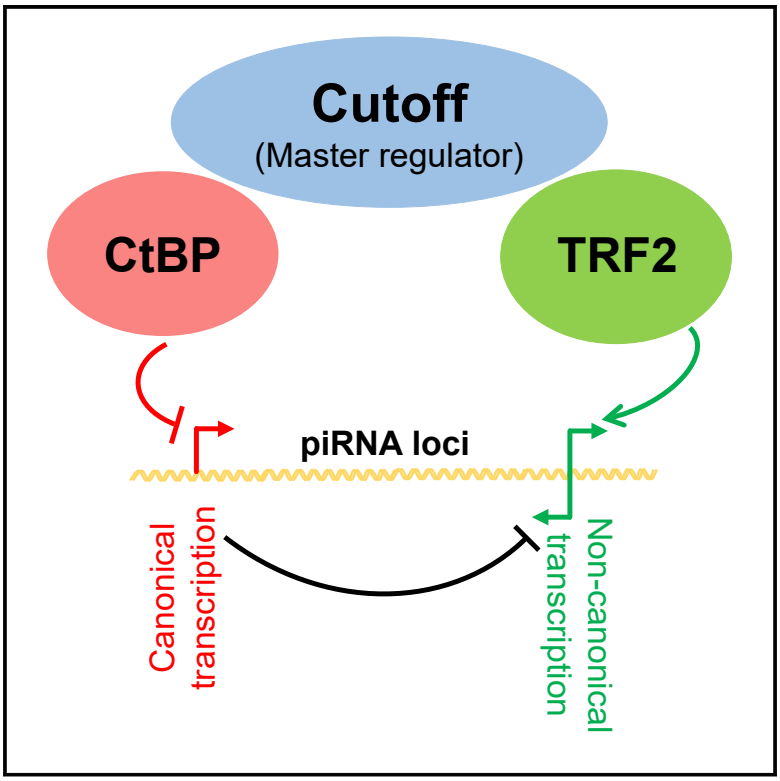
Article

Cell Reports

Adaptive Evolution Targets a piRNA Precursor Transcription Network

Graphical Abstract



Authors

Swapnil S. Parhad, Tianxiong Yu, Gen Zhang, Nicholas P. Rice, Zhiping Weng, William E. Theurkauf

Correspondence

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

In Brief

Parhad et al. use cross-species complementation to determine the functional impact of adaptive evolution. These studies show that adaptive evolution of the piRNA pathway protein Cutoff, which is required for transposon silencing and genome maintenance, targets interactions with conserved canonical and non-canonical transcription factors that regulate piRNA precursor expression.

Highlights

- Adaptive evolution produces a dominant-negative allele of the piRNA gene *cuff*
- Cutoff balances interlinked canonical and non-canonical piRNA cluster transcription
- CtBP suppresses canonical transcription of both piRNA clusters and transposons

Parhad et al., 2020, Cell Reports *30*, 2672–2685
February 25, 2020
https://doi.org/10.1016/j.celrep.2020.01.109

CellPress

OPEN ACCESS
CellPress

Cell Reports
Article

Adaptive Evolution Targets a piRNA Precursor Transcription Network

Swapnil S. Parhad,[1,4] Tianxiong Yu,[2,3] Gen Zhang,[1] Nicholas P. Rice,[1] Zhiping Weng,[2,*] and William E. Theurkauf[1,5,*]
[1]Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA
[2]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA
[3]School of Life Sciences and Technology, Tongji University, Shanghai 200092, China
[4]Present address: Department of Cell Biology, Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA
[5]Lead Contact
*Correspondence: zhiping.weng@umassmed.edu (Z.W.), william.theurkauf@umassmed.edu (W.E.T.)
https://doi.org/10.1016/j.celrep.2020.01.109

SUMMARY

In *Drosophila*, transposon-silencing piRNAs are derived from heterochromatic clusters and a subset of euchromatic transposon insertions, which are bound by the Rhino-Deadlock-Cutoff complex. The HP1 homolog Rhino binds to Deadlock, which recruits TRF2 to promote non-canonical transcription from both genomic strands. Cuff function is less well understood, but this Rai1 homolog shows hallmarks of adaptive evolution, which can remodel functional interactions within host defense systems. Supporting this hypothesis, *Drosophila simulans* Cutoff is a dominant-negative allele when expressed in *Drosophila melanogaster*, in which it traps Deadlock, TRF2, and the conserved transcriptional co-repressor CtBP in stable complexes. Cutoff functions with Rhino and Deadlock to drive non-canonical transcription. In contrast, CtBP suppresses canonical transcription of transposons and promoters flanking the major germline clusters, and canonical transcription interferes with downstream non-canonical transcription and piRNA production. Adaptive evolution thus targets interactions among Cutoff, TRF2, and CtBP that balance canonical and non-canonical piRNA precursor transcription.

INTRODUCTION

Transposable elements (TEs) are major genome components that can induce mutations and facilitate ectopic recombination, but transposons have also been co-opted for normal cellular functions, and transposon mobilization has rewired transcriptional networks to drive evolution (Ayarpadikannan and Kim, 2014; Chuong et al., 2017; Hedges and Deininger, 2007; Horváth et al., 2017; Jangam et al., 2017; Piacentini et al., 2014). Species survival may therefore depend on a balance of transposon silencing and activation. The Piwi interacting RNA (piRNA) pathway transcriptionally and post-transcriptionally silences transposons in the germline (Biémont and Vieira, 2006; Canapa et al., 2015; Ghildiyal and Zamore, 2009; Parhad and Theurkauf, 2019; Weick and Miska, 2014). However, how this pathway is regulated is not completely understood.

In *Drosophila*, piRNAs are produced from heterochromatic clusters composed of complex arrays of nested transposon fragments, which appear to provide genetic memory of past genome invaders (Bergman et al., 2006; Brennecke et al., 2007). Adaptation to new genome invaders is proposed to involve transposition into a cluster, which leads to sequence incorporation into precursors that are processed into trans-silencing anti-sense piRNAs (Khurana et al., 2011; Parhad and Theurkauf, 2019). However, a subset of isolated transposon insertions also produce sense and anti-sense piRNAs (Mohn et al., 2014), providing an independent adaptation mechanism and epigenetic memory of genome invaders. Expression of piRNAs from these loci is disrupted by *piwi* mutations (Mohn et al., 2014), but Piwi-bound piRNAs map to all insertions, not just the subset that function in piRNA biogenesis. The mechanism that defines these "mini-cluster" thus remains to be determined.

In *Drosophila*, germline piRNA clusters and transposon mini-clusters are bound by the RDC complex, which is composed of the HP1 homolog Rhino (Rhi), which co-localizes with the linker protein Deadlock (Del) and the Rai1 homolog Cutoff (Cuff) (Chang et al., 2019; Chen et al., 2016; Le Thomas et al., 2014; Mohn et al., 2014; Pane et al., 2011; Parhad et al., 2017; Yu et al., 2015; Zhang et al., 2014, 2018). The three components of the RDC are co-dependent for localization to clusters and essential to germline piRNA production. Rhi is composed of chromo, hinge, and shadow domains (Vermaak et al., 2005). The chromo domain binds to H3K9me3-modified histones, and the shadow domain binds Del, which recruits Moonshiner (Moon) and TATA box binding protein-related factor 2 (TRF2), promoting "non-canonical" transcription from both genomic strands (Andersen et al., 2017; Le Thomas et al., 2014; Mohn et al., 2014).

The third RDC component, Cuff, was identified in a screen for female sterile mutations (Schüpbach and Wieschaus, 1989) and found to encode a homolog of the decapping exonuclease Rai1 required for transposon silencing and piRNA biogenesis (Chen et al., 2007; Pane et al., 2011). Critical residues in the catalytic pocket of Rai1 are not conserved in Cuff, but sidechains that bind the RNA backbone are retained, suggesting that Cuff may be an RNA 5′ end-binding protein (Pane et al., 2011; Zhang et al., 2014). Intriguingly, germline piRNAs in *Drosophila* are preferentially produced from unspliced transcripts, and *cuff*

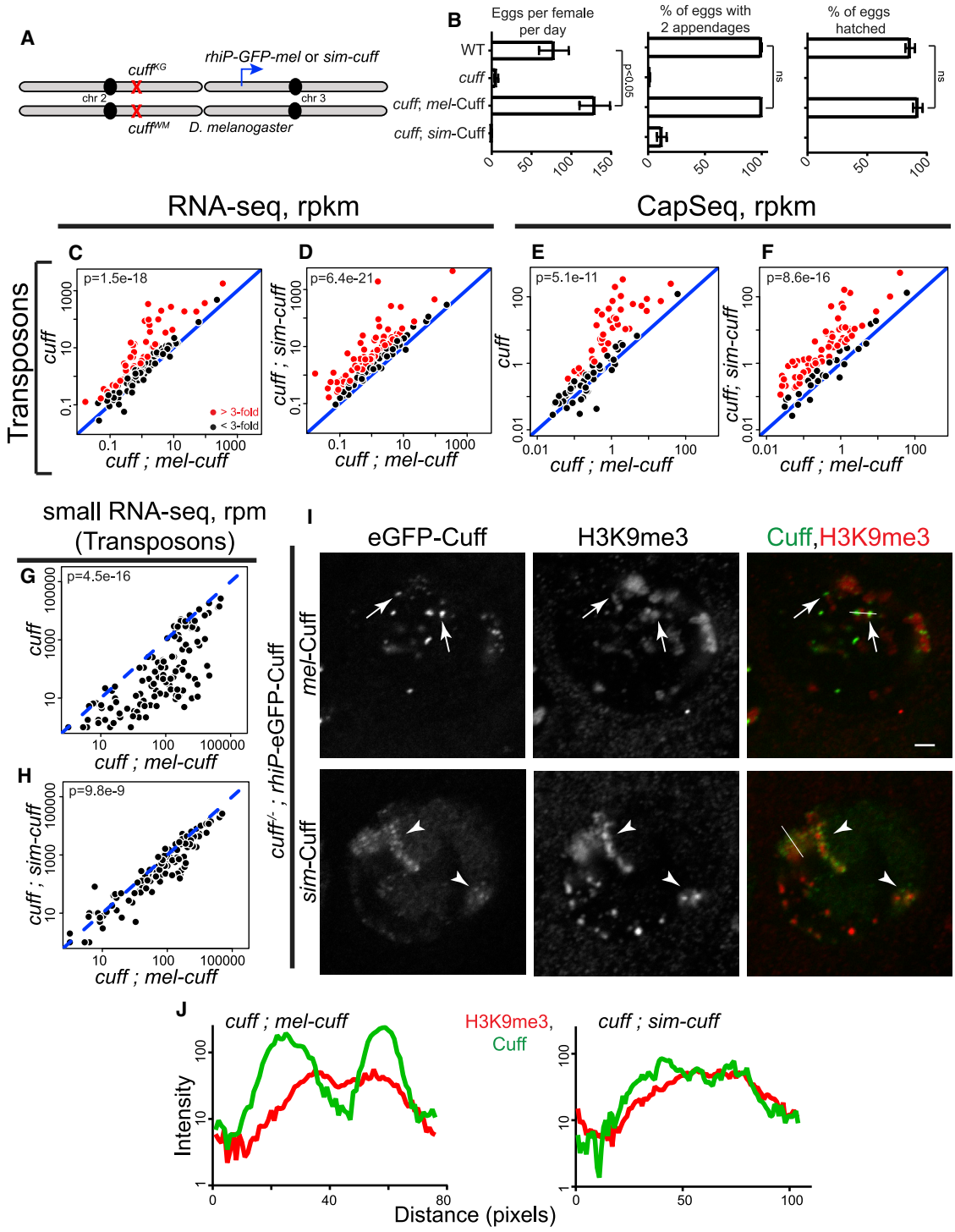


Figure 1. *sim*-Cuff Does Not Complement *D. melanogaster cuff* Mutations

(A) Genetic complementation strategy. The *sim-cuff* or *mel-cuff* genes were expressed in *D. melanogaster cuff* mutants using the germline-specific *rhi* promoter and assayed for phenotypic rescue.

