo Research	Cat# R1015
Agencourt AMPure XP	Beckman Coulter	Cat# A63880
Deposited Data		
Scaffold files for Mass Spec results	This study	https://doi.org/10.17632/j4yktssk9g.1
High throughput Sequencing	This study	SRP151054
Experimental Models: Organisms/Strains		
D. melanogaster/uap56^{28}	Zhang et al., 2012a	N/A
D. melanogaster/uap56^{sz15}	Zhang et al., 2012a	N/A
D. melanogaster/thoc7^{nl05792}	Harvard Exelixis stock collection	Stock# d05792
D. melanogaster/Df(3L)BSC128	Bloomington *Drosophila* Stock Center	Stock# 9293
D. melanogaster/thoc5^{e00906}	Harvard Exelixis stock collection	Stock# e00906
D. melanogaster/thoc5^{1}	Moon et al., 2011	N/A
D. melanogaster/rhino^{2}	Klattenhoff et al., 2009	N/A
D. melanogaster/rhino^{KG}	Klattenhoff et al., 2009	N/A
D. melanogaster/uap56 promoter > UAP56Venus	Zhang et al., 2012a	N/A
D. melanogaster/uap56 promoter > UAP56^{sz15}Venus	Zhang et al., 2012a	N/A
D. melanogaster/thoc5 promoter > Thoc5-GFP	Moon et al., 2011	N/A
D. melanogaster/vasa promoter > GFP-nls	Zhang et al., 2014	N/A
D. melanogaster/nanos promoter > Gal4	William Theurkauf lab	N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
D. melanogaster/w^{1}	William Theurkauf lab	N/A
Oligonucleotides		
Random primers	ThermoFisher Scientific	Cat# 48190011
Primers for ChIP-qPCR, see Table S5	Klattenhoff et al. and This study	N/A
Software and Algorithms		
Prism 7	GraphPad Prism	https://www.graphpad.com/
Image Studio Lite	LI-COR	https://www.licor.com/bio/products/software/image_studio_lite/
RStudio	N/A	https://www.rstudio.com/
ImageJ	N/A	https://imagej.nih.gov/ij/
Scaffold	Proteome Software	http://www.proteomesoftware.com/products/scaffold/
UCSC Genome Browser	Kent et al., 2002	https://genome.ucsc.edu/cgi-bin/hgGateway
Bowtie	Langmead et al., 2009	N/A
BEDTools	Quinlan and Hall, 2010	N/A
TopHat	Trapnell et al., 2009	N/A
BWA	Li and Durbin, 2009	N/A
MACS2	Zhang et al., 2008	N/A
Picard Tools	N/A	http://broadinstitute.github.io/picard/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact William Theurkauf (william.theurkauf@umassmed.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All flies were raised at 25°C on cornmeal medium. All experiments were performed on ovaries from 2–4 days old female *Drosophila* melanogaster raised in the presence of yeast paste. Ovaries were dissected at room temperature in Robb's buffer (100 mM HEPES ph 7.4, 100 mM sucrose, 55 mM Potassium Acetate, 40 mM Sodium Acetate, 10 mM glucose, 1.2 mM $MgCl_2$, and 1 mM $CaCl_2$).

METHOD DETAILS

Immunofluorescent staining and image analysis

Fixation and immuno-staining of *Drosophila* ovaries was performed with Buffer A protocol as described previously (Theurkauf, 1994). The images were acquired with Leica TCS SP8 confocal microscope. The line scan analyses were done in ImageJ (Schneider et al., 2012).

Immuno-precipitation, western blotting, proteomics, and RNA IP

Immuno-precipitation (IP) from ovary lysate was performed as described previously (Parhad et al., 2017). For GFP or Venus tagged proteins IP, GFP-Trap A agarose beads (ChromoTek)

and ANTI-FLAG M2 agarose beads (Sigma-Aldrich) were used, respectively. For antibody based IP, the antibody was first conjugated to Magnetic Dynabeads protein A/G (Invitrogen) in Citric phosphate buffer (7.10 g Na_2HPO_4, 11.5g Citric acid in 1 l water, ph 5.6) for 2 hours at room temperature with rotation. For RNA isolation, beads were resuspended in RTL buffer and the supernatant was processed using the RNeasy mini kit (QIAGEN). The proteomics samples were prepared and processed as described previously (Parhad et al., 2017). The western blots were detected by LI-COR Odyssey Infrared Imaging System. The fluorescent band intensities were quantified using Image Studio Lite (LI-COR).

