A dual histone code specifies the binding of heterochromatin protein Rhino to a subset of piRNA source loci

Abdou Akkouche[2,*], Emma Kneuss[1,*], Susanne Bornelöv[1,*], Yoan Renaud[2], Evelyn L. Eastwood[1], Jasper van Lopik[1], Nathalie Gueguen[2], Mingxuan Jiang[1], Pau Creixell[1], Stéphanie Maupetit-Mehouas[2], Benjamin Czech Nicholson[1,#], Emilie Brasset[2,#], Gregory J. Hannon[1,#]

[1] Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge CB2 0RE, UK
[2] iGReD, Université Clermont Auvergne, CNRS, INSERM, Faculté de Médecine, 63000 Clermont-Ferrand, France

* These authors contributed equally
[#] To whom correspondence should be addressed: BCN (ben.nicholson@cruk.cam.ac.uk), EB (emilie.brasset@uca.fr), GJH (greg.hannon@cruk.cam.ac.uk)

Keywords: Epigenetics, *Drosophila*, piRNA pathway, transposon silencing, Heterochromatin

ABSTRACT

Animal germ cells deploy a specialized small RNA-based silencing system, called the PIWI-interacting RNA (piRNA) pathway, to prevent unwanted expression of transposable elements and maintain genome integrity. In *Drosophila* germ cells, the majority of piRNA populations originate from dual-strand piRNA clusters, genomic regions highly enriched in transposon fragments, via an elaborate protein machinery centred on the heterochromatin protein 1 homolog, Rhino. Although Rhino binds to peptides carrying trimethylated H3K9 in vitro, it is not fully understood why in vivo only a fraction of H3K9me3-decorated heterochromatin is occupied by Rhino. Recent work uncovered that Rhino is recruited to a subset of piRNA clusters by the zinc finger protein Kipferl. Here we identify a Kipferl-independent mode of Rhino targeting that is dependent on the histone H3 lysine 27 methyltransferase Enhancer of Zeste and the presence of H3K9me3 and H3K27me3 marks. At Kipferl-independent sites, we find that Rhino, through its dimeric chromodomain, specifically binds to loci marked by both H3K9me3 and H3K27me3. These results expand our understanding of the characteristic binding profile of the heterochromatin protein Rhino. Our work reveals a role for dual histone modifications in defining the binding specificity of a chromatin protein.

INTRODUCTION

Transposable elements (TEs) are genetic sequences, present in nearly all organisms, that have the ability to move and insert themselves into different positions within a genome. While TEs can contribute to genetic diversity and evolution, their uncontrolled activity in gonadal cells can have detrimental effects, including the potential to dramatically reduce fertility. Thus, in animal gonads, a dedicated mechanism known as the PIWI-interacting RNA (piRNA) pathway plays a crucial role in silencing and controlling active TEs to safeguard germline integrity (Brennecke et al. 2007; Czech et al. 2018; Ozata et al. 2019). The piRNA machinery utilises 23- to 30-nt small RNA molecules, which associate with PIWI family proteins to recognize and target TE transcripts for post-transcriptional degradation or co-transcriptional silencing. In *Drosophila* ovaries, the majority of precursor transcripts that give rise to piRNAs originate from specific, transposon-rich genomic regions called piRNA clusters, with the remaining referred to as piRNA source loci. While all piRNA source loci in somatic follicle cells resemble canonical transcription units, in the germline most piRNA clusters are expressed non-canonically and produce piRNAs from both genomic strands (Klattenhoff et al. 2009; Goriaux et al. 2014; Mohn et al. 2014; Zhang et al. 2014; Chen et al. 2016; Andersen et al. 2017).

Rhino (Rhi), a homologue of Heterochromatin Protein 1a (HP1a), is a germline-specific piRNA pathway factor essential for piRNA production, TE repression, and fertility (Klattenhoff et al. 2009). Both Rhi and HP1a belong to a distinct group of chromatin-binding proteins containing chromodomains (CDs). Among CD-containing proteins, different unique binding properties exist. For instance, *Drosophila* HP1a is primarily associated with H3K9me2/3-enriched regions such as constitutive heterochromatin, whereas the developmental regulator Polycomb (Pc), another chromodomain protein, specifically binds to H3K27me3 marks often found in facultative heterochromatin (Franke et al. 1992; Fischle et al. 2003; Bernstein et al. 2006). In line with Rhi being evolutionary related to HP1a, several studies have shown that Rhi localises to discrete genomic regions, typically characterised by repeat-rich content and the presence of

H3K9me3 marks (Klattenhoff et al. 2009; Le Thomas et al. 2014; Mohn et al. 2014; Zhang et al. 2014).

At the genomic regions it binds, Rhi acts as nucleator for a multi-protein complex that triggers non-canonical transcription from heterochromatic regions and facilitates the export of the resulting piRNA precursors from the nucleus (Klattenhoff et al. 2009; Pane et al. 2011; Le Thomas et al. 2014; Mohn et al. 2014; Zhang et al. 2014; Chen et al. 2016; Hur et al. 2016; Andersen et al. 2017; Zhang et al. 2018; ElMaghraby et al. 2019; Kneuss et al. 2019; Zhang et al. 2021). Although it was demonstrated that Rhi binds the repressive histone mark H3K9me3 in vitro (Le Thomas et al. 2014; Mohn et al. 2014; Yu and Huang 2015), Rhi immunofluorescence staining and previously published ChIP-seq data indicate that in vivo it associates only with a fraction of the genomic regions enriched in H3K9me3 marks (Mohn et al. 2014), hence raising questions about how its specificity in vivo is determined. A recent study identified the zinc finger domain protein Kipferl (Kipf) as essential cofactor for Rhi in the female germline, with the interaction between the two proteins being mediated through Rhi's chromodomain (Baumgartner et al. 2022). Kipf was shown to recruit Rhi to dedicated heterochromatic loci that exhibit enrichment in a guanosine-rich DNA motif, suggesting that sequence content within these regions could contribute to the binding specificity of Rhi. However, while the interaction between Kipf and Rhi is critical for the recruitment of Rhi to numerous genomic loci, prominent Rhi-dependent piRNA producing regions, such as cluster *42AB* and *38C*, show little to no dependence on Kipf (Baumgartner et al. 2022). These observations suggest that additional cues are critical in specifying or contributing to Rhi's chromatin binding pattern. Here we uncover an unexpected, Kipf-independent mechanism for Rhi targeting that relies on dual decoding of H3K27me3 and H3K9me3 by the Rhino chromodomain to identify major piRNA source loci.

RESULTS

The histone methyltransferase E(z) is required for TE silencing in *Drosophila* germ cells

Because Kipf-mediated targeting cannot explain Rhi binding with all piRNA loci in the genome, we sought to search for additional candidate factors that could affect transposon silencing in germ cells through a focused reverse-genetic screen in *Drosophila* ovaries. We specifically targeted chromatin proteins as well as histone modifiers and probed the effects of their depletion on the expression of germline-specific TEs by reverse transcription followed by quantitative PCR (qRT-PCR) (**Fig. 1a**). Knockdown of 32 of these genes in germ cells resulted in transposon de-repression and highlighted the potential role of different protein complexes in transposon regulation (e.g., COMPASS complex, NSL complex) (**Fig. S1**).

As expected from previous reports (Rangan et al. 2011), depletion of the H3K9me3 methyltransferase Eggless (Egg) resulted in a strong deregulation of germline transposons. Unexpectedly, however, we also found that the depletion of multiple subunits of the Polycomb repressive complex 2 (PRC2), namely *E(z), su(z)12 and caf1-55*, resulted in the de-repression of germline transposons. Germline-specific depletion of the histone methyltransferase Enhancer of Zeste (E(z)), resulted in upregulated transposon levels (**Fig. S1**). In *Drosophila*, E(z) is responsible for trimethylation of lysine 27 on histone H3 (H3K27me3) and is essential for Polycomb-mediated chromatin silencing (Margueron and Reinberg 2011; DeLuca et al. 2020). In several other organisms, E(z) has been linked to transposon control (Walter et al. 2016; Frapporti et al. 2019; Zhao et al. 2019; Montgomery et al. 2020; Hisanaga et al. 2023), and the H3K27me3 mark was proposed to act as a backup system for transposon silencing in low DNA methylation contexts in flowering plants and in mammals (Mathieu, Probst and Paszkowski 2005; Weinhofer et al. 2010; Deleris et al. 2012; Saksouk et al. 2014; Walter et al. 2016; Rougee et al. 2021). Intriguingly, in ciliates, the Polycomb protein Ezl1 is targeted to transposable elements via interaction with a PIWI protein and this mediates H3K9me3 and H3K27me3 methylation at these loci (Frapporti et al. 2019; Xu et al. 2021; Miro-Pina et al. 2022).

