ARTICLE

doi:10.1038/nature23482

A heterochromatin-dependent transcription machinery drives piRNA expression

Peter Refsing Andersen[1], Laszlo Tirian[1], Milica Vunjak[1] & Julius Brennecke[1]

Nuclear small RNA pathways safeguard genome integrity by establishing transcription-repressing heterochromatin at transposable elements. This inevitably also targets the transposon-rich source loci of the small RNAs themselves. How small RNA source loci are efficiently transcribed while transposon promoters are potently silenced is not understood. Here we show that, in *Drosophila*, transcription of PIWI-interacting RNA (piRNA) clusters—small RNA source loci in animal gonads—is enforced through RNA polymerase II pre-initiation complex formation within repressive heterochromatin. This is accomplished through Moonshiner, a paralogue of a basal transcription factor IIA (TFIIA) subunit, which is recruited to piRNA clusters via the heterochromatin protein-1 variant Rhino. Moonshiner triggers transcription initiation within piRNA clusters by recruiting the TATA-box binding protein (TBP)-related factor TRF2, an animal TFIID core variant. Thus, transcription of heterochromatic small RNA source loci relies on direct recruitment of the core transcriptional machinery to DNA via histone marks rather than sequence motifs, a concept that we argue is a recurring theme in evolution.

Eukaryotic genome integrity relies on repression of transcription and recombination at transposon insertions and other repeats through heterochromatin formation[1]. In plants, fungi, and animals, sequence-specific heterochromatin formation depends on small RNA pathways[2,3]. These act through RNA-induced silencing complexes composed of an Argonaute protein and a small guide RNA. While small RNA-mediated silencing allows repression of transposable elements throughout the genome, it poses an inherent paradox: how do the transposon-rich small RNA source loci escape transcriptional silencing to sustain ongoing small RNA biogenesis?

In animals, the central genome defence small RNA pathway is the piRNA pathway. It acts in gonads and targets transposons at the transcriptional and post-transcriptional levels via PIWI-clade Argonautes bound to 22–30 nucleotide (nt)-long piRNAs[4,5]. piRNAs originate from transposon-rich genomic loci termed piRNA clusters. In *Drosophila melanogaster*, most piRNA clusters are bidirectionally transcribed and yield piRNAs from both genomic strands[6,7] (also termed 'dual-strand' clusters). For this reason, such clusters are always targeted by the piRNAs they produce and, indeed, bidirectional piRNA clusters exhibit signatures of transcriptional silencing, such as histone3 lysine9 tri-methylation (H3K9me3)[6,8]. How this silencing is compatible with transcription of small RNA precursors is not understood. A key molecule for piRNA cluster transcription is Rhino, a heterochromatin protein-1 (HP1) paralogue that is specifically enriched at bidirectional piRNA loci[6,9,10]. However, how Rhino licenses transcription at piRNA clusters remains unknown.

Transcription by RNA polymerase II (Pol II) is facilitated by basal transcription factors, which direct the stepwise assembly of the pre-initiation complex on core promoter sequences[11]. The first step in this assembly is the positioning of the basal transcription factor complex TFIID with its central component TBP on the core promoter DNA. At this stage TFIIA stabilizes the binding of TFIID–TBP to DNA resulting in a 'committed' complex[12,13]. Recruitment of TFIID–TBP to promoters is mediated by transcription factors that bind DNA motifs in enhancer and promoter regions. Given that heterochromatin restricts DNA accessibility, the transcription of small RNA loci, particularly transcription initiation, must follow alternatives routes.

Here we uncover a pathway that enables transcription initiation within heterochromatin, resulting in the production of piRNA precursors. Central to this pathway is a TFIIA–TFIID variant complex that acts specifically at Rhino-bound piRNA clusters. It involves CG12721, a germline-specific TFIIA-L paralogue, which we name Moonshiner for its activity under the transcriptional 'prohibition' of heterochromatin. Moonshiner interacts with the Rhino-associated protein Deadlock and activates transcription by recruiting TRF2 to chromatin. Our data show that piRNA precursors in *Drosophila* originate via widespread transcription initiation within piRNA clusters, which is mediated by a coupling between heterochromatic histone marks and the Pol II pre-initiation complex.

Transcription initiation within piRNA clusters

Drosophila bidirectional piRNA clusters are transcribed by Pol II, yet lack discernible promoters and are enriched in H3K9me3 marks. Two models of how Pol II transcribes these loci have been proposed[6]. In one, Pol II enters the loci by read-through transcription from flanking genes (Fig. 1a, left). Indeed, bidirectional piRNA clusters are often flanked by transcribed genes pointing towards the cluster. Furthermore, the Rhino-associated protein Cutoff possesses transcription anti-termination function[6,14]. In the other model, Pol II transcribes piRNA loci via pervasive internal transcription initiation (Fig. 1a, right).

We tested the read-through model by deleting the promoters of *Pld*, which flanks *cluster42AB*, the largest bidirectional piRNA locus (Fig. 1b). In homozygous *Pld*-Δpromoter flies, *cluster42AB* piRNA levels are not changed (see also ref. 14). Instead, ectopic small RNAs are now produced within the *Pld* locus, an effect that is amplified by expression of a *Pld* cDNA *in trans* (Fig. 1b, Extended Data Fig. 1a and Supplementary Note 1). This suggests that *cluster42AB* spreads into the promoter-less *Pld* locus, resulting in bidirectional transcription of small RNA precursors. Indeed, Rhino occupancy within *Pld* is elevated in *Pld*-Δpromoter flies (Fig. 1c). We obtained similar results at *cluster80F* (Extended Data Fig. 1b). In conclusion, bidirectional piRNA cluster expression does not rely on read-through transcription. Instead, flanking transcription units delimit piRNA clusters.

To test the internal initiation model, we searched for signatures of transcription initiation within piRNA clusters. We determined transcription start sites at nucleotide resolution by Cap sequencing (Cap-seq)[15] (Extended Data Fig. 1c). This uncovered more than

[1]Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna Biocenter (VBC), Dr. Bohrgasse 3, 1030 Vienna, Austria.

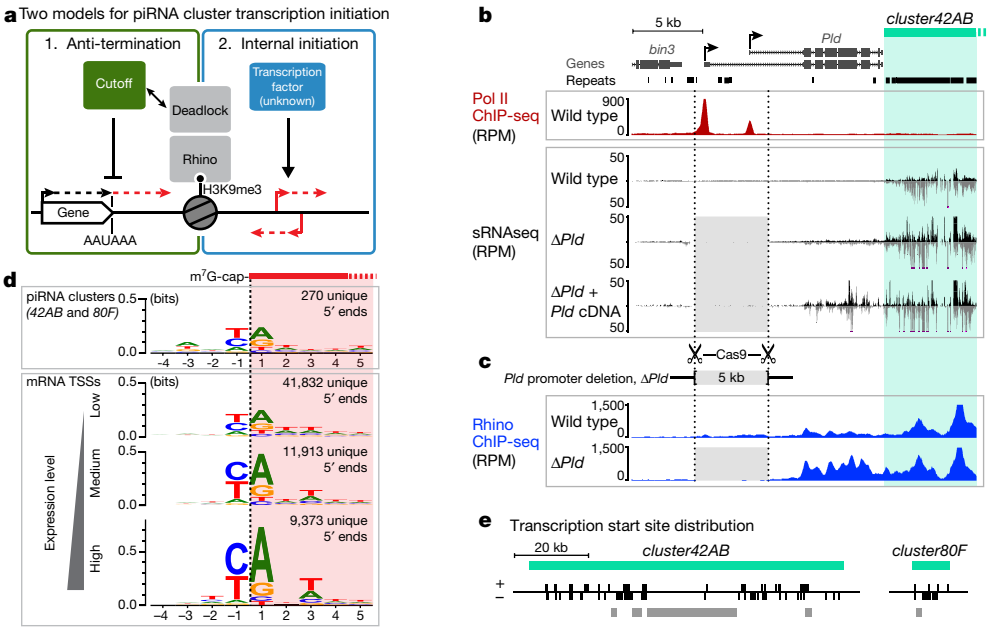


Figure 1 | Heterochromatic piRNA source loci utilize internal transcription initiation sites. a, Two models for transcription initiation at Rhino-dependent piRNA source loci: read-through from flanking genes aided by Cutoff (left) or internal initiation (right). **b**, **c**, Genome browser panels showing investigation of flanking promoter dependency for the centromere-proximal part of *cluster42AB*. Shown are piRNA levels, Rhino occupancy, and Pol II occupancy. Light grey shading indicates the promoter deletion; RPM, reads per million. **d**, DNA sequence motif at 5′ ends of capped *cluster42AB*- and *cluster80F*-derived RNAs compared with that of mRNA 5′ ends (binned by expression strength). TSSs, transcription start sites. **e**, Distribution of transcription start sites with 'YR'-motif inside *cluster42AB/80F* (grey bars: regions with low mappability).

200 putative transcription start sites within *cluster42AB* and *cluster80F* and an additional ~500 in all other Rhino-occupied loci. These are enriched for 'YR' dinucleotides at the −1/+1 positions (Fig. 1d and Extended Data Fig. 1d), a signature of the initiator element, a central core promoter motif that is bound by TFIID during pre-initiation complex assembly[16,17]. When cloning RNA 5′ ends with mono-phosphate groups instead of a cap structure from the same RNA sample, the known piRNA biogenesis signatures of uridine and adenosine residues at positions +1 and +10, respectively, emerged[7,18] (Extended Data Fig. 1e). This shows that the YR signature is not a feature of piRNA processing intermediates. Almost 60% of the putative *cluster42AB/80F* transcription start sites harbour the YR motif (Extended Data Fig. 1f), and these are distributed on both strands over the entire clusters (Fig. 1e). Taken together, our data reveal widespread Pol II transcription initiation within heterochromatic piRNA clusters.

An alternative TFIIA–TRF2 complex at piRNA clusters

To identify factors required for transcription of heterochromatic piRNA clusters, we searched a transposon de-repression screen[19] for hits with links to transcription initiation. On the basis of iterative PSI-BLAST searches, CG12721—an uncharacterized protein that we name Moonshiner—stood out as a potential paralogue of TFIIA-L (Fig. 2a), the large subunit of the TFIIA complex. In contrast to the ubiquitously expressed *TFIIA-L* and *TFIIA-S* factors, *moonshiner* is specifically expressed in ovaries (Extended Data Fig. 2a).

Moonshiner shares two regions of homology with TFIIA-L: an N-terminal α-helical region, which in TFIIA-L facilitates the interaction with TFIIA-S[20,21], and a second region that is part of the middle region of TFIIA-L (Fig. 2a and Extended Data Fig. 2b). In agreement with this, Moonshiner interacts with TFIIA-S, but not with TFIIA-L (Extended Data Fig. 2c). This suggests that Moonshiner and TFIIA-S form an alternative TFIIA complex involved in piRNA cluster expression. In support of this, TFIIA-S also scored in the transposon de-repression screen[19]. Moonshiner lacks the C-terminal β-roll domain, which in TFIIA-L interacts with TBP[20,21] (Fig. 2a). Consistent with this, the Moonshiner/TFIIA-S complex does not interact with TBP (Extended Data Fig. 2d).

The function of TFIIA is to stabilize the binding of TBP/TFIID to promoter DNA. To elucidate what alternative function Moonshiner may serve, we characterized its *in vivo* protein interactome. We generated flies expressing Moonshiner with a localization and affinity purification (LAP) tag (3×Flag–V5–GFP). We immunoprecipitated LAP–Moonshiner from ovary lysates and determined co-purifying

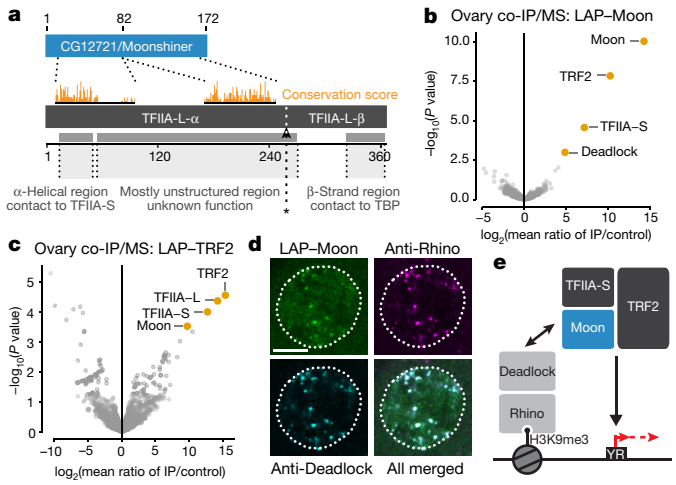


Figure 2 | The TFIIA-L paralogue Moonshiner localizes to Rhino domains and forms an alternative TFIIA–TRF2 complex. a, Schematic alignment showing homology between CG12721/Moonshiner and TFIIA-L; orange bars: conservation score (see Extended Data Fig. 2b). Also shown are TFIIA-L parts with known interaction partners and the Taspase1 cleavage site (asterisk). **b**, **c**, Volcano plots showing enrichment values and corresponding significance levels (for proteins co-purifying with LAP–Moonshiner (n (biological replicates) = 6) or LAP–TRF2 (n = 3) from ovary lysates (four most significantly enriched proteins are labelled). **d**, Localization of LAP–Moonshiner, Rhino, and Deadlock within an ovarian germline nucleus (see also Extended Data Fig. 3f). **e**, Model summarizing the identified protein interactions in the context of Rhino-dependent piRNA cluster transcription.

proteins by quantitative mass spectrometry. The three most enriched proteins were Moonshiner, TFIIA-S, and the short isoform of TRF2 (Fig. 2b and Extended Data Fig. 3a, b). TRF2 is an animal TBP paralogue that is essential for early embryogenesis[22–26] and fertility[27,28]. In contrast, TFIIA-L and TBP were not enriched. We substantiated these findings with a reciprocal experiment using LAP–TRF2, which resulted in co-purification of TRF2 with TFIIA-L, TFIIA-S, and Moonshiner (Fig. 2c and Extended Data Fig. 3c). Moonshiner also interacts with TRF2 in Schneider cells (Extended Data Fig. 3d). Taken together, Moonshiner forms an alternative TFIIA–TBP complex in ovaries consisting of Moonshiner, TFIIA-S and TRF2.

The next most enriched protein co-purifying with Moonshiner was Deadlock, which directly interacts with Rhino[6] (Fig. 2b). This revealed a molecular connection between Moonshiner/TFIIA-S/TRF2 and Rhino. To substantiate this, we asked whether Moonshiner, like Rhino and Deadlock, is enriched at bidirectional piRNA loci[6,9,10]. Indeed, LAP–Moonshiner, which is specifically expressed in germline cells, is concentrated in nuclear foci that are also positive for Rhino and Deadlock (Fig. 2d and Extended Data Fig. 3e, f). Furthermore, Moonshiner's localization to nuclear foci, but not its overall level, depends on Rhino and Deadlock (Extended Data Fig. 3g, h). In contrast, Rhino localization to nuclear foci did not depend on Moonshiner (Extended Data Fig. 3g).

In summary, our data suggest that Moonshiner forms an alternative TFIIA–TRF2 complex at bidirectional piRNA clusters via an interaction with Deadlock, a binding partner of the HP1 variant Rhino (Fig. 2e).

Moonshiner drives piRNA cluster transcription

The model in Fig. 2e predicts that loss of Moonshiner should result in defective transcription of Rhino-dependent piRNA clusters. To test this, we generated *moonshiner* mutant flies (Extended Data Fig. 4a). These flies are viable, contain ovaries with normal morphology, but are sterile. We first sequenced piRNA populations from *moonshiner* mutants. This showed that >90% of *cluster80F* piRNAs and ~80% of *cluster42AB* piRNAs depend on Moonshiner, while the Rhino-independent piRNA clusters *20A* and *flamenco* are also Moonshiner-independent (Fig. 3a and Extended Data Fig. 4b). The *moonshiner* and *rhino* mutants also show similar reductions in piRNAs mapping to individual transposons as well as similar increases in transposon mRNA levels resulting from de-repression (Extended Data Fig. 4c). We note that loss of Rhino results in a stronger phenotype than that of Moonshiner loss, indicating that Rhino serves other functions in addition to recruiting Moonshiner. Importantly, while Rhino deposition undergoes slight redistribution in *moonshiner* mutants, Rhino levels at *cluster80F* remain unchanged despite the strong loss of piRNA production (Extended Data Fig. 4d). Thus, the observed piRNAs losses are probably a direct consequence of Moonshiner loss, rather than an indirect result of perturbed Rhino occupancy.