High-throughput sequencing

Strand specific RNA-seq libraries were constructed as described previously (Zhang et al., 2012b) with modification in the rRNA depletion procedure using enzymatic digestion of rRNA by Hybridase™ Thermostable RNase H (Epicenter) with a comprehensive mixture of antisense rRNA oligos (Fu et al., 2018). The small RNaseq library is constructed as detailed previously (Li et al., 2009) with 2S rRNA depletion as described in (Zhang et al., 2011). The ChIPseq libraries were prepared as described previously (Zhang et al., 2014), with ovaries from 120 females per ChIP. RNaseq and ChIPseq libraries were paired-end sequenced, and small RNaseq libraries were single-end sequenced on the Nextseq 500 platform (Illumina).

ChIP-qPCR

The qPCR was performed using the QuantiTect SYBR Green PCR Kit (QIAGEN) in Step ONE plus real time PCR system (Applied Biosystem). PCR primer sequences were described previously (Klattenhoff et al., 2009; Zhang et al., 2012a) and presented in Table S5. Raw fluorescent amplification intensity from ChIP qPCR was used in DART-qPCR to estimate relative abundance (Peirson et al., 2003). Fold enrichment for precipitates was relative to input, and statistics and graphing was performed using Prism 7 (GraphPad).

Bioinformatics Analysis

The bioinformatics analysis was performed as described previously (Li et al., 2009; Parhad et al., 2017; Zhang et al., 2014). RNA-seq: The raw RNaseq reads were mapped to *Drosophila* genome (dm3) using TopHat 2.0.8 (Trapnell et al., 2009) with default parameters. The gene annotation was obtained from Flybase r5.50. The piRNA clusters annotation was described previously (Brennecke et al., 2007). The transcript abundance (rpkm: Reads Per Kilobase per Million mapped reads) was counted by BEDTools (Quinlan and Hall, 2010) and normalized to total number of mapped reads, after excluding rRNA mapping reads. The transposon family expression was calculated by proportionally collecting the reads that were aligned to the genomic annotation of transposons. We did differentially expression analysis of transposons and protein-coding genes with edgeR (Robinson et al., 2010) between *thoc7*$^{d/Df}$ and wild-type from two biological replicates. The replicated RNaseq from wild-type were published previously and deposited in NCBI Sequence Read Archive: SRP111075 (Parhad et al., 2017). The raw small RNaseq reads were mapped to *Drosophila* genome (dm3) using bowtie (Langmead et al., 2009) by allowing no mismatches after removing the 3^0 end linker sequence. We counted unique mapped reads for piRNA clusters. The read counts for piRNA clusters were normalized to the number of reads mapping to miRNA.

The raw ChIPseq reads were mapped to *Drosophila* genome (dm3) using by BWA 0.6.2 (Li and Durbin, 2009). The duplicated reads were discarded as PCR duplication by Picard tools (http://broadinstitute.github.io/picard). We called Rhi domains in each genotype from uniquely mapped ChIPseq reads using MACS2 (Zhang et al., 2008) with following parameters: macs2 callpeak -q 0.01 –ratio 2. For each replicate, Rhi domains within one kilo base pairs were merged. The Rhi domains present in both biological replicates were kept as high confidence Rhi domains for further analysis. Rhi domains mapping to mitochondrial genome were discarded. To quantify the long RNA and small RNA abundance in the Rhi domains across genotypes, the Rhi domain from each genotype are merged and assigned back to each genotype, which is used to quantify the long RNA and small RNA abundance.

QUANTIFICATION AND STATISTICAL ANALYSIS

The statistical method and sample size are reported in the Figures and corresponding legends. The statistical tests for quantitative western and ChIP-qPCR were t test for at least three biological replicates. The statistical tests for deep sequencing data were Wilcoxon Rank-Sum test. The minimal *p*-value reported was 2.2e-16. We used R and Prism 7 (GraphPad) to do the statistical test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We would like to thank the members of the Theurkauf and Weng labs for their insightful discussions and comments throughout the project; Elisa Izaurralde (Max Planck Institute) for anti-Hpr1 and anti-Tho2 antibodies; Yun Doo Chung (University of Seoul) for anti-Thoc7 antibody, *thoc5*^{1}, and Thoc5GFP transgenic stocks; the Bloomington and Harvard Exelixis stock centers for fly strains; and John Leszyk and the Mass Spectrometry Facility of the University of Massachusetts Medical School (UMass Med) for the proteomics service. This work was supported by NIH grants R01 HD049116 and P01 HD078253 to W.E.T.