We noted that depletion of E(z) using a strong Gal4 driver (UAS-Dcr2 transgene combined with nos-Gal4, referred to as D2G4) caused rudimentary ovaries. To exclude the possibility that the observed TE de-repression was due to pleiotropic oogenesis defects, we used different Gal4 drivers, namely nos-Gal4 which is active from the germarium through stage 2, then inactive between stages 3 and 6 of oogenesis but reactivated at later stages (Van Doren, Williamson and Lehmann 1998), and TOsk-Gal4 which is expressed immediately after germline cyst formation (ElMaghraby et al. 2022), as well as two independent E(z)-RNAi lines, all of which resulted in nearly normal ovarian morphology (**Fig. S2a**), and Rhi localisation as measured by immunofluorescence (**Fig. S2b**). mRNA-seq from ovaries depleted of *E(z)* using the nos-Gal4 driver revealed a strong upregulation of 13 germline TE families (fold change [FC]>4, padj<0.05), including *gypsy12*, *burdock* and *copia* (**Fig. 1b)**. Transposon de-repression results were validated by RNA-FISH, with *burdock*, *gypsy12* and *copia* RNA levels strongly elevated in germ cells of ovaries depleted of *E(z)* (**Fig. 1c**). Overall, however, the effects on TEs that we observed for *E(z)* knockdown were less severe compared to ovaries depleted of the piRNA factor Rhi, which affected 53 TE families (FC>4, padj<0.05), suggestive of partial contributions of E(z) to Rhi function (**Fig. S2c**; data from (Baumgartner et al. 2022)).

To distinguish between compromised global chromatin silencing and a specific role in the piRNA pathway, we performed RNA-FISH for precursor transcripts derived from the dual-strand piRNA clusters *42AB* (the major, Kipf-independent piRNA producer in the germline) and *80F* (Kipf-dependent), and the unistrand cluster *20A*. Germline-specific knockdown of *E(z)* revealed a strong loss in signal for precursors derived from *42AB*, but not for transcripts originating from *80F* and *20A*, suggesting a potential role of E(z) in the transcription of Kipf-independent dual-strand piRNA clusters (**Fig. 1d, Fig. S2d**). Next, we analysed the effects of germline-specific knockdown of *E(z)* on piRNA precursors more globally using ribo-depleted RNA-seq. To measure transcription across entire piRNA clusters, we divided the genome into 1kb bins and quantified the signal per bin, similar to

previous work (Mohn et al. 2014). Upon germline-specific E(z) depletion, we observed a strong reduction in precursor transcript levels across the dual-strand cluster *42AB*, and moderately reduced levels for *38C*, whereas the unistrand piRNA clusters *20A* or *flam* were essentially unchanged (**Fig. 1e**). Strikingly, precursor levels for Kipf-dependent cluster *80F* were increased upon *E(z)* knockdown. qRT-PCR analysis confirmed the results obtained by RNA-seq (**Fig. S2e**). Of note and as reported previously (Klattenhoff et al. 2009; Le Thomas et al. 2014; Mohn et al. 2014), Rhi-depleted ovaries showed a global reduction of precursor transcripts that originate from dual-strand clusters but not unistrand *20A* and *flam* (**Fig. S2f**). These data suggest a more nuanced role of E(z) in TE silencing in the germline as compared to Rhi.

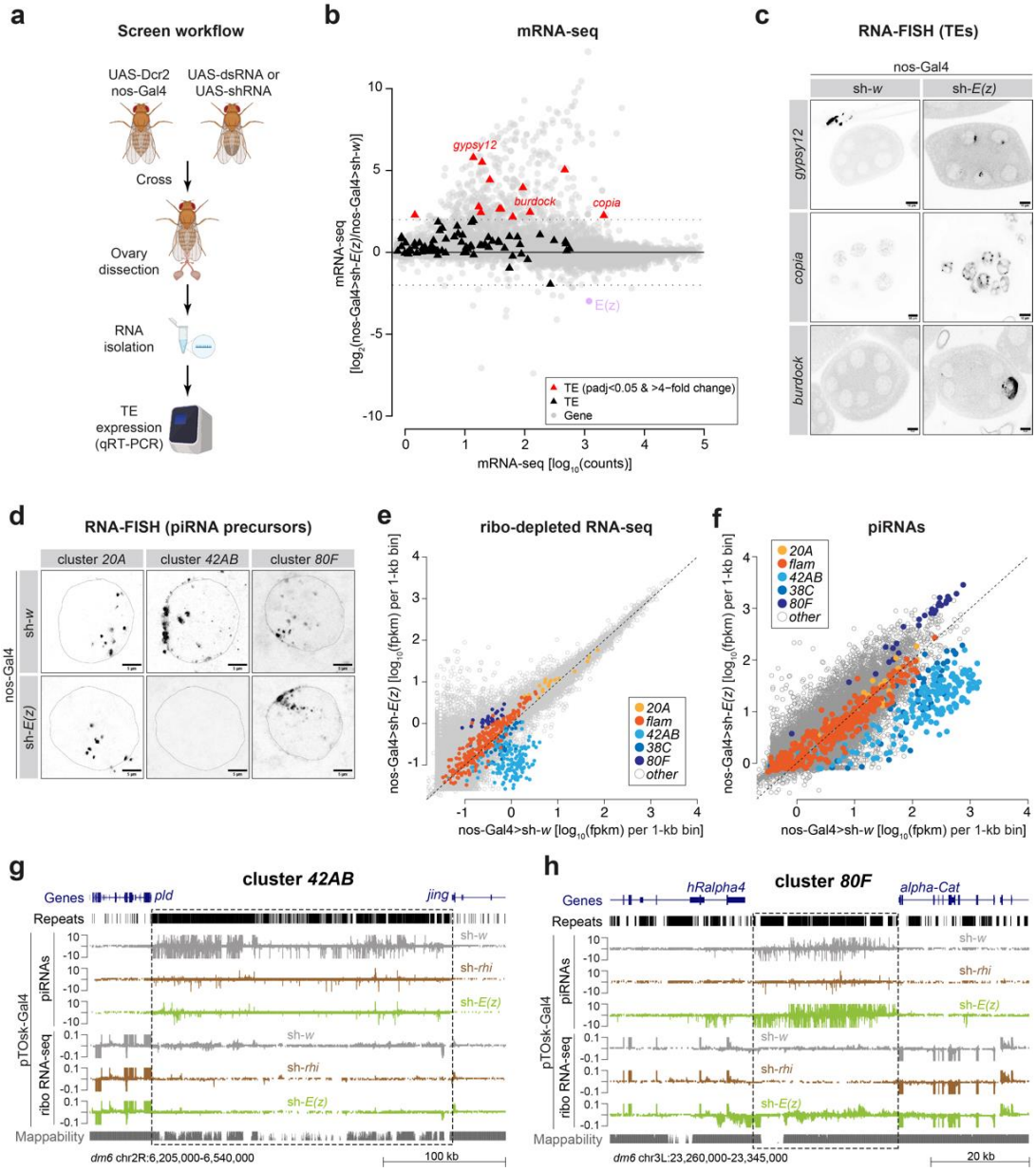

Figure 1: E(z) is required for TE silencing in *Drosophila* germ cells and affects piRNA production

