

The HP1 Homolog Rhino Anchors a Nuclear Complex that Suppresses piRNA Precursor Splicing

Zhao Zhang,^{1,3,4} Jie Wang,^{2,6} Nadine Schultz,¹ Fan Zhang,¹ Swapnil S. Parhad,¹ Shikui Tu,² Thom Vreven,² Phillip D. Zamore,^{3,4,5} Zhiping Weng,^{2,*} and William E. Theurkauf^{1,*}

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

²Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, 368 Plantation Street, Worcester MA 01605, USA

³Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 364 Plantation Street, Worcester MA 01605, USA

⁴RNA Therapeutics Institute, University of Massachusetts Medical School, 368 Plantation Street, Worcester MA 01605, USA

⁵Howard Hughes Medical Institute

⁶Present address: Department of Biochemistry, University at Buffalo, NY State Center of Excellence in Bioinformatics and Life Sciences, 701 Ellicott Street, Buffalo, NY 14203, USA

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

<http://dx.doi.org/10.1016/j.cell.2014.04.030>

SUMMARY

piRNAs guide an adaptive genome defense system that silences transposons during germline development. The *Drosophila* HP1 homolog Rhino is required for germline piRNA production. We show that Rhino binds specifically to the heterochromatic clusters that produce piRNA precursors, and that binding directly correlates with piRNA production. Rhino colocalizes to germline nuclear foci with Rai1/DXO-related protein Cuff and the DEAD box protein UAP56, which are also required for germline piRNA production. RNA sequencing indicates that most cluster transcripts are not spliced and that *rhino*, *cuff*, and *uap56* mutations increase expression of spliced cluster transcripts over 100-fold. LacI::Rhino fusion protein binding suppresses splicing of a reporter transgene and is sufficient to trigger piRNA production from a trans combination of sense and antisense reporters. We therefore propose that Rhino anchors a nuclear complex that suppresses cluster transcript splicing and speculate that stalled splicing differentiates piRNA precursors from mRNAs.

INTRODUCTION

Transposons and other repetitive elements are major genome constituents that can mobilize and induce DNA breaks and insertional mutations (McClintock, 1950; Bennetzen, 2000; Beck et al., 2010). The germline transmits the inherited genetic complement, and PIWI-interacting RNAs (piRNAs) have a conserved role in suppressing transposon expression and maintaining genome integrity during germline development (Khurana and

Theurkauf, 2010; Siomi et al., 2011; Guzzardo et al., 2013). The 23–30 nt long piRNAs bear a 5' monophosphate, a 3' terminal 2'-O-methyl group, and bind to PIWI clade Argonautes (Aubergine, Piwi, and Ago-3 in *Drosophila*) (Grivna et al., 2006; Girard et al., 2006; Aravin et al., 2006; Vagin et al., 2006; Lau et al., 2006; Horwich et al., 2007; Saito et al., 2007). piRNAs bound to PIWI proteins can guide cleavage of complementary targets, which contributes to transposon silencing and generates the precursors of sense-strand piRNAs that direct cleavage of antisense precursors and drive a “ping-pong amplification” cycle (Brennecke et al., 2007; Gunawardane et al., 2007).

The primary piRNAs that initiate the ping-pong cycle are derived from clusters of nested transposon fragments, which generally reside in subtelomeric or pericentromeric heterochromatin (Brennecke et al., 2007). In *Drosophila*, the majority of these heterochromatic domains produce piRNAs from both genomic strands, and piRNAs mapping uniquely to these “dual-strand clusters” are the dominant species in germline cells. A distinct set of clusters produce unique piRNAs from only one genomic strand, and piRNAs from these “uni-strand clusters” dominate in the somatic follicle cells that surround the germline (Brennecke et al., 2007; Malone et al., 2009). It is unclear how the piRNAs precursors produced by dual-strand and uni-strand clusters are distinguished from gene transcripts.

The rapidly evolving heterochromatin protein 1 (HP1) homolog Rhino (Rhi) is required for production of primary piRNAs from the germline-specific dual-strand clusters (Klattenhoff et al., 2009). Here, we show that Rhi binds specifically to dual-strand clusters and that binding correlates with germline piRNA production. Significantly, we also present evidence that Rhi functions with the Rai1/DXO-related protein Cutoff (Cuff) and the DEAD box protein UAP56 to suppress cluster transcript splicing, and that Rhi binding suppresses splicing of a reporter transgene and is sufficient to drive piRNA biogenesis from transgenic reporters that express complementary transcripts. Stalled splicing intermediates

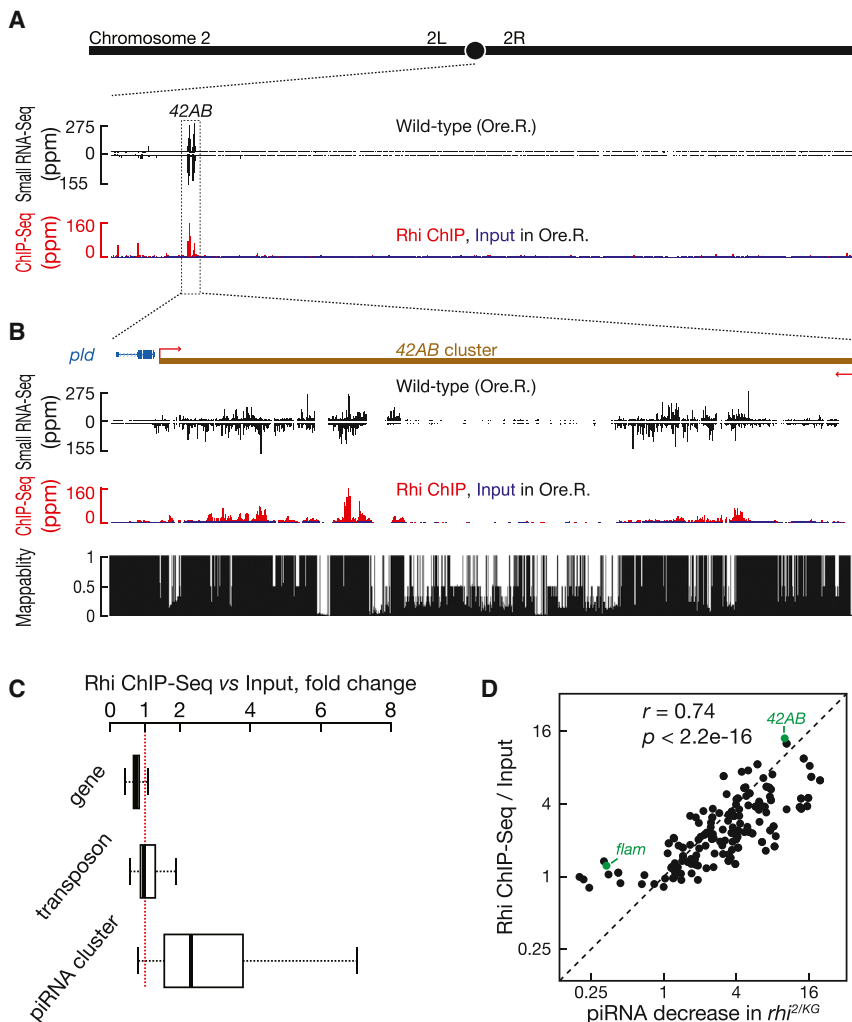


Figure 1. Rhi Binding Correlates with piRNA Production

(A) piRNA and Rhi ChIP-seq signal across the right arm of chromosome 2.

(B) piRNA and Rhi ChIP-seq signal across the 42AB piRNA cluster, which produces ~30% of fly ovary piRNAs.

(C) Boxplots showing Rhi enrichment at protein coding genes, transposons and piRNA clusters. Outliers are not shown.

(D) Rhi binding correlates with Rhi-dependent piRNA production. Scatter plots showing Rhi-binding enrichment (y axis) relative to reduction in piRNA production in *rhi*^{2/KG} mutants (x axis). Each point represents a distinct piRNA cluster. The major germline piRNA cluster at 42AB and the major somatic cluster (*flam*) are indicated in green. See also Figures S1 and S2.

are precursors for transposon-silencing siRNAs in the pathogenic yeast *Cryptococcus* (Dumesic et al., 2013). Stalled splicing may therefore have a conserved function in differentiating potentially deleterious RNAs from gene transcripts and produce the precursors of transsilencing small RNAs that guide host-defense systems.