REFERENCES

Andersen PR, Tirian L, Vunjak M, and Brennecke J (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. Nature 549, 54–59. [PubMed: 28847004]

Aravin AA, Hannon GJ, and Brennecke J (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318, 761–764. [PubMed: 17975059]

Belancio VP, Hedges DJ, and Deininger P (2008). Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res. 18, 343–358. [PubMed: 18256243]

Bergman CM, Quesneville H, Anxolabe´ he` re D, and Ashburner M (2006). Recurrent insertion and duplication generate networks of transposable element sequences in the Drosophila melanogaster genome. Genome Biol. 7, R112. [PubMed: 17134480]

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, and Hannon GJ (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128, 1089–1103. [PubMed: 17346786]

Chen YA, Stuwe E, Luo Y, Ninova M, Le Thomas A, Rozhavskaya E, Li S, Vempati S, Laver JD, Patel DJ, et al. (2016). Cutoff suppresses RNA polymerase II termination to ensure expression of piRNA precursors. Mol. Cell 63, 97–109. [PubMed: 27292797]

Chi B, Wang Q, Wu G, Tan M, Wang L, Shi M, Chang X, and 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. [PubMed: 23222130]

Eberl DF, Lorenz LJ, Melnick MB, Sood V, Lasko P, and Perrimon N (1997). A new enhancer of position-effect variegation in Drosophila melanogaster encodes a putative RNA helicase that binds chromosomes and is regulated by the cell cycle. Genetics 146, 951–963. [PubMed: 9215899]

Fu Y, Wu P-H, Beane T, Zamore PD, and Weng Z (2018). Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. BMC Genomics 19, 531. [PubMed: 30001700]

Gatfield D, Le Hir H, Schmitt C, Braun IC, Ko¨ cher T, Wilm M, and Izaurralde E (2001). The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in Drosophila. Curr. Biol 11, 1716–1721. [PubMed: 11696332]

Ghildiyal M, and Zamore PD (2009). Small silencing RNAs: an expanding universe. Nat. Rev. Genet 10, 94–108. [PubMed: 19148191]

Girard A, Sachidanandam R, Hannon GJ, and Carmell MA (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature 442, 199–202. [PubMed: 16751776]

Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, and Siomi MC (2007). A slicer-mediated mechanism for repeat-associated siRNA 50 end formation in Drosophila. Science 315, 1587–1590. [PubMed: 17322028]

Hedges DJ, and Deininger PL (2007). Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. Mutat. Res. 616, 46–59. [PubMed: 17157332]

Hur JK, Luo Y, Moon S, Ninova M, Marinov GK, Chung YD, and Aravin AA (2016). Splicing-independent loading of TREX on nascent RNA is required for efficient expression of dual-strand piRNA clusters in Drosophila. Genes Dev. 30, 840–855. [PubMed: 27036967]

Iwasaki YW, Siomi MC, and Siomi H (2015). PIWI-interacting RNA: its biogenesis and functions. Annu. Rev. Biochem 84, 405–433. [PubMed: 25747396]

Jagut M, Mihaila-Bodart L, Molla-Herman A, Alin MF, Lepesant JA, and Huynh JR (2013). A mosaic genetic screen for genes involved in the early steps of Drosophila oogenesis. G3 (Bethesda) 3, 409–425. [PubMed: 23450845]

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, and Haussler D (2002). The human genome browser at UCSC. Genome Res. 12, 996–1006. [PubMed: 12045153]

Khurana JS, Xu J, Weng Z, and Theurkauf WE (2010). Distinct functions for the Drosophila piRNA pathway in genome maintenance and telomere protection. PLoS Genet. 6, e1001246. [PubMed: 21179579]

Kim VN, Han J, and Siomi MC (2009). Biogenesis of small RNAs in animals. Nat. Rev. Mol. Cell Biol 10, 126–139. [PubMed: 19165215]

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

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 dualstrand clusters. Cell 138, 1137–1149. [PubMed: 19732946]

Langmead B, Trapnell C, Pop M, and Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25. [PubMed: 19261174]

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, and Kingston RE (2006). Characterization of the piRNA complex from rat testes. Science 313, 363–367. [PubMed: 16778019]

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. [PubMed: 25085419]

Li H, and Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. [PubMed: 19451168]

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. [PubMed: 19395009]

Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, and Hannon GJ (2009). Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell 137, 522–535. [PubMed: 19395010]

McClintock B (1950). The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. U S A 36, 344–355. [PubMed: 15430309]

Meignin C, and Davis I (2008). UAP56 RNA helicase is required for axis specification and cytoplasmic mRNA localization in Drosophila. Dev. Biol 315, 89–98. [PubMed: 18237727]

Mohn F, Sienski G, Handler D, and Brennecke J (2014). The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. Cell 157, 1364–1379. [PubMed: 24906153]

Moon S, Cho B, Min SH, Lee D, and Chung YD (2011). The THO complex is required for nucleolar integrity in Drosophila spermatocytes. Development 138, 3835–3845. [PubMed: 21828100]

Muerdter F, Olovnikov I, Molaro A, Rozhkov NV, Czech B, Gordon A, Hannon GJ, and Aravin AA (2012). Production of artificial piRNAs in flies and mice. RNA 18, 42–52. [PubMed: 22096018]