(B) Bar graphs showing number of eggs laid per female per day, percentage of eggs with two appendages, and percentage of hatched eggs produced by OrR (wild-type [WT] control), *cuff* mutants, and *cuff* mutants expressing either *mel-cuff* or *sim-cuff*. Error bars show standard deviation of three biological replicates, with a minimum of 500 embryos scored per replicate, except for *cuff* mutants and *cuff* mutants rescued by *sim-cuff*, for which average of 230 and 23 eggs were scored, respectively.

(legend continued on next page)

mutations significantly increase piRNA precursor splicing, and 5′ cap binding by the nuclear cap binding complex (CBC) promotes splicing. Together, these findings suggest that that Cuff competes with the CBC for binding to capped cluster transcripts, suppressing splicing and promoting piRNA biogenesis. However, tethering Cuff to a reporter transcript increases read-through transcription (Chen et al., 2016), consistent with suppression of transcription termination. The molecular function for Cuff in piRNA biogenesis thus remains enigmatic.

All three RDC genes are rapidly evolving under positive selection, suggesting that adaptive evolution of the complex is driven by a genetic conflict with the transposons the piRNA pathway silences, but other mechanisms are possible (Blumenstiel et al., 2016; Lee and Langley, 2012; Parhad and Theurkauf, 2019; Simkin et al., 2013). We previously found that rapid evolution has modified the Rhi-Del interface, producing orthologs that function as mutant alleles when moved across species (Parhad et al., 2017; Yu et al., 2018). Analysis of these cross-species incompatibilities defined an interaction between the Rhi shadow domain and Del that prevents ectopic assembly of piRNA cluster chromatin. Crosses between *Drosophila melanogaster* and *Drosophila simulans*, which are sibling species, lead to hybrid sterility and are important model for genetic control of reproductive isolation (Barbash, 2010). Significantly, sterile hybrids between these species phenocopy piRNA pathway mutations (Kelleher et al., 2012). Adaptive evolution of piRNA pathway genes may therefore contribute to reproductive isolation and speciation (Barbash, 2010; Kelleher et al., 2012; Parhad and Theurkauf, 2019; Parhad et al., 2017).

These findings also suggest that cross-species analysis of rapidly evolving genes may provide a powerful genetic approach to structure-function analysis. Here we apply this approach to *cuff*. These studies indicate that adaptive evolution has targeted direct or indirect interactions among Cuff, the Del-TRF2 non-canonical transcriptional complex, and the transcriptional co-repressor C-terminal binding protein (CtBP). CtBP was first identified as a host binding partner of Adenovirus E1A and subsequently implicated in diverse developmental pathways and cancer (Boyd et al., 1993; Chinnadurai, 2002; Dcona et al., 2017; Mani-Telang et al., 2007; Schaeper et al., 1995; Stankiewicz et al., 2014). We show that *Drosophila* CtBP suppresses canonical transcription from promoters in transposon terminal repeats and from promoters flanking two major germline piRNA clusters. Significantly, in both contexts, activation of canonical transcription interferes with downstream non-canonical transcription and piRNA production. Adaptive evolution has therefore targeted interactions between Cuff and two transcription regulators, which coordinately control germline piRNA expression.

RESULTS

D. simulans cuff Is a Dominant Separation of Function Allele in *D. melanogaster*

The heterochromatic clusters that produce germline piRNA precursors in *Drosophila* are bound by the HP1 homolog Rhi, which anchors a complex containing a group of proteins that control transcription and processing of piRNA precursors. Rhi binds directly to Del, which recruits TRF2 through the linker protein Moon (Andersen et al., 2017). This complex promotes non-canonical transcription from both strands. Del also interacts with Cuff, and this Rai1 homolog suppresses cluster transcript splicing and transcription termination. Adaptive evolution has remodeled an interface between Rhi and Del that helps define cluster location (Parhad et al., 2017). Strikingly, *cuff*, *moon*, and *Trf2* are also evolving very rapidly (Figure S1B), suggesting that the chromatin-bound protein complex that drives piRNA precursor production is engaged in a genetic conflict.

Adaptive evolution, as opposed to genetic drift, is predicted to alter functionally important domains. To determine if *cuff* evolution has altered functional domains, we expressed GFP-tagged *D. simulans* Cuff (*sim*-Cuff) and GFP-tagged *D. melanogaster* Cuff *(mel*-Cuff*)* in *D. melanogaster cuff* mutants and assayed phenotypic rescue. Both Cuff variants were expressed using the germline-specific *rhi* promoter and were integrated into the same chromosomal location, using PhiC31-mediated transformation (Figure 1A). Direct visualization of GFP signal in egg chambers, using identical imaging conditions, indicates that *sim*-Cuff and *mel*-Cuff are both nuclear, and suggest that the proteins are expressed at comparable levels (Figure 1I). Direct analysis of protein production was not possible because Cuff is expressed at low levels and neither fusion protein could be detected by western blotting. However, comparable levels of the two fusion proteins were recovered after affinity purification, assayed using mass spectrometry (see below). The two fusion proteins thus appear to be expressed at comparable levels. Mutations in *cuff* lead to female sterility and production of eggs with dorsal appendage defects, which reflect disruption of D-V patterning in response to genome instability (Klattenhoff et al., 2007). The *mel-cuff* transgene restored D-V patterning and hatching, but the *sim-cuff* transgene failed to rescue hatching or embryo patterning and was comparable with the null allelic combination by these biological measures (Figure 1B).

To determine if *sim*-Cuff supports transposon silencing (Chen et al., 2007; Pane et al., 2011), we used CapSeq (Gu et al., 2012) and strand-specific RNA sequencing (RNA-seq) (Zhang et al., 2012b) to assay steady-state expression of transposons and genes. The *mel-cuff* transgene restored transposon silencing, but overall transposon expression was comparable with the null allelic combination on rescue with the *sim-cuff* transgene

(C–H) Scatterplots showing comparisons of RNA-seq (C and D), CapSeq (E and F), and small RNA-seq signal (G and H) for transposon families in *cuff* mutant or *cuff* mutant expressing *sim-cuff* versus *cuff* mutant expressing *mel-cuff*. Each point on the scatterplots shows RPKM (long RNAs) or RPM (small RNAs) for a transposon family in ovaries of the indicated genotype. For transposons, anti-sense piRNA abundance is plotted. Diagonal represents x = y. Points in red show y/x > 3. p value for differences obtained using Wilcoxon test.

(I and J) Localization of GFP-tagged Cuff with respect to H3K9me3-marked chromatin in germline nuclei of *cuff* mutants expressing *rhi* promoter-driven *mel*-Cuff or *sim*-Cuff. Color assignments for merged images shown on top. Arrowheads and arrows denote locations of *mel*-Cuff and *sim*-Cuff foci, respectively. Identical imaging conditions were used for all panels. Scale bar, 2 μm. Fluorescence intensities calculated across the white lines in the merged images (I) are shown in (J).

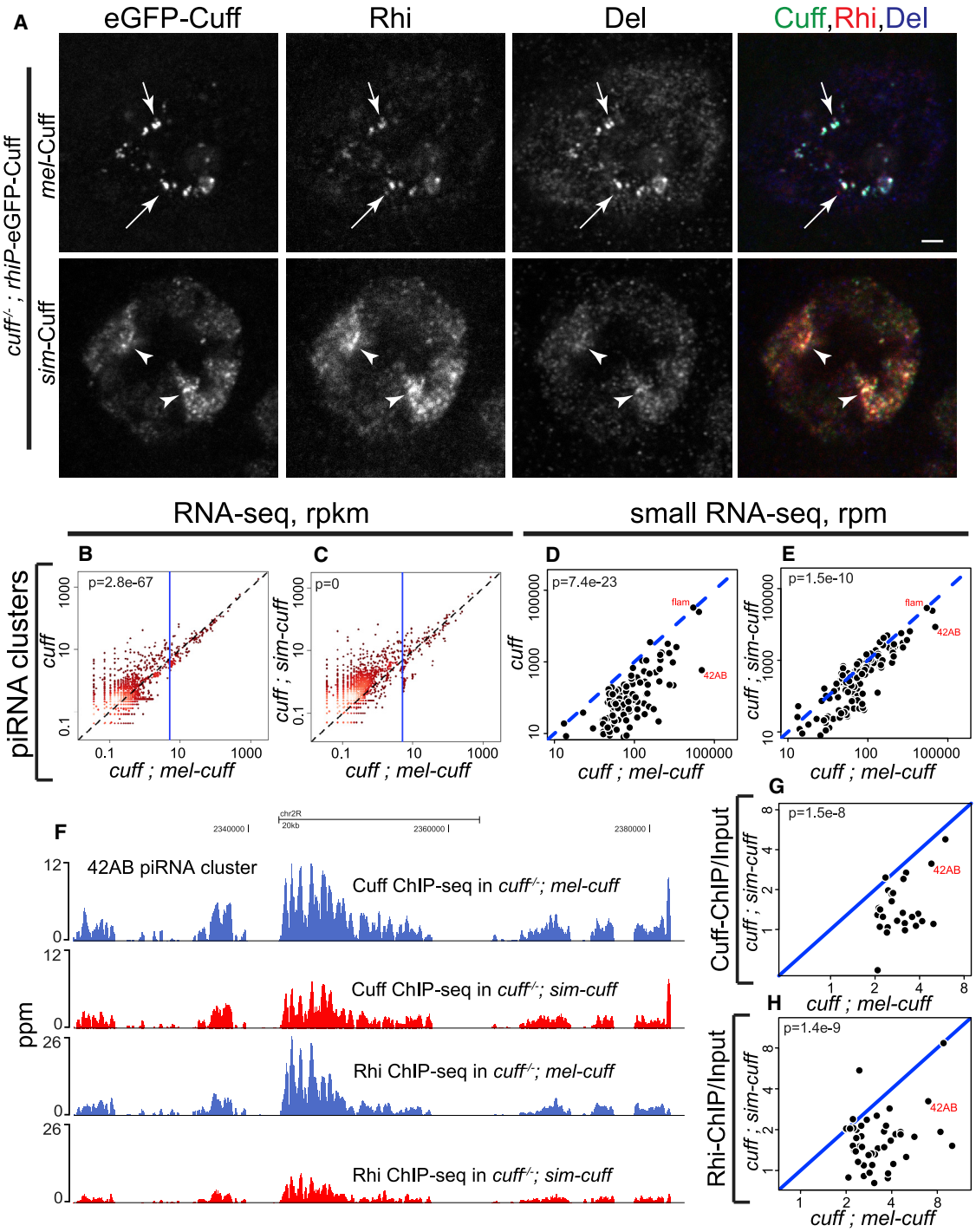


Figure 2. *sim*-Cuff Disrupts RDC Localization

(A) Localization of GFP-tagged Cuff with respect to Rhi and Del in the germline nuclei of *cuff* mutants expressing *rhi* promoter-driven *mel*-Cuff or *sim*-Cuff. Color assignments for merged images shown on top. Arrows and arrowheads denote locations of *mel*-Cuff and *sim*-Cuff foci, respectively. Scale bar, 2 μm.

(B–E) Scatterplots showing comparisons of RNA-seq signal (B and C) and small RNA-seq signal (D and E) at piRNA clusters in ovaries with genotypes *cuff* mutant or *cuff* mutant expressing *sim-cuff* versus *cuff* mutant expressing *mel-cuff*. In (B) and (C), each point on the scatterplots shows RPKM value for a 1 kb piRNA clusters bin. In (D) and (E), each point shows RPM value for an entire cluster. Diagonal represents x = y. p value for differences obtained using Wilcoxon test.

(F) Genome Browser view of GFP-Cuff (top) and Rhi (bottom) ChIP-seq profiles at 42AB piRNA cluster in the ovaries of *cuff* mutants expressing either *mel-cuff* (blue) or *sim-cuff* (red).

(legend continued on next page)

(Figures 1C–1F, S2B, and S2D). Surprisingly, a number of transposon families were more highly expressed in *cuff* mutant expressing *sim-cuff* than in the *cuff* null mutant combination (Figures S2A–S2D). Cuff is required for piRNA biogenesis, and small RNA-seq showed that the *mel-cuff* transgene restored transposon and cluster mapping piRNA expression (Figures 1G and 2D). We anticipated that *sim-cuff* would also fail to support piRNA expression, but median transposon and cluster mapping piRNA levels were restored to 45% and 70% of control levels by the *sim-cuff* transgene, and many clusters and transposons showed essentially wild-type piRNA profiles (Figures 1H, S2E, 2E, and S3C). The *D. simulans* ortholog is therefore a partial separation-of-function allele in *D. melanogaster*, which supports significant piRNA expression but not transposon silencing.