a, Schematic showing the in vivo screen workflow. Male flies expressing dsRNA or shRNA constructs under the inducible UAS promoter were crossed to females expressing the germline-specific nos-GAL4 driver and a *Dcr2* transgene. F1 offspring were transferred to new vials and fertility was determined by counting eggs laid and hatched. RNA was extracted from ovaries of F1 offspring and transposon levels were measured by qPCR. **b,** MA plot showing counts per gene (grey) and transposons (black) in mRNA-seq libraries from ovaries depleted of white (control) and E(z) using the nos-Gal4 driver. TEs with padj<0.05 and >4x fold change (FC) are shown in red, E(z) is indicated in purple. **c,** Confocal images showing RNA-FISH signal for the indicated TEs (*burdock*, *copia*, and *gypsy12*) in the indicated nos-Gal4 mediated knockdowns (scale bar: 10 µm). **d,** Confocal images showing RNA-FISH signal for transcripts derived from piRNA clusters *42AB*, *80F* and *20A* in control and *E(z)* depleted ovaries using the nos-Gal4 driver (scale bar: 5 µm). **e,** Scatter plot depicting normalized ribo-depleted RNA levels (fkpm) of 1-kb bins in ovaries with nos-Gal4 driven *E(z)* knockdown versus control (average of three replicate experiments each). **f,** Scatter plot depicting normalized piRNA levels of uniquely mapping piRNAs in 1-kb bins in ovaries with nos-Gal4 driven *E(z)* knockdown versus control (two replicates each). **g,** UCSC genome browser tracks displaying the piRNA cluster *42AB.* Levels of piRNAs and ribo-depleted RNAs from ovaries with the indicated knockdowns using the pTOsk-Gal4 driver (average of three replicate experiments each) are shown along with tracks displaying genes, repeats and mappability. Dashed line indicates approximate cluster boundaries. **h,** as in g but showing cluster *80F*.

Next, we sequenced piRNA populations from *Drosophila* ovaries in which E(z) or Rhi were depleted specifically in germ cells and compared these to control knockdowns (targeting the *white* gene) (**Table S1**). In agreement with the reduction of precursor transcripts from dual-strand clusters *38C* and *42AB* upon *E(z)* knockdown (**Fig. 1e**), we detected a severe reduction in mature piRNA levels produced from those loci. As expected from their precursor expression, piRNA levels derived from the *80F* cluster were more abundant, while piRNAs from unistrand clusters, such as cluster *20A* and *flam*, were unaffected upon *E(z)* germline knockdown (**Fig. 1f**). The reduction in precursor and piRNA levels observed at cluster *42AB* was similar for ovaries depletion for E(z) and Rhi (**Fig. 1g**). In contrast, transcript levels and piRNA expression were enhanced at the *80F* locus upon *E(z)* knockdown, whereas there was a pronounced reduction of these RNA populations in the *rhi* knockdown (**Fig. 1h**).

Ovaries depleted of E(z) also showed a reduction in antisense piRNAs targeting individual germline-specific transposons, which could explain the observed increase in transposon mRNA levels (**Fig. S2g**). As expected, piRNAs against transposons that are mainly active in the somatic compartment of the ovary (e.g., *Tabor*, *gypsy*, *ZAM*), were not affected upon E(z) depletion, likely due to their origin from the soma-specific *flam* cluster (**Fig. S2g,h**). Altogether, our results reveal that the histone methyltransferase E(z) is required for TE silencing and its depletion affects piRNA production in germ cells from a subset of dual-strand clusters.

H3K27me3 and H3K9me3 marks co-exist on E(z)-dependent piRNA source loci

To elucidate why certain dual-strand piRNA sources are

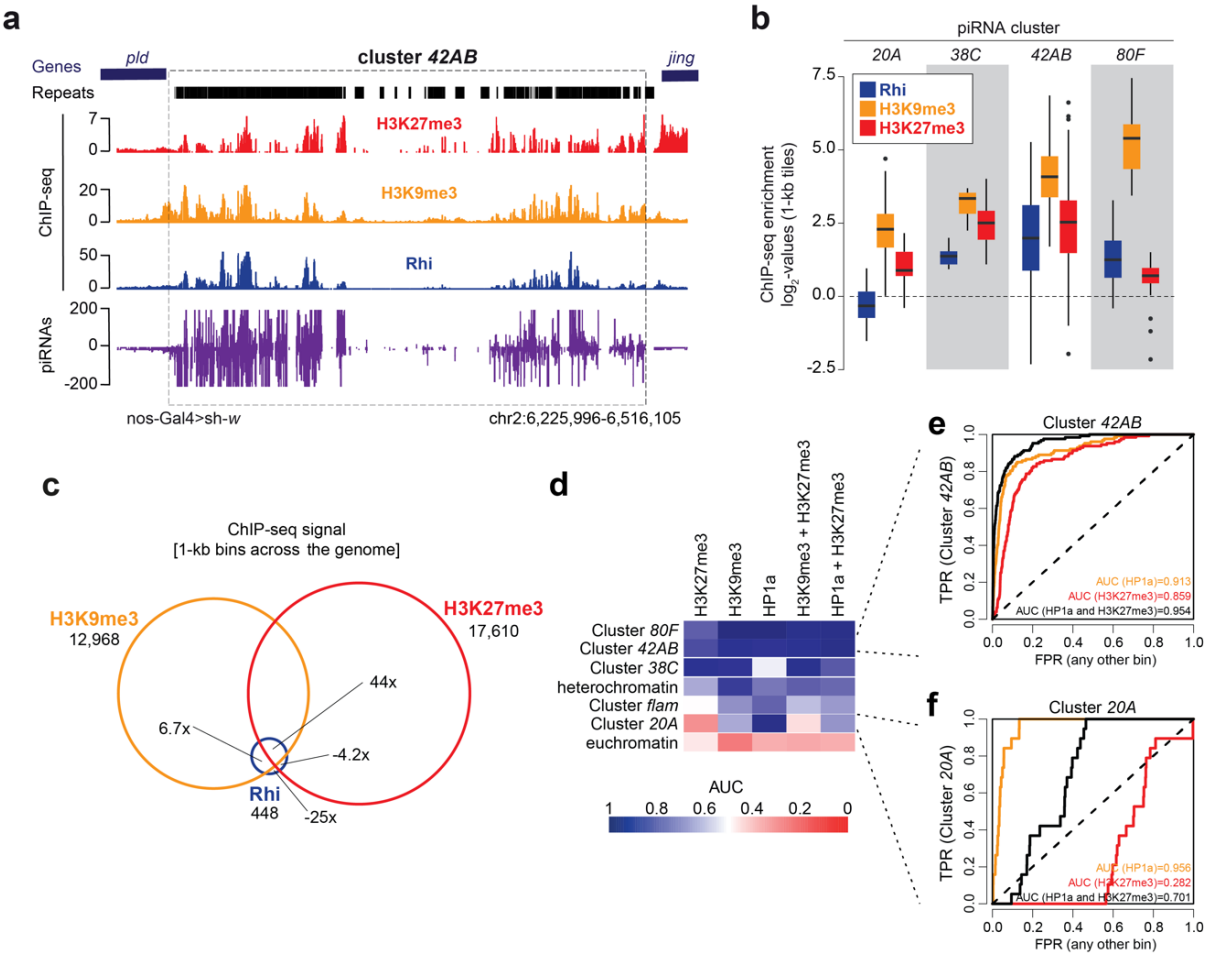

Figure 2: H3K27me3 is present on certain piRNA source loci

a, UCSC genome browser tracks of the region comprising piRNA cluster *42AB* displaying profiles of Rhi (ChIP-seq, blue), H3K27me3 (ChIP-seq, red), H3K9me3 (ChIP-seq, yellow), uniquely mapping piRNAs (purple), and tracks indicating genes and repeats. ChIP-seq signal are depicted as coverage per million reads, piRNA coverage normalized to miRNA reads). **b,** Boxplot showing average log2-fold ChIP-seq enrichment over input of H3K27me3, H3K9me3, and Rhi at the indicated piRNA clusters (n=1-kb bins) in control ovaries. Box plots show median (centre line), with interquartile range (box) and whiskers indicate at most 1.5x interquartile range). **c,** Venn diagram showing overlap between Rhi, H3K9me3, and H3K27me3 ChIP-seq signal across genomic 1kb bins. Signal was considered to be present if the bin overlapped a MACS2 broad peak (q<0.05, broad-cutoff<0.1) across two pooled biological replicates. **d,** Heatmap displaying predictive performance measured as area under ROC curve (AUC) across different chromatin marks (top) and target regions and (left). The regions are ordered according to their AUC for HP1a+H3K27me3. **e,** Line graph comparing the ability to identify dual-strand piRNA cluster *42AB* based on the strength of individual features (H3K27me3 and HP1a) or combinations (mean signal for H3K27me3 and HP1a). **f,** as in e but showing unistrand cluster *20A*.