Pane A, Jiang P, Zhao DY, Singh M, and Schu€pbach T (2011). The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline. EMBO J. 30, 4601–4615. [PubMed: 21952049]

Parhad SS, Tu S, Weng Z, and Theurkauf WE (2017). Adaptive evolution leads to cross-species incompatibility in the piRNA transposon silencing machinery. Dev. Cell 43, 60–70.e5. [PubMed: 28919205]

Peirson SN, Butler JN, and Foster RG (2003). Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. Nucleic Acids Res. 31, e73. [PubMed: 12853650]

Proudfoot NJ, Furger A, and Dye MJ (2002). Integrating mRNA processing with transcription. Cell 108, 501–512. [PubMed: 11909521]

Quinlan AR, and Hall IM (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. [PubMed: 20110278]

Reed R, and Cheng H (2005). TREX, SR proteins and export of mRNA. Curr. Opin. Cell Biol 17, 269–273. [PubMed: 15901496]

Rehwinkel J, Herold A, Gari K, Ko¨ cher T, Rode M, Ciccarelli FL, Wilm M, and Izaurralde E (2004). Genome-wide analysis of mRNAs regulated by the THO complex in Drosophila melanogaster. Nat. Struct. Mol. Biol 11, 558–566. [PubMed: 15133499]

Robinson MD, McCarthy DJ, and Smyth GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. [PubMed: 19910308]

Schneider CA, Rasband WS, and Eliceiri KW (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671–675. [PubMed: 22930834]

Schwanha¨ usser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, and Selbach M (2011). Global quantification of mammalian gene expression control. Nature 473, 337–342. [PubMed: 21593866]

Siomi MC, Miyoshi T, and Siomi H (2010). piRNA-mediated silencing in Drosophila germlines. Semin. Cell Dev. Biol 21, 754–759. [PubMed: 20080197]

Siomi MC, Sato K, Pezic D, and Aravin AA (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat. Rev. Mol. Cell Biol 12, 246–258. [PubMed: 21427766]

Theurkauf WE (1994). Immunofluorescence analysis of the cytoskeleton during oogenesis and early embryogenesis. Methods Cell Biol. 44, 489–505. [PubMed: 7707968]

Trapnell C, Pachter L, and Salzberg SL (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111. [PubMed: 19289445]

Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, and Zamore PD (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science 313, 320–324. [PubMed: 16809489]

Viphakone N, Hautbergue GM, Walsh M, Chang CT, Holland A, Folco EG, Reed R, and Wilson SA (2012). TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. Nat. Commun. 3, 1006. [PubMed: 22893130]

Yu B, Cassani M, Wang M, Liu M, Ma J, Li G, Zhang Z, and Huang Y (2015). Structural insights into Rhino-mediated germline piRNA cluster formation. Cell Res. 25, 525–528. [PubMed: 25613572]

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, and Liu XS (2008). Modelbased analysis of ChIP-seq (MACS). Genome Biol. 9, R137. [PubMed: 18798982]

Zhang Z, Xu J, Koppetsch BS, Wang J, Tipping C, Ma S, Weng Z, Theurkauf WE, and Zamore PD (2011). Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. Mol. Cell 44, 572–584. [PubMed: 22099305]

Zhang F, Wang J, Xu J, Zhang Z, Koppetsch BS, Schultz N, Vreven T, Meignin C, Davis I, Zamore PD, et al. (2012a). UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell 151, 871–884. [PubMed: 23141543]

Zhang Z, Theurkauf WE, Weng Z, and Zamore PD (2012b). Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. Silence 3, 9. [PubMed: 23273270]

Zhang Z, Wang J, Schultz N, Zhang F, Parhad SS, Tu S, Vreven T, Zamore PD, Weng Z, and Theurkauf WE (2014). The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. Cell 157, 1353–1363. [PubMed: 24906152]

Highlights

- UAP56 and THO complex binding distinguishes cluster transcripts from pre-mRNAs
- The HP1 homolog Rhino promotes assembly of piRNA precursor complexes
- Assembly of piRNA precursor complexes restricts Rhino to cluster heterochromatin