RESULTS

Rhino Marks Dual-Strand piRNA Clusters

The HP1 family protein Rhino (Rhi), also referred to as HP1D, is required for transposon silencing and piRNA production from heterochromatic dual-strand clusters (Klattenhoff et al., 2009). In fission yeast, HP1 appears to bind centromeric transcripts and direct these RNAs to the degradation machinery (Keller et al., 2012). We therefore assayed for cluster transcript binding by Rhi, using coimmunoprecipitation and qPCR conditions that yield significant cluster transcript enrichment by UAP56 (Zhang et al., 2012a). We did not detect significant cluster transcript binding in these studies (F.Z. and W.T., unpublished data). By contrast, our earlier Rhi chromatin immunoprecipitation-qPCR

(ChIP-qPCR) experiments showed significant enrichment for two regions in the major 42AB piRNA cluster (Klattenhoff et al., 2009). While we cannot rule out RNA binding by Rhi, these observations suggest that this HP1 family protein interacts directly with chromatin. We therefore determined the genome-wide distribution of Rhi by ChIP-sequencing (ChIP-seq). These studies indicate that Rhi is enriched at essentially all of the germline-specific dual-strand clusters but shows only background binding to uni-strand clusters and protein coding genes (Figures 1 and S1A available online).

Visual inspection of genome browser profiles revealed a striking correlation

between unique Rhi ChIP-seq and unique piRNA profiles across several of the major germline clusters (Figures 1A, 1B, and S1A). We therefore compared Rhi binding to cluster chromatin, assayed by ChIP-seq, to fold reduction in piRNA production in *rhi* mutants, assayed by small RNA-seq. Eleven piRNA clusters, including the two major uni-strand clusters (Cluster 2 and *flamenco*), showed no decrease in piRNA expression in *rhi*^{2/KG} mutants and showed only background Rhi binding by ChIP-seq (Figures 1D and S1A). For the remaining 131 piRNA clusters, by contrast, Rhi binding was significantly correlated with fold reduction in piRNA expression in *rhi* mutants (Figure 1D, Pearson correlation coefficient $r = 0.74$; $p < 2.2 \times 10^{-16}$). ChIP-seq signal of a matched preimmune serum control did not correlate with piRNA expression ($r = 0.19$, $p = 0.03$; Figure S1B).

The correlation between Rhi binding and piRNA expression raised the possibility that piRNAs direct Rhi to chromatin, perhaps through a process analogous to siRNA guided centromeric heterochromatin assembly in *S. pombe* (Verdel et al., 2004; Iida et al., 2008; Grewal, 2010). We therefore performed Rhi ChIP-seq in ovaries mutant for *armi* (*armi*^{1/armi}^{72.1}), which

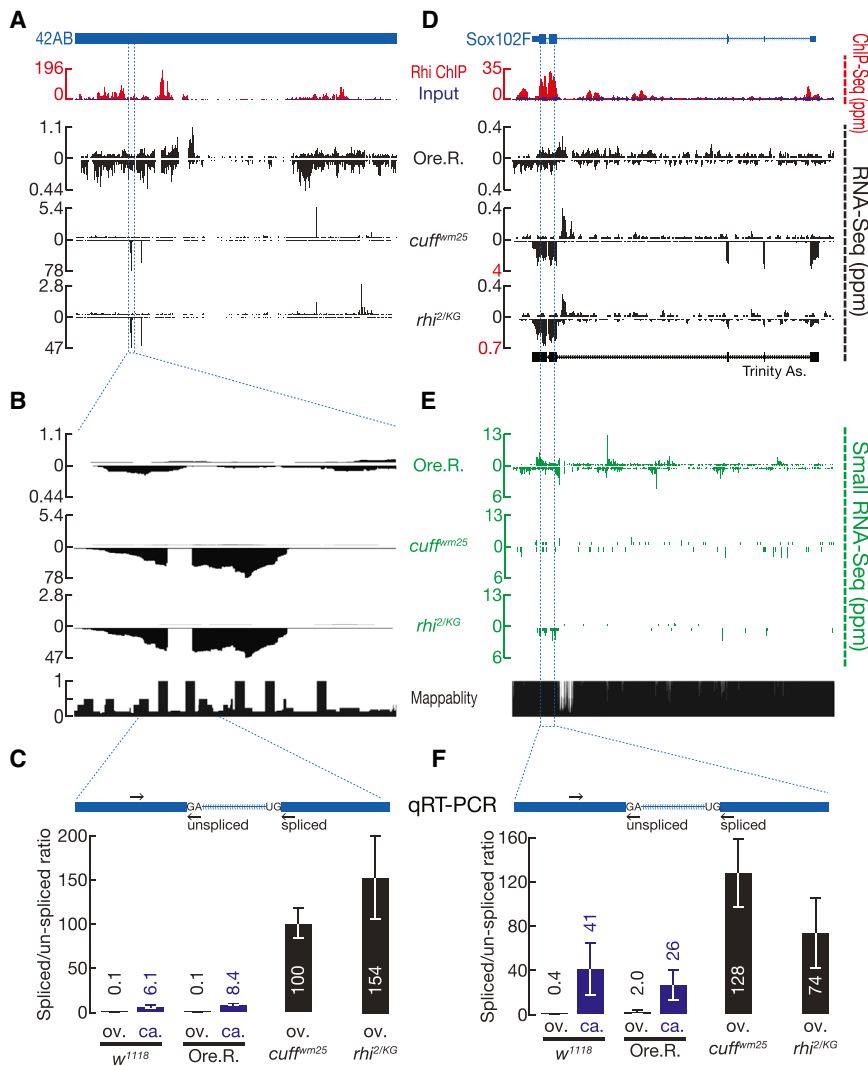


Figure 2. Rhi and Cuff Suppress Splicing at the 42AB and *sox102F* Clusters

(A and D) Rhi ChIP-seq and RNA-seq profiles for the 42AB cluster (A) and *sox102F* locus (D) are shown. For ChIP-seq, Rhi signal (red) is superimposed on input (blue). RNA-seq is shown for Oregon R (Ore. R.) control, *cuff*, and *rhi* mutants. In Ore. R. controls, signal is spread over both 42AB and *sox102F*. In *rhi* and *cuff* mutants, by contrast, signal shifts to distinct peaks (Note that data are scaled to avoid peak clipping). At the *sox102F* locus, the boundaries correspond to annotated splice sites in the mature somatic transcript, and de novo transcript assembly from these data yields the annotated gene structure (Trinity As.)

(B) A high-resolution expansion of the indicated region of 42AB shows that the peak is interrupted by a region with little signal, defined by very sharp boundaries characteristic of intron removal.

(C and F) qRT-PCR quantify the splicing efficiency at 42AB (C) and *sox102F* (F) loci. The diagrams show the putative introns (blue) and the position of primers (arrow) used to assay unspliced and spliced transcripts at 42AB (C) and *sox102F* (F). Both spliced and unspliced transcripts are amplified using the same forward primers. Reverse primers for unspliced transcripts span the splice site and for spliced transcripts span the mature junction. Bar graphs show the ratio of spliced to unspliced RNAs in ovary (ov.) and carcass (ca.) in two different control strains (*w¹¹¹⁸* and Ore. R.) and in *cuff* and *rhi* mutants. In ovaries, splicing at 42AB and *sox102F* increases over 80-fold in both *cuff* and *rhi* mutants. The *sox102F* locus is expressed in somatic tissue present in the carcass, and the transcripts are spliced. Data are mean \pm SD for three independent biological samples.

(E) piRNA production from *sox102F* locus in Ore.R., *cuff* and *rhi* mutants.