dependent on E(z) and how this relates to Rhi action at these loci, we examined the genome-wide distribution of H3K27me3, which is catalysed by E(z), and H3K9me3, which has been associated with piRNA clusters and Rhi recruitment, by performing ChIP-seq and CUT&RUN in *Drosophila* ovaries (**Fig. S3a**). These experiments showed that a number of loci, including the dual-strand piRNA cluster *42AB*, in addition to being decorated by H3K9me3, are also decorated with H3K27me3 (**Fig. 2a,b**). A similar,albeit less strong, pattern of H3K9me3 and H3K27me3 was observed at other piRNA source loci, including the *ey*/*Sox102F* region on chromosome 4 (**Fig. S3b**). Notably, we found that Rhi preferentially occupied regions decorated both by H3K9me3 and H3K27me3 over those only carrying H3K9me3 (44-fold enrichment vs 6.7-fold enrichment) (**Fig. 2c**). To determine whether one of the two marks, or both, could best predict piRNA clusters, we divided the genome into 1kb bins and ranked these bins based on their level of histone mark enrichment. To enable comparisons between different marks, we calculated their classification performance as area under the ROC curve (AUC) (**Fig. S3c**). We used published HP1a ChIP-seq data (Zenk et al. 2021) as a proxy for H3K9me3, as we noticed that the H3K9me3 antibody we used appeared to capture some H3K27me3 signal (see **Supplementary Note 1**). Based on analysis of the AUC, we found that H3K27me3 is an accurate predictor of dual-strand piRNA cluster location (AUC 0.867), similar to the predictive power of HP1a binding (AUC 0.842) but combining both marks further improved classification performance as a predictor of dual-strand piRNA clusters (AUC 0.927) (**Fig. S3d**). Of note, we observed that H3K27me3 enrichment varied between piRNA source loci in agreement with E(z) dependence in piRNA production: clusters *42AB* and *38C* showing strong enrichment in H3K27me3 (**Fig. 2d,e**), whereas other clusters such as *20A* and *80F* whose piRNA production was not disrupted by E(z) depletion displayed lower levels of H3K27me3 (**Fig. 2d,f**). These observed correlations between H3K27me3/H3K9me3 co-occupancy and the production of piRNAs points towards a role for both H3K27me3 and H3K9me3 in the specification of piRNA cluster identity.

Rhi binding to a subset of piRNA source loci depends on the histone methyltransferase E(z)

Cluster *42AB* is the major source of Rhi-dependent piRNAs in germ cells (**Fig. 3a**), and our data thus far support piRNA production from this cluster being dependent on E(z). Prior work implicates Kipf in guiding Rhi to a subset of piRNA source loci that includes cluster *80F*, which we find is E(z)-independent (**Fig. 1e,f**), but not *42AB* (Baumgartner et al. 2022). We therefore hypothesized that Rhino binding to piRNA source loci that are independent of Kipf may depend on E(z). To test this hypothesis, we compared piRNA production from Rhi-dependent loci (n=632, ≥2-fold piRNA reduction after Rhi loss) in Kipf or E(z)-depleted ovaries. We observed that E(z), but not Kipf, is required for efficient piRNA production from cluster *42AB* and piRNA-producing loci on chr4, whereas cluster *80F* and most heterochromatic regions are E(z)-independent but require Kipf (**Fig. 3b,c**). Other clusters, such as *38C*, depend weakly on both Kipf and E(z) (**Fig. 3b**). Together, this suggests that E(z) and Kipf may provide two distinct modes of Rhi recruitment to separate subsets of piRNA source loci.

To determine whether E(z) directs Rhi binding to selected loci, we next performed Rhi ChIP-seq in E(z)-depleted ovaries. Looking specifically at piRNA clusters, we found that upon nos-Gal4-mediated E(z) depletion, Rhi enrichment was increased at Kipf-dependent piRNA source loci and reduced at clusters that depend on E(z) (**Fig. 3d**). We next asked whether depletion of E(z) affects the global genomic distribution of Rhi. In agreement with the largely unchanged localization of Rhi by immunofluorescence (**Fig. S2b**), we found that Rhi was not globally redistributed between euchromatin and heterochromatin upon knockdown of *E(z)* (**Fig. 3e, left**). Instead, we detected a strong reduction of Rhi binding at specific regions, including the major piRNA cluster *42AB* and the *ey*/*Sox102F* region on chromosome 4 (**Fig. 3e, Fig. S4a**), whereas other loci, such as the *80F* cluster, accumulated more Rhi than in control ovaries (**Fig. 3e**). Notably, the observed changes in Rhi occupancy following E(z) depletion (**Fig. 3f**) followed the same pattern as piRNA precursor levels, which showed reduced levels for *42AB* and an enrichment at cluster *80F* (**Fig. 1e,f**). Moreover, the observed differences in Rhi binding do not appear to be a consequence of changes in H3K9me3, as H3K9me3 levels remain essentially unaltered genome-wide (**Fig. S4b**).

We next analysed published ChIP-seq data (Baumgartner et al. 2022) for Rhi and Kipf occupancy across the genome. Based on the Rhi ChIP-seq, we identified 1,629 high-confidence Rhi binding sites using MACS2 (q<0.05, >4-fold enrichment). Out of these, 1,166 Rhi peaks were lost in Kipf-KD and were denoted as Kipf-dependent, while 463 peaks were retained and denoted as Kipf-independent (**Fig. S4c**). In agreement with two separate modes of Rhi recruitment, we observed no reduction in H3K27me3 levels or Rhi occupancy at Kipf-dependent Rhi peaks following E(z) depletion, in line with these loci lacking H3K27me3 marks and suggesting that Rhi is recruited by Kipf in an H3K27me3-independent manner. In contrast, we observed a broad loss of both H3K27me and Rhi occupancy at Kipf-independent Rhi peaks upon E(z) depletion (**Fig. 3g**).

Our mRNA-seq data indicated an up to 6-fold reduction in mRNA levels of Moonshiner (Moon), a paralogue of TFIIA required for piRNA production from the majority of Rhi-dependent piRNA source loci (Andersen et al. 2017), in E(z) depleted ovaries (**Table S2**). We also observed reduced Rhi binding at Kipf-independent peaks in published ChIP-seq from Moon-KD (Andersen et al. 2017) (**Fig. S4d**). It is thus a possibility that Rhi loss upon E(z) depletion may in part be driven by the reduced Moon levels. However, we consider secondary effects driven by Moon reduction unlikely since its loss was reported to lead to a complete loss of piRNAs from cluster *80F* and an increase in *38C* derived piRNAs (Andersen et al. 2017), opposite to what is observed in our data. We