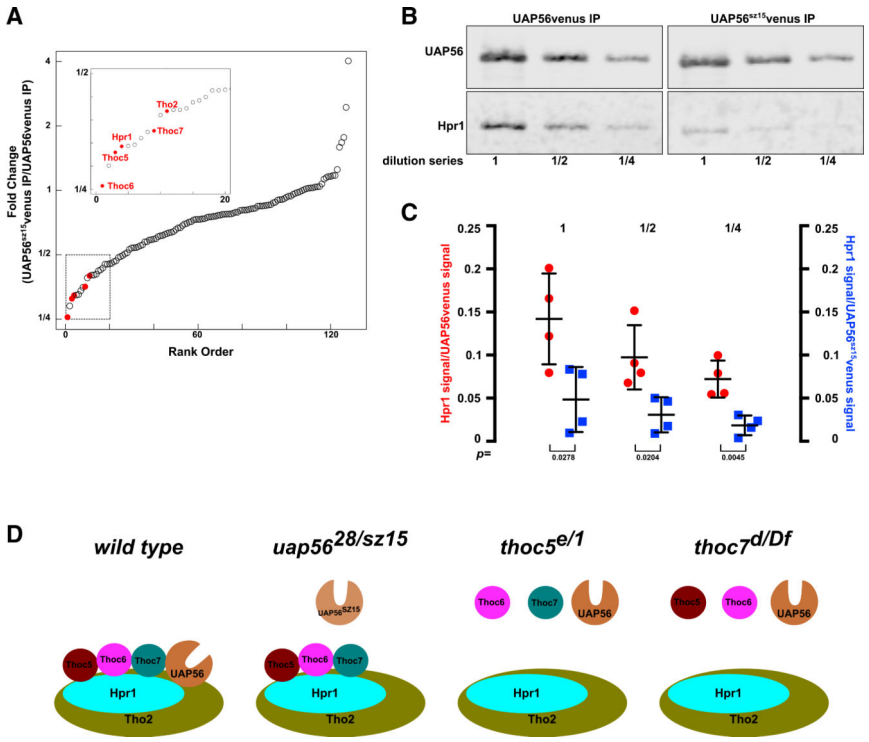


Figure 1. UAP56-THO Complex Interactions Are Required for piRNA Biogenesis

The *uap56*^{sz15} point mutation disrupts piRNA biogenesis. We used IP-mass spectrometry to identify proteins showing altered binding to UAP56^{sz15} protein relative to wild-type.

(A) Rank-order plot of fold change in protein abundance in UAP56^{sz15}venus precipitates relative to UAP56venus controls. Average fold change was calculated from three biological replicates. The inset shows the top 20 proteins with the highest reduction in binding to UAP56^{sz15}venus.THO subunits are labeled in red.

(B) Western blot for the THO complex protein Hpr1 in UAP56venus and UAP56^{sz15}venus precipitates from wild-type *Drosophila* ovary lysates. UAP56venus and UAP56^{sz15}venus were detected using rabbit anti-GFP, and Hrp1 was detected with ratanti-Hrp1. The blot was sequentially probed for Hpr1 and GFP, which were detected with distinct fluorescent secondary antibodies. A series of dilutions of each precipitated sample was analyzed.

(C) Plot showing Hpr1 signal normalized to the corresponding Venus tag from four biological replicates. The p values were calculated from t tests.

(D) Schematic representation of TREX complex integrity in wild-type and mutants (*uap56*$^{28/sz15}$, *thoc5*$^{e/1}$, and *thoc7*$^{d/Df}$), on the basis of IP-mass spectrometry.

See also Figure S1 and Tables S1–S3.