See also Figures S3, S4, and S5.

is required for production of piRNAs from most dual-strand piRNA clusters (Malone et al., 2009). Rhi binding in the *armi* mutants and matched controls were essentially identical ($r = 0.63\sim 0.84$, $p < 2.2 \times 10^{-16}$, Figure S2, shown for the 42AB cluster in Figure S2A), suggests that Rhi localization to clusters is independent of piRNA production. However, maternal piRNAs could localize Rhi to clusters during early development, and this distribution could be epigenetically propagated to the adult stage by a piRNA-independent mechanism. It is also possible that the low level of piRNAs expression in *armi* mutants is sufficient to localize Rhi. The mechanisms that localize Rhi and define cluster location thus remain to be determined.

Cluster Transcript Splicing

Rhi colocalizes to nuclear foci with the piRNA pathway proteins Cuff and UAP 56, mutations in *rhi* disrupt localization of both proteins, and mutations in *cuff* and *uap56* disrupt Rhi localization (Pane et al., 2011; Zhang et al., 2012a). The Rhi protein appears to directly associate with chromatin (Klattenhoff et al., 2009),

while Cuff and UAP56 are related to well-characterized RNA-binding proteins, and cluster transcripts coimmunoprecipitate with UAP56 (Pane et al., 2011; Zhang et al., 2012a). These three proteins may therefore have distinct molecular functions at a related step in piRNA precursor biogenesis.

qPCR assays for two regions of the 240 kb 42AB piRNA cluster showed reduce RNA expression in *rhi* mutants (Klattenhoff et al., 2009), suggesting that transcription or transcript stability may be decreased. To extend these findings, we used strand-specific paired-end RNA sequencing (RNA-seq) to profile the transcriptome in *rhi* mutants and in ovaries mutant for *cuff* and *uap56*. These studies did not show a consistent reduction in reads mapping to clusters in *rhi* mutants or in ovaries mutant for *cuff* or *uap56* (Figure S3). However, visual inspection of several prominent clusters showed that all three mutations produce similar shifts in RNA-seq signal to a few well-defined peaks (Figure 2A). The regions previously assayed by qPCR in *rhi* mutants fall between these peaks, explaining the apparent discrepancy with our early findings (Klattenhoff et al., 2009).

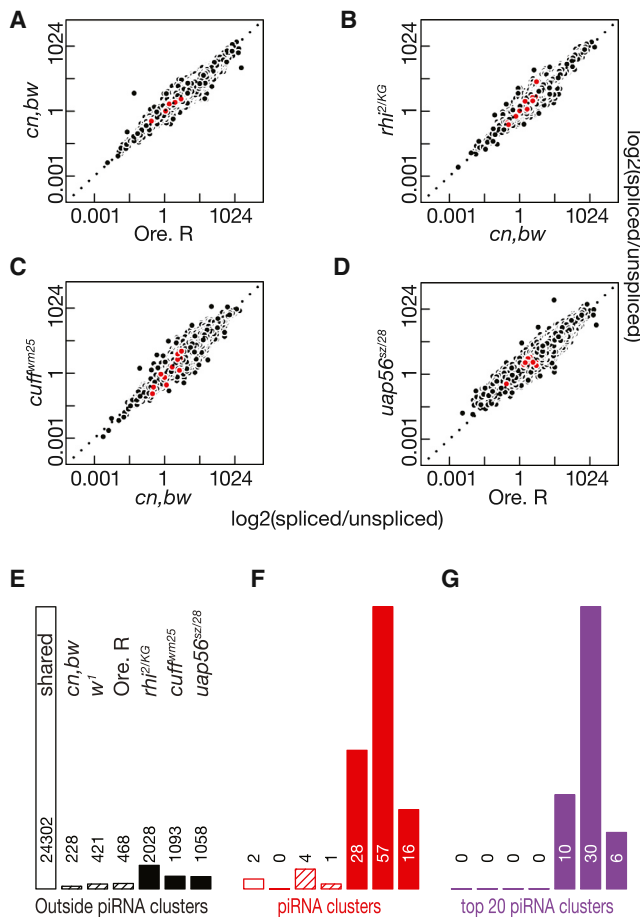


Figure 3. *rhi*, *cuff*, and *uap56* Mutations Do Not Alter Global Splicing Efficiency, but Lead to Splicing of Novel Cluster Introns

(A–D) Scatter plots showing splicing efficiency at introns shared between the two indicated genotypes. Each point is one intron. Cluster mapping introns are in red and introns mapping outside clusters in black.

(E–G) Bar graphs quantifying shared and genotype-specific introns. Introns outside of piRNA clusters are in black, introns mapping to all clusters are in red, and introns mapping to the top 20 clusters are in purple. Introns shared by the *cn,bw*, *w¹* and Oregon R control strains are indicated by the open bars. Mutant-specific introns are in solid bars. Hatched bars indicate introns specific to individual control strains. For each set of bar graphs, the genotypes are ordered as in (E), and the number of introns detected is above or within the bar.

RNA-seq reads that cross-splice junctions map to two genomic locations separated by the intron length. In genome browser views, these split reads produce signal profiles that are interrupted by sharply defined gaps. Surprisingly, the cluster peaks in *rhi*, *cuff*, and *uap56* mutants are often interrupted by sharply defined gaps that precisely map to consensus splice donor and acceptor sites. Figure 2B shows an example in the 42AB cluster. qPCR studies confirmed that RNAs crossing this unique donor-acceptor site junction increase over 100-fold in both *cuff* and *rhi* mutants (Figure 2C). The spliced peak in 42AB is in the sense orientation of a *gypsy12* mobile element that could be activated in the mutant strains, and increased expression of the spliced RNA could therefore reflect activation

of an active element. We therefore assayed a second putative intron in a chromosome 4 cluster that is antisense to the telomeric transposon TART. Utilization of this intron cannot be explained by expression of the active element. qPCR confirmed that spliced transcripts mapping to this intron increase in *rhi*, *cuff*, and *uap* mutants (F.Z. and W.T., unpublished data). Mutations in *rhi*, *cuff*, and *uap56* thus lead to accumulation of spliced transcripts from two major germline piRNA clusters.

To analyze splicing across the transcriptome, we computationally identified all split reads mapping to consensus splice donor and acceptor sites in three control strains (Oregon R, *w¹*, and *cn,bw*) and in *rhi*, *cuff*, and *uap56* mutants. To filter potential sequencing artifacts, we restricted our analysis to introns defined by a minimum of ten reads mapping across the splice junctions and with a minimum splicing entropy of two (the entropy cutoff controls for potential PCR amplification artifacts; (Graveley et al., 2011)). To calculate splicing efficiency, we divided the number of split reads (defining spliced RNAs) by the number of reads crossing the corresponding splice sites (defining unspliced RNAs). As shown in the scatter plots in Figures 3A–3D, *rhi*, *cuff*, and *uap56* did not alter global splicing efficiency for protein coding genes (black points) or for the rare cluster mapping introns (red points). Consistent with these observations, the mutations did not alter global gene expression (Figure S3).

Splicing efficiency was determined for introns that were utilized in both mutant and control strains, but the cluster introns that we detected by visual inspection were often spliced in one or more mutant strain, but not in any of the control strains. We therefore quantified introns that were used in only the control or mutant strains. We detected over 24,000 introns that were used in all of the strains, but only 228 to 468 introns were used in one or more of the control strains and none were used of the mutant strains (Figures 3E and 3F, black hatched bars). Five of these introns mapped to piRNA clusters, with the balance mapping to protein coding genes. However, none of cluster introns, and only 20 of the genic introns, were used in all of the control strains. Rhi, Cuff, and UAP56 thus do not promote splicing of clusters transcripts, or of the vast majority of genic pre-mRNAs.