Cuff, Rhi, and Del associate with peri-centromeric piRNA clusters and localize to cytologically distinct nuclear foci that are frequently adjacent to large domains of constitutive heterochromatin, marked by H3K9me3 (Mohn et al., 2014; Parhad et al., 2017). Consistent with the data presented above, the control *mel*-Cuff:GFP fusion localizes to foci adjacent to these H3K9me3 domains. In contrast, *sim*-Cuff:GFP broadly co-localizes with H3K9me3 and to distinct foci embedded within these domains (Figures 1I and 1J). To determine if *sim*-Cuff disrupts localization of other RDC components, we labeled ovaries expressing the Cuff:GFP fusions for Rhi and Del (Figure 2A). Both proteins colocalized with both *mel*-Cuff and *sim*-Cuff, indicating that *sim*-Cuff recruits the RDC to bulk heterochromatin.

To assay RDC localization at the genome level, we performed chromatin immunoprecipitation sequencing (ChIP-seq) for Cuff and Rhi in *cuff* mutants expressing *mel*-Cuff or *sim*-Cuff. As shown in the Genome Browser view in Figure 2F, the *sim*-Cuff fusion shows reduced binding to the 42AB cluster relative to the *mel*-Cuff control, and this is accompanied by reduced Rhi binding. The scatterplots in Figures 2G and 2H compare Cuff and Rhi ChIP-seq enrichment at clusters, on rescue with *sim-cuff* (y axis) relative to the *mel-cuff* control (x axis). Rescue with *sim-cuff* leads to reduced cluster binding by Cuff and Rhi across the genome. Consistent with our cytological observations (Figures 1I, 1J, and 2A), *sim*-Cuff also shows enhanced binding to two A/T rich repeats associated with constitutive heterochromatin (Figure S2F).

D. simulans Cuff Traps a Cluster Transcription Complex

To identify protein interactions that are altered by amino acid substitutions in the *D. simulans* ortholog, we expressed GFP-tagged *sim*-Cuff and *mel*-Cuff in wild-type *D. melanogaster* ovaries, affinity-purified the fusion proteins using GFP-Trap beads, and identified differentially bound proteins using mass spectrometry (see STAR Methods). To quantify differences in binding, we calculated the ratio of iBAQ values relative to the GFP tag (Figures 3A and 3B). Under our precipitation conditions, which do not use cross-linkers, known piRNA pathway proteins did not co-precipitate with *mel*-Cuff (Figure 3A). However, Cuff co-localizes with Del and interacts with Del in yeast two-hybrid assays (Mohn et al., 2014). Together, these observations suggest that Cuff directly interacts with Del, but binding is relatively weak and does not survive our immunoprecipitation protocol. In striking contrast, Del was the second most abundant protein, following Cuff itself, in precipitates of *sim*-Cuff (Figures 3A and 3B). In addition, TRF2, which interacts with Del through the Moon, was the fourth most abundant co-precipitating protein. We did not identify Moon in *sim*-Cuff or *mel*-Cuff, as the low molecular weight of the protein makes detection using mass spectrometry difficult. Substitutions in the *sim*-Cuff protein thus stabilize a complex with *D. melanogaster* Del and TRF2, which is likely to include Moon.

These findings suggest that *sim*-Cuff could sequester essential piRNA biogenesis factors in stable complexes, inhibiting function. To test this hypothesis, we overexpressed *sim*-Cuff in wild-type females and assayed fertility, piRNA production, and gene and transposon expression. Relatively modest 2.6-fold overexpression of *sim*-Cuff, using the germline-specific *rhi* promoter, did not alter fertility (Figure 1B). However, 45-fold overexpression of *sim*-Cuff, using the *UASp* promoter and germline-specific *nanos*-Gal4 driver, induced maternal-effect lethality and embryonic dorsal appendage defects, which are characteristic of piRNA pathway mutations (Figures 3E and S5A). In contrast, overexpression of *mel*-Cuff did not compromise hatch rate or embryo patterning (Figures 3E and S5A). The somatic follicle cells that surround the developing *Drosophila* oocyte express piRNAs, which are produced through a Cuff-independent mechanism. Mutations that disrupt this somatic piRNA pathway arrest oogenesis and lead to production of rudimentary ovaries (Lin and Spradling, 1997). However, the phenotype induced by *sim*-Cuff overexpression in the germline and soma, using an *Act5C*-Gal4 driver, was identical to the phenotype induced by germline-specific overexpression. The *sim*-Cuff protein thus appears to disrupt a germline-specific function.

To determine if *sim*-cuff overexpression disrupts transposon silencing and piRNA biogenesis, we performed small and long RNA-seq. These studies show that *sim*-Cuff overexpression disrupts transposon silencing (Figure S5C) but produces only a modest reduction in transposon and cluster mapping piRNAs (Figure S5D). Overexpression of *sim*-Cuff in wild-type thus triggers genetically dominant defects in fertility, transposon silencing, and piRNA biogenesis, which are nearly identical to the recessive defects observed on rescue of *cuff* mutants with *sim-cuff* (Figures 1D, 1H, S5C, and S5D).

To gain insight into the molecular basis for this unusual combination of phenotypes, we immuno-precipitated the overexpressed *sim*-Cuff and *mel*-Cuff proteins and identified associated proteins using mass spectrometry. As observed with the *rhino* promoter-driven fusions, TRF2 co-precipitated with overexpressed *sim*-Cuff but not with the *mel*-Cuff control (Figure S4A). In addition, CtBP consistently showed enhanced binding to *sim*-Cuff relative to *mel*-Cuff (Figure S4A). CtBP is a conserved transcriptional co-repressor, initially identified as an adenovirus E1A binding protein, and subsequently implicated

(G and H) Scatterplots showing comparisons of ChIP/Input values for GFP-Cuff (G) and Rhi (H) at piRNA clusters in ovaries with genotypes *cuff* mutant expressing *sim-cuff* versus *mel-cuff*. The clusters with prominent Cuff or Rhi binding (RPKM > 2) in *cuff* mutant with *mel-cuff* control were used for analysis. Diagonal represents x = y. p value for differences obtained using Wilcoxon test.

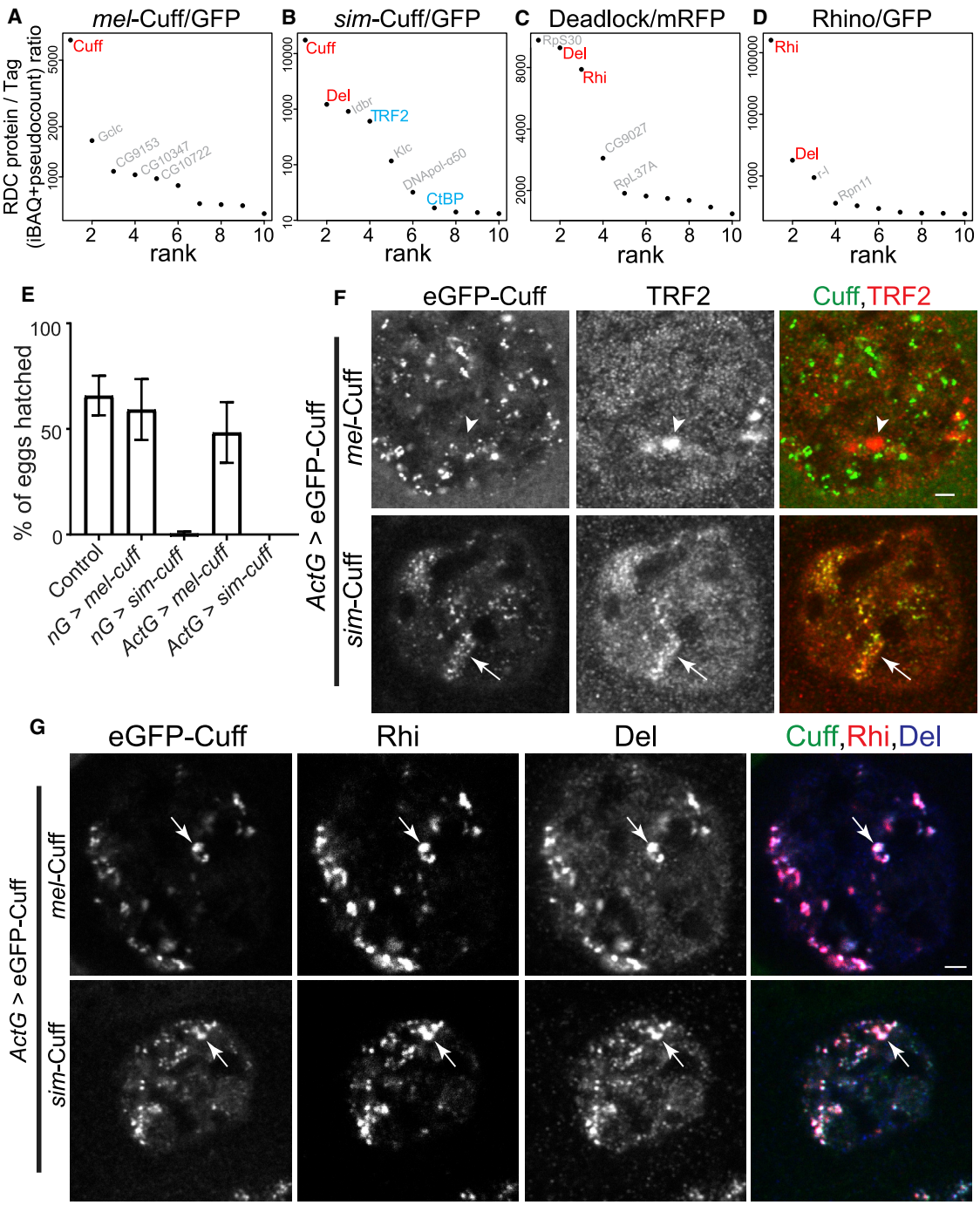


Figure 3. *D. simulans* Cuff Traps Transcription Factors and Acts as a Dominant Negative

(A–D) Mass spectrometric analysis of *mel*-Cuff (A), *sim*-Cuff (B), Del (C), and Rhi (D) binding proteins. Graphs show ratios of iBAQ value of a bound protein in a RDC protein IP versus tag control IP ranked by ratio values. RDC components are shown in red, TRF2 and CtBP in blue.

(E) Bar graphs showing percentages of hatched eggs produced by control (w^1; *Sp*/CyO) and flies overexpressing either *mel-cuff* or *sim-cuff* by either *nanos*-Gal4 (nG) or *Act5C*-Gal4 (*Act*-Gal4) drivers. Error bars show standard deviation of three biological replicates, with a minimum of 200 embryos scored per replicate, except for *nanos*-Gal4-driven *sim-cuff*, for which an average of 50 eggs were scored.

(F) Localization of overexpressed GFP-tagged Cuff with respect to TRF2 in the germline nuclei of *Act*-Gal4-driven *mel*-Cuff or *sim*-Cuff. Color assignments for merged images shown on top. Arrowheads and arrows denote locations of TRF2 foci. Scale bar, 2 μm.

(G) Localization of overexpressed GFP-tagged Cuff with respect to Rhi and Del in the germline nuclei of *Act*-Gal4-driven *mel*-Cuff or *sim*-Cuff. Color assignments for merged images shown on top. Arrows denote locations of RDC complex foci. Scale bar, 2 μm.