Consistent with a transcriptional defect at piRNA clusters in *moonshiner* mutants, steady-state piRNA precursor levels were severely reduced at *cluster80F* and *cluster42AB* but not at *cluster20A* and *flamenco* (Fig. 3b, c, top, and Extended Data Fig. 4e). In contrast, steady-state levels of mRNAs were hardly changed in *moonshiner* mutants. While not excluding additional roles for Moonshiner in processes other than the piRNA pathway, these data argue against a broad gene expression role of this TFIIA-L paralogue (Extended Data Fig. 4f). To directly probe for a transcriptional defect, we determined Pol II occupancy using chromatin immunoprecipitation followed by sequencing (ChIP-seq). This revealed loss of Pol II specifically at Rhino-dependent piRNA clusters in *moonshiner* mutants, mirroring the reductions in piRNA and precursor RNA levels (Fig. 3b, c, bottom, and Extended Data Fig. 4e). We finally assessed piRNA cluster transcription in nurse cell nuclei by quantitative fluorescent *in situ* hybridization (FISH) (Extended Data Fig. 4g–i). We observed a pronounced drop in *cluster42AB* signal in *rhino* and *moonshiner* mutants compared with wild type (Fig. 3d). In sum, loss of Moonshiner results in defective transcription of bidirectional piRNA clusters in the developing ovary.

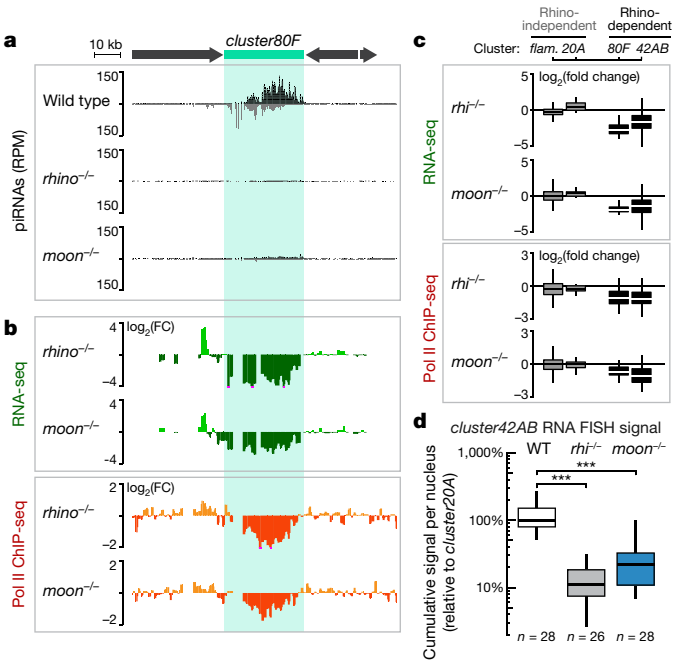


Figure 3 | Rhino-bound piRNA clusters require Moonshiner for their efficient transcription. a, Genome browser panel showing *cluster80F* piRNA levels from ovaries with indicated genotype. **b**, piRNA precursor abundance (RNA-seq) and Pol II occupancy (ChIP-seq) at *cluster80F* in wild-type ovaries and the corresponding changes in *rhino* or *moonshiner* mutant ovaries (log_2(fold change) calculated for 1-kb windows). **c**, Boxplots showing piRNA precursor abundance (top) and Pol II occupancy (bottom) for indicated piRNA clusters in *rhino* or *moonshiner* mutant ovaries relative to wild type (log_2(fold change) of 1-kb windows; box plots display median (line), first and third quartiles (box), and highest/lowest value within 1.5 × interquartile range (whiskers)); *flam.*, *flamenco*. **d**, Quantification of *cluster42AB* RNA FISH signal in germline nuclei of ovaries with indicated genotype relative to that of *cluster20A* (boxplots as in **c**; $***P < 0.0001$; Mann–Whitney–Wilcoxon tests; *n*, number of germline nuclei analysed).

To answer whether Moonshiner exerts its function within the identified variant TFIIA–TRF2 complex, we generated flies with germline-specific depletion of TFIIA-S or TRF2. These flies are sterile, display de-repression of several transposons, and their ovaries contain strongly reduced levels of piRNAs derived specifically from bidirectional clusters (Extended Data Figs 5 and 6 and Supplementary Note 2). We conclude that piRNA production from Rhino-dependent piRNA loci requires Moonshiner, TFIIA-S, and TRF2, presumably acting together in a complex that stimulates transcription initiation.

DNA-encoded promoters can bypass Moonshiner function

Our data are consistent with Moonshiner being required for efficient transcription at all Rhino-dependent piRNA loci. For some transposons, however, piRNA levels are very different in *moonshiner* versus *rhino* mutants (Extended Data Fig. 7a). To understand this discrepancy, we compared *moonshiner* and *rhino* mutant piRNA profiles genome-wide. Most Rhino-dependent loci are also Moonshiner-dependent, confirming that Moonshiner is essential for bidirectional piRNA precursor transcription (Fig. 4a). However, some loci—while strongly dependent on Rhino—produce piRNAs independently of Moonshiner, often even at elevated levels. Most of these map to *cluster38C1* and *38C2*. These clusters harbour prominent Pol II peaks at their boundaries[6], a pattern that—besides the distal part of *cluster42AB* —is not found at other Rhino-dependent clusters.

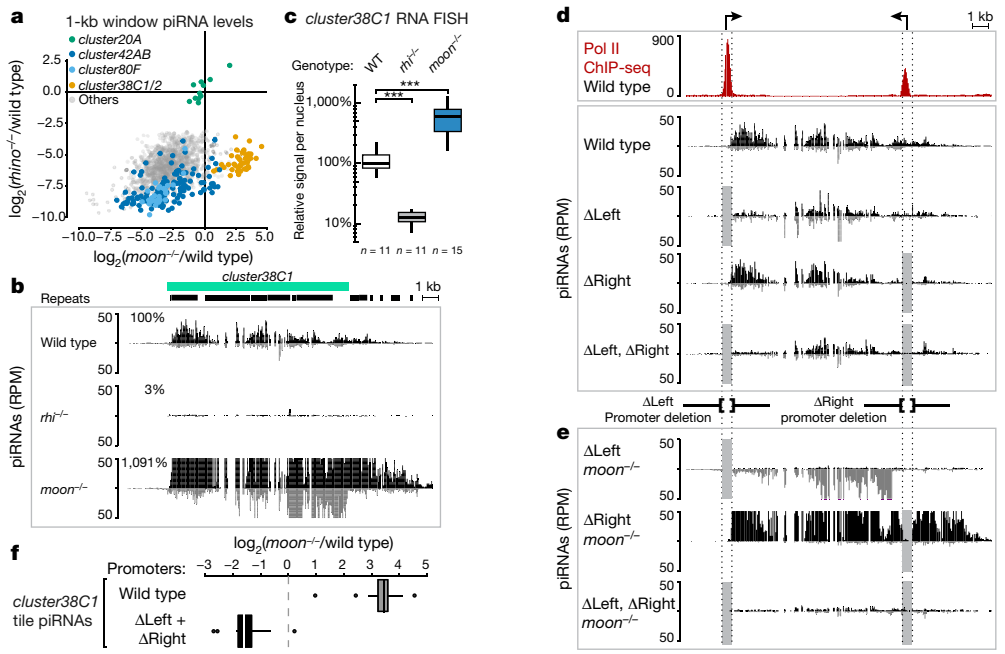


Figure 4 | Endogenous piRNA cluster promoters bypass Moonshiner-dependent transcription initiation. a, $\log_2$(fold changes) of piRNAs mapping uniquely to Rhino-dependent genomic 1-kb tiles in *rhino* versus *moonshiner* mutants (relative to wild type). Tiles from major piRNA clusters are coloured (*cluster20A* tiles serve as Rhino-independent control group). **b**, Genome browser panel showing *cluster38C1* piRNA levels from ovaries with indicated genotype. **c**, Quantification of *cluster38C1* RNA FISH signal in germline nuclei relative to that of *cluster20A* (***$P < 0.0001$; Mann–Whitney–Wilcoxon tests; *n*, number of germline nuclei analysed; boxplots as in Fig. 3c). **d**, Pol II occupancy and piRNA levels at *cluster38C1* in ovaries with indicated genotypes. ΔLeft and ΔRight indicate *cluster38C1* promoter deletions (light grey boxes). **e**, As **d**, but from *moonshiner* mutant ovaries. **f**, Boxplot (defined as in Fig. 3c) displaying $\log_2$(fold changes) in *cluster38C1* piRNA levels ($n = 12$ 1-kb windows) in *moonshiner* mutant compared with wild-type ovaries when both *cluster38C1* promoters are wild type or deleted.

To elucidate this Moonshiner-independent piRNA production, we investigated *cluster38C1* in detail. While Rhino loss results in a near-complete collapse of piRNAs from *cluster38C1*, loss of Moonshiner results in greater than tenfold higher piRNA levels (Fig. 4b). piRNA levels also increase in ovaries depleted of TFIIA-S or TRF2 (Extended Data Fig. 7b). Quantitative RNA FISH revealed that the increased piRNA production is caused by elevated transcription of *cluster38C1* in *moonshiner* mutants (Fig. 4c and Extended Data Fig. 7c).

To directly test the involvement of the flanking promoters in *cluster38C1* transcription, we used CRISPR/Cas9 (ref. 29) to precisely delete them. This leads to substantial reductions in promoter-proximal piRNA levels, mainly on the strand transcribed by the respective promoter. However, piRNA production more distal to the deleted promoters is hardly changed (Fig. 4d). Similarly, in flies that lack both *cluster38C1* promoters, piRNAs at the cluster boundaries are considerably reduced while piRNA production from the central region is only mildly affected. This points to an alternative mechanism of transcription initiation from within the cluster.

We hypothesized that Moonshiner is responsible for this promoter-independent transcription. To test this, we generated flies harbouring the various promoter deletions in a *moonshiner* mutant background. Both single promoter deletions of the *38C1* cluster display nearly exclusive unidirectional piRNA profiles, which initiate just downstream of the non-modified promoter (Fig. 4e). These results predict that piRNA production from the double-promoter-deleted cluster should be Moonshiner-dependent. Indeed, loss of Moonshiner results in roughly fourfold reduced piRNA levels. We see very similar results for the distal ~20 kb of *cluster42AB*, which also harbours a flanking promoter[6] (Extended Data Fig. 7d) causing Moonshiner-independent piRNA production for promoter-proximal *cluster42AB* tiles (Fig. 4a). We conclude that, in the absence of flanking promoters, Moonshiner-independent loci such as *cluster38C1* become Moonshiner-dependent just like all other Rhino-dependent loci (Fig. 4e, f).

These findings allow three conclusions: (1) piRNA precursors transcribed from bidirectional clusters can be 10–15 kb in length (Fig. 4b, e); (2) as Pol II does not elongate into *cluster38C1* from flanking promoters in the absence of Rhino, at least one other effector protein must act at piRNA clusters (Fig. 4b); (3) Moonshiner specifically stimulates transcription initiation as DNA-encoded promoters can replace its function at Rhino-dependent piRNA loci.

Moonshiner functions to recruit TRF2

Canonical TFIIA stabilizes the binding of TBP onto core promoters[30]. We therefore tested whether ectopic recruitment of TRF2 stimulates transcription in Schneider cells, which lack Moonshiner expression. Recruiting additional TRF2 to the known TRF2-driven histone H1 core promoter does not elevate transcription of a reporter, while recruiting TRF2 to the same promoter carrying a mutation that disrupts its endogenous activity[31] results in an approximate sixfold stimulation of transcription (Extended Data Fig. 8a, b). Transcription is also stimulated (two- to sixfold) upon TRF2 recruitment to ten randomly chosen 150-nt piRNA cluster fragments. This resembles the approximate tenfold stimulation of cluster transcription by Moonshiner observed *in vivo*, suggesting that Moonshiner stimulates transcription from a broad range of DNA sequences by recruiting TRF2 to chromatin.

Moonshiner levels and its localization to Rhino foci are unchanged in ovaries depleted of TFIIA-S or TRF2 (Extended Data Fig. 8c, d), yet these flies phenotypically resemble *moonshiner* mutants. This supports a model where recruitment of TRF2 to piRNA clusters is Moonshiner's main function. We tested this hypothesis *in vivo* by recruiting TRF2 to the Rhino-interactor Deadlock using a single-chain anti-GFP nanobody[32] (Fig. 5a), thereby bypassing the requirement for Moonshiner. We engineered flies to express Deadlock fused to the GFP-nanobody, which enabled specific recruitment of GFP-tagged proteins to Rhino domains in germline nuclei (Fig. 5b). We then combined expression of Deadlock–GFP-nanobody and GFP–TRF2 fusion proteins to recruit TRF2 to piRNA source loci in

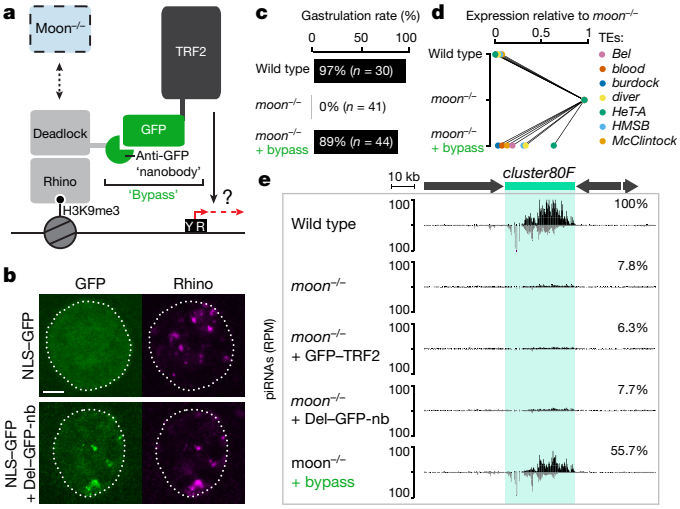


Figure 5 | Moonshiner stimulates heterochromatic transcription by recruiting TRF2 to Rhino domains. a, Schematic of the bypass experiment where GFP–TRF2 is recruited directly to Rhino domains via a Deadlock–GFP-nanobody fusion protein. **b**, Localization of nuclear localization signal (NLS)–GFP and Rhino in control (top) or in germline nuclei expressing the Deadlock–GFP-nanobody (nb) fusion protein (bottom); scale bar, 5 μm. **c**, Percentages of embryos with the indicated genotype displaying successful gastrulation (*n*, number of embryos assayed; see also Extended Data Fig. 8f). **d**, Steady-state levels of transposable element (TE) mRNAs in ovaries with indicated genotype relative to their level in *moonshiner* mutants (average of three biological replicates; for details see Extended Data Fig. 8h). **e**, Profiles of *cluster80F* piRNA in ovaries with indicated genotypes.

a Moonshiner-independent fashion (Extended Data Fig. 8e). The *moonshiner* mutant flies harbouring the two bypass transgenes are fertile, with nearly 90% of their laid eggs developing beyond gastrulation and 37% hatching into larva, several of which develop into adult flies (Fig. 5c and Extended Data Fig. 8f, g). Moreover, mRNA levels of transposons that are strongly de-repressed in *moonshiner* mutants are largely restored to wild-type levels in 'bypass' females (Fig. 5d and Extended Data Fig. 8h; only weak rescue for the telomeric *HeT-A* element). Similarly, we see rescue of piRNAs mapping to *cluster80F* or *cluster42AB* in 'bypass' females, while all other genetic combinations that lack Moonshiner display the loss of piRNAs characteristic for *moonshiner* mutants (Fig. 5e and Extended Data Fig. 8i). In agreement with the partial rescue in transposable element silencing and fertility, piRNA levels derived from clusters as well as transposon-targeting piRNAs in general do not return to wild-type levels in bypass flies (Extended Data Fig. 8i, j). Taken together, the high congruence between the rescue at the developmental and molecular level strongly supports a model where Moonshiner stimulates transcription within heterochromatin via recruitment of TRF2 to Rhino-decorated piRNA source loci.