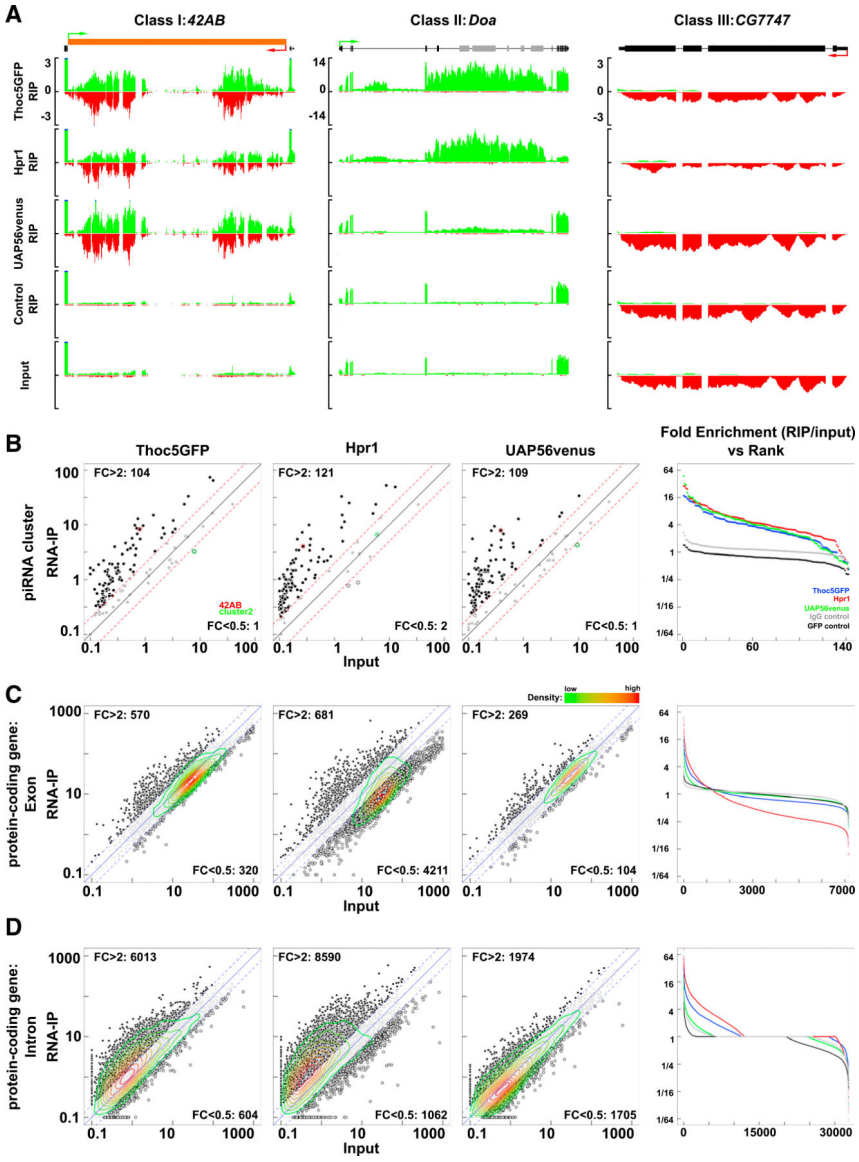


Figure 2. UAP56 and THO Binding Is Unique to piRNA Cluster Transcripts

(A) Genome browser views of Thoc5GFP, Hpr1,and UAP56venus RIP-seq signal, with IP and input controls, over the *42AB* piRNA cluster (class I), a large intron in the *Doa* gene (class II), and mature spliced RNA from the *CG7747* gene (class III). Germline cluster transcripts fall into class I and bind to UAP56 and THO complex subunits. A subset of introns define class II and bind strongly to THO but show weak binding to UAP56. Class III includes mature mRNAs and somatic piRNA cluster transcripts.

(B) Scatterplots showing normalized piRNA clustertranscripts abundance ($\log_{10}$[RPKM + 0.1]) in precipitates of Thoc5GFP, Hpr1, and UAP56venus, relative to input. The last panel on the right shows the rank-ordered fold enrichment (RIP/ input) for each cluster in the experimental precipitates and in GFP and non-specific antibody controls (as indicated in panel).

(C) Scatter and rank-order plot, as described in (B), for exon mapping transcript enrichment. Color-coded contour lines reflecting data point density are overlaid on each scatterplot.
(D) Scatter with contour lines and rank-order plots for intron mapping RNA-seq read RIP enrichment.
See also Figure S2.