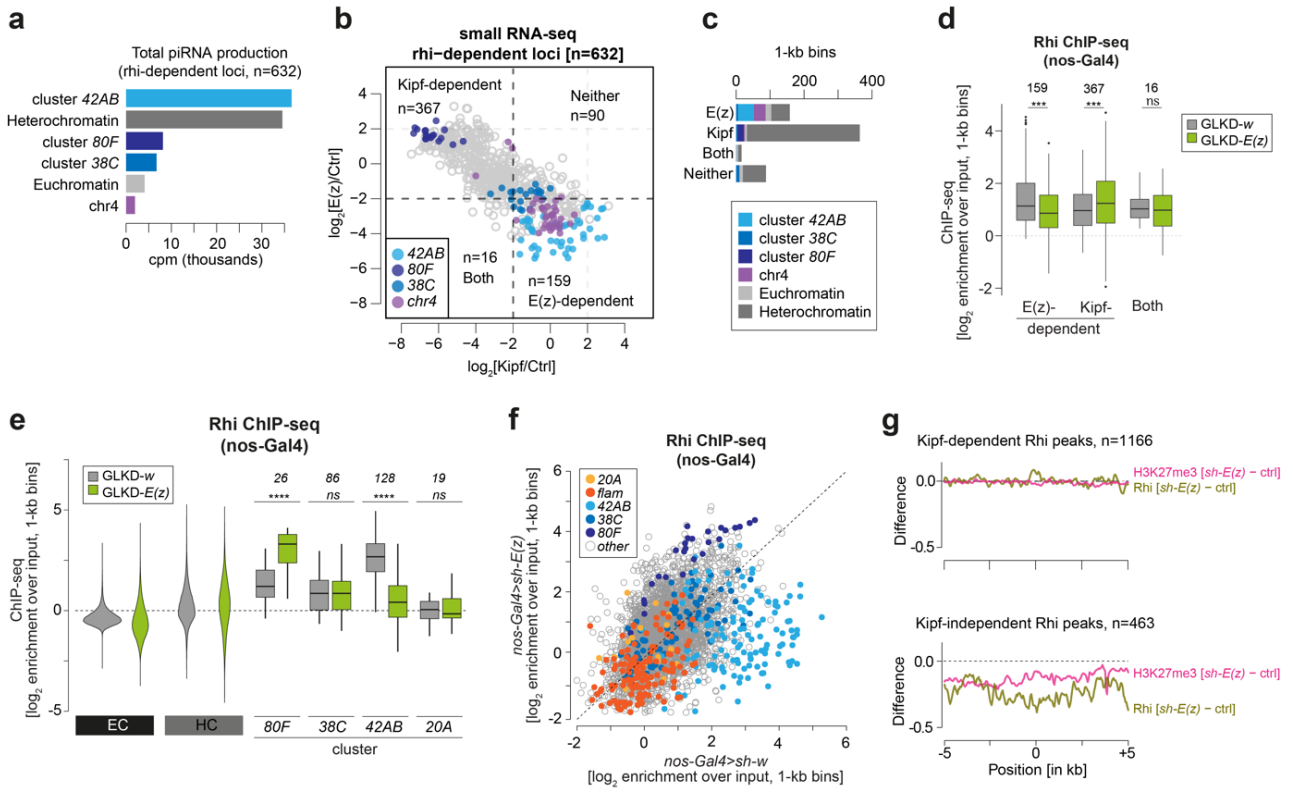

Figure 3: Rhi binding at Kipf-independent loci requires E(z)
a, Bar graph showing genomic origin of Rhi-dependent piRNAs. The analysis was done across 1-kb bins (n=632) that displayed Rhi-dependency (>2-fold mean piRNA loss upon Rhi depletion) across several depletion strategies (pTOsk-Gal4- or MTD-Gal4-driven knockdown, and *rhi* knockout; 1-2 replicates each). **b**, Scatter plots showing log2-fold mean changes in levels of uniquely mapping piRNAs in 1-kb bins relative to control across several depletion strategies (nos-Gal4- or pTOsk-Gal4-driven knockdown of *E(z)*, 2 replicates each; MTD-Gal4-driven knockdown of *kipf* or *kipf* knockout, 1 replicate each). Only Rhi-dependent regions are shown as defined in a. Bins were categorized into E(z)-dependent and/or Kipf-dependent if displaying at least a 4-fold mean loss of piRNAs. **c**, Bar graph showing genomic origin for 1-kb bins per E(z)/Kipf-dependency category as defined in b. **d**, Boxplots showing Rhi ChIP-seq enrichment per E(z)/Kipf-dependency category as defined in b. **** corresponds to P<0.0001 based on Wilcoxon signed-rank test. **e**, Density plots (left) and boxplots (right) showing average log2-fold Rhi enrichment by ChIP-seq from ovaries with nos-Gal4 driven *E(z)* knockdown versus control (average of two replicate experiments each) in heterochromatin (HC) and euchromatic chromosome arms (EC), quantified across 1-kb bins (excluding piRNA clusters). Rhi occupancy at indicated piRNA clusters is shown as boxplot quantification (n depicts 1-kb bins analysed for each cluster). **** corresponds to P<0.0001 based on Wilcoxon signed-rank test. Box plots show median (centre line), with interquartile range (box) and whiskers indicating at most 1.5x interquartile range). **f**, Scatter plot of genomic 1-kb bins contrasting average log2-fold ChIP-seq enrichment of Rhi in ovaries with nos-Gal4 driven *E(z)* knockdown versus control (average of two replicate experiments each). **g**, Metaplot showing mean difference in H3K27me3 (pink) and Rhi (yellowish green) ChIP-seq signal upon nos-Gal4-driven *E(z)* knockdown across Rhi peaks categorised as either Kipf-dependent or not (see Methods).

therefore attribute the reduction in Rhi binding at Kipf-independent sites to loss of H3K27me3 upon *E(z)* knockdown.

Overall, our results provide evidence that H3K9me3 and H3K27me3 co-exist on a subset of piRNA source loci and appear to be a prerequisite for Rhi recruitment at these regions. Our results indicate that two non-redundant mechanisms influence Rhi binding to different piRNA source loci, namely E(z)-dependent and Kipf-dependent recruitment.

Rhi binds a H3K9me3/H3K27me3 combinatorial code in S2 cells

To decisively investigate whether Rhi binding in Kipf-independent regions is driven by H3K9me3/H3K27me3 dual marked chromatin domains, we made use of S2 cells. To assess the role of the combinatorial histone marks in this system, which lacks an endogenous piRNA pathway (Eastwood et al. 2021), we took advantage of well-characterized chromatin reader domains with known binding specificities. Recent work showed that synthetic dual-chromatin reader domains for specific histone marks can be used to detect chromatin modifications, including combinatorial histone codes, through the use of different binders linked in tandem (Villasenor et al. 2020). Rhi, like other HP1a family proteins, contains a chromodomain (CD) that binds H3K9me3-modified histone tails at its N-terminus and a chromoshadow domain at the C-terminal region, connected together via the hinge region (Vermaak, Henikoff and Malik 2005; Le Thomas et al. 2014; Mohn et al. 2014; Yu and Huang 2015). Using CUT&RUN in S2 cells, we characterised the binding patterns of dual-CD

constructs similar in design to Villasenor and colleagues (see methods for details). Specifically, we generated dual-CD constructs to recognize H3K9me3 (CD^{2xCBX1}), H3K27me3 (CD^{2xPc}), or dually H3K9me3/H3K27me3-decorated domains (a bivalent chromodomain fusion $CD^{1xCBX1-1xPc}$) (**Fig. 4a, Fig. S5a**). We confirmed that CD^{2xCBX1} bound to regions decorated in H3K9me3 (**Fig. 4b,c**), the CD^{2xPc} associated with regions enriched in H3K27me3 (**Fig. 4b,d**), and the bivalent $CD^{1xCBX1-1xPc}$ reader exhibited preferential enrichment at loci marked by both H3K9me3 and H3K27me3 modifications (**Fig. 4b,e**).