Discussion

Here we identify a heterochromatin-dependent transcription machinery in *Drosophila* that allows piRNA precursor production despite potent silencing of transposon-encoded promoters and enhancers. We show that Moonshiner-dependent transcription, which cannot rely on recognition of DNA motifs because of their inaccessibility in heterochromatin, achieves locus specificity through Rhino, an HP1 variant that binds H3K9me3 marks at piRNA clusters (Figs 2 and 5). Thereby the cell allows transcription of transposon-rich loci into piRNA precursors while transcription of the same loci into functional transposon mRNAs is suppressed via heterochromatin-mediated exclusion of sequence-specific transcription factors.

Small RNA source loci embedded in heterochromatin and transcribed on both genomic strands are also a hallmark of genome defence pathways in plants and fungi. In fission yeast, a 'passive' mode of small RNA expression has been proposed, where Pol II transcribes small RNA precursors from pericentromeric regions during G1/S phases when heterochromatin is less condensed[33,34]. In contrast, an active recruitment mode with conceptual similarities to the Moonshiner pathway occurs in plants. Here SHH1, a reader of H3K9me marks, recruits the plant-specific RNA polymerase IV to heterochromatin to transcribe small RNA precursors[35,36]. Although SHH1 and Rhino both bind H3K9me residues, the two proteins are unrelated, suggesting that specification of small RNA source locus transcription via heterochromatin readers has evolved independently in animals and plants. Also in plants, small RNA precursor transcription initiates at 'YR' initiator sites dispersed on both genomic strands[37]. Whether Moonshiner-mediated transcription, like that of plant Pol IV, depends on collaboration with nucleosome remodellers to access heterochromatic target loci is unclear. The reported interaction of TRF2 with the NURF chromatin remodelling complex[38] supports this possibility. The recurring evolution of small RNA source locus transcription specified by chromatin marks rather than DNA sequence suggests that this constitutes a common alternative mode of transcriptional activation. The DNA inaccessibility of heterochromatin is thereby transformed into a specificity mark for non-canonical transcription activation (Extended Data Fig. 9). We note that the major *Drosophila* somatic piRNA cluster, *flamenco*, is transcribed from a single defined enhancer-driven promoter and avoids piRNA-mediated silencing because of the antisense orientation of the vast majority of the contained transposons[7,39]. The production of plant siRNAs from Pol IV transcripts initiates a positive feedback loop: siRNA-mediated targeting leads to DNA methylation, which in turn increases H3K9 methylation, thereby bringing in SHH1 and Pol IV[3,35]. In a similar fashion, production of Moonshiner-dependent piRNA precursors leads to generation of Piwi-bound piRNAs, which in turn guide H3K9 methylation and thereby Rhino recruitment[6]. This explains how Piwi-mediated transcriptional silencing 'transforms' active transposon insertions into heterochromatic piRNA source loci with bidirectional transcription.

Rhino and the associated factors Deadlock and Cutoff are required for transcription of dual-strand piRNA clusters. Owing to its ability to inhibit co-transcriptional processes such as splicing and transcription termination, Cutoff has been suggested to be the main effector of this complex[6,10,14]. Such an inhibition of termination is supported by our data on *cluster38C1*, where transcription from defined promoters results in 10–15 kb transcripts in a Rhino-, Deadlock-, and Cutoff-dependent manner (Fig. 4b, e). Cutoff also interacts with the transcription/export (TREX) complex, which orchestrates several co-transcriptional processes and which is required for transcription of Rhino-dependent piRNA source loci[40,41]. Together with the identification of Moonshiner/TRF2 as piRNA cluster transcription initiation factors, this suggests that Rhino acts as a molecular hub for several effector proteins that stimulate different (co)-transcriptional processes. Though Rhino is not conserved outside drosophilids, data from mouse studies support a conserved role of TRF2 in transcription of germline heterochromatin (Supplementary Note 3 and refs 42–52). In summary, we uncover the molecular mechanism by which heterochromatic piRNA loci are transcribed in *Drosophila* and propose that the identified coupling of chromatin readers to basal transcription factors is a recurring theme in eukaryotic heterochromatin biology.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 April; accepted 14 July 2017.
Published online 23 August 2017.

1. Fedoroff, N. V. Transposable elements, epigenetics, and genome evolution. *Science* **338,** 758–767 (2012).
2. Castel, S. E. & Martienssen, R. A. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat. Rev. Genet.* **14,** 100–112 (2013).

3. Holoch, D. & Moazed, D. RNA-mediated epigenetic regulation of gene expression. *Nat. Rev. Genet.* **16,** 71–84 (2015).
4. Siomi, M. C., Sato, K., Pezic, D. & Aravin, A. A. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* **12,** 246–258 (2011).
5. Czech, B. & Hannon, G. J. One loop to rule them all: the ping-pong cycle and piRNA-guided silencing. *Trends Biochem. Sci.* **41,** 324–337 (2016).
6. Mohn, F., Sienski, G., Handler, D. & Brennecke, J. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell* **157,** 1364–1379 (2014).
7. Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128,** 1089–1103 (2007).
8. Le Thomas, A. *et al.* Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev.* **28,** 1667–1680 (2014).
9. Klattenhoff, C. *et al.* The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* **138,** 1137–1149 (2009).
10. Zhang, Z. *et al.* The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell* **157,** 1353–1363 (2014).
11. Sainsbury, S., Bernecky, C. & Cramer, P. Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16,** 129–143 (2015).
12. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56,** 549–561 (1989).
13. Papai, G. *et al.* TFIIA and the transactivator Rap1 cooperate to commit TFIID for transcription initiation. *Nature* **465,** 956–960 (2010).
14. Chen, Y. A. *et al.* Cutoff suppresses RNA polymerase II termination to ensure expression of piRNA precursors. *Mol. Cell* **63,** 97–109 (2016).
15. Gu, W. *et al.* CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as *C. elegans* piRNA precursors. *Cell* **151,** 1488–1500 (2012).
16. Kaufmann, J. & Smale, S. T. Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes Dev.* **8,** 821–829 (1994).
17. Purnell, B. A., Emanuel, P. A. & Gilmour, D. S. TFIID sequence recognition of the initiator and sequences farther downstream in *Drosophila* class II genes. *Genes Dev.* **8,** 830–842 (1994).
18. Gunawardane, L. S. *et al.* A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. *Science* **315,** 1587–1590 (2007).
19. Czech, B., Preall, J. B., McGinn, J. & Hannon, G. J. A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway. *Mol. Cell* **50,** 749–761 (2013).
20. Geiger, J. H., Hahn, S., Lee, S. & Sigler, P. B. Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science* **272,** 830–836 (1996).
21. Tan, S., Hunziker, Y., Sargent, D. F. & Richmond, T. J. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature* **381,** 127–134 (1996).
22. Dantonel, J. C., Quintin, S., Lakatos, L., Labouesse, M. & Tora, L. TBP-like factor is required for embryonic RNA polymerase II transcription in *C. elegans*. *Mol. Cell* **6,** 715–722 (2000).
23. Kaltenbach, L., Horner, M. A., Rothman, J. H. & Mango, S. E. The TBP-like factor CeTLF is required to activate RNA polymerase II transcription during *C. elegans* embryogenesis. *Mol. Cell* **6,** 705–713 (2000).
24. Kopytova, D. V. *et al.* Two isoforms of *Drosophila* TRF2 are involved in embryonic development, premeiotic chromatin condensation, and proper differentiation of germ cells of both sexes. *Mol. Cell. Biol.* **26,** 7492–7505 (2006).
25. Müller, F., Lakatos, L., Dantonel, J., Strähle, U. & Tora, L. TBP is not universally required for zygotic RNA polymerase II transcription in zebrafish. *Curr. Biol.* **11,** 282–287 (2001).
26. Veenstra, G. J., Weeks, D. L. & Wolffe, A. P. Distinct roles for TBP and TBP-like factor in early embryonic gene transcription in *Xenopus*. *Science* **290,** 2312–2315 (2000).
27. Martianov, I. *et al.* Late arrest of spermiogenesis and germ cell apoptosis in mice lacking the TBP-like TLF/TRF2 gene. *Mol. Cell* **7,** 509–515 (2001).
28. Zhang, D., Penttila, T. L., Morris, P. L., Teichmann, M. & Roeder, R. G. Spermiogenesis deficiency in mice lacking the *Trf2* gene. *Science* **292,** 1153–1155 (2001).
29. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).
30. Yokomori, K. *et al. Drosophila* TFIIA directs cooperative DNA binding with TBP and mediates transcriptional activation. *Genes Dev.* **8,** 2313–2323 (1994).
31. Isogai, Y., Keles, S., Prestel, M., Hochheimer, A. & Tjian, R. Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev.* **21,** 2936–2949 (2007).
32. Rothbauer, U. *et al.* Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nat. Methods* **3,** 887–889 (2006).
33. Kloc, A., Zaratiegui, M., Nora, E. & Martienssen, R. RNA interference guides histone modification during the S phase of chromosomal replication. *Curr. Biol.* **18,** 490–495 (2008).
34. Chen, E. S. *et al.* Cell cycle control of centromeric repeat transcription and heterochromatin assembly. *Nature* **451,** 734–737 (2008).
35. Law, J. A. *et al.* Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* **498,** 385–389 (2013).
36. Law, J. A., Vashisht, A. A., Wohlschlegel, J. A. & Jacobsen, S. E. SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. *PLoS Genet.* **7,** e1002195 (2011).
37. Zhai, J. *et al.* A one precursor one siRNA model for Pol IV-dependent siRNA biogenesis. *Cell* **163,** 445–455 (2015).
38. Hochheimer, A., Zhou, S., Zheng, S., Holmes, M. C. & Tjian, R. TRF2 associates with DREF and directs promoter-selective gene expression in *Drosophila*. *Nature* **420,** 439–445 (2002).
39. Goriaux, C., Desset, S., Renaud, Y., Vaury, C. & Brasset, E. Transcriptional properties and splicing of the flamenco piRNA cluster. *EMBO Rep.* **15,** 411–418 (2014).
40. Hur, J. K. *et al.* Splicing-independent loading of TREX on nascent RNA is required for efficient expression of dual-strand piRNA clusters in *Drosophila*. *Genes Dev.* **30,** 840–855 (2016).
41. Zhang, F. *et al.* UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell* **151,** 871–884 (2012).
42. Zeidler, M. P., Yokomori, K., Tjian, R. & Mlodzik, M. *Drosophila* TFIIA-S is up-regulated and required during Ras-mediated photoreceptor determination. *Genes Dev.* **10,** 50–59 (1996).
43. Ni, J.-Q. *et al.* A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. *Nat. Methods* **8,** 405–407 (2011).
44. Schmitz, M. L., Stelzer, G., Altmann, H., Meisterernst, M. & Baeuerle, P. A. Interaction of the COOH-terminal transactivation domain of p65 NF-κ B with TATA-binding protein, transcription factor IIB, and coactivators. *J. Biol. Chem.* **270,** 7219–7226 (1995).
45. Lieberman, P. M. & Berk, A. J. A mechanism for TAFs in transcriptional activation: activation domain enhancement of TFIID-TFIIA--promoter DNA complex formation. *Genes Dev.* **8,** 995–1006 (1994).
46. Kobayashi, N., Boyer, T. G. & Berk, A. J. A class of activation domains interacts directly with TFIIA and stimulates TFIIA-TFIID-promoter complex assembly. *Mol. Cell. Biol.* **15,** 6465–6473 (1995).
47. Aoyagi, N. & Wassarman, D. A. Genes encoding *Drosophila melanogaster* RNA polymerase II general transcription factors: diversity in TFIIA and TFIID components contributes to gene-specific transcriptional regulation. *J. Cell Biol.* **150,** F45–F50 (2000).
48. Duttke, S. H. C., Doolittle, R. F., Wang, Y.-L. & Kadonaga, J. T. TRF2 and the evolution of the bilateria. *Genes Dev.* **28,** 2071–2076 (2014).
49. Martianov, I. *et al.* Distinct functions of TBP and TLF/TRF2 during spermatogenesis: requirement of TLF for heterochromatic chromocenter formation in haploid round spermatids. *Development* **129,** 945–955 (2002).
50. Oyama, T. *et al.* Cleavage of TFIIA by Taspase1 activates TRF2-specified mammalian male germ cell programs. *Dev. Cell* **27,** 188–200 (2013).
51. Martianov, I., Velt, A., Davidson, G., Choukrallah, M.-A. & Davidson, I. TRF2 is recruited to the pre-initiation complex as a testis-specific subunit of TFIIA/ALF to promote haploid cell gene expression. *Sci. Rep.* **6,** 32069 (2016).
52. Brancorsini, S., Davidson, I. & Sassone-Corsi, P. TIPT, a male germ cell-specific partner of TRF2, is chromatin-associated and interacts with HP1. *Cell Cycle* **7,** 1415–1422 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K. Meixner for experimental support, D. Handler and D. Jurczak for bioinformatics help, P. Duchek and J. Gokcezade for generating CRISPR-edited and transgenic flies, K. Mechtler and his team for mass spectrometry, T. Lendl for RNA FISH quantification, A. Schleiffer and M. Novatchkova for Moonshiner phylogenetic analysis, the Vienna Biocenter Core Facilities Next Generation Sequencing unit for deep sequencing, M. Elmaghraby for the Deadlock antigen, the Max F. Perutz Laboratories monoclonal facility for the Deadlock antibody, and the Vienna Drosophila RNAi Center, TRiP, and Bloomington stock centers for flies. We thank A. Ordonez, D. Handler, F. Mohn, F. Muerdter, M. Bühler, and especially O. Wueseke (Impulse Science) and Life Science Editors for comments on the manuscript. This work was supported by the Austrian Academy of Sciences and the European Community (ERC grant 260711EU and ERC-2015-CoG-682181). P.R.A. is supported by fellowships from the Alfred Benzon Foundation and the Novo Nordisk Foundation.

Author Contributions P.R.A. performed the experiments except the genetic bypass and promoter deletion experiments (both L.T.), and the S2 cell-based protein interaction assays (M.V.). P.R.A. and J.B. analysed the data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to J.B. (julius.brennecke@imba.oeaw.ac.at) or P.R.A. (peter.andersen@imba.oeaw.ac.at).

Reviewer Information *Nature* thanks E. Brasset, T. Juven-Gershon and P. Zamore for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Fly husbandry. A complete list of fly strains with genotypes, identifiers, and original sources can be found in Supplementary Table 1. All flies were kept at 25 °C. For ovary dissection, flies aged 2–6 days were given fresh food with yeast for 2 days and then dissected after brief immobilization by CO_2 anaesthesia (blinding and randomization not applied). All fly strains used in the study (see Supplementary Table 1) are available via VDRC (http://stockcenter.vdrc.at/control/main).

Generation of transgenic fly strains. LAP–Moonshiner transgenic flies were obtained by inserting an N-terminal LAP tag (3× Flag–V5–GFP) into a Pacman clone (CH322-15N16) containing the *moonshiner* gene locus via bacterial recombineering[53]. The Pacman transgene was then inserted into the attP2 landing site (FlyBase identifier FBti0040535) and the transgene was verified to rescue the sterility phenotype of homozygous *moonshiner* frameshift mutations ($moon^{-/-}$). LAP–TRF2 transgenic flies were generated by insertion of a TRF2 germline expression construct (*nanos* promoter and *vasa* 3′ untranslated region; short isoform of TRF2; Extended Data Fig. 3a) into the attP40 landing site (FlyBase identifier FBti0114379). Fly strains harbouring short hairpin RNA (shRNA) expression cassettes for germline knockdown were created by cloning shRNAs (shRNA construct cloning oligonucleotide sequences are listed in Supplementary Table 2) into the Valium-20 or Valium-22 vector modified with a white selection marker[43]. The LacZ sensor flies for *HeT-A* were generated by replacing the target fragment in the *Burdock* sensor with a 700-base-pair fragment of the *HeT-A* transposon. *Burdock* and *gypsy* LacZ sensor flies are described in refs 54 and 55, respectively.