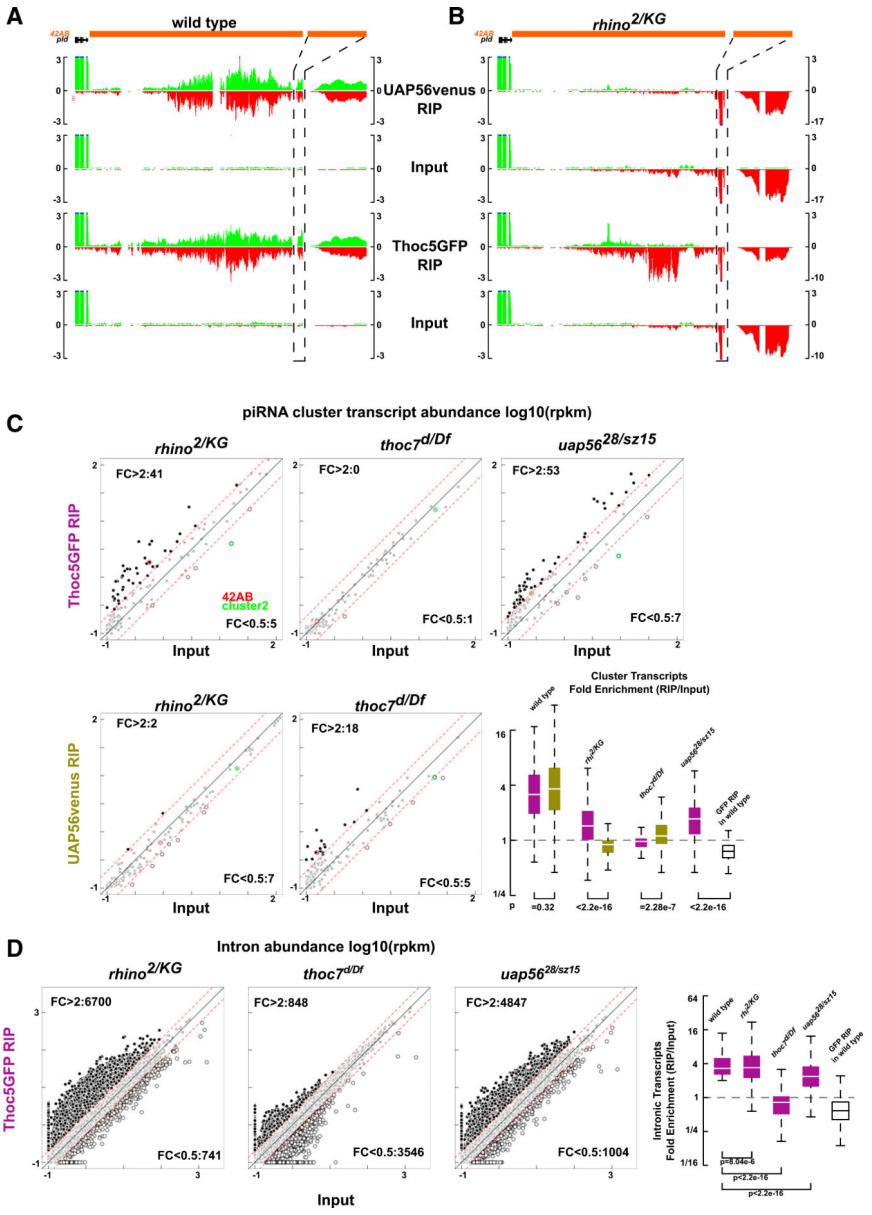


Figure 3. Rhi Promotes UAP56 Binding to Cluster Transcripts

(A and B) Genome browser view of the left end of the *42AB* piRNA cluster.

(A) RIP-seq signal for UAP56venus and Thoc5GFP from wild-type ovaries.

(B) RIP-seq signal for UAP56venus and Thoc5GFP from *rhi*$^{2/KG}$ ovaries.

The dashed boxes in (A) and (B) indicate a partial *gypsy12* element in *42AB* that is spliced in *rhi*$^{2/KG}$. This region is expanded on the right side of each panel. In wild-type, the unspliced transcripts from this region bind to UAP56venus and Thoc5GFP. In *rhi*$^{2/KG}$, unspliced transcripts bind to Thoc5GFP but not to UAP56venus, while the spliced transcripts do not bind either UAP56venus or Thoc5GFP.

(C) Scatterplots comparing cluster transcript abundance ($\log_{10}$[RPKM + 0.1]) in Thoc5GFP RIP (top row) and UAP56venus RIP (bottom row) relative to the corresponding inputs, from *rhi*$^{2/KG}$, *thoc7*$^{d/Df}$, and *uap56*$^{28/sz15}$ ovaries. The box plot summarizes cluster transcript fold

enrichment (RIP/input) in wild-type and mutants. The p values were calculated using Wilcoxon rank-sum tests. GFP RIP served as non-specific control.

(D) Scatterplots of intronic transcript abundance($\log_{10}$[RPKM + 0.1]) in Thoc5GFP RIP relative to input, from *rhi*$^{2/KG}$, *thoc7*$^{d/Df}$, and *uap56*$^{28/sz15}$ ovaries. The box plot summarizes the fold enrichment (RIP/input) in the mutants for the 14% of introns that are enriched by more than 2-fold in both Hpr1 and Thoc5GFP in wild-type (Figure 3). The p values were calculated using Wilcoxon rank-sum tests. GFP RIP served as non-specific control.

See also Figure S3.