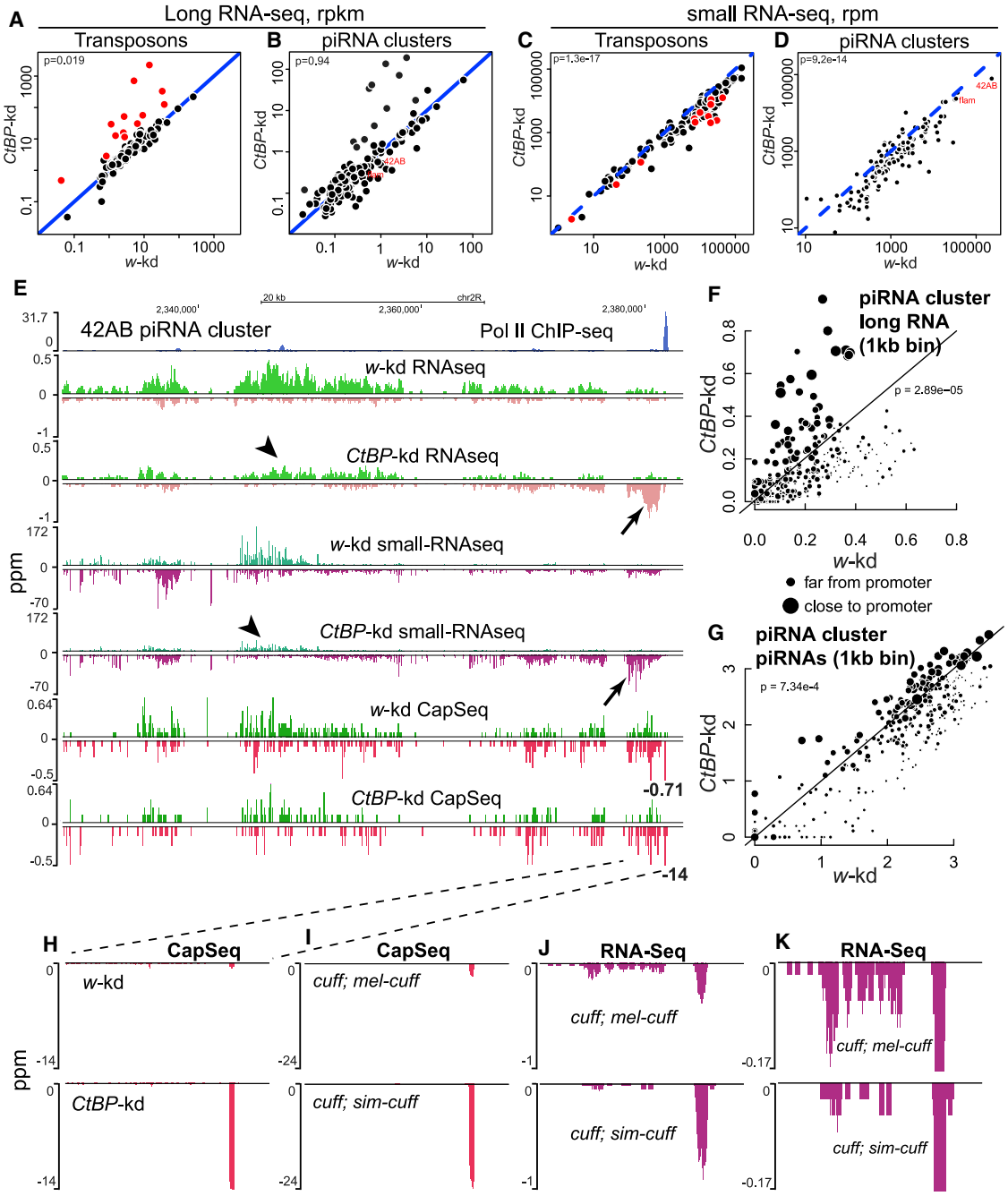


Figure 4. CtBP Suppresses Canonical Transcription at piRNA Clusters

(A and B) Scatterplots showing comparisons of RNA-seq signal for transposons (A) and piRNA clusters (B) in *CtBP*-kd versus *w*-kd ovaries. TEs with more than 3-fold overexpression in *CtBP*-kd versus *w*-kd as shown in red.

(C and D) Scatterplots showing comparisons of small RNA-seq signal for transposons (C) and piRNA clusters (D) in *CtBP*-kd versus *w*-kd ovaries. For transposon mapping plots, only anti-sense piRNAs are shown. Red points denote piRNA abundance for TEs that are overexpressed in *CtBP*-kd (A). Each point on the scatterplots shows RPKM or RPM value for a transposon family or a piRNA cluster. Diagonal represents x = y. p value for differences obtained using Wilcoxon test.

(E) Genome Browser view of RNA-seq (top), small RNA-seq (middle), and CapSeq (bottom) profiles at 42AB piRNA cluster from *w*-kd and *CtBP*-kd ovaries. Pol II ChIP-seq peak in *nanos*-Gal4-driven *mel*-Cuff ovaries marks the cluster promoter (blue). Arrows and arrowheads show the increase in canonical transcripts and decrease in non-canonical transcripts respectively after *CtBP*-kd. CapSeq profiles are saturated at promoters. The peak heights of CapSeq promoters are denoted by numbers next to the peaks.

(legend continued on next page)

in cancer and control of developmentally regulated genes (Boyd et al., 1993; Schaeper et al., 1995; Stankiewicz et al., 2014).

To determine if stable binding to *sim*-Cuff alters the distribution of these transcription factors, we immuno-localized TRF2 and CtBP in *cuff* mutants expressing low levels of *mel*-Cuff or *sim*-Cuff and in wild-type ovaries overexpressing *mel*-Cuff or *sim*-Cuff. In wild-type and *cuff* mutants expressing *mel*-Cuff, TRF2 localized to a few large nuclear domains, which did not overlap with RDC foci (Figures S4C and S4D). These large domains may represent histone repeats, which are regulated by TRF2 (Isogai et al., 2007). In *cuff* mutants expressing *sim*-Cuff, in contrast, TRF2 was displaced from these large foci (Figure S4D), and in wild-type ovaries overexpressing *sim*-Cuff, TRF2 colocalized with the overexpressed fusion protein (Figure 3F). Available primary antibodies did not allow direct co-localization of TRF2 with *sim*-Cuff, Del, and Rhi. However, in *sim-Cuff* overexpression background, *sim*-Cuff, Del, and Rhi co-localize (Figure 3G). Overexpression of *sim*-Cuff thus drives TRF2 into nuclear foci with the RDC. CtBP, in contrast, accumulates in the nucleus but does not localize to foci, in all of these backgrounds (data not shown). These cytological observations suggest that *sim*-Cuff associates with TRF2, Rhi, and Del nuclear foci, and with CtBP, in a distinct from, which is dispersed in the nucleus.

CtBP Inhibits Canonical Transcription of piRNA Clusters and Transposons

TRF2 linked to Del through Moon drives non-canonical cluster transcription (Andersen et al., 2017). The role of CtBP in the piRNA pathway, in contrast, has not been previously described. *CtBP* null mutants are lethal (Poortinga et al., 1998), so we used RNAi to knock down *CtBP* specifically in the germline. To confirm specificity, we used three different *CtBP* knockdown (*CtBP*-kd) lines, and a *white* knockdown (*w*-kd) control. The VDRC KK107313 line showed the strongest knockdown efficiency (Figure S6A), and the data obtained using this line are shown. The vast majority (89.5%) of eggs produced by control *w*-kd females hatch. In contrast, *CtBP*-kd reduced the hatch rate to 0.5% (Figure S6B), and RNA-seq revealed significant overexpression of 13 transposon families, but only modest changes in gene expression, including expression of known piRNA pathway genes (Figures 4A and S6C). This pattern is typical of piRNA pathway mutations. However, small RNA-seq showed that *CtBP*-kd produced only subtle reductions in cluster and transposon mapping piRNAs (Figures 4C and 4D). piRNA precursor transcripts also showed only modest reductions (Figure 4B). *CtBP*-kd, like *sim*-Cuff overexpression, thus disrupts transposon silencing without blocking piRNA biogenesis. These findings suggest that binding to *sim*-Cuff may inhibit CtBP, contributing to dominant sterility.

Most germline piRNA clusters are transcribed from internal non-canonical sites, but the right end of the 42AB cluster and both ends of the 38C cluster are transcribed from canonical promoters, which are marked by prominent RNA Pol II and TATA binding protein (TBP) ChIP-seq peaks (Figures 4E, S6E, 6A, and 6B). *CtBP*-kd produced relatively modest changes in total cluster transcript and piRNA levels, but long RNA and piRNA distributions near the promoters flanking the 42AB and 38C germline clusters were altered (Figures 4E and S6E). Close to the right end of 42AB, *CtBP*-kd produced a significant increase in minus strand long RNAs and piRNAs and a corresponding decrease in long RNAs and piRNAs from both strands in regions further downstream. A similar pattern was observed at both ends of 38C, where plus-strand long RNAs and piRNAs increased at the left flank, while minus strand long RNAs and piRNAs increase at the right flank (Figure S6E). To quantify these observations, we divided the 42AB and 38C clusters into 1 kb bins and generated a scatterplot comparing expression in each bin in *w*-kd and *CtBP*-kd, with point size decreasing with increasing distance from the flanking promoters (Figures 4F and 4G). For both long RNAs and piRNAs, *CtBP*-kd increased expression in bins close to the canonical promoters (large points), and decreased expression in bins away from promoters, which is driven by non-canonical transcription (small points). In contrast, the 80F cluster lacks flanking canonical promoters, and *CtBP*-kd did not change long RNA or small RNA expression across this cluster (Figure S6F). *Trf2* and *moonshiner* knockdown also increase piRNAs adjacent to the canonical promoter at the 42AB cluster, but these knockdowns result in global reduction in non-canonical piRNAs, including 80F cluster (Andersen et al., 2017). These findings suggest that TRF2 and Moon could function with CtBP to control canonical transcription. Alternatively, non-canonical transcription promoted by these proteins could inhibit canonical transcription.

To directly investigate the impact of CtBP on transcription initiation, we used CapSeq to quantify capped transcripts. On *CtBP*-kd, we observed a pronounced increase in capped transcripts associated with promoters flanking 42AB and 38C clusters. Significantly, we observed a similar increase in *cuff* mutant ovaries expressing *sim*-Cuff (Figures 4E, 4H, and 4I). *CtBP*-kd and replacement of *mel*-Cuff with the *sim*-Cuff ortholog thus activate canonical promoters flanking 42AB and 38C, which is associated with reduced non-canonical transcription from downstream sequences.

Heterochromatic clusters are the major source of germline piRNAs in *Drosophila* ovaries, but a subset of isolated euchromatic transposons function as ''mini piRNA clusters'' and are bound by Rhi and produce sense and anti-sense piRNAs (Figures 5A and 5B). Because transposon mobilization generates nearly identical insertions, internal sequences cannot be mapped to integration sites. However, Rhi spreads into flanking unique sequences from these insertions, leading to non-canonical transcription and piRNA production, resulting in a characteristic ''butterfly'' piRNA profile. To identify these loci, we first used paired-end genome sequencing to map all euchromatic

(F and G) Scatterplots showing comparisons of RPM values for 1 kb bins of piRNA clusters, which have RNA Pol II and TBP promoter peaks, for RNA-seq (F) and small RNA-seq (G) in *CtBP*-kd versus *w*-kd. The bins close to promoters are shown by large circles and ones farther away by small circles. p value for differences obtained using Wilcoxon test.

(H–K) Genome Browser views of CapSeq or RNA-seq signals at 42AB promoter for *CtBP*-kd versus *w*-kd (H) and *cuff* mutants expressing either *mel-cuff* or *sim-cuff* (I–K). (J) and (K) show RNA-seq profiles at different scales.