Generation of mutant fly strains. Frameshift mutant alleles of *moonshiner* were generated as described in ref. 56 by injection of pDCC6 plasmids modified to express *moonshiner*-targeting guide-RNAs using the oligonucleotides given in Supplementary Table 2.

To generate promoter deletions, homology arms of approximately 1 kb were cloned into pHD-dsRed (Addgene) by Gibson assembly and co-injected with pCFD4 (Addgene) containing two single-guide RNA expression cassettes into *y*,*w*, *ZH2A(Act5C-Cas9)* embryos. Removal of the dsRed cassette was done by crossing to an hs-Cre strain. After stock establishment, homozygous flies were screened by PCR and sequenced for the presence of the targeted deletion, the loss of the wild-type allele, and for the lack of vector backbone integrations. To delete the *cluster38C1* right promoter in the *cluster38C1* Δleft background, a similar vector was generated with flanking FRT sites and a *white* selection marker. The vectors were injected into *actin>Cas9; 38C1 Δleft promoter* embryos. The selection cassette was removed by crossing to an hsFLP stock.

To delete the *cluster42AB* right promoter, two FRT insertions flanking the promoter were generated by oligonucleotide-directed DNA repair after gRNA induced cuts. The two FRT insertions were brought *in trans* and the promoter deletion was triggered by crossing to a hsFLP strain.

The *rhino* mutant fly strains were generated by removal of the entire *rhino* open reading frame using ends-out homologous recombination[57].

***Drosophila* Schneider 2 (S2) cell culture.** *Drosophila* Schneider 2 (S2) cells (in-house stock regularly tested to be virus- and mycoplasma-free) were grown at 25 °C in S2 cell media supplemented with 10% fetal bovine serum (Thermo Fisher Scientific).

X-gal staining of *Drosophila* ovaries. Dissected ovaries from flies subjected to control or germline knockdown were fixed in 0.5% glutaraldehyde/PBS for 15 min at room temperature and then rinsed twice in PBS. The fixed ovaries were then incubated in staining solution (10 mM PBS, 1 mM $MgCl_2$, 150 mM NaCl, 3 mM potassium ferricyanide, 3 mM potassium ferrocyanide, 0.1% Triton X-100, 0.1% X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside)) at room temperature with rotation for 2 h (*HeT-A* and *Burdock* sensors or overnight (*gypsy* sensor).

Scoring of fly embryogenesis and hatching rates. To quantify the correct start of embryogenesis, non-virgin females were kept together with *w1118* males for 2 days. One- to 3-hour-old embryos were bleached, formaldehyde fixed, and stained with 4′,6-diamidino-2-phenylindole (DAPI) according to standard procedures. Embryos with hundreds or thousands of regularly spaced nuclei resembling embryonic stages 3–7 were scored as 'normal'. Embryos laid by *moonshiner* mutant females had usually five or fewer irregular DAPI foci, a phenotype scored as 'arrested'. From the same cages, eggs were collected overnight and the hatching rate was counted 30 h later in numbers as practically feasible.

Protein co-immunoprecipitation from S2 cell lysates. S2 cells were seeded at $\sim 1 \times 10^6$ cells per millilitre and transfected using FuGENE with plasmids harbouring Act5C-driven expression cassettes of the described tagged proteins. These plasmids were cloned by insertion of the transgene open reading frame (for TRF2, the short isoform was used; Extended Data Fig. 3a) into the pAcM_empty expression vector driven by the *Drosophila* Act5C promoter. Forty-eight hours after transfection, cells were collected by centrifugation and pellets were snap-frozen in liquid nitrogen and stored for later processing. S2 cell pellets were resuspended in 50 µl S2 lysis buffer for S2 cells (LBS2) (30 mM HEPES pH 7.4, 150 mM NaCl, 2 mM $MgCl_2$, 0.5% Triton X-100, 5 mM DTT) and rotated for 20 min at 4 °C. Lysate was then cleared by centrifugation for 10 min at 16,000*g* (4 °C) and protein concentrations measured using Bradford reagent. For each immunoprecipitation, 100 µl lysate at $\sim 1\ \mu g\ \mu l^{-1}$ total protein was incubated for 2 h at 4 °C with 20 µl Flag M2 Magnetic Beads. The beads were then washed three times for 10 min in immunoprecipitation washing buffer (IPWB) (30 mM HEPES pH 7.4, 300 mM NaCl, 2 mM $MgCl_2$, 0.5% Triton X-100, 5 mM DTT) and co-purifying proteins were eluted by a 5-min incubation at 95 °C in 50 µl 1× SDS buffer.

Western blot analysis of co- immunoprecipitations from S2 cell lysates. Western blotting was done according to standard protocols. Briefly, protein samples were resolved by SDS–polyacrylamide gel electrophoresis (PAGE) and transferred to 0.45 µm nitrocellulose membranes (Bio-Rad) before blotting overnight with primary antibodies (Supplementary Table 3) in PBX (0.01% Triton X-100 in 1× PBS). After three washes with PBX, incubation with HRP-coupled secondary antibodies, and three more washes in PBX, the membranes were incubated with Clarity Western ECL Blotting Substrate (Bio-Rad) and imaged using a ChemiDoc MP imaging system (Bio-Rad).

Protein co-immunoprecipitation from ovary lysates. For each sample, roughly 200 ovary pairs were dissected and immediately transferred to ice-cold PBS. Each ovary sample was then homogenized with 20 strokes using a douncer (tight pestle) in 1 ml ovary protein lysis buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 10% glycerol, 1 mM DTT, 1 mM PefaBloc, 0.2% NP-40). The homogenate was then transferred to clean 1.5 ml low-retention tubes and incubated on ice for 15 min with occasional inversion. The lysate was then cleared by centrifugation for 5 min at 16,000 *g*. To each cleared lysate sample, 20 µl of a solution of anti-Flag M2 magnetic beads diluted to 1 µl beads per 5 µl total volume with beads buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl) were added. Samples were then incubated for 3 h at 4 °C with rotation and subsequently washed four times for 10 min in ovary protein lysis buffer followed by six quick rinses in co-immunoprecipitation wash buffer (20 mM HEPES pH 7.4, 150 mM NaCl, 2 mM $MgCl_2$). Most of the wash buffer was then removed and the pelleted magnetic beads were stored at 4 °C until processing for mass spectrometry analysis.

Mass spectrometry analyses. Co-immunoprecipitated proteins were subjected to on-bead digestion with LysC and elution with glycine before digestion with Trypsin. The resulting peptides were analysed using a Dionex UltiMate 3000 HPLC RSLC nano system coupled to a Q Exactive mass spectrometer equipped with a Proxeon nanospray source (Thermo Fisher Scientific). Peptides were eluted using a flow rate of 230 nl min^{-1} and a binary 3 h gradient, respectively 225 min, and the data were acquired with the mass spectrometer operated in data-dependent mode with tandem mass spectrometry scans of the 12 most abundant ions. For peptide identification, the RAW files were loaded into Proteome Discoverer (version 2.1.0.81, Thermo Scientific) and the created spectra were searched using MSAmanda version 1.0.0.6186 (ref. 58) against *D. melanogaster* reference translations retrieved from Flybase (dmel_all-translation-r6.06). An in-house-developed tool, Peakjuggler, was used for the peptide and protein quantification (IMP/IMBA/GMI Protein Chemistry Facility; http://ms.imp.ac.at/?goto=peakjuggler). Using custom R scripts, average enrichments between bait and control immunoprecipitation experiments were calculated. Adjusted *P* values were calculated using the limma R package[59].

Quantitative reverse transcription PCR analysis of transposon expression. Five to ten pairs of freshly dissected ovaries were homogenized in TRIzol reagent followed by RNA purification according to the manufacturer's protocol. One microgram of total RNA was digested with RQ1 RNase-Free DNase (Promega) and then reverse transcribed using random hexamer primers and Superscript II (Invitrogen) following standard protocols. cDNA was then used as template for RT–qPCR quantification of transposon and mRNA abundances (for primers see Supplementary Table 2).

Luciferase reporter assays. Plasmids for luciferase reporter assays (Supplementary Table 4) were cloned as described in ref. 60 by inserting the open reading frames of GFP or TRF2 (short isoform; Extended Data Fig. 3a) into pAGW-GAL4-DBD_empty and by replacing the developmental core promoter (dCP) of pGL3_4xUAS_UPS_hkCP with 150-base-pair cluster fragments amplified using the oligonucleotides indicated in Supplementary Table 2. For plasmid transfections, 1×10^5 S2 cells were seeded in 100 µl S2 cell medium in 96-well plates. For each sample, six replicate wells were seeded and the cells were allowed to settle for 4 h. The S2 cells were then co-transfected with three plasmids using FuGene HD Transfection Reagent (Promega). Each well was transfected with a total of 80 ng plasmid in the following mixture: 5 ng pUbi_RL, which drives ubiquitous expression of *Renilla* firefly luciferase as a transfection and viability control; 25 ng pGL3 reporter vector containing individual putative core promoters; 50 ng pAct5C vector expressing

Gal4-DNA binding domain (DBD) fused to either GFP or Trf2 (pAGW-GAL4-DBD_GFP/TRF2S). Forty-eight hours later, the transfected cells were washed with PBS and lysed in 40 µl 1× passive lysis buffer (Dual-Luciferase Reporter Assay System, Promega). Firefly and *Renilla* luciferase activity was measured on a Synergy H1 plate reader (BioTek). For analyses, firefly luciferase activity was normalized to that of *Renilla* and averaged over technical replicates. Average values from five such biological replicates were then calculated and analysed for statistical differences between GAL4-DBD–GFP and GAL4-DBD–TRF2 tethering for each reporter construct by two-tailed *t*-tests (for calculations, see figure source data).

Immunofluorescence staining of ovaries. Five to ten ovaries were dissected into ice-cold PBS and then immediately fixed by incubation in IF fixing buffer (4% paraformaldehyde, 0.3% Triton X-100, 1× PBS) for 20 min at room temperature. The fixed ovaries were then washed three times for 10 min in PBX (0.3% Triton X-100, 1× PBS) and blocked with BBX (0.1% BSA, 0.3% Triton X-100, 1× PBS) for 30 min. Blocked ovaries were incubated overnight at 4 °C with antibodies diluted in BBX followed by three washes in PBX. Subsequently, the ovaries were incubated with fluorophore-coupled secondary antibodies overnight at 4 °C and washed three times in PBX, with a second wash done with DAPI added to the PBX to stain DNA. The samples were imaged on a Zeiss LSM-780 Axio Imager confocal microscope and the resulting images processed using FIJI/ImageJ[61]. Rabbit anti-Rhino antibodies are described in ref. 6.

RNA FISH. Five to ten ovary pairs were dissected into ice-cold PBS and fixed in formaldehyde solution (4% formaldehyde, 0.15% Triton X-100 in PBS) for 20 min at room temperature with agitation. The fixed ovaries were then washed three times for 10 min in 0.3% Triton X-100/PBS and permeabilized overnight at 4 °C in 70% ethanol. For probe hybridization, permeabilized ovaries were first rehydrated for 5 min in RNA FISH wash buffer (10% (v/w) formamide in 2× SSC). Subsequently, the ovaries were resuspended in 50 µl hybridization buffer (10% (v/w) dextran sulfate and 10% (v/w) formamide in 2× SSC), and 0.5 µl 25 µM Stellaris RNA probe set (for probe sequences see Supplementary Table 5) was added followed by an overnight incubation at 37 °C with rotation. The ovaries were then rinsed twice with RNA FISH wash buffer and rotated for 1 h at room temperature in a solution of wheat germ agglutinin-coupled Alexa Fluor 488 conjugate (WGA-488) at a final concentration of 5 ng µl^{-1} in RNA FISH wash buffer. Ovaries were then washed for 30 min at room temperature in RNA FISH wash buffer, incubated for 10 min in a DAPI/2xSSC solution, and finally washed twice for 10 min in 2× SSC buffer. The wash buffer was then carefully removed and each ovary sample was resuspended in one drop (~40 µl) of Prolong Diamond mounting medium before mounting on microscopy slides. Mounted samples were allowed to equilibrate for at least 24 h before imaging on a Zeiss LSM 780 confocal microscope equipped with an Airyscan detector. Each germline nucleus was imaged with a ×40 oil lens in a *Z*-stack of 120 planes with 150 nm step size. The image stack was subsequently subjected to Airyscan image processing with standard settings. The quantification analysis was performed fully automated using Definiens Developer Suite XD. The nucleus was segmented in three dimensions on the DAPI channel, and the borders were refined using a DoG-filtered version of the WGA-488 signal (proxy for nuclear membrane). Within the nucleus, the genomic loci were segmented on channel 1 (*cluster20A* RNA FISH in the far-red channel) and channel 2 (*cluster42AB* or *cluster38C1* RNA FISH in the red channel). A band-pass filter was applied to shape out the loci and reduce differences in intensities for segmentation. Larger clusters were segmented into individual spots by detecting seed points on local maxima. RNA FISH signal from cluster transcripts is observed both inside the nucleus (representing transcription loci[6]) and in the cytoplasmic nuclear peripheral region, the nuage. Therefore, to quantify specifically the transcriptional output of piRNA source loci, only loci objects within the nucleus were counted. Objects touching the borders with more than 25% surface area were excluded from the analysis as these may not have represented transcriptional foci. Segmented loci were then resized to the full width of half maximum to approximate the real extent. Number, size, and intensities per cell and per channel were exported for analysis and plotting in R. Statistical differences between genotype groups were tested using non-parametric Wilcoxon rank sum tests.

Defining and curating 1-kb genomic windows. The genome 1-kb tiles were generated as previously described[6]. Briefly, we split the main chromosomes of *D. melanogaster* dm6 (r6.10) genome into non-overlapping 1-kb tiles. A mappability score was then given to each tile on the basis of estimation using mapping of synthetic short reads of 25-nt length.

ChIP-seq. With minor modifications, ChIP was performed as described in ref. 62. Briefly, ~200 pairs of ovaries were dissected into ice-cold PBS, rinsed once, and cross-linked in 1.8% paraformaldehyde/PBS for 10 min at room temperature. Glycine was then added to quench the cross-linking reaction and the ovaries were washed in PBS followed by homogenization in a glass douncer using a tight pestle. Nuclei were then lysed on ice for 20 min and DNA was sheared for 20 min using a Covaris E220 Ultrasonicator. Nuclear lysates were incubated overnight at 4 °C with antibodies specific to the target epitope. Fifty microlitres of a 1:1 mix of Protein A and Protein G Dynabeads were then added and samples were incubated for 2 h at 4 °C. The beads were then washed multiple times and DNA–protein complexes were eluted and de-cross-linked overnight at 65 °C. RNA and protein were digested by RNase A and proteinase K treatment, respectively, before final DNA purification using ChIP DNA Clean & Concentrator columns (Zymo). ChIP efficiency was assessed by qPCR using part of the immunoprecipitation sample and the remainder was then used to prepare barcoded libraries using the NEBNext Ultra DNA Library Prep Kit for Illumina (NEB) and finally sequenced on a HiSeq2500 (Illumina).

ChIP-seq analysis. ChIP-seq reads were trimmed to high-quality bases 5–45 before mapping to the *D. melanogaster* genome (dm6, r6.10) using Bowtie (release 0.12.9) with 0-mismatch tolerance. Reads were then computationally extended to 300 nt, reflecting an estimated median DNA fragment length. Normalization between samples was done on the basis of the number of genome-unique mapping reads for each sample. Subsequent quantification of reads mapping to 1-kb tiles was done using bedtools, while relative quantification and plotting were done in R (see code availability below). Briefly, Rhino ChIP-seq tile signal was normalized to the estimated mappability scores for each 1-kb window, while for Pol II ChIP-seq normalization was done by quantile normalization using the preprocessCore R package. This normalization was under the assumption that the Pol II occupancy did not change globally in any of the assayed genotypes (justified by the observed completion ovary development in all genotypes). A pseudo-count of 1 was then added to each tile value before calculation of $\log_2$(fold change) values relative to control genotype samples.