We detected between 1,058 and 2,028 genic introns that were utilized in one or more mutant strains but not in any of the control strains (Figure 3E, black bars), but only 107 of these introns were used in all three mutants. Rhi, Cuff, and UAP56 could directly regulate splicing at these introns, but these events are very rare relative to splicing of the 24,000 genic introns that are used in all of the control and mutant strains. In striking contrast, we detected only two cluster mapping introns that were used in all of the strains, and only five additional cluster introns that were used in a one or more control strain but not in the mutants (Figure 3F). In the control strains, most cluster transcripts thus do not appear to be spliced. In *rhi*, *cuff*, and *uap56* mutants, by contrast, we detected 16 to 57 cluster mapping introns (Figure 3F), and these introns were enriched in the top 20 clusters, which make the majority of piRNAs in the ovary (Figure 3G). The most striking increase in cluster transcript splicing was in *cuff* mutants, where we detected 57 introns were not in any of the control strains, and 30 of these introns mapped to the top 20 piRNA clusters. In

addition, five of these cluster-mapping introns were used in *uap56*, *cuff*, and *rhi* mutants. This represents a minimum estimate of cluster introns that are expressed in the mutants because the repeated sequences that make up most of these loci cannot be uniquely mapped. These findings suggest that Rhi, Cuff, and UAP56 suppress splicing of cluster transcripts or destabilize spliced transcripts from these loci. However, the repeated nature of the clusters, noted above, made independent verification of these splicing events difficult. In addition, some of the changes in intron utilization could be secondary to transposon mobilization and genome instability. To overcome these limitations and more rigorously explore *rhi*, *cuff*, and *uap56* function in cluster transcript processing, we exploited a unique germline piRNA cluster that is also a somatic protein coding gene, and developed a transgenic reporter system that allowed us to directly assay Rhi function in wild-type ovaries, in the absence of genome instability.

Rhino, Cuff, and UAP56 Convert a Somatic Protein-Coding Gene to a Germline piRNA Cluster

The *sox102F* locus is largely composed of unique sequences and cDNA and RNA-seq data indicate that this locus produces four distinct spliced primary transcripts in somatic cells (FlyBase Genome Annotators, FlyBase.org; Figure 2D). In the germline, by contrast, we find that this locus produces piRNAs from both genomic strands (Figures 2E and S4). Mutations in *rhi*, *cuff*, and *uap56* disrupt production of these piRNAs (Figures 2E and S4). Consistent with this pattern of piRNA production, RNA sequencing on control ovaries revealed transcripts mapping across both strands of the locus and to both the introns and exons of the somatic gene (Figure 2D). ChIP-seq shows that Rhi binds to this locus (Figures 2D and S4). Intriguingly, the Rhi ChIP signal correlates with somatic exons, despite the absence of splicing, suggesting that RNA processing signals may be recognized in the germline and recruit Rhi to chromatin (Figures 2D and S4). These findings indicate that the *sox102F* locus is a protein coding gene in the soma and a Rhi-dependent piRNA cluster in the ovary.

In ovaries mutant for *rhi*, *cuff*, and *uap56*, spliced reads mapping to annotated *sox102F* donor and acceptor sites increase over 100-fold, and strand-specific qPCR confirmed this increase in spliced pre-mRNA expression (Figures 2D, 2F, and S4). The increase in splicing-specific reads was particularly pronounced in *cuff* mutants, where all of the annotated introns were efficiently excised and de novo transcript assembly generated an mRNA that precisely matches the somatic *sox102F* transcript (Figure 2D). To determine if background mutations contribute to the increase in splicing, we expressed a wild-type *rhi* transgene in *rhi* mutants and used qPCR to assay spliced and unspliced transcript levels at both *sox102F* and 42AB (Figure S5A and data not shown). The transgene fully suppressed expression of the spliced transcripts from this locus (Figure S5A). Transposon mobilization and DNA damage are common to all piRNA pathway mutants and could contribute to the splicing defects. We therefore analyzed cluster splicing in ovaries mutant for *qin*, which disrupt a component of the cytoplasmic nuage required for piRNA amplification and transposon silencing (Zhang et al., 2011). In contrast, mutations in *rhi*, *cuff*, and

uap56, the *qin* mutant combination did not increase spliced transcripts from the *sox102F* locus (Figure S4). The increase in spliced cluster transcripts in *rhi*, *cuff*, and *uap56* mutant ovaries thus do not appear to be caused by background mutations or DNA damage.

Increased expression of spliced cluster transcripts in mutant ovaries could be due to enhanced splicing or preferential stabilization of spliced transcripts. If the mutations increase splicing efficiency, accumulation of spliced transcripts should be coupled to reduced expression of unspliced transcripts. By contrast, preferential stabilization of spliced transcripts would increase expression of processed RNAs without altering unspliced transcript levels. We therefore quantified RNA-seq reads mapping across spliced and unspliced RNAs from the *sox102F* locus (Table S1). In *cn,bw*, Ore. R, and *w¹* control strains, 23, 12, and 15 reads mapped across the splice sites of unprocessed transcripts, and 6, 0, and 6 reads mapped to splice junctions. In *cuff* mutants, only 2 reads mapped across splice sites, and no reads mapping across splice sites were recovered in *rhi* and *uap56* mutants. By contrast, 110 reads mapped across mature splice junctions in *cuff*, and 18 and 5 reads mapped to these junctions in *rhi* and *uap56*. Consistent with these data, qPCR indicated unspliced reads decrease and spliced reads increase in *rhi* mutants (Figure S5B and S5C). Furthermore, we found that expression of a wild-type *rhi* transgene in the mutant background restored expression of unspliced transcripts and suppresses accumulation of spliced transcripts (Figure S5B and S5C). These findings are indirect, but support the hypothesis that *rhi*, *cuff*, and *uap56* function to suppress cluster transcript splicing.

To determine if unspliced transcripts are preferentially processed into piRNAs, we compared expression of spliced and unspliced long RNAs to expression of piRNA reads mapping to the splice junction and splice sites in the *sox102F* locus (Figure S5D). In the control *w¹* strain, RNA sequencing and qPCR assays indicate that the ratio of unspliced to spliced transcripts is approximately 3:1 (Table S1). By contrast, we identified 49 reads, representing 11 species, mapping to *sox102F* splice sites, but no reads mapping to mature junctions. These data, while limited, suggest that unspliced cluster transcripts are preferentially processed into piRNAs.

Rhi Tethering Suppresses Splicing and Promotes piRNA Production from Complementary Transcripts

To determine if Rhi binding is sufficient to suppress splicing and induce piRNA production, we used a transgenic LacO-LacI system to “tether” Rhino to an ectopic locus. For these experiments, we generated transgenic flies harboring a reporter transgene containing 36 LacO DNA-binding sites upstream a truncated *vasa* promoter, which drives EGFP expression in the germline and somatic follicle cells of the ovary. The transcription unit contains the 84B α -*Tubulin* 5' UTR and first intron, followed by an exon encoding EGFP (Figure 4A). We then introduced transgenes carrying LacI or a LacI::Rhi fusion proteins, under the control of inducible UASp promoters. Germline expression was induced using that *nanos*-Gal4-VP16 driver and EGFP expression was assayed by laser scanning confocal microscopy and western blotting (Figures 5A and S6). Germline expression of

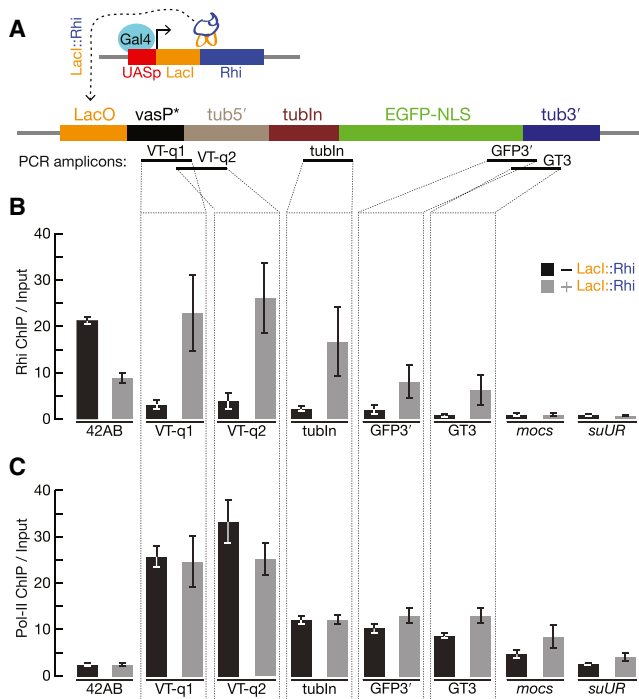


Figure 4. Rhi “Tethering” Leads to Spreading through the Target Transcription Unit, but Does Not Reduce Pol-II Occupancy

(A) Schematic diagram of the experimental design. An inducible UASp promoter was used to drive expression of the DNA-binding protein LacI fused to full-length Rhi, in the presence of a reporter gene containing LacI-binding sites (LacO) upstream of a promoter (truncated vasa promoter) that drives expression of the 84B *alpha tubulin* 5'UTR and first intron fused to EGFP with nuclear localization signal (NLS). PCR amplicons indicates positions assayed for Rhi and Pol-II binding in (B) and (C).