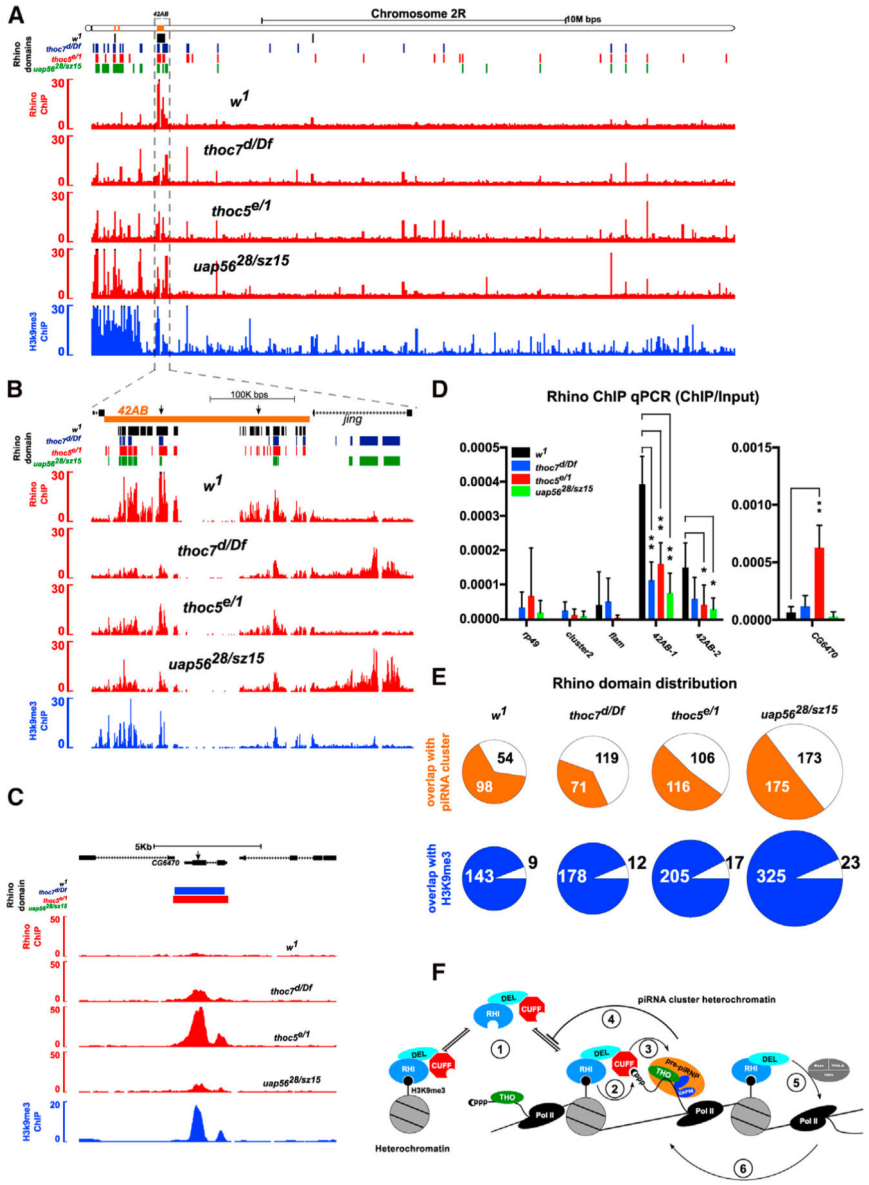


Figure 4. The TREX Complex Restricts Rhi to H3K9me3 Marks on piRNA Clusters

(A–C) Genome browser view of chromosome 2R (A), the piRNA cluster at *42AB* (B), and ectopic Rhi peaks at CG6470 (C). Red tracks are Rhi ChIPseq signal from w^{1}, $thoc7^{d/Df}$, $thoc5^{e/1}$, and $uap56^{28/sz15}$. The blue track is H3K9me3 ChIP-seq signal from w^{1}. The annotated piRNA clusters are highlighted in orange. The genomic positions of computationally defined Rhi ChIP-seq peaks for each genotype are color highlighted in the Rhi domain track. In wild-type, Rhi is largely confined to piRNA clusters. In $thoc7^{d/Df}$, $thoc5^{e/1}$, and $uap56^{28/sz15}$, cluster binding is reduced, and new peaks are present across the chromosome arm.

(D) Rhi ChIP-qPCR for 42AB and CG6470, with *rp49*, *cluster2*, and *flam as* controls. Arrowheads in the gene models in (B) and (C) indicate the location of qPCR primers. Significance of ChIP signal in the mutants relative to w^{1} was determined using t test from four biological replicates. *p < 0.05, **p < 0.001.

(E) Pie charts showing the number of Rhi domains that overlap with annotated piRNA clusters (orange charts) and Rhi domains that overlap with H3K9me3 domains in each genotype (blue charts).

(F) A speculative feedforward mechanism for piRNA cluster heterochromatin assembly. We propose that the RDC, through the Rhi chromo domain, samples H3K9me3 marks throughout the genome, but binding at transcriptional silent chromatin is unstable (1). In contrast, RDC binding to H3K9me3 marks at transcribed piRNA clusters is followed by Cuff association with capped cluster transcripts, which blocks cap binding by the cap binding complex, stalling splicing and stabilizing UAP56 and THO binding (pre-piRNP) (3). Within this chromatin-bound protein-RNA complex, the RDC does not exchange with the soluble pool, driving the complex to H3K9me3 marks on clusters (4). Deadlock then recruits transcription factors (TFIIA-S, Moonshiner, and TRF2) (Andersen et al., 2017) to trigger capped non-canonical transcription on both strands (5), which enhances prepiRNP assembly and RDC localization (6).

See also Figure S4 and Table S4.