RNA-seq. Total RNA was purified further using RNAeasy columns, including an on-column DNase I digest (Qiagen). Five micrograms of purified total RNA were subjected twice to Ribo-Zero rRNA removal using a magnetic Human-Mouse-Rat kit (Illumina). Libraries were then cloned using a NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB), following the recommended kit protocol and sequenced on a HiSeq2500 (Illumina). The modENCODE RNA-seq data[63] presented in Extended Data Fig. 2a were extracted from Flybase.

RNA-seq analysis. RNA-seq reads were trimmed to high-quality bases 5–45 before mapping to the *Drosophila* genome (dm6, r6.10) using STAR[64] or to *D. melanogaster* transposon consensus sequences using SALMON[65]. For genomic mapping by STAR, normalization between samples was done on the basis of the number of genome-unique mapping reads for each sample. Subsequent quantification of reads mapping to 1-kb tiles was done using bedtools, while relative quantification and plotting were done in R. Briefly, RNA-seq tile signal was normalized to the estimated mappability scores for each 1-kb window. A pseudo-count of 1 was then added to each tile value before calculation of $\log_2$(fold change) values relative to control genotype samples.

Cap-seq. Cap-seq was performed on the basis of refs 15 and 66. In brief, 1 µg total RNA isolated from wild-type ovaries was treated with TurboDNase (Thermo Fisher Scientific) and purified using RNA Clean & Concentrator-5 columns (Zymo). 5′-Monophosphorylated RNAs were then digested by Terminator Exonuclease enzyme (EpiCentre) and any remaining 5′ phosphorylated RNAs were dephosphorylated by treatment with calf intestine alkaline phosphatase. Next, 5′ caps were removed by treatment with tobacco acid pyrophosphatase enzyme (EpiCentre; note: the product has been discontinued, but can be replaced by RNA 5′ pyrophosphohydrolase (RppH) from NEB). The 5′ linkers were then ligated to the de-capped RNA 5′ ends and cDNA was generated by reverse transcription using an Illumina-compatible RT primer with eight random 3′ nucleotides to allow random priming. The cDNA libraries were amplified by PCR using KAPA HiFi HotStart Realtime Mix (Peqlab) and sequenced on a HiSeq2500 (Illumina).

Degradome-seq. Degradome-seq for profiling of 5′-monophosphorylated RNA 5′ ends was done using the Cap-seq protocol, but omitting the Terminator Exonuclease, calf intestine alkaline phosphatase, and tobacco acid pyrophosphatase enzymatic reactions.

Cap-seq and degradome-seq analysis. Reads were trimmed by removal of the 5′ linker sequence including the four random nucleotides. Trimmed reads were then mapped to the *Drosophila* genome (dm6, r6.10) using Bowtie (release 0.12.9) with 0-mismatch tolerance. Uniquely mapping reads were collapsed to the 5′-most nucleotide for display of 5′ ends specifically. For analyses of DNA sequence biases around the mapping position, reads mapping either to piRNA clusters or to annotated transcription start sites were extracted and counted, and the DNA sequence surrounding the 5′ end mapping sites was retrieved. These DNA sequences were then analysed by generation of weblogos or by quantification of YR motif occurrence (see also code availability below).

Small RNA-seq. Small RNA libraries were generated as previously described[67]. Briefly, 18- to 29-nt-long small RNAs were purified by preparative PAGE from 20 µg of total ovarian RNA. Next, the 3′ linker (containing four random nucleotides) was ligated overnight using T4 RNA ligase 2, truncated K227Q (NEB), after which the products were recovered by a second PAGE purification. 5′ RNA linkers

with four terminal random nucleotides were then ligated to the small RNAs using T4 RNA ligase (NEB) followed by a third PAGE purification. The cloned small RNAs were then reverse transcribed and PCR amplified before sequencing on a HiSeq2500 (Illumina). All linker and primer sequences are given in Supplementary Table 2.

Small RNA-seq analysis. Small RNA sequencing reads were trimmed by removal of the 3′ linker sequence (AGATCGGAAGAGCACACGTCT), as well as the four random nucleotides at each end. Trimmed reads were then mapped to the *Drosophila* genome (dm6, r6.10) using Bowtie (release 0.12.9) with 0-mismatch tolerance. Genome coverage was calculated and normalized to the number of uniquely mapping microRNA reads (in millions). Reads mapping to rRNA, tRNA, snRNA, and snoRNA were excluded. Subsequent quantification of reads mapping to 1-kb tiles was done using bedtools, while relative quantification and plotting were done in R (see code availability below). Briefly, small RNA-seq tile signal was normalized to the estimated mappability scores for each 1-kb window. A pseudo-count of 1 was then added to each tile value before calculation of $\log_2$(fold change) values relative to control genotype samples.

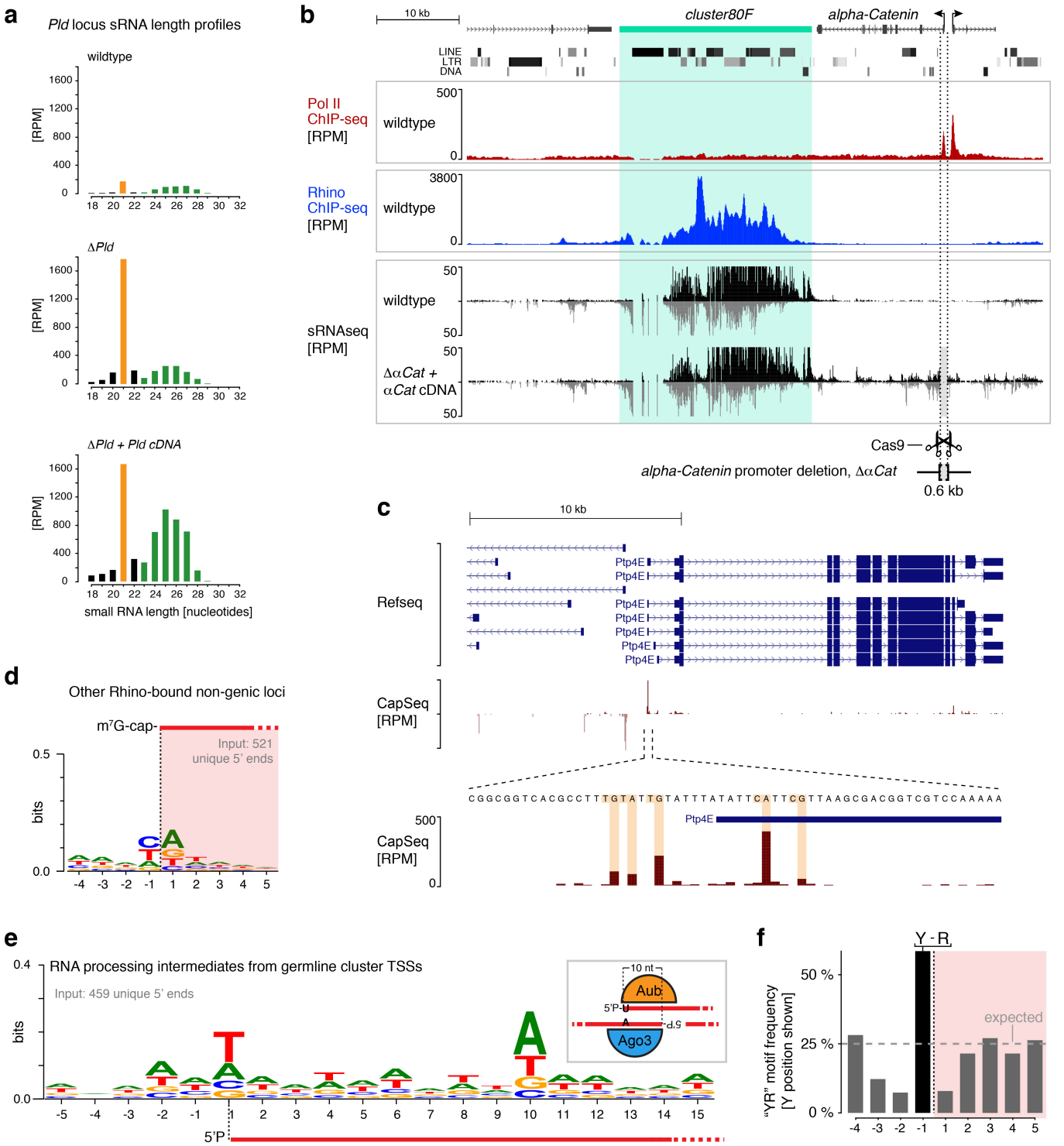
Distant homology searches. An iterative NCBI-PSIBLAST (version 2.4.0+) search with the *D. melanogaster* conserved region of Moonshiner (amino acids 9–168) first identified all *Drosophila* orthologues in round 1. In the following iteration, numerous transcription initiation factor IIA subunit 1 (TFIIA-L) proteins were hit significantly; among these were *Dendroctonus ponderosae* XP_019771835.1 (region 15–165, Expect (*E*) value 6×10^{-07}), *Tribolium castaneum* XP_969067.1 (region 15–160, *E* value 2×10^{-05}), and *D. melanogaster NP_476995.1* (region 138–244, *E* value 0.001). In round 3, after incorporation of insect TFIIA-L proteins into the PSSM model (default inclusion threshold of 0.002), the *Mus musculus* GTF2A1L hit significantly (NP_076119.2, region 12–83, *E* value 0.007). Unlike the arthropod TFIIA-L family that covers both Moonshiner domains, vertebrate GTF2A1 hits lie mainly in the amino-terminal domain. In an alternative search strategy, a hidden Markov model[68] with conserved regions of Moonshiner orthologues significantly identified arthropod TFIIA-L proteins such as *Aedes aegypti* XP_001652503.1 (regions 13–53 and 160–243, *E* value 1.5×10^{-07}).

Orthologue identification and alignment. An NCBI-BLASTP[69] search with the *D. melanogaster* Moonshiner protein (172 amino acids) within the NCBI non-redundant protein database identified orthologues solely in the *Drosophila* genus with significant *E* values below 1×10^{-10}. The sequences were aligned with MAFFT[70] (mafft-linsi, version 7.305b), visualized in Jalview[71], and the secondary structure was predicted with JPRED[72]. The relevant Moonshiner and TFIIA-L sequence accessions can be found in Supplementary Table 6. Two conserved domains could be identified: the amino-terminal domain covers *D. melanogaster* residues 9–64, is characterized by two distinctive α-helices, and is separated from the carboxy-terminal domain (residues 91–168) by a compositionally biased region, rich in proline and lysine residues[73].

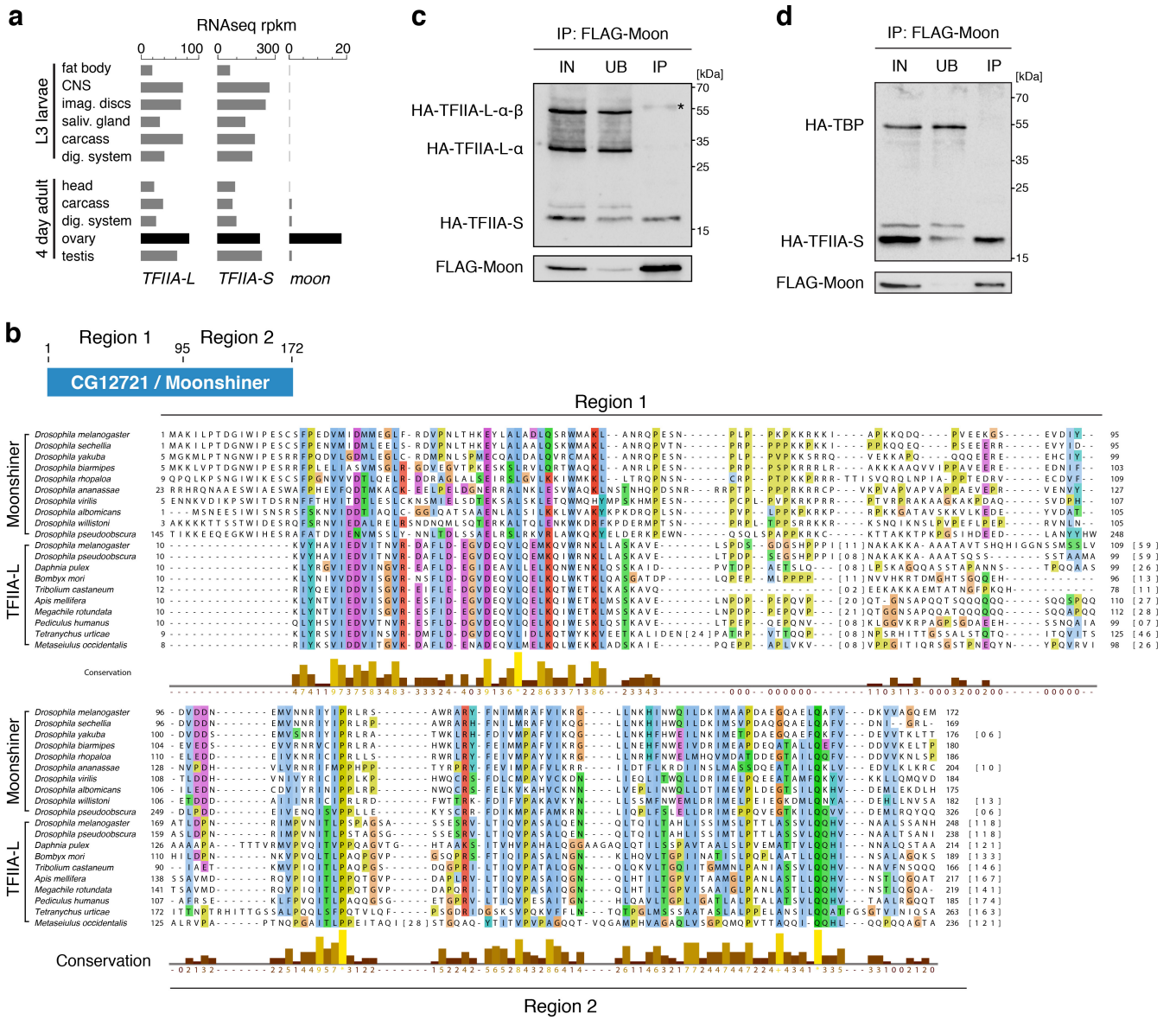
Plotting and data visualization. Data visualization and statistical analyses were done using R[74] in conjunction with the following software packages: ggplot2[75], reshape[76], scales[77], and preprocessCore[78]. The University of California, Santa Cruz (UCSC) genome browser[79,80] was used to explore sequencing data as well as to prepare the genome browser panels shown in the individual figures.

Data and software availability. The main scripts used for the presented analyses as well as raw confocal image files are available upon request from the corresponding authors or from https://gitlab.com/Andersen_Moonshiner_2017. All sequencing data produced for this publication have been deposited in the NCBI Gene Expression Omnibus under accession number GSE97719. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[81] partner repository under data set identifier PXD005026.