(B) Fold enrichment by Rhi ChIP across the reporter in the absence (black) or presence of LacI::Rhi. The LacI::Rhi lead to Rhi binding through the transcription unit.

(C) Fold enrichment by RNA pol-II ChIP across the reporter, bars as indicated for (B). RNA polymerase binding across the transcription unit is not altered in the presence of the LacI::Rhi. The 42AB locus is used as a positive control for Rhi binding. The *mocs* and *suUR* loci are located downstream of reporter construct in the genome. We do not detect Rhi spreading or changes in RNA Pol-II at these loci.

Data are mean \pm SD for three independent biological samples.

the LacI control did not block EGFP accumulation in the nurse cells or in somatic follicle cells (Figure 5A). Immunolabeling confirmed that LacI was expressed in the nurse cells (Figure 5A). Germline expression of the LacI::Rhi fusion, by contrast, silenced EGFP in the nurse cells, but not in the somatic follicle cells (Figure 5A). Rhi mediated EGFP silencing was confirmed by western blotting (Figure S6). Rhi binding upstream of the reporter promoter thus silences expression of the downstream transcription unit.

HP1a recruits the methyl transferase that modifies histone H3 to generate HP1a-binding sites, which leads to heterochromatin spreading (Danzer and Wallrath, 2004). Rhi is an HP1a homolog, and we speculated that LacI::Rhi binding upstream of the reporter gene promoter may lead to Rhi spreading into the downstream transcription unit. We therefore assayed Rhi bind-

ing at sites through the transgene reporter using ChIP and qPCR. In the absence of LacI::Rhi fusion protein, we observed background binding of Rhi through the transgene. However, expression of the LacI::Rhi fusion was linked to significant Rhi binding through the transcription unit, with the highest levels near the LacO-binding sites at 5' end of the gene (Figure 4B). To determine if LacI::Rhi binding and Rhi spreading reduces transcription, we used ChIP-qPCR to measure total RNA polymerase II (Pol II) binding at the target transgene (Figure 4C). Essentially identical levels of Pol II were observed in the absence or presence the LacI::Rhi fusion, suggesting that silencing is posttranscriptional. Consistent with this hypothesis, qPCR revealed a 6-fold reduction in the ratio of spliced to unspliced reporter transcripts in the presence of the LacI::Rhi fusion (Figures 5B and 5C). Rhi binding thus appears to suppress pre-mRNA splicing.

To assay piRNA production, we sequenced small RNAs from ovaries carrying the target transgene and expressing either the LacI control or the LacI::Rhi fusion. With both combinations, we detected only very low levels of 23 to 30 nt putative piRNAs mapping to the reporter transgene (Figures 6 and S7). Rhi is specifically required for piRNA production from dual-strand clusters, and the target transgene is transcribed from only one strand. We therefore constructed a second transgene with a promoter driving expression of antisense target sequences, integrated this gene into the same chromosomal locus as the sense-strand reporter, and generated females carrying a trans combination of sense and antisense reporters. Small RNA sequencing showed that expression of the LacI::Rhi fusion, but not the LacI control, triggered production of 23 to 30 nt small RNAs from both strands of the reporter construct, including the intron, EGFP, and LacO-binding sites (Figure S7).

piRNAs and endo-siRNAs bear a 2'-O-methyl group at their 3' termini, which protects them from oxidation (Horwich et al., 2007; Saito et al., 2007). miRNAs and nonspecific RNA degradation products, by contrast, don't carry this modification and are rendered unclonable by oxidation. We therefore performed deep sequencing on oxidized RNA samples, which are enriched for 21 nt siRNAs and 23–30 nt piRNAs. As shown in Figure 6, LacI::Rhi tethering to the transcombination of reporter genes triggered production of oxidation resistant small RNA with a size distribution typical of endogenous piRNAs. During the P-M hybrid dysgenesis, de novo production of *P element* piRNAs increase with adult female age (Khurana et al., 2011). Reporter-specific piRNAs also increased as flies carrying LacI::Rhi were aged from 2–4 days to 12–14 days (Figure 6). Total endogenous ovary piRNAs, by contrast, did not change (Table S2B). LacI::Rhi binding is therefore sufficient to drive de novo production of primary piRNAs from a trans combination of reporters producing complementary RNAs.

DISCUSSION

Primary piRNA Production from Dual-Strand and Uni-Strand piRNA Clusters

The piRNA pathway has an evolutionarily conserved role in transposon control during germline development and is essential for transmission of the inherited genetic complement. In the

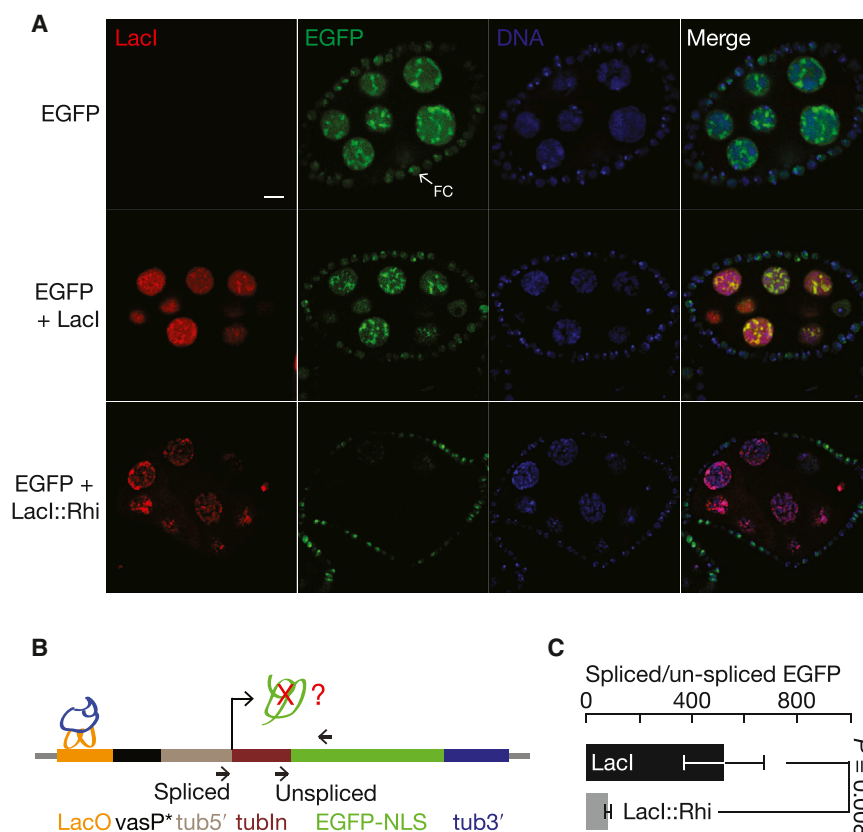


Figure 5. Tethering Rhi Suppresses EGFP Expression and Splicing

(A) Germline expression of LacI (red) does not suppress EGFP expression (green). By contrast, expression of LacI::Rhi (red) suppresses EGFP accumulation in the nurse cell nuclei (green). Note that the EGFP reporter is expressed in both the germline and surrounding somatic follicle cells (arrows, FC). The fusion does not expressed in the follicle cells, and EGFP expression in these cells is not reduced. The bar in the top right is 10 μ m, and applies to all panels.

(B) Splicing at the target locus. The diagram shows the target transgene and indicates that position of LacI or LacI::Rhi binding (LacO) and the primers used to assay both spliced and unspliced transcripts by qRT-PCR.

(C) Bar graph showing the ratio of spliced to unspliced target in the presence of LacI (black) or LacI::Rhi (gray). LacI::Rhi binding lead to a significant reduction in splicing efficiency ($p = 0.008$), data are mean \pm SD for three independent biological samples.