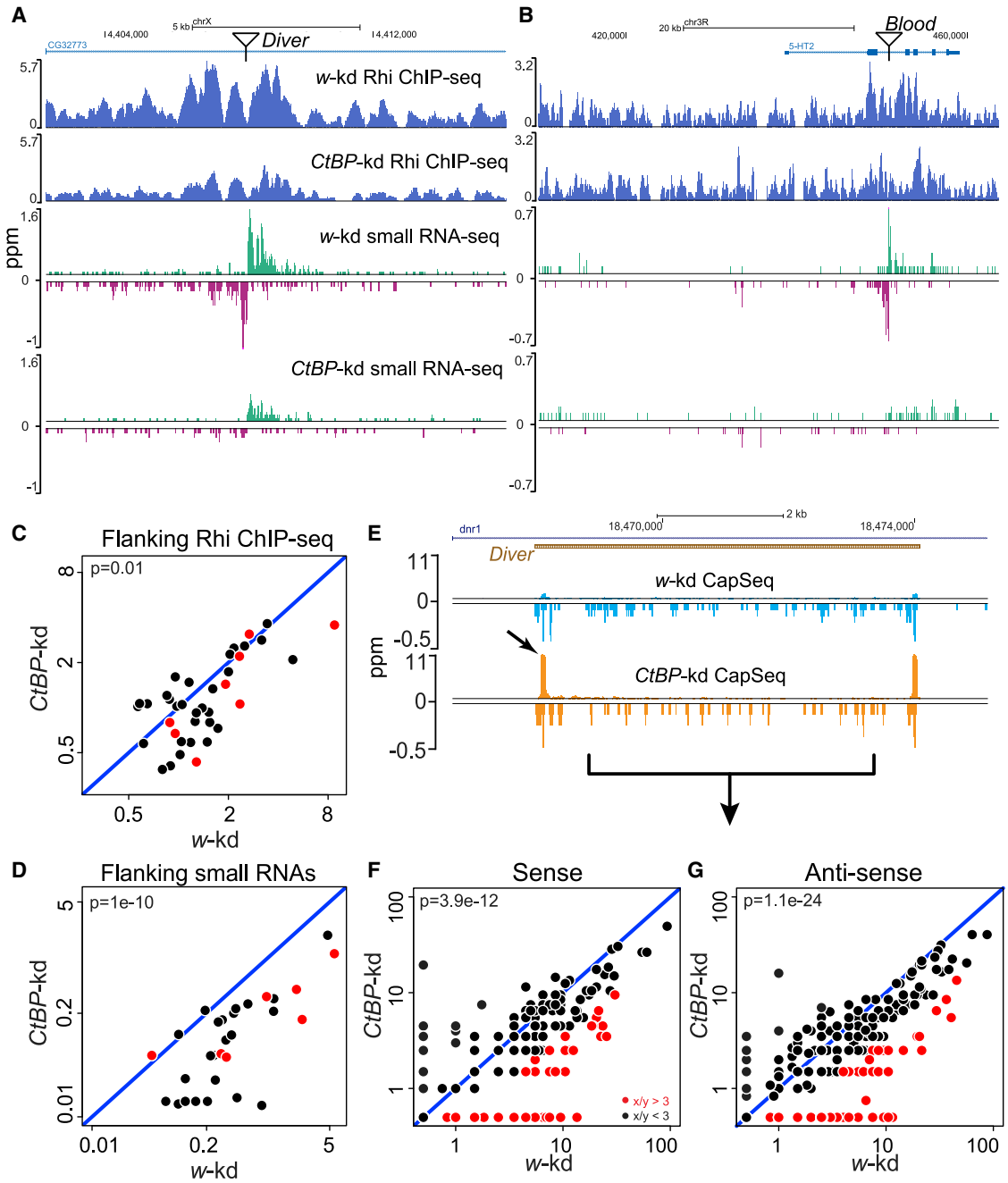


Figure 5. CtBP Suppresses Canonical Transcription of Dispersed Transposon Insertions

(A and B) Genome Browser views of Rhi ChIP-seq and small RNA-seq profiles flanking dispersed transposons, *Diver* (A) and *Blood* (B), in *CtBP*-kd and *w*-kd. The transposon insertion is shown at the top.

(C and D) Scatterplots showing comparisons of RPM values of Rhi ChIP-seq (C) and small RNAs (D), 0.5 kb upstream and downstream of new transposons in *CtBP*-kd versus *w*-kd. The transposons insertions were identified by genomic sequencing with TEMP (Zhuang et al., 2014), and the graphs show the values for new TEs (not present in the reference genome), which have both flanking piRNAs and Rhi signal. Red points denote expression of TEs overexpressed in *CtBP*-kd, as shown in Figure 4A. p value for differences obtained using Wilcoxon test.

(E) Genome Browser view of CapSeq signal at *Diver* insertion in *CtBP*-kd versus *w*-kd. Arrow shows increased CapSeq signal at *Diver* 5′ end in *CtBP*-kd. The signal shows all *Diver* insertion mapping reads and are not specific to this insertion.

(F and G) Scatterplots showing comparisons of CapSeq signal for 1 kb bins mapping to transposons present outside clusters, (bins at 5′ and 3′ ends are excluded, to remove canonical transcription peaks) for *CtBP*-kd versus *w*-kd. (F) shows sense strand and (G) shows anti-sense strand initiation. Points in red show x/y > 3. p value for differences obtained using Wilcoxon test.

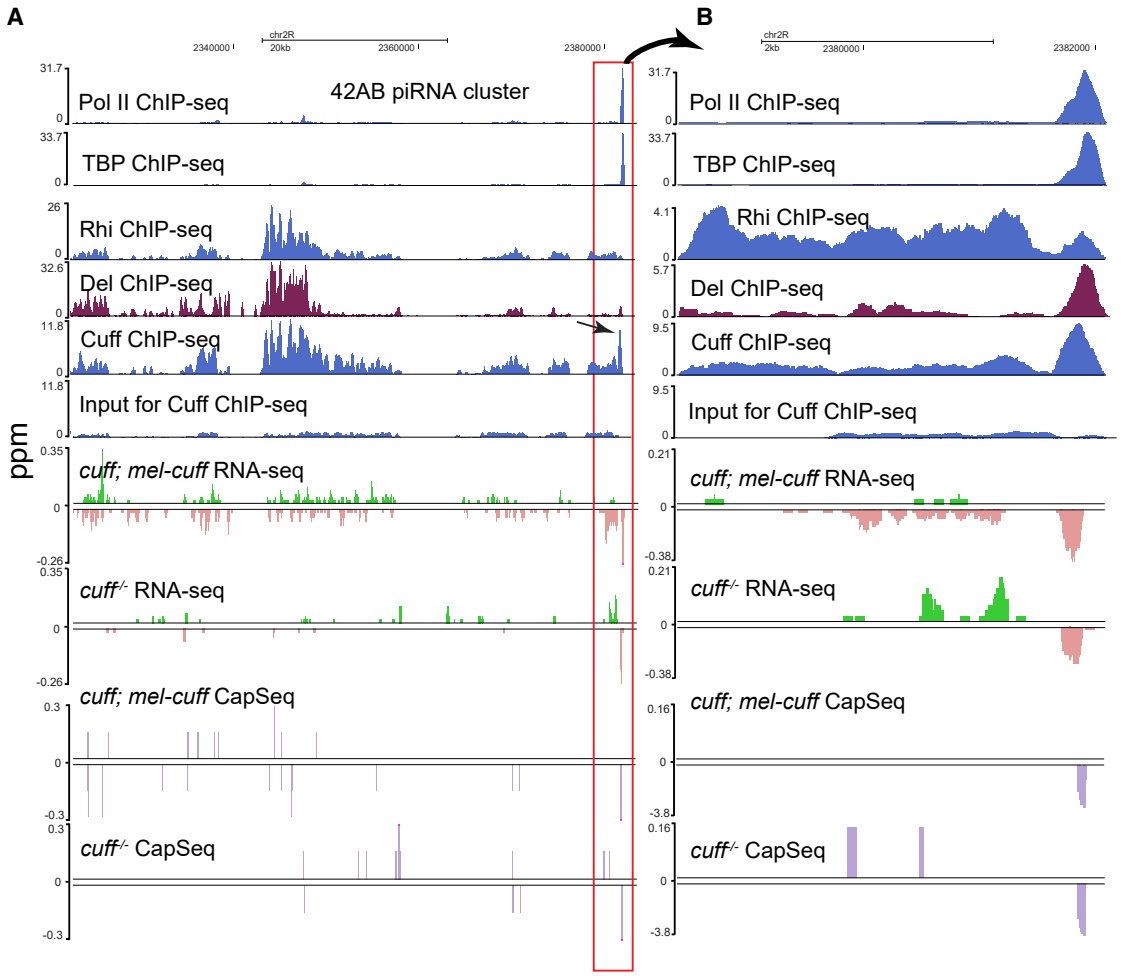


Figure 6. Role of Cuff in piRNA Cluster Transcription

(A and B) Genome Browser views at 42AB cluster. (A) Right side of the 42AB cluster, proximal to the flanking canonical promoter, showing Pol II, TBP (TATA binding protein), Rhi, Del, and Cuff ChIP-seq, and RNA-seq and CapSeq signals. Rhi, Del, and Cuff localize throughout the clusters, while Cuff and Del also show peaks that correspond to the flanking canonical promoter, marked by Pol II (arrow). (B) Zoomed-in view of the promoter region for all the tracks in (A). All the ChIP-seq tracks are auto-scaled, except for input track. RNA-seq and CapSeq profiles shown in *cuff* mutants and *cuff* mutants expressing *mel-cuff*.

transposon insertions and then identified the subset of insertions with flanking Rhi ChIP-seq peaks and divergently expressed piRNAs. Figures 5A and 5B show examples of *Diver* and *Blood* insertions that function as mini-clusters in the control *w*-kd line. In both cases, *CtBP*-kd reduced Rhi binding and triggered a near collapse of flaking piRNA expression. The scatterplots in Figures 5C and 5D summarize data for all of the new piRNA producing insertions identified by genomic DNA sequencing, showing that this loss of Rhi and piRNA production extends across the genome. CapSeq shows that the loss of Rhi binding and piRNA production is also associated with significant increases in canonical transcription from promoters within the long terminal repeats (LTRs) of the inserted elements (Figure 5E). In contrast, transcription initiation from within the transposons, which appears to reflect non-canonical transcription, is reduced for both sense and anti-sense strands (Figures 5F and 5G). CtBP thus suppresses canonical transcription from promoters linked to clusters and euchromatic transposon insertions. In both contexts, increased canonical transcription is associated with reductions in both non-canonical transcription and piRNA production.

Cuff Associates with Canonical and Non-canonical Transcription Sites

These data, with previous studies, link Cuff to factors that regulate canonical and non-canonical transcriptions of piRNA source loci. Further supporting this link, ChIP-seq shows that endogenous Cuff localizes with Pol II and TBP at canonical promoters flanking major germline clusters and confirms that Cuff co-localizes with Rhi and Del at sites of non-canonical transcription in the body of piRNA clusters (Figures 6A; Figure 6B is a zoomed-in view of the canonical transcription start in Figure 6A). Cuff, Rhi, and Del are co-dependent for cluster binding (Chen et al., 2016; Mohn et al., 2014). Consistent with these studies, long RNA and CapSeq indicate that *cuff* mutations reduce transcription from both strands of internal cluster sequences (Figure 6A), and ChIP-seq

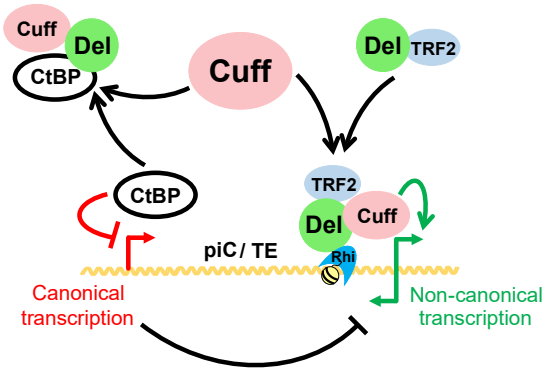


Figure 7. Model for a Transcriptional Network Balancing Canonical and Non-canonical piRNA Precursor Transcription

piRNAs are generated from both piRNA clusters and dispersed transposon insertions, which act as "mini-clusters." At both locations, Rhi binds to H3K9me3 histone marks and recruits Del, TRF2, and Cuff proteins, through direct or indirect interactions, to initiate non-canonical transcription from both strands. Non-canonical transcription (green lines) is inhibited by canonical transcription (red lines), and CtBP represses canonical transcription, regulating non-canonical transcription and piRNA production.

indicates that this correlates with reduced Rhi binding to 42AB and other germline piRNA clusters (Figures S7A and S7B). In contrast, *cuff* mutants did not reduce CapSeq signal associated with the canonical promoters flanking the 42AB (Figure 6B) or 38C clusters (Figure S7C). However, the transcripts from these canonical promoters are terminated shortly after initiation (Figure 6B), and tethering *Cuff* to the 3′ end of a reporter transcript enhances read-through transcription (Chen et al., 2016). These findings suggest that endogenous Cuff suppresses termination of transcription from canonical promoters flanking the major germline clusters but does not directly control transcription initiation from these promoters. In contrast, rescue of *cuff* mutants with the *sim*-Cuff ortholog leads to 7.7- and 2.3-fold expression of capped transcripts from the 42AB and 38C promoters, respectively (Figures 4I and S7C). As CtBP-kd also increases initiation from these promoters, we speculate that this increase is due *to sim*-Cuff binding to CtBP, leading to partial inhibition.