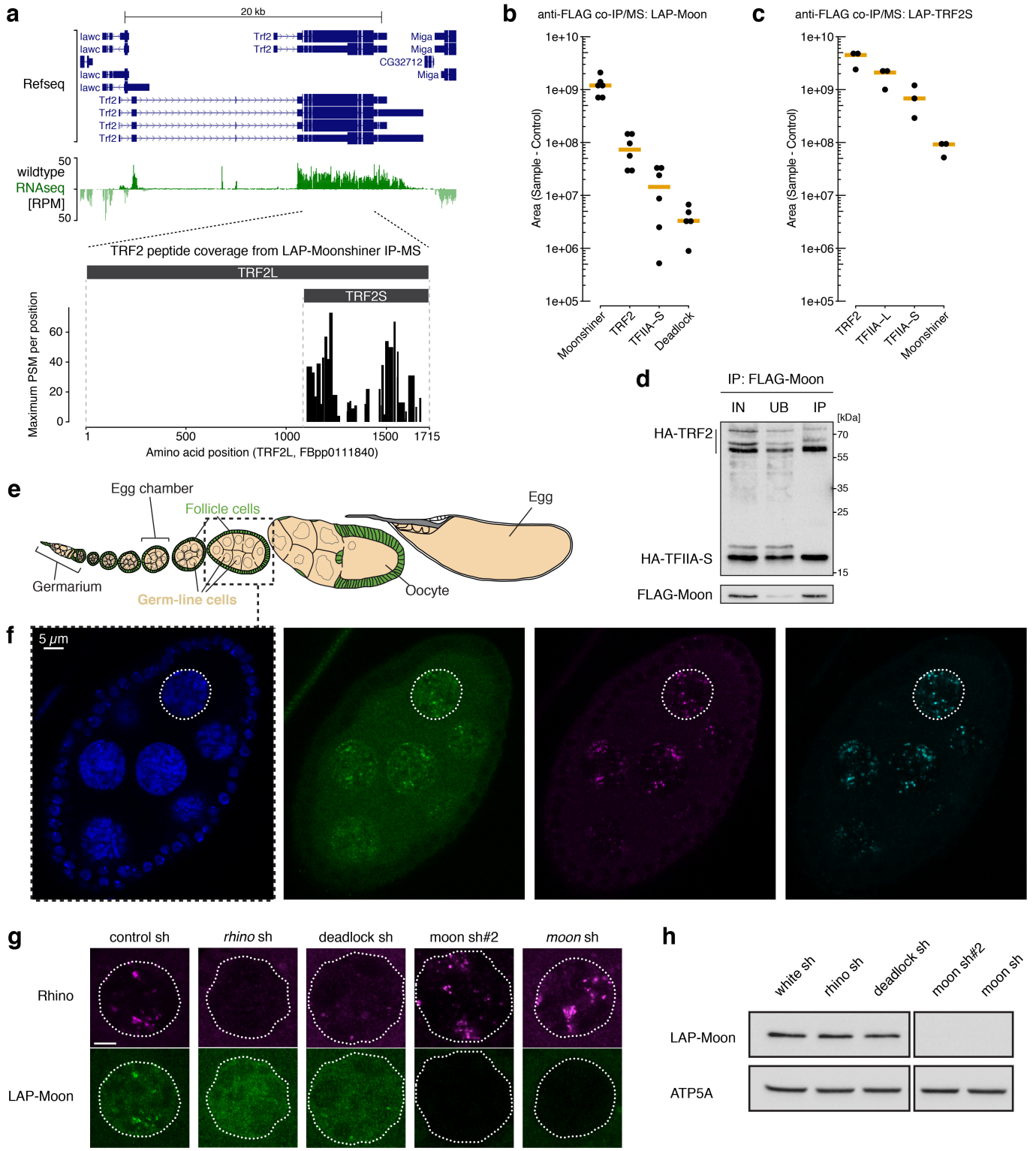
53. Venken, K. J. T. *et al.* Versatile P[acman] BAC libraries for transgenesis studies in *Drosophila melanogaster*. *Nat. Methods* **6,** 431–434 (2009).
54. Handler, D. *et al.* The genetic makeup of the *Drosophila* piRNA pathway. *Mol. Cell* **50,** 762–777 (2013).
55. Sarot, E., Payen-Groschêne, G., Bucheton, A. & Pélisson, A. Evidence for a *piwi*-dependent RNA silencing of the gypsy endogenous retrovirus by the *Drosophila melanogaster flamenco* gene. *Genetics* **166,** 1313–1321 (2004).
56. Gokcezade, J., Sienski, G. & Duchek, P. Efficient CRISPR/Cas9 plasmids for rapid and versatile genome editing in *Drosophila*. *G3 (Bethesda)* **4,** 2279–2282 (2014).
57. Maggert, K. A., Gong, W. J. & Golic, K. G. Methods for homologous recombination in *Drosophila*. *Methods Mol. Biol*. **420,** 155–174 (2008).
58. Dorfer, V. *et al.* MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *J. Proteome Res.* **13,** 3679–3684 (2014).
59. Ritchie, M. E. *et al. limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. **43,** e47 (2015).
60. Stampfel, G. *et al.* Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature* **528,** 147–151 (2015).
61. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9,** 676–682 (2012).
62. Lee, T. I., Johnstone, S. E. & Young, R. A. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat. Protocols* **1,** 729–748 (2006).
63. Brown, J. B. *et al.* Diversity and dynamics of the *Drosophila* transcriptome. *Nature* **512,** 393–399 (2014).
64. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
65. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14,** 417–419 (2017).
66. Nechaev, S. *et al.* Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* **327,** 335–338 (2010).
67. Jayaprakash, A. D., Jabado, O., Brown, B. D. & Sachidanandam, R. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res*. **39,** e141 (2011).
68. Eddy, S. R. Accelerated profile HMM Searches. *PLoS Comput. Biol*. **7,** e1002195 (2011).
69. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. **25,** 3389–3402 (1997).
70. Katoh, K. & Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* **9,** 286–298 (2008).
71. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).
72. Cole, C., Barber, J. D. & Barton, G. J. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res*. **36,** W197–W201 (2008).
73. Wootton, J. C. & Federhen, S. Analysis of compositionally biased regions in sequence databases. *Methods Enzymol*. **266,** 554–571 (1996).
74. R Core Team. R: A language and environment for statistical computing (R Foundation for Statistical Computing, 2016).
75. Wickham, H. *ggplot2* (Springer, 2016).
76. Wickham, H. Reshaping data with the reshape package. *J. Stat. Softw.* **21** (12), 1–20 (2007).
77. Wickham, H. scales: scale functions for visualization. R package version 0.4.0. https://CRAN.R-project.org/package=scales (2016).
78. Bolstad, B. M. preprocessCore: a collection of pre-processing functions. R package version 1.28.0. https://github.com/bmbolstad/preprocessCore (2013).
79. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12,** 996–1006 (2002).
80. Raney, B. J. *et al.* Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* **30,** 1003–1005 (2014).
81. Vizcaíno, J. A. *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44** (D1), D447–D456 (2016).



Extended Data Figure 1 | Characterization of transcription initiation events at piRNA clusters. a, Size profile histograms of small RNAs mapping to the *Pld* gene locus from ovaries with indicated genotypes. siRNAs (21 nt) are highlighted in orange and piRNAs (23–29 nt) are highlighted in green. **b**, UCSC genome browser panels showing *cluster80F* for which flanking promoter dependency was investigated by deletion of the promoter region of *alpha Catenin*. Shown are Pol II occupancy (red), Rhino occupancy (blue), and piRNA levels (black/grey). Flanking transcription units are shown in grey, light grey shading indicates the experimental promoter deletion. As *alpha Catenin* is an essential gene, a cDNA rescue transgene was expressed from another locus. **c**, UCSC genome browser panels showing the Cap-seq profile at the promoter of a canonical gene. **d**, DNA sequence motif at 5′ ends of capped RNAs mapping to Rhino-bound genomic loci (Rhino ChIP-seq reads per kilobase per million mappers > 300; *cluster80F* and *42AB* excluded) outside known transcription units. **e**, DNA sequence motif at 5′ ends of 5′-monophosphorylated RNAs mapping to *cluster42AB* or *cluster80F*. The schematic to the right shows how the 'ping-pong' amplification loop involving Aub- and Ago3-mediated cleavages gives rise to the observed sequence biases at positions +1 and +10. **f**, Histogram of the 'YR' dinucleotide occurrence around *cluster42AB* and *cluster80F* transcription start sites (expected chance occurrence 25%).

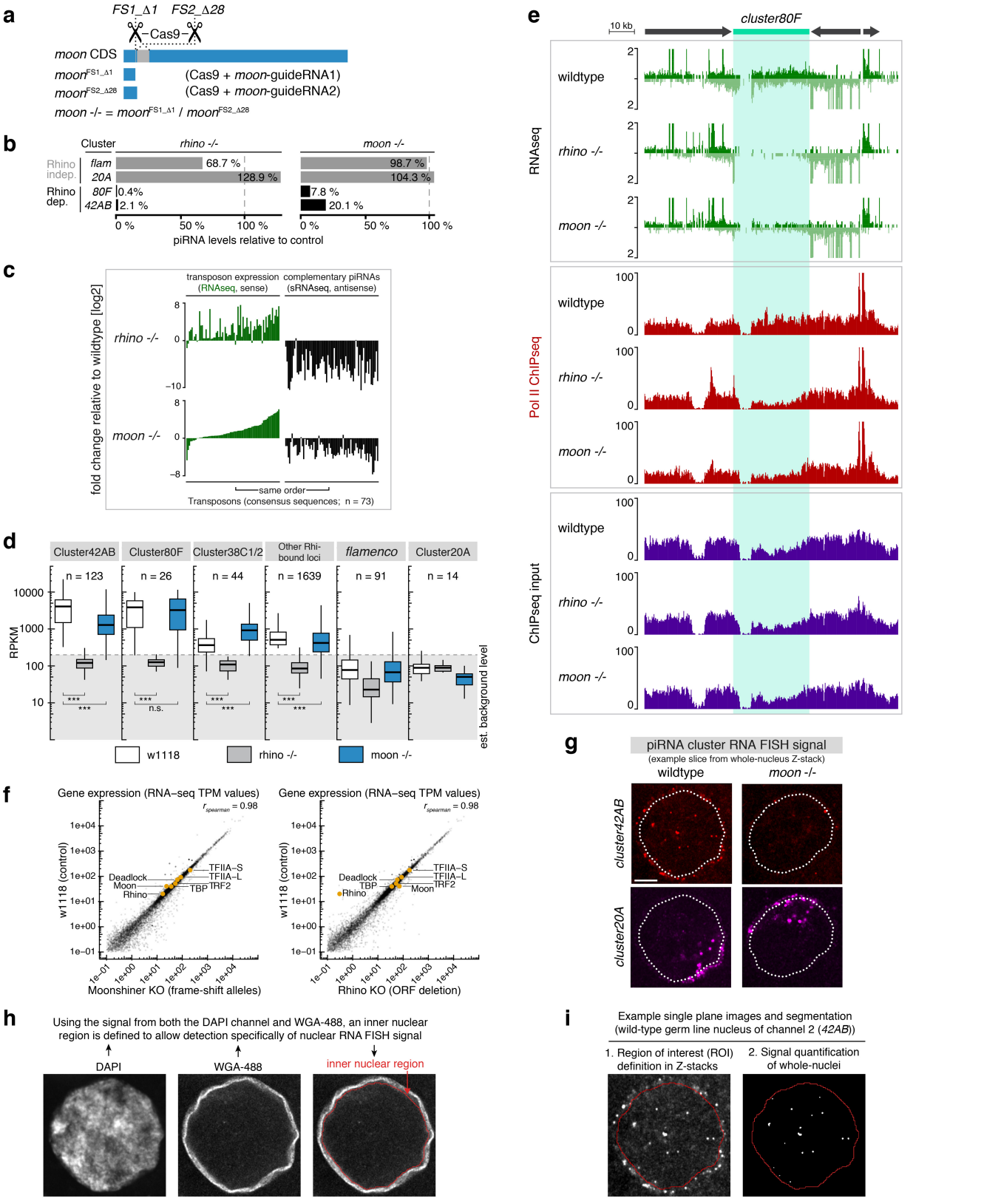


Extended Data Figure 2 | CG12721/Moonshiner is a germline-specific TFIIA-L paralogue. a, Expression levels of indicated genes in larval/adult tissues on the basis of modENCODE RNA-seq data; rpkm, reads per kilobase per million mappers. **b**, The top schematic denotes the two regions of homology shown in Fig. 2a. Shown below is the amino-acid sequence alignment of these two regions from drosophilid species (Moonshiner) and selected insect species (TFIIA-L). The alignment was created using JalView with standard ClustalX colour coding and conservation score calculation. **c**, **d**, Western blot analyses of Flag–Moonshiner co-immunoprecipitation from lysates of S2 cells transfected with indicated expression constructs (IN, input; UB, unbound; IP, immunoprecipitate; asterisk, signal from anti-Flag heavy chain).



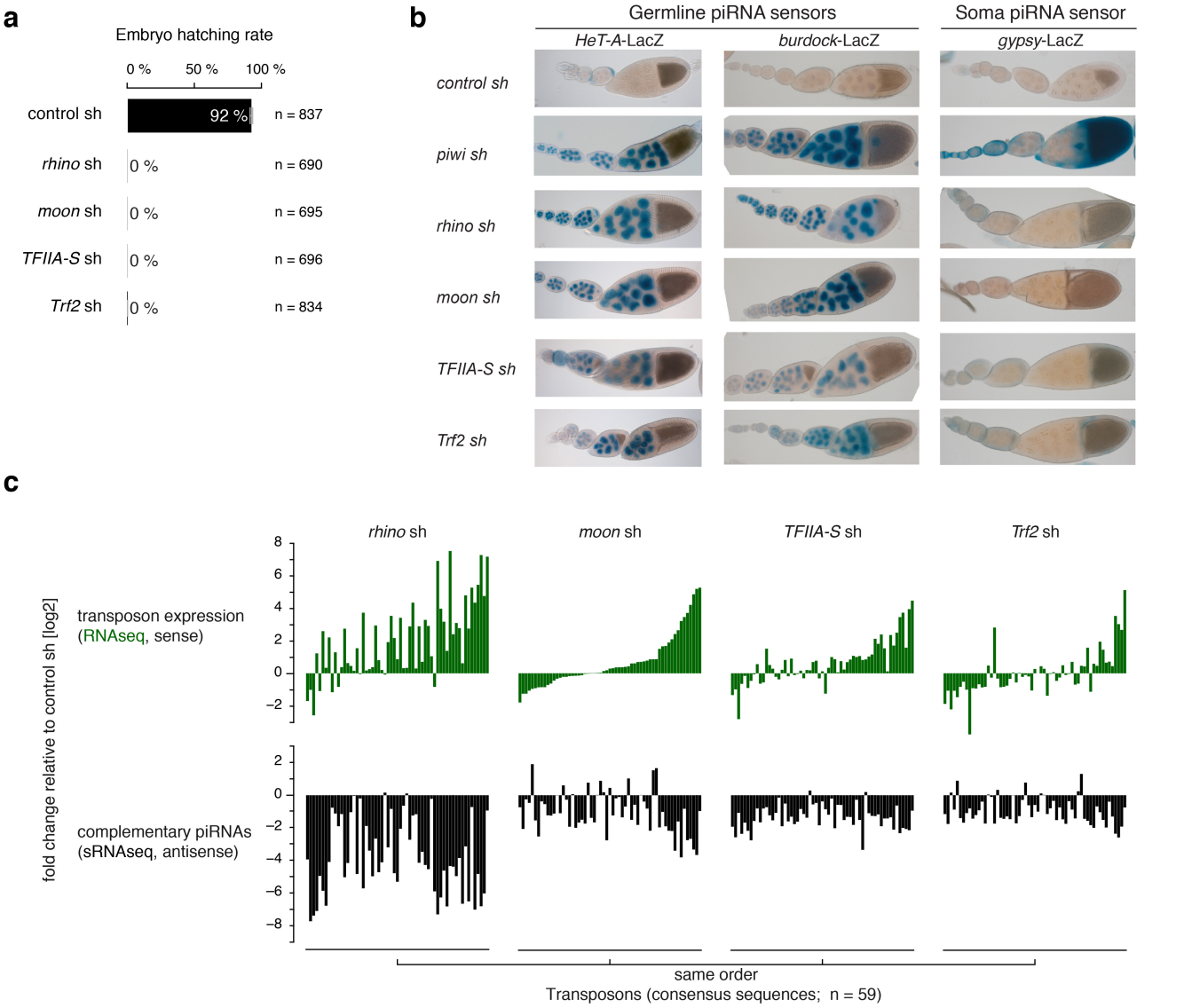
Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Moonshiner forms an alternative TFIIA–TRF2 complex enriched at piRNA clusters. a, TRF2 isoform characterization by total wild-type ovary RNA-seq (top) and LAP–Moonshiner co-immunoprecipitation mass spectrometry (bottom). The identified TRF2 peptides show that Moonshiner is in complex only with the shorter TRF2 isoform. We therefore specifically investigated this isoform, also known as TRF2S, in the remainder of the study. **b**, **c**, Absolute peptide peak intensities for the main protein interactors identified in Fig. 2b, c. Peak area intensities are displayed as immunoprecipitation values subtracted that of the paired control immunoprecipitation experiment. On the basis of this, we conclude that TFIIA-S and TRF2 are robust Moonshiner interactors (supportive of an alternative TFIIA–TRF2 complex), while only a small fraction of Moonshiner is bound to Deadlock. Furthermore, the data show that, in ovaries, TRF2 interacts predominantly with canonical TFIIA, but also clearly with Moonshiner. Black dots represent individual biological replicate values. Orange bars show median values. **d**, Western blot analyses as Extended Data Fig. 2d, but addressing interaction with HA–TRF2 (lower bands probably represent TRF2 decay intermediates). **e**, Schematic of a developing *Drosophila* ovariole with germline cells in beige and somatic support cells in green. Confocal images were typically taken from egg chambers of stage 7 (highlighted by a dashed box). **f**, Whole egg chamber confocal image stained for DNA (DAPI; blue), LAP–Moonshiner (GFP auto-fluorescence; green), Rhino (magenta), and Deadlock (cyan). The circled nucleus is shown in Fig. 2d. **g**, Fluorescence images of nurse cell nuclei (depleted for indicated factors using sh-lines) indicating levels and localization of Moonshiner and Rhino (scale bar, 5 μm). **h**, Western blot showing levels of LAP–Moonshiner in ovaries where the indicated factors were depleted in the germline via sh-lines (ATP synthase serves as loading control).