See also Figure S6.

Drosophila ovary, unique piRNAs are concentrated in “clusters” composed of complex arrays of nested transposon fragments that are generally localized to pericentromeric or subtelomeric heterochromatin (Brennecke et al., 2007). These loci fall into two classes, based on strand bias. Clusters that produce piRNAs from both genomic strands (dual-strand clusters) are dominant in the germline, while clusters that are expressed on only one genomic strand (uni-strand clusters) produce most of the piRNAs in somatic follicle cells that surround the germline (Brennecke et al., 2007; Malone et al., 2009). Primary piRNAs from dual-strand clusters, bound to PIWI proteins, appear to drive a ping-pong cycle that amplifies the silencing RNA pool (Brennecke et al., 2007; Gunawardane et al., 2007). Primary piRNAs that initiate the amplification cycle, by definition, are produced by a ping-pong-independent mechanism. Similarly, ping-pong amplification is not required for production of piRNAs from uni-strand clusters. These observations suggest a simple model in which primary piRNAs from uni-strand and dual-strand clusters are produced by a common mechanism, and dual-strand clusters are equivalent to convergently transcribed uni-strand clusters. However, uni-strand cluster piRNA production is independent of *rhi*, *uap56*, and *cuff*, which are essential for production of piRNAs that map uniquely to dual-strand clusters (Klattenhoff et al., 2009; Pane et al., 2011; Zhang et al., 2012a). In addition, we show that Rhi-dependent piRNA production from an ectopic locus requires a combination of transgenes expressing complementary transcripts (Figures 6

and S7). Primary piRNA production by dual-strand and uni-strand clusters thus appear to proceed by distinct mechanisms. These findings also suggest that piRNA production by dual-strand clusters requires complementary precursors. The role of complementary RNAs in the germline piRNA biogenesis pathway, however, remains to be determined.

Does Stalled Splicing Distinguish piRNA Precursors from mRNAs?

piRNA pathway mutations increase expression of a subset of transposons by over 200-fold, but do not alter germline gene expression (Klattenhoff et al., 2009; Zhang et al., 2012a; Li et al., 2009; Czech et al., 2013; Handler et al., 2013). This remarkable specificity is almost certainly essential to gamete production, but how piRNA precursors are differentiated from mRNAs is not understood. The vast majority of protein coding pre-mRNAs are efficiently spliced, exported from the nucleus, and translated in the cytoplasm. By contrast, splicing is suppressed at a transgene inserted into the *Drosophila* X-TAS piRNA cluster, (Muerdter et al., 2012), and our transcriptome wide studies indicate that rapidly evolving HP1 homolog Rhi, the Rai1-related protein Cuff, and the DEAD box protein UAP56 suppress splicing at resident consensus donor and acceptor sites in germline clusters. This is most clearly illustrated at the *sox102F* locus, which produces efficiently spliced pre-mRNAs in the soma but is the source of piRNAs from unspliced primary transcripts in the germline (Figure S5D). Significantly, accumulation of both unspliced transcripts and piRNAs requires *rhi*, *cuff*, and *uap56*, and tethering a LacI::Rhi fusion to a intron-containing reporter transgene suppresses splicing and is sufficient to trigger de novo piRNA production from a trans combination of sense and

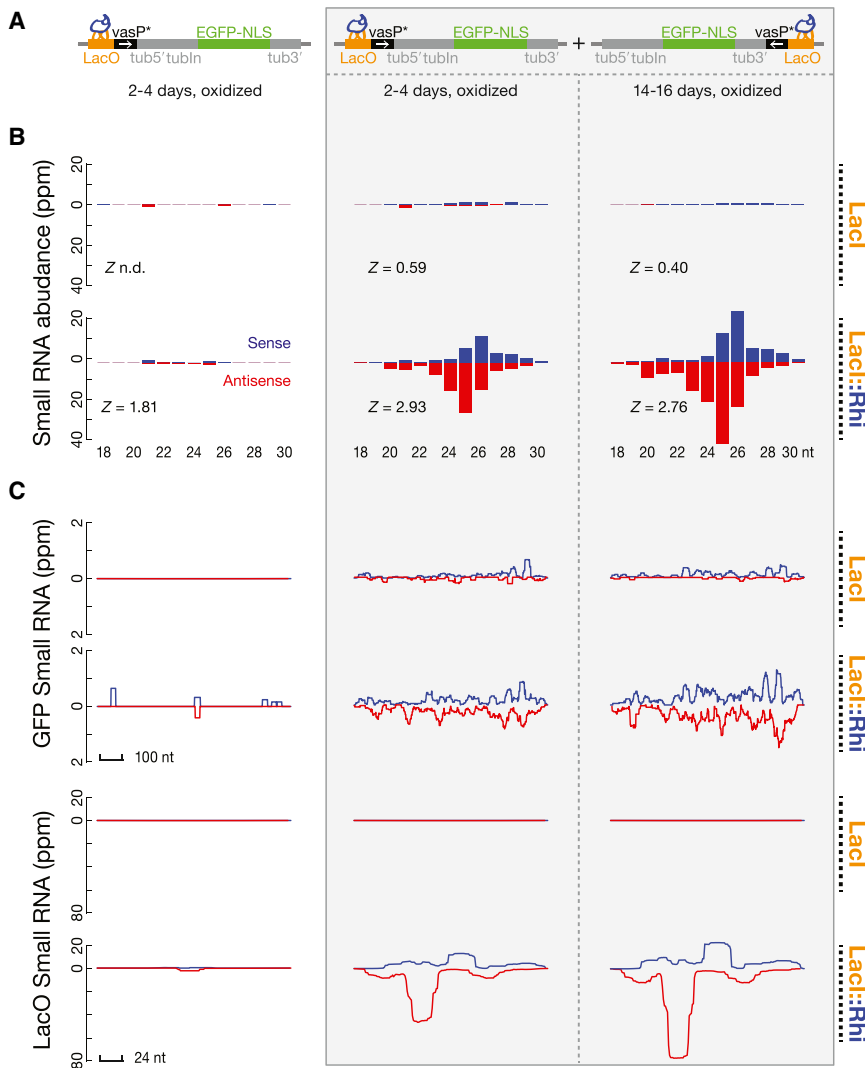


Figure 6. Rhi Binding to Complementary Transcription Units Triggers piRNA Production

(A) Diagrams show sense-strand reporter and combination of sense and antisense reporters, indicating positions of LacI or LacI::Rhi binding and position and orientation of promoters (vasP).

(B) Length distribution of the small RNAs mapping to the reporter constructs. Blue indicates sense-strand species and red indicates antisense species. Z scores indicate the significance of the 10 nt overlap between sense and antisense piRNAs (Ping-Pong signature). Z score = 1.96 corresponds to p value = 0.05. Too few piRNAs were detected with the LacI control for the Z score to be determined (indicated as n.d.). piRNAs carry a 3' end modified and therefore are resistant to oxidation. Both unoxidized (Figure S7) and oxidized RNAs (shown) were used to prepare libraries for sequencing. LacI::Rhi binding to the sense-strand reporter did not lead to significant production of oxidation resistant species between 23 and 30 nt. By contrast, LacI::Rhi binding to the combination of sense and antisense reporters trigger production of oxidation resistant species showing length distributions characteristic of mature piRNAs. piRNAs from opposite strands showed a weak bias toward a 10 nt overlap, which is typical of primary piRNAs produced by a ping-pong-independent mechanism.

(C) Distribution of Small RNA reads over EGFP and the LacO repeats in the presence of LacI or LacI::Rhi. Sense signal is in blue and antisense signal is in red.

See also Figure S7.

antisense transgenes (Figures 6 and S7). We therefore propose that Rhi functions with Cuff and UAP56 to suppress cluster transcript splicing and that the stalled splicing intermediates are the precursors for primary piRNAs.