Next, we tested whether the association of a Rhi chromodomain homodimer (CD^{2xRhi}) recapitulated the binding patterns observed for the bivalent H3K9me3/H3K27me3 reader $CD^{1xCBX1-1xPc}$ (**Fig. 4e**). Strikingly, CD^{2xRhi} binding was essentially indistinguishable from that of the bivalent $CD^{1xCBX1-1xPc}$ genome-wide (**Fig. 4f, Fig. S5b**) and within heterochromatic regions (**Fig. 4g, Fig. S5c**) where $CD^{1xCBX1-1xPc}$ binding strongly correlates with H3K27me3 levels. In contrast, Rhi chromodomain homodimers containing a F48A point mutation in the aromatic cage ($CD^{2xRhi[F48A]}$) essentially recapitulated the IgG control (**Fig. 4b,f-g**). Although CD^{2xRhi} binding was strong already at loci with only H3K9me3 (Odds ratio [OR] 5.9, Fisher's exact test), the strongest CD^{2xRhi} binding was observed at loci marked with both H3K9me3 and H3K27me3 (OR 19), whereas binding was depleted at loci only marked with H3K27me3 (OR -4.1) (**Fig. 4h**). Of note, regions marked by both H3K9me3 and H3K27me3 showed no presence of the PRC1 deposited mark H2AK118ub generally associated with the presence of H3K27me3 (Margueron and Reinberg 2011) (**Fig. 4b,d**). These results indicate that Rhi recruitment depends on H3K27me3 but is independent of H2AK118ub. Taken together, these results show that, similar to $CD^{1xCBX1-1xPc}$, CD^{2xRhi} binds to regions decorated in both H3K9me3 and H3K27me3.

Unlike the chromodomain of HP1a, the Rhi CD has been reported to form a homodimer (Yu et al. 2015). To investigate whether Rhi chromodomain dimers have the capacity to bind a histone H3 tail carrying dual-modified K9me3/K27me3, we modelled the binding of the Rhi CD dimer to a single 40 amino acid histone H3 tail peptide carrying simultaneous tri-methylation of lysine 9 and lysine 27 using MD simulations (**Fig. S5d**). In support of our binding pattern results, protein modelling suggested an ability of Rhi chromodomain dimers to bind a H3K9me3/H3K27me3 dual-modified histone tail.

Kipf-independent Rhi binding depends on E(z)

To directly test the importance of H3K27me3 in specifying Rhi chromodomain binding, we performed CUT&RUN in S2 cells for CD^{2xCBX1}, CD^{2xPc}, $CD^{1xCBX1-1xPc}$, and CD^{2xRhi} upon control [*siRen*] or E(z) knockdown [*siE(z)*]. Principle component analysis (PCA) showed high levels of similarities between genome-wide CD^{2xPc} binding and H3K27me3 distribution and between genome-wide H3K9me3 distribution and CD^{2xCBX1} binding. PCA also confirmed in control S2 cells, as we observed in earlier experiments (**Fig. 4f,g**), that the binding pattern of CD^{2xRhi} more closely resembles that of the $CD^{1xCBX1-1xPc}$ binding pattern over that of CD^{2xCBX1} or CD^{2xPc} (**Fig. 4i**). In the *E(z)* knockdown cells, binding of CD^{2xPc} was strongly affected while CD^{2xCBX1} and H3K9me3 distribution remained mostly unchanged (**Fig. 4i**).

Next, to further determine the impact of H3K27me3 on Rhi binding, we divided the genome onto 1kb bins. Bins were categorised into six different groups based on the degree of H3K27me3 loss upon depletion of E(z). We then plotted for each group the fold-change in binding levels of the different chromodomains as assessed by CUT&RUN in *siE(z)* compared to *siRen*. While loss of H3K27me3 strongly reduced CD^{2xPc} binding, no major genome-wide effects were seen for CD^{2xCBX1}, $CD^{1xCBX1-1xPc}$ or CD^{2xRhi} (**Fig. S5e**, upper panels). This reflects the ability of CD^{2xPc} to bind all H3K27me3-decorated regions genome-wide, whereas only a subset of H3K27me3-decorated regions, those also marked with H3K9me3 can be bound by $CD^{1xCBX-1xPc}$ and CD^{2xRhi}. To identify H3K9me3 and H3K27me3 dual marked regions affected by the loss of E(z), we focused on CD^{2xCBX1}-bound regions (as a proxy for H3K9me3) and observed that depletion of E(z) resulted in reduced binding of both CD^{2xRhi} and $CD^{1xCBX1-1xPc}$ in these H3K9me3-enriched bins (**Fig. S5e**, lower panels).

We next identified 449 $CD^{1xCBX1-1xPc}$ binding sites where its binding was significantly reduced (DiffBind, padj<0.05) after H3K27me3 loss upon E(z) depletion. Notably, these regions displayed a similar loss of CD^{2xRhi} binding following H3K27me3 depletion, despite retaining similar H3K9me3 levels (**Fig. 4j, Fig. S5f**). These results indicate that binding of $CD^{1xCBX1-1xPc}$ and CD^{2xRhi} at these H3K9me3/H3K27me3-decorated loci require the presence of both H3K27me3 and H3K9me3 marks, and that H3K9me3 alone is not sufficient for binding. Taken together, these findings indicate that a combinatorial H3K9me3/H3K27me3 histone code, recognized by the Rhi chromodomain, underlies Rhi binding specificity at a subset of piRNA clusters.

The combinatorial histone code H3K9me3/H3K27me3 at major piRNA source loci is conserved in Drosophilids

It has previously been shown by Rhi ChIP-seq that the *Drosophila simulans* Rhi CD has similar binding specificity as those of *Drosophila melanogaster* when expressed in *D. melanogaster* ovaries (Parhad et al. 2017). Based on this, we hypothesized that Rhi association with H3K9me3/H3K27me3 dual domains could be conserved across species. This is further supported by the high sequence conservation of the Rhino CD across *Drosophila* species (Vermaak, Henikoff and Malik 2005). In particular, aromatic residues in the CD, important for binding to methylated lysine residues, as well as amino acids reported to mediate Rhi CD dimerization, show high conservation, suggesting that Rhi chromodomains in Drosophilids all have the potential to bind methylated histone tails and to homodimerize (**Fig. S6a**).

To test this, we assessed by CUT&RUN the presence of H3K9me3 and H3K27me3 at piRNA producing loci in ovaries from *Drosophila simulans* (MRCA 5.6 MYA), *Drosophila erecta* (MRCA 8.2 MYA), *Drosophila yakuba*