DISCUSSION

Adaptive evolution is a hallmark of genes engaged in a genetic conflict (Daugherty and Malik, 2012), which typically leads to co-evolution of host-pathogen gene pairs that encode interacting proteins (Elde and Malik, 2009). However, pathogens can also produce mimics that target interactions within host defense systems (Daugherty and Malik, 2012), raising the possibility that adaptation can also remodel interaction between host proteins. Supporting the possibility, adaptive evolution has remodeled an interface between the Rhi and Del, which are core components of the host transposon defense machinery (Parhad et al., 2017; Yu et al., 2018). These adaptive changes prevent gene function across closely relates species and define an interaction that is required to restrict the RDC to piRNA clusters, which defines the specificity of the transposon silencing machinery. These findings suggest that adaptive evolution targets important functional domains, which can be functionally analyzed using cross-species complementation. Here we apply this approach to the third RDC component, *cuff*, and show that adaptive evolution targets interactions between this Rai1 homolog and proteins that coordinate canonical and non-canonical piRNA cluster transcription and piRNA biogenesis.

sim-Cuff Captures piRNA Precursor Transcription Factors

Transposon silencing piRNAs are derived from heterochromatic clusters and a subset of euchromatic transposon insertions, and Cuff co-localizes with Rhi and Del at these piRNA source loci (Mohn et al., 2014). Rhi binds to H3K9me3 marks and recruits Del. Del, in turn, binds Moon, which recruits TRF2 to initiate non-canonical transcription from both genomic strands (Andersen et al., 2017; Le Thomas et al., 2014; Mohn et al., 2014; Yu et al., 2015). In contrast, our data suggest that Cuff coordinates canonical and non-canonical cluster expression. We show that the *D. simulans cuff* ortholog fails to rescue *D. melanogaster cuff* mutations and leads to dominant sterility when overexpressed in wild-type flies (Figures 1, 2, and 3). Significantly, these phenotypes are associated with stable binding to Del, TRF2, and CtBP. As noted above, Del and TRF2 function in non-canonical transcription of piRNA clusters (Andersen et al., 2017). CtBP is a conserved transcriptional co-repressor, first identified as a host factor that binds to Adenovirus E1a, and subsequently shown to function in numerous developmental pathways (Chinnadurai, 2003, 2007). CtBP does not directly interact with DNA, but binds sequence specific transcription factors and recruits histone-modifying enzymes (Chinnadurai, 2003, 2007). We show that CtBP-kd activates canonical promoters linked to piRNA source loci (Figure 4). Adaptive evolution has therefore remodeled interactions between Cuff and factors that control both canonical and non-canonical transcription of piRNA precursors loci.

Dominant phenotypes can result from mutations that produce new interactions or functions (neomorphic mutations) and assembly of complexes that are not formed by wild-type proteins (Jeffery, 2011). However, our findings, with previous studies, suggest that substitutions in sim-Cuff stabilize normally transient complexes with both TRF2 and CtBP. In *D. melanogaster*, Cuff and Del do not co-precipitate, but the proteins co-localize to nuclear foci, interact in two-hybrid assays, and are co-dependent for association with piRNA clusters (Mohn et al., 2014). Del, in turn, co-precipitates with TRF2 and Moon, and all three proteins are required for non-canonical cluster transcription (Andersen et al., 2017), but TRF2 does not normally accumulate at clusters (Figure 3F). In contrast, overexpression of *sim*-Cuff drives TRF2 co-localization with the RDC (Figure 3). Similarly, ChIP-seq shows that Cuff and Del localize to canonical promoters that are suppressed by CtBP, but CtBP does not accumulate at these promoters (S.S.P. and W.E.T., unpublished data). Substitutions in the *sim*-Cuff ortholog thus appear to stabilize normally transient associations with Del and TRF2 and with CtBP.

The majority of *Drosophila* germline clusters are transcribed from internal non-canonical initiation sites and do not have flanking canonical promoters. CtBP-kd does not significantly alter long RNA or piRNA expression from these loci. However, canonical promoters flank the right side of the 42AB cluster and both

ends of the 38C cluster, and CtBP-kd increases transcription from these canonical promoters, which is associated with reduced transcription and piRNA production from downstream regions, (Figures 4F and 4G). We cannot directly assay non-canonical transcription at most transposon insertions that produce piRNAs, as the inserted sequences are repeated, but CtBP-kd increases canonical transcription of transposons and is linked to collapse of piRNAs mapping to sequences flanking these insertions (Figure 5). In addition, deletion of the promoters flanking 42AB and 38C leads to spreading of piRNA production into flanking domains (Andersen et al., 2017). Together, these findings indicate that canonical transcription directly or indirectly represses non-canonical transcription and piRNA production.

A piRNA Precursor Transcription Network

On the basis of these findings, we propose that Cuff coordinates canonical and non-canonical piRNA precursor transcription (Figure 7). By stabilizing Rhi, Del, Moon and TRF2, Cuff promotes non-canonical transcription. By contrast, Cuff appears to function with CtBP to control canonical transcription. Rescue of *cuff* mutants with *sim*-Cuff, which shows enhanced binding to CtBP, is phenocopied by CtBP-kd: both lead to increased canonical transcription (CapSeq; Figures 4H and 4I). Formation of stable complexes with sim-Cuff thus appears to inhibit CtBP, activating canonical transcription and reducing downstream non-canonical transcription. Normally, the interaction between Cuff and CtBP is weak and free CtBP suppresses canonical promoters, while Cuff functions with Del-TRF2 to drive of non-canonical transcription. We speculate that this balance may be altered in response to stress or environmental signals, which can activate transposons (Maze et al., 2011; Miousse et al., 2015; Nätt and Thorsell, 2016). Intriguingly, CtBP is also an NADH/NAD binding protein (Fjeld et al., 2003; Jack et al., 2011), suggesting that the balance between canonical and non-canonical piRNA precursor transcriptions may be regulated in response to metabolic state.

The RDC proteins Moon and TRF2 are required for piRNA precursor transcription, and all of these factors are rapidly evolving (Figure S1B). By contrast, CtBP is conserved from flies to humans (Chinnadurai, 2002; Rabenstein et al., 1999), and a putative human oncogene (Dcona et al., 2017; Stankiewicz et al., 2014). The data presented here, with our earlier analysis of Rhi and Del (Parhad et al., 2017), indicate that rapid evolution has modified multiple interactions between rapidly evolving proteins in the piRNA biogenesis, and association of these proteins with a highly conserved transcriptional co-repressor. Rapidly evolving genes with specialized functions are frequently the most accessible to phenotype-based forward genetic approaches in model systems, and linking these specialized genes to conserved pathways can be a challenge. The studies reported here indicate that cross-species studies can help define these links, bridging the gap between genetically tractable model organisms and human biology.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Generation of transgenic flies
  - Fertility assays
  - RT-qPCR
  - Immuno-staining
  - Immuno-precipitation
  - Small RNA-seq
  - RNA-seq
  - CapSeq
  - ChIP-seq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Bioinformatics analysis
  - Mass spectrometry Proteomic Analysis
  - Analysis of RT-qPCR data
- STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.01.109.

ACKNOWLEDGMENTS

We would like to thank the members of Weng and Theurkauf labs and the UMass RNA Biology community for their insightful discussions and critical comments; Trudi Schüpbach for *cuff* stocks; Julius Brennecke for Del antibody; James Kadonaga for TRF2 antibody; Bloomington, Vienna Drosophila Resource Center (VDRC), and University of California, San Diego (UCSD), *Drosophila* stock centers for fly stocks; and John Leszyk and the UMass Proteomics facility for mass spectrometry. This work was supported by the National Institute of Child Health and Human Development (R01HD049116 and P01HD078253).

AUTHOR CONTRIBUTIONS

S.S.P., W.E.T., and Z.W. conceived the project. S.S.P. performed all the experiments, except *CtBP*-kd and *w*-kd RNA-seq and small RNA-seq, performed by G.Z. and N.P.R., respectively. S.S.P. and T.Y. performed bioinformatics analysis. S.S.P. and W.E.T. wrote the paper with help from Z.W. and the other authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 17, 2019
Revised: October 23, 2019
Accepted: January 30, 2020
Published: February 25, 2020

REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Andersen, P.R., Tirian, L., Vunjak, M., and Brennecke, J. (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. Nature *549*, 54–59.

Ayarpadikannan, S., and Kim, H.S. (2014). The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases. Genomics Inform. *12*, 98–104.

Barbash, D.A. (2010). Ninety years of Drosophila melanogaster hybrids. Genetics *186*, 1–8.

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

Biémont, C., and Vieira, C. (2006). Genetics: junk DNA as an evolutionary force. Nature *443*, 521–524.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc. Natl. Acad. Sci. U S A *104*, 3312–3317.

Blumenstiel, J.P., Erwin, A.A., and Hemmer, L.W. (2016). What drives positive selection in the *Drosophila* piRNA machinery? The genomic autoimmunity hypothesis. Yale J. Biol. Med. *89*, 499–512.

Boyd, J.M., Subramanian, T., Schaeper, U., La Regina, M., Bayley, S., and Chinnadurai, G. (1993). A region in the C-terminus of adenovirus 2/5 E1a protein is required for association with a cellular phosphoprotein and important for the negative modulation of T24-ras mediated transformation, tumorigenesis and metastasis. EMBO J. *12*, 469–478.

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

Canapa, A., Barucca, M., Biscotti, M.A., Forconi, M., and Olmo, E. (2015). Transposons, genome size, and evolutionary insights in animals. Cytogenet. Genome Res. *147*, 217–239.

Chang, T.H., Mattei, E., Gainetdinov, I., Colpan, C., Weng, Z., and Zamore, P.D. (2019). Maelstrom represses canonical polymerase II transcription within bi-directional piRNA clusters in Drosophila melanogaster. Mol. Cell *73*, 291–303.e6.

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.

Chen, Y.A., Stuwe, E., Luo, Y., Ninova, M., Le Thomas, A., Rozhavskaya, E., Li, S., Vempati, S., Laver, J.D., Patel, D.J., et al. (2016). Cutoff suppresses RNA polymerase II termination to ensure expression of piRNA precursors. Mol. Cell *63*, 97–109.

Chinnadurai, G. (2002). CtBP, an unconventional transcriptional corepressor in development and oncogenesis. Mol. Cell *9*, 213–224.

Chinnadurai, G. (2003). CtBP family proteins: more than transcriptional corepressors. BioEssays *25*, 9–12.

Chinnadurai, G. (2007). Transcriptional regulation by C-terminal binding proteins. Int. J. Biochem. Cell Biol. *39*, 1593–1607.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. Nat. Rev. Genet. *18*, 71–86.

Daugherty, M.D., and Malik, H.S. (2012). Rules of engagement: molecular insights from host-virus arms races. Annu. Rev. Genet. *46*, 677–700.

Dcona, M.M., Morris, B.L., Ellis, K.C., and Grossman, S.R. (2017). CtBP—an emerging oncogene and novel small molecule drug target: advances in the understanding of its oncogenic action and identification of therapeutic inhibitors. Cancer Biol. Ther. *18*, 379–391.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Elde, N.C., and Malik, H.S. (2009). The evolutionary conundrum of pathogen mimicry. Nat. Rev. Microbiol. *7*, 787–797.

Fjeld, C.C., Birdsong, W.T., and Goodman, R.H. (2003). Differential binding of NAD+ and NADH allows the transcriptional corepressor carboxyl-terminal binding protein to serve as a metabolic sensor. Proc. Natl. Acad. Sci. U S A *100*, 9202–9207.