Cuff is a homolog of Rai1/DXO, which binds and degrades mRNAs carrying incomplete cap structures (Jiao et al., 2013). This would appear to support a role for Cuff in destabilizing spliced cluster transcripts, but the residues required for the 5' to 3' exonuclease activity of Rai1/DXO are not conserved in Cuff (Chen et al., 2007; Pane et al., 2011). Murine DXO has been cocrystallized with an uncapped RNA, and with a cap analog. The crystal structures reveal the protein residues that interact with the RNA backbone, and indicate that the cap is bound in a pocket in the interior of the protein (Jiao et al., 2013). We aligned the sequences of *Drosophila* Cuff with the murine DXO. Twenty one percent of the positions are identical, and conserved amino acids are present throughout the entire alignment. Overall protein fold of Cuff is therefore highly likely to resemble DXO. We therefore used the Modeler and I-TASSER

algorithms to build a homology model of Cuff based on the Murine DXO structures (Roy et al., 2010; Zhang, 2008; Sali and Blundell, 1993). Essentially, all of the RNA-binding interactions are preserved in the homology model of Cuff (Figure 7A). We therefore propose that Cuff is not a catalytically active nuclease, but binds to uncapped RNA ends as they emerge from Pol2 or to newly capped cluster transcripts.

Cap binding by the nuclear Cap Binding Complex (CBC) is required for splicing and polyadenylation (Izaurralde et al., 1994; Topisirovic et al., 2011). We therefore propose that Cuff binds to Rhi and cotranscriptionally associates with cluster transcripts, preventing capping or recognition of capped RNAs by the CBC, which blocks splicing (Figure 7B). We previously showed that UAP56 immunoprecipitation significantly enriches for cluster transcripts, not for mRNAs, and that RNAs mapping to the major 42AB cluster are the most highly enriched species in the immunoprecipitated pool (Zhang et al., 2012a). The point mutation in *uap56* that specifically blocks piRNA biogenesis disrupts a salt bridge predicted to stabilize the ATP and RNA bound form of the protein (Zhang et al., 2012a). These observations suggest that stable cluster transcript binding by UAP56 is required for piRNA biogenesis. Mutations in the yeast

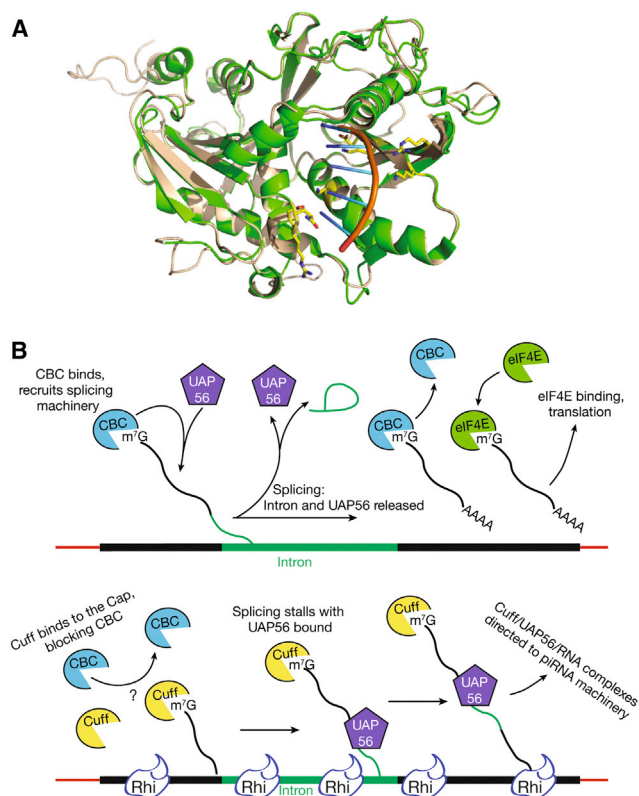


Figure 7. Predicted Cuff Structure and a Model for Rhi Complex Function in piRNA Biogenesis

(A) Predicted overlap of homology modeled *Drosophila* Cuff (wheat) and mouse DXO (Green), a murine mRNA decapping enzyme (Jiao et al., 2013). Catalytic residues are not conserved in Cuff, but residues that interact with the RNA backbone and bases are preserved (yellow).

(B) Top: Cotranscriptional binding by the cap-binding complex (CBC) promotes efficient pre-mRNA splicing, through a process that requires transient binding by UAP56. Subsequent cap binding by eIF4E promotes translation. Bottom: Cuff localizes to clusters through Rhi, and we speculate that cotranscriptional loading on capped cluster transcripts competes with CBC and stalls splicing with UAP56 bound. Stable cluster transcript-UAP56 complexes are directed to the piRNA biogenesis machinery.

cap-binding complex lead to arrest at an early step in the splicing pathway, with the U2 snRNP bound to primary transcripts (Görnemann et al., 2005). UAP56 was identified as a binding partner of the U2 snRNP protein U2AF65 (Fleckner et al., 1997). Cuff binding to capped cluster transcripts may therefore prevent cap recognition by the CBC, arresting splicing with UAP56 stably bound. This aberrant stable complex could differentiate piRNA precursors from pre-mRNAs (Figure 7). While this model is highly speculative, it makes several clear predictions and should therefore serve as a useful starting point to additional studies.

Adaptation to transposon invasion by the piRNA pathway appears to be initiated by insertion of the invading element into a cluster (Khurana et al., 2011). This speculate model, with the observation that Rhi can spread from anchor sites, suggest an adaptation model in which Rhi spreads into active transposons that insert into clusters, leading to Cuff binding to capped transcripts from transposon promoters. This would block processing

and promote production of new piRNAs, thus coordinately silence the inserted element and produce the transilience species that control dispersed active elements.

Studies in the pathogenic yeast *Cryptococcus* provide evidence for a direct link between stalled splicing and transposon silencing by the siRNA pathway (Dumesic et al., 2013). Dumesic et al. (2013) showed that splicing factors associate with the *Cryptococcus* siRNA biogenesis machinery and siRNAs are produced from unspliced transposon transcripts. In addition, intron removal reduces siRNA production, and splice site mutations that reduce splicing efficiency increase siRNA production (Dumesic et al., 2013). Furthermore, recent genome-wide screens have implicated splicing factors in transposon silencing, and the splicing and small RNA-silencing pathways appears to be coevolving (Czech et al., 2013; Handler et al., 2013; Muerdter et al., 2013; Tabach et al., 2013). These findings, with the studies reported here, suggest that stalled splicing generates a conserved molecular signature for potentially deleterious RNAs, which directs these transcripts to small silencing RNA biogenesis pathways. Retrotransposons and retroviruses encode essential spliced transcripts, but splicing must be suppressed to produce full-length genomic RNAs. This novel feature of the retroviral life cycle may have driven evolution of silencing systems that use stalled splicing as a hallmark of pathogenic RNAs.

EXPERIMENTAL PROCEDURES

General Methods

RNA isolation, small RNA library construction and sequencing data analysis, immunoblotting, immunostaining, and quantitative RT-PCR were performed as described (Zhang et al., 2011). Figures were generated using Excel (Microsoft, Redmond, WA, USA), IgorPro (WaveMetrics, Lake Oswego, OR, USA), and Adobe Photoshop and Illustrator (Adobe systems, San Jose CA, USA). Table S2 reports the statistics for the ChIP-seq, RNA-seq, and small RNA-seq data generated in this study. Table S3 reports primer sequences for ChIP-qPCR and qRT-PCR. PCR primers used to clone the LacI-binding domain, Rhi open reading frame and LacO repeats are detailed along with the Supplemental Information. The sources of the published deep sequencing data used in this study are summarized in Table S5. Table S6 lists the antibody information. Unless otherwise specified, p values were calculated from at least three independent biological replicates using a two-tailed, two-sample unequal variance t test (Excel, Microsoft).

Drosophila Stocks

All flies were raised at 25°C. Table S4 summarizes the published fly alleles used in this study. Transgenic flies for tethering Rhi to EGFP locus and for rescuing suppress splicing were made as described in supplemental information.

ChIP-Seq

Detailed protocols are provided in Supplemental information. Briefly, ovaries were crosslinked and sonicated to generate ~150 bp fragments. After immunoprecipitation with antibodies to Rhi or control serum, crosslinking was reversed and enriched DNA was purified and subjected to library cloning by End-Repair, A-tailing, Adaptor ligation, and PCR amplification (Zhang et al., 2012b). Antibodies to Rhi are described in (Klattenhoff et al., 2009).