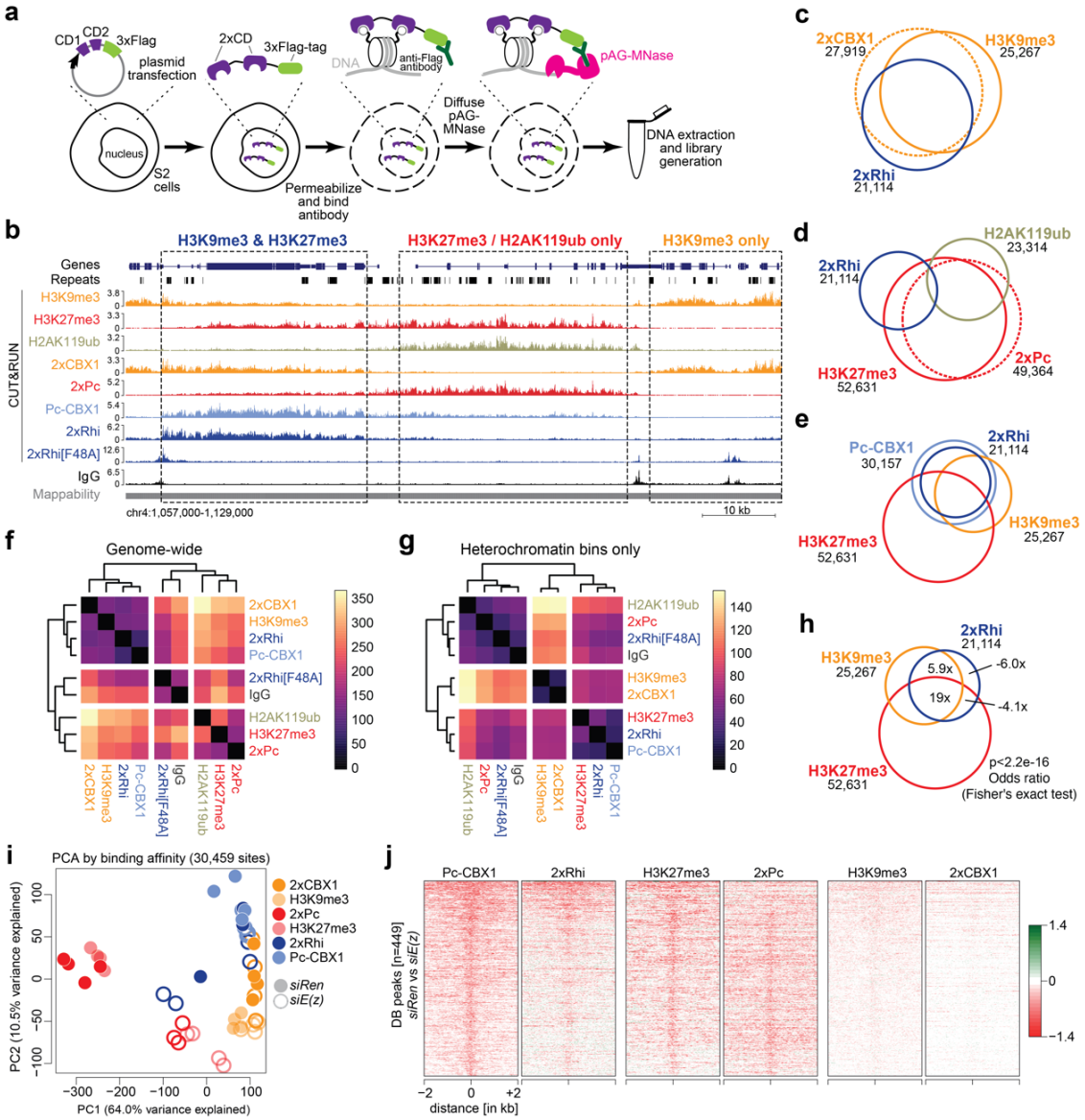

Figure 4: An *in vitro* system that recapitulates Kipf-independent Rhi binding

a, Schematic showing the experimental workflow for the chromodomain binding assay. **b**, UCSC genome browser tracks of a 72 kb region on chromosome 4 displaying the binding profiles of the indicated histone modifications and chromodomain constructs measured by CUT&RUN (counts per million [cpm] across pooled replicates). Dashed boxes indicate representative areas with either both H3K9me3 and H3K27me3, only H3K27me3, or only H3K9me3, respectively. The mappability of 100 nt reads is indicated below the tracks. **c,** Venn diagram showing overlap between CD^{2xCBX1} binding, H3K9me3, and CD^{2xRhi} binding across 125,499 genomic 1kb bins. Signal was considered to be present if the bin overlapped a MACS2 broad peak (q<0.05, broad-cutoff<0.1) present in at least two biological replicates. **d,** Same as c, but showing CD^{2xRhi} binding, H2AK118ub, H3K27me3, and CD^{2xPc} binding. **e,** Same as c, but showing $CD^{1xCBX1-1xPc}$ and CD^{2xRhi} binding, H3K27me3, and H3K9me3. **f**, Heatmap and hierarchical clustering (Euclidian distance) of CUT&RUN signal detected for the indicated chromodomain constructs and histone modifications [log10(fpkm)]. Signal shown as mean across biological replicates. Individual replicates shown in **Fig. S5a**. **g**, Same as f, but across a subset of the genome classified as heterochromatin (10,180 out of 125,499 1kb bins). Individual replicates shown in **Fig. S5b**. **h,** Same as c, but showing H3K9me3, CD^{2xRhi} binding, and H3K27me3. Odds ratios for CD^{2xRhi} binding to H3K9me3, H3K27me3, both or neither were calculated using Fisher's exact test. **i,** Principal component analysis (PCA) plot of the binding affinities of the indicated chromodomain fusion proteins expressed in S2 cells in control condition (solid circle) and E(z) knockdown (circle outline). **j,** Heatmaps showing the log2-fold change of *E(z)* knockdown compared to control knockdown for indicated constructs and histone modifications across Pc-CBX1 differentially bound regions.

(MRCA 11.4 MYA), and *Drosophila ananassae* (MRCA 33.9 MYA) (**Fig. 5a**). We compared our CUT&RUN data for H3K9me3 and H3K27me3 marks with small RNA-seq of ovaries from these species (Lopik et al. 2023). We divided the genomes into 10kb bins that were then classified onto 15 groups according to piRNA levels present in each species. We observed that at least for *Drosophila melanogaster* and *Drosophila yakuba*, bins with high levels of piRNAs also showed strong signals of H3K9me3 and H3K27me3 marks, whereas the absence or low levels of