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

Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. Nat. Rev. Genet. *10*, 94–108.

Goldman, N., and Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. *11*, 725–736.

Gu, W., Lee, H.C., Chaves, D., Youngman, E.M., Pazour, G.J., Conte, D., Jr., and Mello, C.C. (2012). CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as C. elegans piRNA precursors. Cell *151*, 1488–1500.

Hedges, D.J., and Deininger, P.L. (2007). Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. Mutat. Res. *616*, 46–59.

Horváth, V., Merenciano, M., and González, J. (2017). Revisiting the relationship between transposable elements and the eukaryotic stress response. Trends Genet. *33*, 832–841.

Isogai, Y., Keles, S., Prestel, M., Hochheimer, A., and Tjian, R. (2007). Transcription of histone gene cluster by differential core-promoter factors. Genes Dev. *21*, 2936–2949.

Jack, B.H., Pearson, R.C., and Crossley, M. (2011). C-terminal binding protein: a metabolic sensor implicated in regulating adipogenesis. Int. J. Biochem. Cell Biol. *43*, 693–696.

Jangam, D., Feschotte, C., and Betrán, E. (2017). Transposable element domestication as an adaptation to evolutionary conflicts. Trends Genet. *33*, 817–831.

Jeffery, C.J. (2011). Proteins with neomorphic moonlighting functions in disease. IUBMB Life *63*, 489–494.

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

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.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Klattenhoff, C., Bratu, D.P., McGinnis-Schultz, N., Koppetsch, B.S., Cook, H.A., and Theurkauf, W.E. (2007). Drosophila rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. Dev. Cell *12*, 45–55.

Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, 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.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

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

Lee, Y.C., and Langley, C.H. (2012). Long-term and short-term evolutionary impacts of transposable elements on Drosophila. Genetics *192*, 1411–1432.

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

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.

Lin, H., and Spradling, A.C. (1997). A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. Development *124*, 2463–2476.

Mani-Telang, P., Sutrias-Grau, M., Williams, G., and Arnosti, D.N. (2007). Role of NAD binding and catalytic residues in the C-terminal binding protein corepressor. FEBS Lett. *581*, 5241–5246.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 3.

Maze, I., Feng, J., Wilkinson, M.B., Sun, H., Shen, L., and Nestler, E.J. (2011). Cocaine dynamically regulates heterochromatin and repetitive element unsilencing in nucleus accumbens. Proc. Natl. Acad. Sci. USA *108*, 3035–3040.

McKim, K.S., Joyce, E.F., and Jang, J.K. (2009). Cytological analysis of meiosis in fixed Drosophila ovaries. Methods Mol. Biol. *558*, 197–216.

Miousse, I.R., Chalbot, M.C., Lumen, A., Ferguson, A., Kavouras, I.G., and Koturbash, I. (2015). Response of transposable elements to environmental stressors. Mutat. Res. Rev. Mutat. Res. *765*, 19–39.

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.

Nätt, D., and Thorsell, A. (2016). Stress-induced transposon reactivation: a mediator or an estimator of allostatic load? Environ. Epigenet. *2*, dvw015.

Pane, A., Jiang, P., Zhao, D.Y., Singh, M., and Schüpbach, T. (2011). The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline. EMBO J. *30*, 4601–4615.

Parhad, S.S., and Theurkauf, W.E. (2019). Rapid evolution and conserved function of the piRNA pathway. Open Biol. *9*, 180181.

Parhad, S.S., Tu, S., Weng, Z., and Theurkauf, W.E. (2017). Adaptive evolution leads to cross-species incompatibility in the piRNA transposon silencing machinery. Dev. Cell *43*, 60–70.e5.

Piacentini, L., Fanti, L., Specchia, V., Bozzetti, M.P., Berloco, M., Palumbo, G., and Pimpinelli, S. (2014). Transposons, environmental changes, and heritable induced phenotypic variability. Chromosoma *123*, 345–354.

Poortinga, G., Watanabe, M., and Parkhurst, S.M. (1998). Drosophila CtBP: a Hairy-interacting protein required for embryonic segmentation and hairy-mediated transcriptional repression. EMBO J. *17*, 2067–2078.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rabenstein, M.D., Zhou, S., Lis, J.T., and Tjian, R. (1999). TATA box-binding protein (TBP)-related factor 2 (TRF2), a third member of the TBP family. Proc. Natl. Acad. Sci. U S A *96*, 4791–4796.

Schaeper, U., Boyd, J.M., Verma, S., Uhlmann, E., Subramanian, T., and Chinnadurai, G. (1995). Molecular cloning and characterization of a cellular phosphoprotein that interacts with a conserved C-terminal domain of adenovirus E1A involved in negative modulation of oncogenic transformation. Proc. Natl. Acad. Sci. U S A *92*, 10467–10471.

Schüpbach, T., and Wieschaus, E. (1989). Female sterile mutations on the second chromosome of Drosophila melanogaster. I. Maternal effect mutations. Genetics *121*, 101–117.

Schwanhäusser, 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.

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

Stankiewicz, T.R., Gray, J.J., Winter, A.N., and Linseman, D.A. (2014). C-terminal binding proteins: central players in development and disease. Biomol. Concepts *5*, 489–511.

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. *34*, W609–W612.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Vanderweyde, T., Apicco, D.J., Youmans-Kidder, K., Ash, P.E.A., Cook, C., Lummertz da Rocha, E., Jansen-West, K., Frame, A.A., Citro, A., Leszyk, J.D., et al. (2016). Interaction of tau with the RNA-binding protein TIA1 regulates tau pathophysiology and toxicity. Cell Rep. *15*, 1455–1466.

Vermaak, D., Henikoff, S., and Malik, H.S. (2005). Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in Drosophila. PLoS Genet. *1*, 96–108.

Weick, E.M., and Miska, E.A. (2014). piRNAs: from biogenesis to function. Development *141*, 3458–3471.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. *13*, 555–556.

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.

Yu, B., Lin, Y.A., Parhad, S.S., Jin, Z., Ma, J., Theurkauf, W.E., Zhang, Z.Z., and Huang, Y. (2018). Structural insights into Rhino-Deadlock complex for germline piRNA cluster specification. EMBO Rep. *19*, e45418.

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.

Zhang, Z., Wang, J., Schultz, N., Zhang, F., Parhad, S.S., Tu, S., Vreven, T., Zamore, P.D., Weng, Z., and Theurkauf, W.E. (2014). The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. Cell *157*, 1353–1363.

Zhang, G., Tu, S., Yu, T., Zhang, X.O., Parhad, S.S., Weng, Z., and Theurkauf, W.E. (2018). Co-dependent assembly of Drosophila piRNA precursor complexes and piRNA cluster heterochromatin. Cell Rep. *24*, 3413–3422.e4.

Zhuang, J., Wang, J., Theurkauf, W., and Weng, Z. (2014). TEMP: a computational method for analyzing transposable element polymorphism in populations. Nucleic Acids Res. *42*, 6826–6838.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
GFP Booster-ATTO488 (Immuno-staining, 1:200)	ChromoTek	Cat# gba488-100, RRID: AB_2631434
Rabbit anti-Del (Immuno-staining, 1:1000)	Julius Brennecke lab	RRID: AB_2568875
Rabbit anti-TRF2 (Immuno-staining, 1:500)	James Kadonaga lab	N/A
Guinea pig anti-Rhi (Immuno-staining, 1:1000)	Klattenhoff et al., 2009	RRID: AB_2568331
Rabbit anti-H3K9me3 (Immuno-staining, 1:1000)	abcam	Cat# ab8898, RRID:AB_306848
Mouse anti-RNA Pol II (for ChIP)	abcam	RRID:AB_2268549
Rabbit anti-TBP (for ChIP)	James Kadonaga lab	N/A
Rabbit anti-GFP (for ChIP)	ChromoTek	Cat# PABG1-100, RRID:AB_2749857
Chemicals, Peptides, and Recombinant Proteins		
Superscript III	ThermoFisher Scientific	Cat# 18080-085
dNTP mix	NEB	Cat# N0447L
Terminator 5′-Phosphate-Dependent Exonuclease	Lucigen	Cat# TER51020
CIP (Calf Intestinal Alkaline Phasphatase)	NEB	Cat# M0290L
DNaseI	NEB	Cat# M0303L
Tobacco Decapping Enzyme	Enzymax	Cat# 87
T4 RNA ligase	Ambion, Invitrogen	Cat# AM2141
RNase OUT	ThermoFisher Scientific	Cat# 10777-019
TURBO DNase	ThermoFisher Scientific	Cat# AM2238
dUTP mix	Bioline	Cat# BIO-39041
RNaseH	ThermoFisher Scientific	Cat# 18021-071
DNA polymerase I	NEB	Cat# M0209S
T4 DNA polymerase	NEB	Cat# M0203L
Klenow DNA polymerase	NEB	Cat# M0210S
T4 PNK	NEB	Cat# M0201L
Klenow 3′ to 5′ exo	NEB	Cat# M0212L
T4 DNA ligase	Enzymatics Inc.	Cat# L6030-HC-L
UDG	NEB	Cat# M0280S
Phusion Polymerase	NEB	Cat# M0530S
T4 RNA Ligase 2, truncated K227Q	NEB	Cat#M0351L
16% formaldehyde	Ted Pella Inc	Cat# 18505
Gateway® LR Clonase® Enzyme mix	ThermoFisher Scientific	Cat# 11791019
In-Fusion® HD Cloning Kit	Clontech	Cat# 639648
Critical Commercial Assays		
mirVANA miRNA isolation kit	ThermoFisher Scientific	Cat# AM1560
Dynabeads® Protein G	ThermoFisher Scientific	Cat# 10004D
Dynabeads® Protein A	ThermoFisher Scientific	Cat# 10002D
GFP-Trap®_A beads	Chromotek	Cat# gta-100
RNeasy Mini Kit	QIAGEN	Cat# 74104
RNA Clean & Concentrator-5	Zymo Research	Cat# R1015
ZR small-RNA PAGE Recovery Kit	Zymo Research	Cat# R1070
Ribo-Zero Gold rRNA removal kit	Illumina	Cat# MRZG12324
Deposited Data		
High throughput Sequencing	This study	NCBI SRA: PRJNA517772
Raw data	This study	Mendeley Data: https://doi.org/10.17632/6nd35djt9p.1

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Rhi and Del IP Mass Spectrometry Proteome	Parhad et al., 2017	N/A
Del ChIP-seq	Mohn et al., 2014	N/A
Experimental Models: Organisms/Strains		
D. melanogaster: *rhiP* > GFP-*mel*-Cuff	This study	N/A
D. melanogaster: *rhiP* > GFP-*sim*-Cuff	This study	N/A
D. melanogaster: *UASp* > GFP-*mel*-Cuff	This study	N/A
D. melanogaster: *UASp* > GFP-*sim*-Cuff	This study	N/A
D. melanogaster: $cuff^{KG/WM}$	Chen et al., 2007	N/A
D. melanogaster: $cuff^{QQ/WM}$	Chen et al., 2007	N/A
D. melanogaster: Oregon-R	William Theurkauf lab	N/A
D. melanogaster: Act5C > Gal4	William Theurkauf lab	N/A
D. melanogaster: nanos > Gal4	William Theurkauf lab	N/A
D. melanogaster: vasP > GFP-nls	Zhang et al., 2014	N/A
D. melanogaster: UAS-Dcr2;nos-Gal4	Bloomington	Cat # 25751
D. melanogaster: w-RNAi-kd	VDRC	Cat # GD30033
D. melanogaster: CtBP-RNAi-kd	VDRC	Cat # KK107313
D. melanogaster: CtBP-RNAi-kd	VDRC	Cat # GD37609
D. melanogaster: CtBP-RNAi-kd	VDRC	Cat # GD37608
Oligonucleotides		
Sequences given in Method details	Integrated DNA Technologies (IDT)	N/A
Random primers	ThermoFisher Scientific	Cat# 48190011
Recombinant DNA		
pENTR/D-TOPO®	ThermoFisher Scientific	Cat# K2400-20
Drosophila gateway vector: attB-pPGW	Parhad et al., 2017	N/A
Drosophila gateway vector: *rhiP*-attB-pPGW	Parhad et al., 2017	N/A
Software and Algorithms		
GraphPad Prism	https://www.graphpad.com/scientific-software/prism/	N/A
RStudio	https://rstudio.com/	N/A
Adobe Creative Suite 6	Adobe Systems Inc.	N/A
Scaffold	http://www.proteomesoftware.com/products/scaffold/	N/A
UCSC Genome Browser	https://genome.ucsc.edu/cgi-bin/hgGateway	N/A
Microsoft Office	Microsoft	N/A
cutadapt	Martin, 2011	N/A
Bowtie	Langmead et al., 2009	N/A
Bowtie2	Langmead and Salzberg, 2012	N/A
BEDTools	Quinlan and Hall, 2010	N/A
TopHat	Trapnell et al., 2009	N/A
STAR	Dobin et al., 2013	N/A
Hisat2	Kim et al., 2015	N/A
HTSeq	Anders et al., 2015	N/A
BWA	Li and Durbin, 2009	N/A
PAML	Goldman and Yang, 1994; Yang, 1997	N/A
PAL2NAL	Suyama et al., 2006	N/A
TEMP	Zhuang et al., 2014	N/A