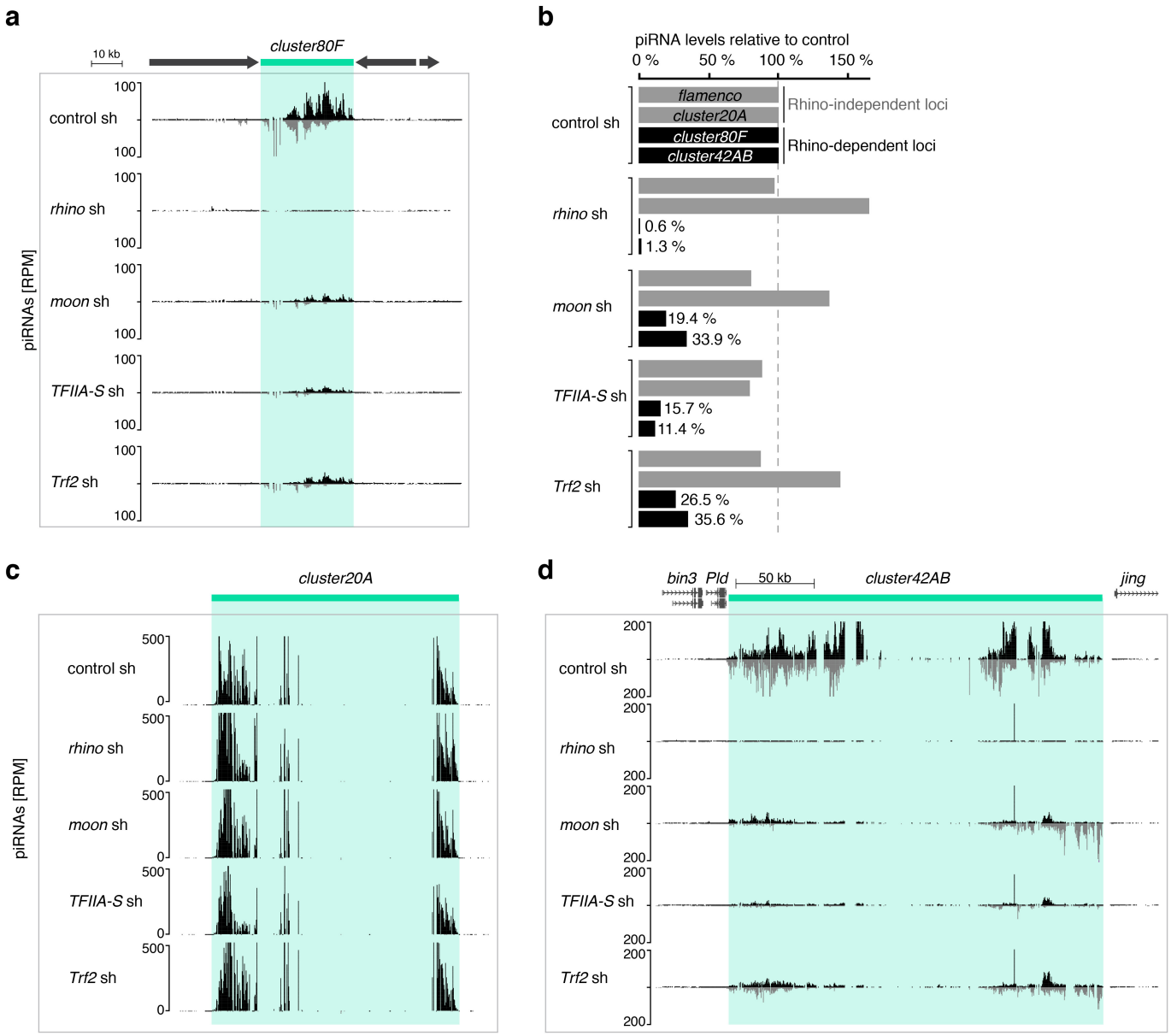


Extended Data Figure 4 | See next page for caption.

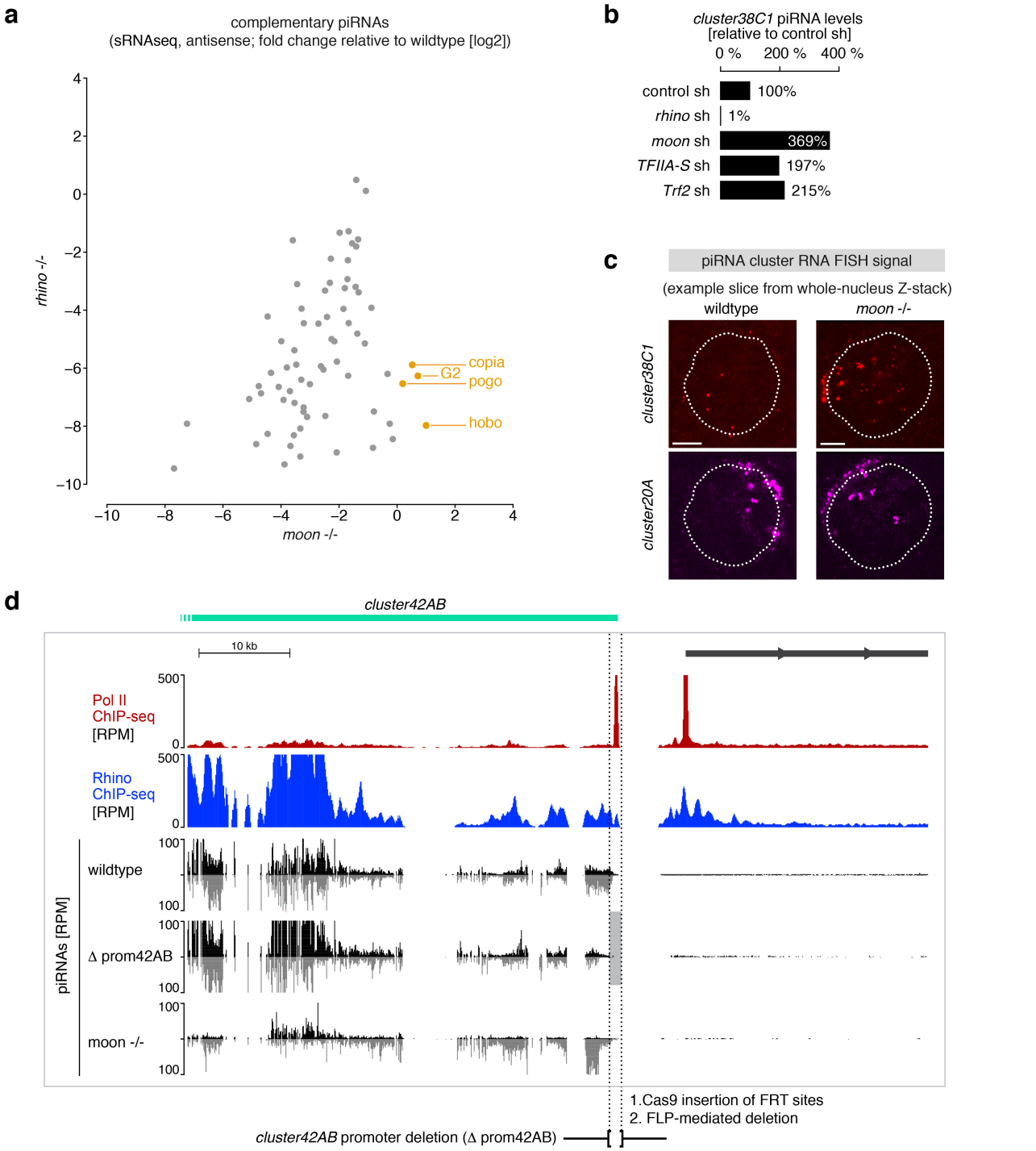
Extended Data Figure 4 | Moonshiner mutants reveal highly specific function at Rhino-bound piRNA clusters. a, Schematic of the *moonshiner* frameshift alleles generated by CRISPR/Cas9. **b**, piRNA levels from ovaries with indicated genotype (relative to wild type) mapping uniquely to indicated piRNA clusters. **c**, Left: deregulation of steady-state transposon transcript levels (RNA-seq; sense only) in ovaries of the indicated mutants fly strains. Right: changes in corresponding piRNA levels (antisense only). The *y* axis values show $\log_2$(fold change) of transcripts per million values relative to wild type. Each bar represents one transposon consensus sequence ($n = 73$; shown are only transposons with minimum expression of RNA-seq transcripts per million > 5 in any library). Sorting of transposons in all panels is identical. The plotted values are available as figure source data. **d**, Rhino occupancy at indicated major piRNA clusters as well as all other Rhino-bound loci is shown as boxplot quantification ($n = 1$-kb windows analysed for each group) of Rhino ChIP-seq read coverage in the indicated genotypes. Boxplots are defined as in Fig. 3c; ***$P < 0.0001$ based on Mann–Whitney–Wilcoxon non-parametric tests. **e**, Genome browser panel showing read coverage at *cluster80F* of the data underlying the $\log_2$(fold change) tracks shown in Fig. 3b. Shown are RNA-seq (green), Pol II ChIP-seq (red), and ChIP-seq input samples (purple) generated from the indicated genotypes. **f**, RNA-seq transcripts per million values for canonical genes compared between control and *moonshiner*$^{-/-}$ (left) or *rhino*$^{-/-}$ (right); key genes related to Moonshiner biology are highlighted in orange. Abbreviation $r_{Spearman}$ denotes the Spearman correlation coefficient for each data set pair. **g**, Representative confocal images underlying the quantitative RNA FISH-based detection of piRNA precursors from *cluster20A* (Rhino-independent) and *cluster42AB* (Rhino-dependent) in germline nuclei of wild-type and *moonshiner* mutant ovaries. **h**, Example confocal images of germline nuclei stained for of DNA (DAPI) and nuclear pore complexes (wheat germ agglutinin, WGA-488), which were used to define the nuclear region in whole-nucleus *Z*-stack images acquired in parallel with images of RNA FISH signal. **i**, Example single-plane images of dual-channel RNA FISH quantification of whole-germline nuclei. RNA FISH signal within the nuclear regions (left, segmented using DAPI and WGA-488 signal) was used to define regions of interest (right), representing active sites of piRNA cluster transcription[6]. Signal in the foci was subsequently quantified for whole nuclei.



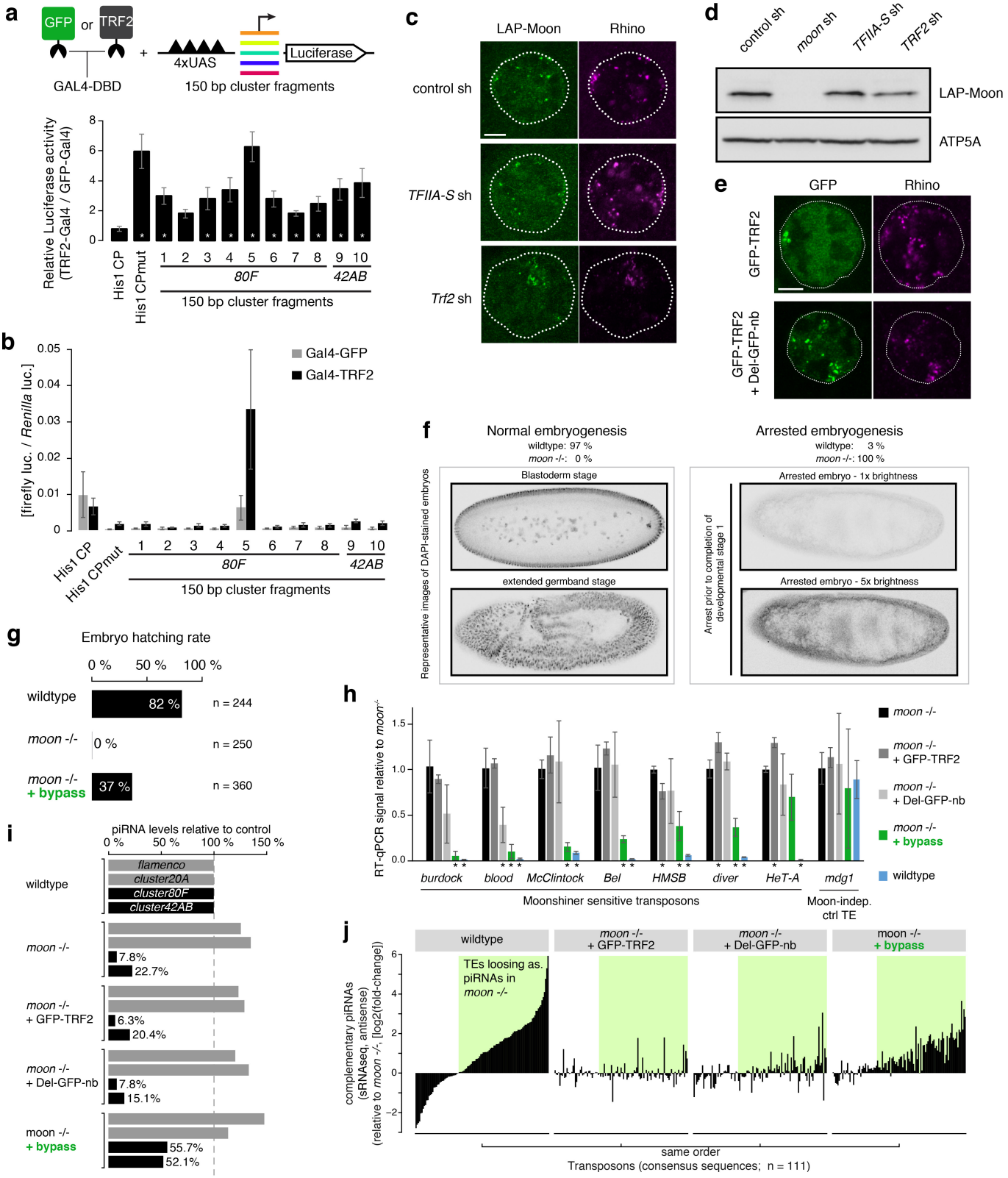
Extended Data Figure 5 | Depletion of Moonshiner, TFIIA-S, or Trf2 activates transposon expression. a, Percentages of eggs hatching into larvae laid by females expressing sh-constructs against the indicated target genes in their germline cells. Error bars, s.e.m. from four independent countings; *n*, the sum of counted eggs (see also figure source data). **b**, Ovarioles from flies expressing indicated piRNA sensors and indicated germline knockdown constructs (sh-lines) stained for β-galactosidase with X-gal. **c**, Top: deregulation of steady-state transposon transcript levels (sense only; compared with control ovaries) in ovaries expressing the indicated germline knockdown constructs. Each bar represents one transposon consensus sequence ($n = 59$; shown are only transposons with minimum expression of RNA-seq transcripts per million > 5 in any library). Bottom: changes in corresponding piRNA levels (antisense only). Sorting of transposons in all panels is identical. For plotted values see figure source data.