Bioinformatics Analysis of Splicing

Strand-specific RNA-seq libraries were made as described (Zhang et al., 2012b). RNA-seq reads were aligned to the genome and the transcriptome (Flybase r5.50) using TopHat 2.0.8 (Trapnell et al., 2009) with the parameters “-x 1000 -g 1000-read-mismatches 2-read-edit-dist 2-read-realign-edit-dist 0-segment-length 50-segment-mismatches 2.” Only reads mapping uniquely

were considered in the downstream analysis. BEDTools (Quinlan and Hall, 2010) was used to count the fragments within a transcript or piRNA cluster, and the number of reads per transcript were normalized by the sequencing depth and transcript length. We collapsed introns detected by TopHat from six libraries (three control strains: Ore. R, *cn,bw,w¹*; three mutants: *rhl^{2/KG}*, *cuff^{vm25}*, *uap56^{sz/28}*), then we counted the spliced reads and the unspliced reads across the donor/acceptor sites. The introns with fewer than ten spliced reads in all six libraries were discarded in the analysis. Splicing efficacy was calculated as the ratio reads mapping to mature splice junctions multiplied by two over the sum of reads to the corresponding donor and acceptor sites, a pseudo count ten was used.

ACCESSION NUMBERS

The NCBI trace archives (<http://www.ncbi.nlm.nih.gov/Traces/>) accession number for the sequence data reported in this paper is SRP030460.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.04.030>.

ACKNOWLEDGMENTS

We thank members of the Theurkauf and Weng labs and the UMass RNA biology community for critical feedback and encouragement. This work was supported by a grant R01HD049116 from the National Institute of Child Health and Human Development, NIH, to W.E.T., Z.W. and P.D.Z., and in part by National Institutes of Health grant GM62862 to P.D.Z.

Received: September 27, 2013

Revised: January 15, 2014

Accepted: April 9, 2014

Published: June 5, 2014

REFERENCES

- Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T., et al. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203–207.
- Beck, C.R., Collier, P., Macfarlane, C., Malig, M., Kidd, J.M., Eichler, E.E., Badge, R.M., and Moran, J.V. (2010). LINE-1 retrotransposition activity in human genomes. *Cell* **141**, 1159–1170.
- Bennetzen, J.L. (2000). Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol.* **42**, 251–269.
- Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**, 1089–1103.
- Chen, Y., Pane, A., and Schüpbach, T. (2007). Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in *Drosophila*. *Curr. Biol.* **17**, 637–642.
- Czech, B., Preall, J.B., McGinn, J., and Hannon, G.J. (2013). A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway. *Mol. Cell* **50**, 749–761.
- Danzer, J.R., and Wallrath, L.L. (2004). Mechanisms of HP1-mediated gene silencing in *Drosophila*. *Development* **131**, 3571–3580.
- Dumesic, P.A., Natarajan, P., Chen, C., Drinnenberg, I.A., Schiller, B.J., Thompson, J., Moresco, J.J., Yates, J.R., 3rd, Bartel, D.P., and Madhani, H.D. (2013). Stalled spliceosomes are a signal for RNAi-mediated genome defense. *Cell* **152**, 957–968.
- Fleckner, J., Zhang, M., Valcárcel, J., and Green, M.R. (1997). U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev.* **11**, 1864–1872.
- Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199–202.
- Görnemann, J., Kotovic, K.M., Hujer, K., and Neugebauer, K.M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol. Cell* **19**, 53–63.
- Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., et al. (2011). The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471**, 473–479.
- Grewal, S.I. (2010). RNAi-dependent formation of heterochromatin and its diverse functions. *Curr. Opin. Genet. Dev.* **20**, 134–141.
- Grivna, S.T., Beyret, E., Wang, Z., and Lin, H. (2006). A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.* **20**, 1709–1714.
- Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **315**, 1587–1590.
- Guzzardo, P.M., Muerdter, F., and Hannon, G.J. (2013). The piRNA pathway in flies: highlights and future directions. *Curr. Opin. Genet. Dev.* **23**, 44–52.
- Handler, D., Meixner, K., Pizka, M., Lauss, K., Schmied, C., Gruber, F.S., and Brennecke, J. (2013). The genetic makeup of the *Drosophila* piRNA pathway. *Mol. Cell* **50**, 762–777.
- Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., and Zamore, P.D. (2007). The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.* **17**, 1265–1272.
- Iida, T., Nakayama, J., and Moazed, D. (2008). siRNA-mediated heterochromatin establishment requires HP1 and is associated with antisense transcription. *Mol. Cell* **31**, 178–189.
- Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E., and Mattaj, I.W. (1994). A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78**, 657–668.
- Jiao, X., Chang, J.H., Kilic, T., Tong, L., and Kiledjian, M. (2013). A mammalian pre-mRNA 5' end capping quality control mechanism and an unexpected link of capping to pre-mRNA processing. *Mol. Cell* **50**, 104–115.
- Keller, C., Adaixo, R., Stunnenberg, R., Woolcock, K.J., Hiller, S., and Bühler, M. (2012). HP1(Swi6) mediates the recognition and destruction of heterochromatic RNA transcripts. *Mol. Cell* **47**, 215–227.
- Khurana, J.S., and Theurkauf, W. (2010). piRNAs, transposon silencing, and *Drosophila* germline development. *J. Cell Biol.* **191**, 905–913.
- Khurana, J.S., Wang, J., Xu, J., Koppetsch, B.S., Thomson, T.C., Nowosielska, A., Li, C., Zamore, P.D., Weng, Z., and Theurkauf, W.E. (2011). Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell* **147**, 1551–1563.
- Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., 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.
- Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., and Kingston, R.E. (2006). Characterization of the piRNA complex from rat testes. *Science* **313**, 363–367.
- Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., et al. (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137**, 509–521.
- Malone, C.D., Brennecke, J., Dus, M., Stark, A., McCombie, W.R., Sachidanandam, R., and Hannon, G.J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* **137**, 522–535.
- McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA* **36**, 344–355.

- Muerdter, F., Olovnikov, I., Molaro, A., Rozhkov, N.V., Czech, B., Gordon, A., Hannon, G.J., and Aravin, A.A. (2012). Production of artificial piRNAs in flies and mice. *RNA* 18, 42–52.
- Muerdter, F., Guzzardo, P.M., Gillis, J., Luo, Y., Yu, Y., Chen, C., Fekete, R., and Hannon, G.J. (2013). A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in *Drosophila*. *Mol. Cell* 50, 736–748.
- Pane, A., Jiang, P., Zhao, D.Y., Singh, M., and Schüpbach, T. (2011). The Cut-off protein regulates piRNA cluster expression and piRNA production in the *Drosophila* germline. *EMBO J.* 30, 4601–4615.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* 5, 725–738.
- Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., and Siomi, M.C. (2007). Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. *Genes Dev.* 21, 1603–1608.
- Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815.
- Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* 12, 246–258.
- Tabach, Y., Billi, A.C., Hayes, G.D., Newman, M.A., Zuk, O., Gabel, H., Kamath, R., Yacoby, K., Chapman, B., Garcia, S.M., et al. (2013). Identification of small RNA pathway genes using patterns of phylogenetic conservation and divergence. *Nature* 493, 694–698.
- Topisirovic, I., Svitkin, Y.V., Sonenberg, N., and Shatkin, A.J. (2011). Cap and cap-binding proteins in the control of gene expression. *Wiley Interdiscip Rev RNA* 2, 277–298.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111.
- Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313, 320–324.
- Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* 303, 672–676.
- Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9, 40.
- Zhang, Z., Xu, J., Koppetsch, B.S., Wang, J., Tipping, C., Ma, S., Weng, Z., Theurkauf, W.E., and Zamore, P.D. (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, B.S., Schultz, N., Vreven, T., Meignin, C., Davis, I., Zamore, P.D., et al. (2012a). UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell* 151, 871–884.
- Zhang, Z., Theurkauf, W.E., Weng, Z., and Zamore, P.D. (2012b). Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly(A) selection. *Silence* 3, 9.