LEAD CONTACT AND MATERIALS AVAILABILITY

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). All unique/stable reagents generated in this study are available from the Lead Contact without restriction.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All experiments were performed in 2-4 days old *Drosophila melanogaster* females, except mentioned otherwise. All flies were maintained at 25°C on cornmeal medium. All transgenic lines were generated by ϕC31 integration at 3L-68A4. $cuff^{WM25}$ ($cuff^{WM}$) and $cuff^{QQ37}$ ($cuff^{QQ}$) alleles were obtained from Trudi Schüpbach (Princeton University) (Chen et al., 2007). $cuff^{KG05951}$ ($cuff^{KG}$) was obtained from Bloomington (Stock # 14462). *Act5C*-Gal4 and *nanos*-Gal4 stocks were used from our lab stocks. RNAi knockdown lines were obtained from VDRC.

METHOD DETAILS

Generation of transgenic flies

mel-cuff was cloned from *D. melanogaster* OregonR ovary cDNA and *sim-cuff* from *Drosophila simulans* C167.4 ovary cDNA. The reverse primer for the PCR reaction was used for making cDNA with Superscript III reverse transcriptase (Thermo Fisher Scientific). *mel-cuff* was PCR amplified from cDNA by using forward primer: CAC CAT GAA TTC TAA TTA CAC AAT ATT AAA C and reverse primer: TTA AAC TAT AGA AGA CAT GGT TTG C and cloned into pENTR-D-TOPO vector by directional TOPO cloning kit (Thermo Fisher Scientific). Similarly, *sim-cuff* was PCR amplified from cDNA using forward primer: CAC CAT GAA TTC TAA TTA CAA AAT ATT GAA C and reverse primer: TTA TTG GTA AAC TGT GGA AGA CAT GG and cloned into pENTR-D-TOPO vector. These served as entry vectors for Gateway cloning. The destination vectors *rhiP*-attB-pPGW (for expressing N' GFP tagged proteins under *rhi* promoter) and attB-pPGW (for expressing N' GFP tagged proteins under *UASp* promoter) were used as described in Parhad et al. (2017). The plasmids obtained from LR gateway cloning reaction were sequenced and injected by ϕC31 integration at chromosomal location 3L-68A4 (Bischof et al., 2007).

Fertility assays

2-4 days old flies were maintained on grape juice agar plates for 1 or 2 days. After removing flies, the eggs were counted for fused appendages. The number of hatched eggs were counted after 2 days. The fertility bar graphs indicate mean and standard deviation from 3 biological replicates.

RT-qPCR

RNA was isolated from 2-4 days old female ovaries. Reverse transcription done using Superscript III reverse transcriptase with random primers. qPCR was done by QIAGEN QantiTect® SYBR® Green PCR mix using Applied Biosystems instrument. Primers sequences for CtBP: forward primer: CAA AAA TCT GAT GAT GCC GAA GCG TTC and reverse primer: AGG ATG GGC ATC TCG ATG GAG CAG TC and Rp49: forward primer: CCG CTT CAA GGG ACA GTA TCT G and reverse primer: ATC TCG CCG CAG TAA ACG C.

Immuno-staining

Immuno-staining and image analysis were performed as described in McKim et al. (2009) and Zhang et al. (2012a). In short, 2-4 days old female ovaries were dissected in Robb's buffer, fixed with 4% formaldehyde, washed, overnight incubated with primary antibody, washed, incubated overnight with secondary antibody with the fluorophore, stained with DAPI for DNA labeling and mounted on slide with mounting medium. To enhance the GFP signal, ChromoTek anti-GFP Booster (Atto-488) antibody was added with secondary antibody. Antibodies used: anti-GFP Booster (ChromoTek) at 1:200, guinea pig anti-Rhi (our lab) at 1:1000, rabbit anti-Del (from Julius Brennecke) at 1:1000, rabbit anti-TRF2 (from James Kadonaga) at 1:500, rabbit anti-H3K9me3 (abcam) at 1:1000.

Immuno-precipitation

IP was performed as described in Parhad et al. (2017). Briefly, 2-4 days old female ovaries were dissected in Robb's medium, lysed by homogenization and sonication and centrifuged to get input for IP. Lysis and IP buffer composition: HEPES (pH 7.5) 50mM, NaCl 150mM, $MgCl_2$ 3.2mM, NP-40 0.5%, PMSF 1mM, Proteinase Inhibitor (Roche) 1X. chromotek GFP-Trap®_A beads were used for GFP IP. The lysate was incubated with beads for 3 hours at 4°C and subsequently washed with lysis buffer 4 times. Finally the beads were suspended in SDS-PAGE lysis buffer. The procedure for mass spectrometry is descried in Vanderweyde et al. (2016). Briefly, the IP samples were resolved on a 10% SDS-PAGE gel. The gel pieces were trypsin digested to get the peptides, which were analyzed by LC-MS/MS. Rhi and Del IP data was used from Parhad et al. (2017).

Small RNA-seq

Small RNA libraries were prepared as mentioned in Zhang et al. (2014). In short, total RNA was prepared by mirVANA kit (Ambion). 18-30 nt length small RNAs were size selected by denaturing PAGE gel purification. These were ligated further at 3′ and 5′ ends by adapters. Reverse transcription and then PCR amplification was performed to obtain libraries. Single end sequencing was done by Illumina platform.

RNA-seq

RNA-seq libraries were prepared as described in Fu et al. (2018) and Zhang et al. (2012b, 2018). Briefly, RNA samples were depleted for ribosomal rRNA by Ribo-Zero kit (Illumina) or rRNA digestion by RNaseH (Epicenter), fragmented and reverse transcribed. After dUTP incorporation for strand specificity, end repair, A-tailing, adaptor ligation and PCR amplification was done to obtain libraries. Paired end sequencing was done by Illumina platform.

CapSeq

This method was performed to sequence 5′ ends of transcripts (Gu et al., 2012). In brief, total RNA was sequentially treated with Terminator 5′-Phosphate-Dependent Exonuclease, CIP (Calf Intestinal Alkaline Phasphatase), DNaseI, Tobacco Decapping Enzyme. After adaptor ligation at the 5′ end, reverse transcription (with primer: 5′-GCACCCGAGAATTCCANNNNNNNN-3′) and two rounds of PCR were done. The PCR products were gel purified after each PCR step. Final library was sequenced by Illumina platform by single end sequencing.

ChIP-seq

ChIP-seq was performed by method described in Parhad et al. (2017). In short, the ovaries were dissected in 1X Robb's medium and fixed with 2% formaldehyde and sonicated in Bioruptor for 2 hours. This lysate was centrifuged and the supernatant was used as input for ChIP. The input was precleared with either Dynabeads Protein A or Dynabeads Protein G (Invitrogen) and was added to the Dynabeads conjugated to an antibody and incubated overnight. After washing, the beads were reverse crosslinked, ChIP DNA was purified and libraries were prepared by end repair, A tailing, adaptor ligation and PCR amplification. Illumina platform was used for paired end sequencing.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bioinformatics analysis

Small RNA-seq reads were first fed into cutadapt (Martin, 2011) for 3′end adaptor trimming (adaptor sequence: TGGAATTCTCGGGTGCCAAGGAACTCCAGTCAC_Index_ATCTCGT). The adaptor removed reads were aligned to the *D. melanogaster* genome (dm3) and transposon consensus sequences by bowtie (Langmead et al., 2009) with parameters -v 1 -a–best–strata, after removing rRNA, miRNA, snRNA, snoRNA and tRNA mapping reads. This allowed 1 mismatch during mapping. Flybase r5.50 transcriptome annotations were used. The piRNA cluster coordinates were from Brennecke et al. (2007). The read counting was done using intersectBed module from BEDTools (Quinlan and Hall, 2010) and normalized to microRNAs. Multiple mapping reads are considered while counting reads and apportioned to their map times. For ping-pong analysis on cluster-mapping reads, the overlaps between all pairs of piRNAs that mapped to the opposite genomic strands were calculated, and then the Z-score for the 10-nt overlap was calculated using the 1-9 nt and 11-30 nt overlaps as the background (Li et al., 2009).

STAR (Trapnell et al., 2009) was used to align RNA-seq reads to the genome allowing 2 mismatches. rRNA reads were removed prior to the quantification of genes, piRNA clusters, and transposons expression via Hisat2 (Kim et al., 2015) with default parameters. The mapping results in the SAM format were transformed into sorted and duplication-removed BAM format using SAMtools (Li et al., 2009). The final mapped reads were assigned to protein-coding genes, non-coding RNAs, and piRNA genes using HTSeq (Anders et al., 2015), and the expression levels of these genes, in reads per million unique mapped reads in per thousand nucleotides (RPKM), were calculated using custom bash scripts. RNA-seq reads after removing rRNAs were also mapped to transposon consensus sequences using Hisat2 with default parameters. Then transposon expression levels were calculated using Bedtools.

For ChIP-seq, genome and transposon alignment was done by Bowtie2 (Langmead and Salzberg, 2012) with default parameters. The ChIP-seq signal in each transposon was indicated by the read counts per million total genome mapping reads per kilo base pairs.

CapSeq was processed like RNA-seq except RT primer removing before any alignment via cutadapt. Only 5′end of each read was considered for profile generating and signal calculating. Total uniquely genome mapped reads were used as the normalization factor.

Mass spectrometry Proteomic Analysis

Proteome Discoverer and Mascot Server were used to process the raw data before display on Scaffold Viewer (Proteome Software, Inc.). The proteins were filtered by criteria: Protein threshold: 90%, Min # peptides: 2, Peptide threshold: 90%. Then iBAQ values (Schwanhäusser et al., 2011) were obtained and pseudocount was added. For Cuff IP, *vas* promoter driven GFP-nls was used as a control. Both replicates of *rhi* promoter (*rhiP*) driven Cuff IP mass spectrometry scaffold tables were combined into a single file. To obtain list of proteins binding to Cuff and not GFP control, only proteins below the threshold of 300000 in GFP IP were selected. The proteins which show more than 3 fold enrichment in all the Cuff protein IPs versus GFP control IP were used to make plots, where

the ratios of (iBAQ + psuedocount) values for each identified protein in a Cuff IP versus GFP IP were plotted against their rank. For sim-Cuff graphs, in addition to the above filters, proteins which show more than 3 fold enrichment for sim-Cuff IP versus mel-Cuff IP were plotted. The graphs were made using R. Similar filters and thresholds were used for Rhi and Del IP mass spectrometry data from Parhad et al. (2017).

Analysis of RT-qPCR data

Quantification done using ΔCt method. Rp49 served as the loading control.

STATISTICAL ANALYSIS

The error bars in the bar graphs show standard deviations from 3 biological replicates.

DATA AND CODE AVAILABILITY

Cloned cuff cDNA sequences and Cuff Proteomics data are deposited in Mendeley Data: https://doi.org/10.17632/6nd35djt9p.1. High throughput sequencing data can be accessed from NCBI SRA: PRJNA517772.