Extended Data Figure 6 | piRNA production from Rhino-bound clusters requires Moonshiner, TFIIA-S, and Trf2. a, UCSC genome browser panel showing piRNA profiles at *cluster80F* in ovaries expressing indicated germline knockdown constructs. **b**, Levels of piRNAs (relative to control) mapping uniquely to indicated Rhino-dependent or Rhino-independent piRNA clusters and derived from ovaries depleted of the indicated factors. **c**, **d**, UCSC genome browser panel showing *cluster20A* (**c**) or *cluster42AB* (**d**) piRNA levels from ovaries expressing indicated germline knockdown constructs.

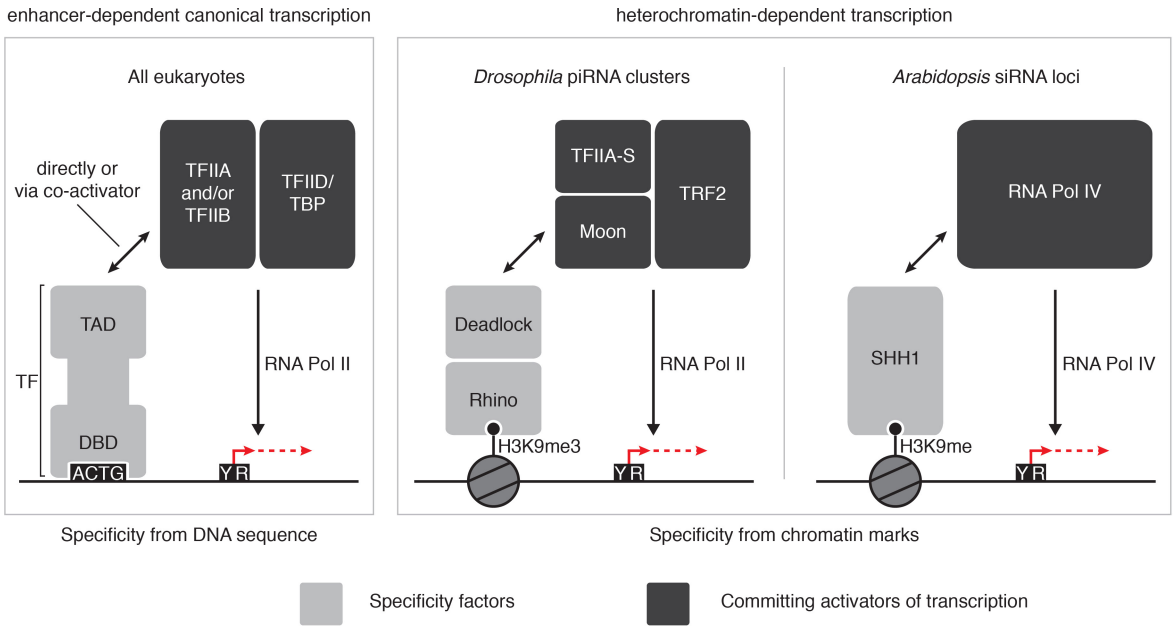


Extended Data Figure 7 | Characterization of Rhino-dependent, but Moonshiner-independent, piRNA production. a, The $\log_2$(fold changes) in levels of piRNAs mapping antisense to transposons are plotted for *rhino* mutants versus *moonshiner* mutants. An outlier group of transposons for which the level of antisense piRNAs is decreased in *rhino* mutants but increased in *moonshiner* mutants is apparent, and elements enriched in *cluster38C1/2* are highlighted in orange. The same transposons are shown as in Extended Data Fig. 4c (n = 73; transposon mRNAs analysed). **b**, Quantification of relative piRNA levels originating from *cluster38C1* in ovaries from flies subjected to the indicated germline knockdowns. Percentages relative to control knockdowns were calculated with the total numbers of piRNA reads mapping uniquely to *cluster38C1*. **c**, Representative confocal images underlying the quantitative RNA FISH-based detection of piRNA precursors from *cluster20A* (Rhino-independent) and *cluster38C1* (Rhino-dependent) in germline nuclei of wild-type and *moonshiner* mutant ovaries (scale bar, 5 μm). **d**, UCSC genome browser panel showing the most distal part of *cluster42AB* for which piRNA production dependency on the right flanking promoter was investigated by deletion of the promoter region. Shown are Pol II occupancy (red), Rhino occupancy (blue), and piRNA levels (black/grey). Flanking transcription units are shown in grey; light grey shading indicates the experimental promoter deletion.



Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | Moonshiner function can be bypassed by directly connecting Deadlock to Trf2. a, Experimental scheme used to recruit GFP or TRF2 to DNA upstream of sequences of interest to test for stimulation of *Luciferase* transcription. Bar diagram shows fold changes in reporter activity upon tethering of TRF2 versus GFP to wild-type or mutant Histone 1 core promoter or to random piRNA cluster fragments (error bars, s.e.m.; $n = 5$ biological replicates; $^*P < 0.05$ based on two-tailed paired *t*-tests). **b**, Firefly luciferase values underlying the relative activities shown in **a**. Firefly luciferase activity was normalized to *Renilla* luciferase activity (transfection and viability control) upon tethering of TRF2 versus GFP to wild-type or mutant Histone 1 core promoter or to ten random piRNA cluster fragments (error bars, s.d. of five biological replicates each with six technical replicates. **c**, Confocal images showing localization of LAP–Moonshiner and Rhino in germline nuclei of ovaries depleted for indicated factors (scale bar, 5 μm). **d**, Western blot showing levels of LAP–Moonshiner in ovaries where the indicated factors were depleted in the germline via sh-lines (ATP synthase serves as loading control). **e**, Confocal images showing localization of germline-expressed LAP–TRF2 and endogenous Rhino in control ovaries (top) or in ovaries expressing the Deadlock–GFP-nanobody fusion protein (bottom) (scale bar, 5 μm). The TRF2 accumulations in wild-type nuclei do not overlap with Rhino foci and instead are reported to be TRF2 accumulations at the repetitive histone loci[31]. We note that TRF2 accumulation at Rhino foci is not visible in wild-type cells, most probably as the levels of this protein are too high to detect this local enrichment, which depends on Moonshiner (a protein expressed at only low levels). **f**, Representative images of DAPI-stained embryos (inverted monochromatic) assessed for progress of early embryogenesis. Left: two images of normal embryo development at the blastoderm stage (top) and at the extended germband stage (after gastrulation; bottom). Right: a typical *moonshiner* mutant embryo arrested early in development (no distinct nuclei are visible; the lower image displays the top image at increased brightness). **g**, Percentages of embryos with the indicated genotype displaying successful hatching. **h**, Relative levels of steady-state transposon mRNAs underlying the panel displayed in Fig. 5d. Bars show mean levels relative to those measured in $moon^{-/-}$ samples. Error bars, s.d. of three biological replicates. $^*P < 0.05$ from two-tailed *t*-tests for difference to *moonshiner* full mutant samples. **i**, Levels of piRNAs mapping uniquely to the indicated clusters (grey, Rhino-independent; black, Rhino-dependent) in the indicated genotypes (values are normalized to the wild-type control levels). **j**, The $\log_2$(fold changes) in levels of piRNAs mapping antisense to transposons are plotted relative to levels in *moonshiner* mutants. The green boxes highlight the set of transposons for which mutation of *moonshiner* results in decreased antisense piRNAs ($n = 111$; transposons with fewer than 100 antisense piRNAs per million were removed from the analyses).



Extended Data Figure 9 | Comparison of canonical enhancer-dependent and heterochromatin-dependent transcription activation pathways. Schematic comparison of canonical enhancer-dependent transcription and transcription of small RNA source loci in *Drosophila* and *Arabidopsis* specified by chromatin marks. Canonical transcription initiation is driven by sequence-specific transcription factor binding to DNA motifs in accessible enhancer and promoter regions, which subsequently leads to positioning of TFIID/TBP onto core promoters (left). In contrast, while Moonshiner-mediated transcription also converges on recruitment of TFIID to DNA, this pathway exclusively utilizes the TBP paralogue TRF2. Furthermore, Moonshiner-mediated transcription gains locus specificity via recognition of heterochromatic histone marks through the HP1 protein Rhino, rather than through DNA motifs, thereby circumventing the transcriptional inhibition imposed by the compact state of heterochromatic DNA (middle). In plants, a conceptually similar pathway has evolved using an entirely different set of proteins (right). Here, the homeodomain protein SHH1 binds H3K9me histone marks and subsequently recruits the Pol IV variant RNA polymerase complex to transcribe small RNA precursors.

nature research

Corresponding author(s): Julius Brennecke

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

Sample sizes were chosen as large as possible while still practically feasible in terms of data collection. Adequate statistics has been applied throughout the manuscript in order to make sure that the observed effects are significant given the reported sample size.

2. Data exclusions

Describe any data exclusions.

no data exclusion in this manuscript.

3. Replication

Describe whether the experimental findings were reliably reproduced.

All attempts at replication were successful, except when caused by technical issues (e.g. material loss during handling leading to low-complexity NGS libraries)

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Not relevant as grouping was not applied.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

No investigator blinding was applied during data acquisition or analyses as the data was mostly analyzed in bulk by (blind) scripts such as for NGS or FISH quantification or blinding was not desirable for data presentation (e.g. blinded loading of western blots).

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	We provide a link to a GitHub depository for the custom code that was used in this manuscript.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	there are no restrictions

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	Antibodies used in this study have either been validated in previous publications (cited in this manuscript) or have been generated for this particular manuscript (anti-Deadlock). Validation of this antibody was by using Deadlock mutants and observing loss of the signal in immuno-fluorescence experiments.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	used cell line in this study: Drosophila melanogaster Schneider cells. This is a standard cell line used in the field.
b. Describe the method of cell line authentication used.	does not apply
c. Report whether the cell lines were tested for mycoplasma contamination.	cells are routinely controlled for mycoplasm infection in the in house facility.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	does not apply

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	This study involved exclusively work with Drosophila melanogaster, a standard invertebrate model organism that does not underlie any ethical restrictions. Standard laboratory procedures have been applied throughout the study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	does not apply

nature research

Corresponding Author: Julius Brennecke

Date: 2017-06-07

ChIP-seq Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data deposition

1. For all ChIP-seq data:

☒ a. Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ b. Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

2. Provide all necessary reviewer access links. (The entry may remain private before publication.)

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=yvcbcqeybvylfkp&acc=GSE97719

3. Provide a list of all files available in the database submission.

ChIPseq_antiRhino_w1118_1_22682_uniq.bw
ChIPseq_antiRhino_Rhino_rhino_KO_22683_uniq.bw
ChIPseq_antiRhino_Rhino_moonshiner_KO_22684_uniq.bw
ChIPseq_antiRhino_Pld_promoter_deletion_43186_uniq.bw
ChIPseq_antiRNAPol2_w1118_1_46327_uniq.bw
ChIPseq_antiRNAPol2_rhino_KO_46330_uniq.bw
ChIPseq_antiRNAPol2_moonshiner_KO_46333_uniq.bw
ChIPseq_input_w1118_1_46326_uniq.bw
ChIPseq_input_rhino_KO_46329_uniq.bw
ChIPseq_input_moonshiner_KO_46332_uniq.bw
ChIPseq_antiRhino_w1118_1_22682.bam
ChIPseq_antiRhino_Rhino_rhino_KO_22683.bam
ChIPseq_antiRhino_Rhino_moonshiner_KO_22684.bam
ChIPseq_antiRhino_Pld_promoter_deletion_43186.bam
ChIPseq_antiRNAPol2_w1118_1_46327.bam
ChIPseq_antiRNAPol2_rhino_KO_46330.bam
ChIPseq_antiRNAPol2_moonshiner_KO_46333.bam
ChIPseq_input_w1118_1_46326.bam
ChIPseq_input_rhino_KO_46329.bam
ChIPseq_input_moonshiner_KO_46332.bam

4. If available, provide a link to an anonymized genome browser session (e.g. UCSC).

UCSC browser-compatible bigwig (.bw) files are included in the GEO submission

Methodological details

5. Describe the experimental replicates.

The ChIPseq experiments are supported by complimentary methods:
- Rhino ChIPseq: supported by Rhino IF to assess the typical accumulation in nuclear foci representing germline piRNA clusters (Mohn et al 2014)
- RNA Pol II ChIPseq: supported by evaluation of piRNA cluster transcription by RNAseq and quantitative RNA FISH analyses.

6. Describe the sequencing depth for each experiment.

ChIPseq_antiRhino_w1118_1_22682.bam: 43.52 million reads with 26.01 million uniquely mapped
ChIPseq_antiRhino_Rhino_rhino_KO_22683.bam: 53.33 million

reads with 41.99 million uniquely mapped
ChIPseq_antiRhino_Rhino_moonshiner_KO_22684.bam: 58.43 million reads with 33.99 million uniquely mapped
ChIPseq_antiRhino_Pld_promoter_deletion_43186.bam: 30.7 million reads with 12.09 million uniquely mapped
ChIPseq_antiRNAPol2_w1118_1_46327.bam: 30.85 million reads with 21.52 million uniquely mapped
ChIPseq_antiRNAPol2_rhino_KO_46330.bam: 28.14 million reads with 20.59 million uniquely mapped
ChIPseq_antiRNAPol2_moonshiner_KO_46333.bam: 30.7 million reads with 23.01 million uniquely mapped
ChIPseq_input_w1118_1_46326.bam: 29.26 million reads with 21.97 million uniquely mapped
ChIPseq_input_rhino_KO_46329.bam: 34.8 million reads with 26.47 million uniquely mapped
ChIPseq_input_moonshiner_KO_46332.bam: 31.37 million reads with 24.01 million uniquely mapped

All libraries were sequence single-end 50 bp

7. Describe the antibodies used for the ChIP-seq experiments.

ChIPseq_antiRhino samples: Anti-Rhino polyclonal antibody produced in Rabbit (Mohn et al 2014)
ChIPseq_antiRNAPol2 sample: anti-RNA Polymerase II, 8WG16 (Abcam, ab819)

8. Describe the peak calling parameters.

Peak calling was not utilized in the study

9. Describe the methods used to ensure data quality.

1. qPCR-based enrichment over input (>100 fold enriched at expected loci relative to genomic background)
2. Visual inspection of data in the genome browser to confirm that the ChIPseq signal in wildtype accumulates as expected based on previous literature
3. For Rhino ChIPseq: inclusion of Rhino ChIPseq from Rhino null flies as a control for background IP signal.

10. Describe the software used to collect and analyze the ChIP-seq data.

ChIPseq reads were trimmed to high quality bases 5-45 before mapping to the Drosophila melanogaster genome (dm6, r6.10) using Bowtie (release 0.12.9) with 0-mismatch tolerance. Reads were then computationally extended to 300 nt, reflecting an estimated median DNA fragment length. Normalization between samples was done based on the number of genome-unique mapping reads for each sample. Subsequent quantification of reads mapping to 1 kb tiles was done using bedtools, while relative quantification and plotting was done in R. Rhino ChIP-seq tile signal was normalized to the estimated mappability scores for each 1 kb window, while for Pol II ChIP-seq normalization was done by quantile normalization using the preprocessCore R package. This normalization is under the assumption the Pol II occupancy does not change globally in any of the assayed genotypes (justified by the observed completion ovary development in all genotypes). A pseudo-count of 1 was then added to each tile value before calculation of log2 fold-change values relative to control genotype samples.