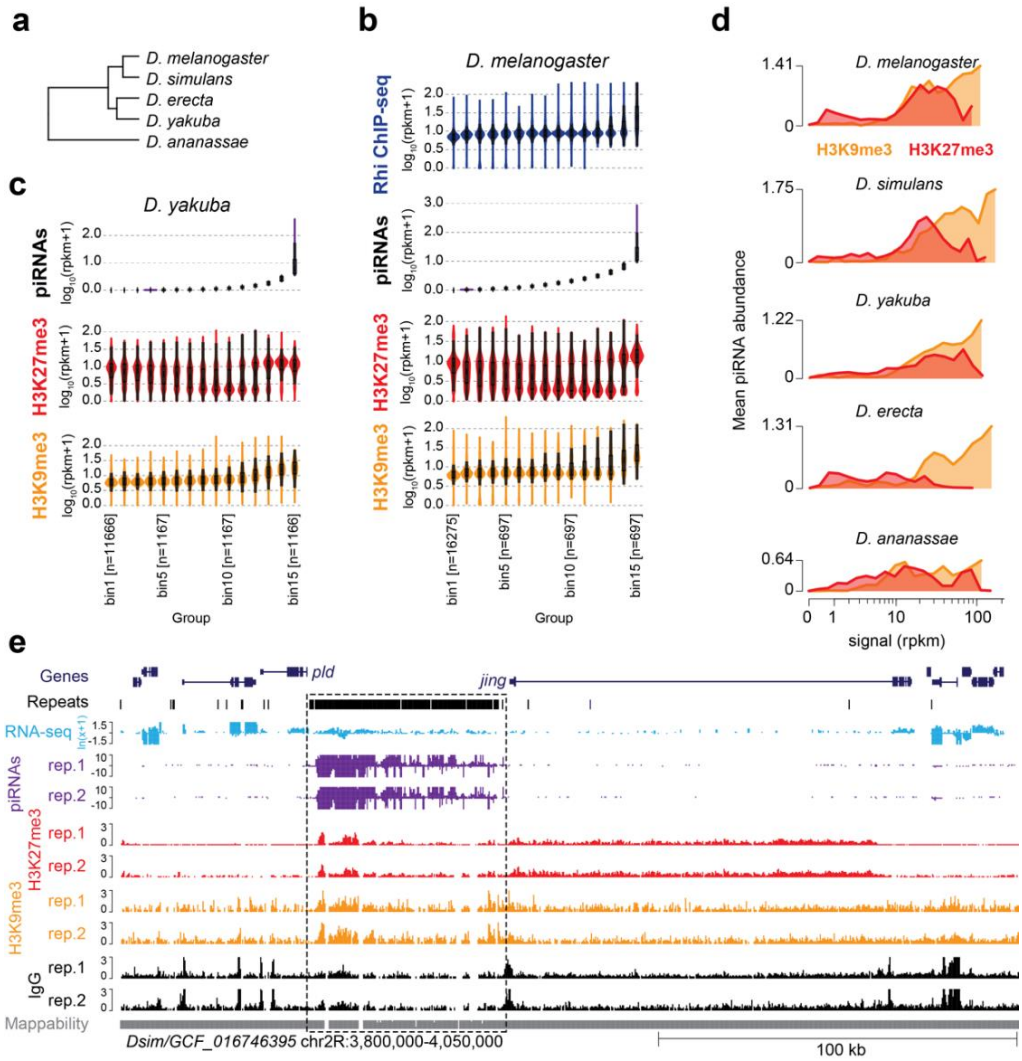

Figure 5: Conservation of Rhi association with H3K9me3 and H3K27me3 in Drosophilids
a, Phylogenetic tree of the five studied *Drosophila* species. **b**, Violin plots illustrating the distribution of piRNAs, H3K9me3 and H3K27me3 CUT&RUN signal, and Rhi ChIP-seq levels [$\log_{10}$(rpkm+1)] in *D. melanogaster* ovaries. 10-kb bins across the genome are distributed into 15 groups. Group 1 contains all 10-kb bins with no piRNA expression. The 14 remaining groups contain an equal number of 10-kb bins. Groups are ranked according to piRNA expression levels with the lowest level of the two strands shown to focus on dual-strand regions. **c**, As in b, but showing piRNA, H3K9me3 and H3K27me3 levels [$\log_{10}$(rpkm+1)] for *Drosophila yakuba* ovaries. **d**, Line graphs showing mean piRNA abundance (n=3 replicates; signal shown as the lowest of the two strands, log10(rpkm+1)) plotted against H3K9me3 signal (orange; reads per kilo base per million mapped reads, rpkm) and H3K27me3 signal (red) for all 10kb bins of the indicated *Drosophila* species (*D. melanogaster*, *D. simulans*, *D. yakuba*, *D. erecta*, and *D. ananassae*). **e**, UCSC genome browser tracks from *D. simulans* displaying ribo-depleted RNA-seq, piRNAs, and CUT&RUN for H3K27me3 and H3K9me3 (counts per million [cpm]) at the *42AB*-syntenic region (dashed box). The mappability of 100 nt reads is indicated below the tracks.

piRNAs correlated mostly with the presence of H3K27me3 and absence of H3K9me3 (**Fig. 5b,c, Fig. S6b**). Genome-wide analyses showed that piRNA producing loci are preferentially decorated with both H3K9me3 and H3K27me3 in *Drosophila simulans*, *Drosophila yakuba*, *Drosophila erecta* and *Drosophila ananassae* (**Fig. 5d**). We also detected both histone marks at the region of the *Drosophila simulans* genome that is syntenic to the *Drosophila melanogaster 42AB* cluster (**Fig. 5e**). These results suggest that the combination of H3K9me3 and H3K27me3 in regions where Rhi binds is a conserved feature in Drosophilids.

In conclusion, our study identifies the H3K27 methyltransferase E(z) as an important regulator of transposon expression in *Drosophila* germ cells. We show that, in addition to the well-described H3K9me3 mark, H3K27m3 is important for guiding Rhi binding in a sequence- and Kipf-independent manner, thus contributing to the definition of a dual-strand piRNA source loci.

DISCUSSION

Chromatin structure and gene expression are strongly influenced by complex patterns of histone modifications, such as methylation, acetylation, phosphorylation, and ubiquitination. It is becoming increasingly clear that these modifications occur simultaneously, creating a diverse range of potential combinations, also referred to as the

combinatorial histone code, and these patterns can be interpreted by various reader/effector proteins.

Here we provide evidence that the co-occurrence of the histone modifications H3K9me3 and H3K27me3, usually thought to occupy distinct domains, plays an important role in marking genomic loci for specific binding of the heterochromatin protein Rhi, a key step to promote piRNA cluster expression. By performing a screen for novel factors important for transposon control, we identified the H3K27me3 methyltransferase E(z) and demonstrated that several Rhi-bound loci that produce vast amounts of piRNAs in germ cells are decorated with both H3K9me3 and H3K27me3 marks. We demonstrate that at these loci Rhi recruitment is independent of the zinc finger factor Kipf and instead requires H3K27me3, deposited by E(z), and H3K9me3, likely through binding of Rhi's chromodomain to both histone modifications. These data demonstrate the existence of two independent modes of Rhi recruitment at distinct subsets of piRNA clusters: that reliant on Kipf-recognized DNA sequence and that dependent on chromatin context as established by E(z). Our results also point to the involvement of additional complexes in transposon regulation, such as the COMPASS and NSL complex. Of note, a recent study has shown that the depletion of germline-specific subunits of the NSL complex (NSL1, NSL2, and NSL3) resulted in reduced piRNA production from telomeric piRNA clusters, thus emphasizing the role of the NSL complex in piRNA precursor transcription within these clusters (Iyer et al. 2023).

This work is the first report and functional exploration of the co-occupancy of H3K9me3 and H3K27me3 modifications at piRNA source loci in *Drosophila*. While H3K9me3 and H3K27me3 histone marks have previously been suggested to be mutually exclusive, our work and that of others demonstrate that they co-exist in a number of contexts, although until now without a clear function (Liu et al. 2011; Ho et al. 2014; Mozzetta et al. 2014; Evans et al. 2016; Walter et al. 2016; Ichihara et al. 2022). Indeed, the co-occurrence of H3K9me3 and H3K27me3 at transposon sequences has been reported in other species. In the ciliate *Paramecium tetraurelia* and in mice the deposition of both marks play a crucial role in maintaining control over a wide range of transposons, thereby ensuring the long-term stability of the genome (Walter et al. 2016; Frapporti et al. 2019). In the ancestral eukaryotic lineage, such as the marine diatom *Phaeodactylum*, H3K27me3 is also enriched at TE sequences with or without cooccurring H3K9me3 (Veluchamy et al. 2015; Deleris, Berger and Duharcourt 2021). Interestingly, *C. elegans* piRNA clusters that give rise to 21U-RNAs are found within H3K27me3-rich genomic regions (Beltran et al. 2019). In addition, recruitment of the *C. elegans* Upstream Sequence Transcription Complex, required for piRNA precursor transcription, is guided by the chromodomain-containing protein UAD-2 which can bind H3K27me3 (Huang et al. 2021). In the absence of H3K27me3, UAD-2 fails to be recruited to piRNA source loci, resembling the effect observed for Rhi upon germline depletion of E(z).

Although E(z) catalysed H3K27me3 is required for Rhi recruitment to a subset of piRNA clusters, the mechanisms by which E(z) is recruited and when during development H3K27me3 is deposited at specific *Drosophila* dual-strand clusters remains to be determined. In *Paramecium*, the Polycomb protein Ezl1 plays an effector function and is targeted directly to transposon insertions via interaction with a PIWI protein and subsequently mediates H3K9me3 and H3K27me3 deposition at these loci (Miro-Pina et al. 2022). *Drosophila* E(z) has been shown to co-immunoprecipitate with Piwi in ovaries (Peng et al. 2016), which hints at a possible role of PIWI proteins in guiding E(z) to deposit H3K27me3 at piRNA clusters. Moreover, the presence of H3K27me3 at these loci in multiple cell types, including germ cells and cells derived from the somatic lineage (e.g., S2 cells that are derived from *Drosophila* embryos), suggests that H3K27me3 is established early during development or possibly deposited maternally (Zenk et al. 2017), and retained somatically as development progresses. A previous study demonstrated divergence in the spatio-temporal expression patterns of Kipf and Rhino, with Kipf exhibiting very low levels in ovarian germline stem cells and cystoblasts (Baumgartner et al. 2022). In light of this observation, we suggest that H3K27me3 may collaborate with H3K9me3 during the early stages of oogenesis, where Rhino functions independently of Kipf. Understanding how and when histone methyltransferases are guided to piRNA source loci and their other targets at which they deposit histone marks is a critical further question to better understand the piRNA system and its interplay with chromatin biology.

While bivalent chromatin, characterised by the presence of both H3K4me3 and H3K27me3, or dual domains have been previously suggested to play important roles in regulating gene expression, the diversity of chromatin reader domains able to interpret either each histone PTM or exhibiting dual recognition of both PTMs in vitro brings additional complexity to the understanding of the chromatin biology (Azuara et al. 2006; Mikkelsen et al. 2007; Du et al. 2012; Gaertner et al. 2012; Musselman et al. 2012; Kumar et al. 2021; Mashtalir et al. 2021; Barral et al. 2022). Proteins with a paired domain, such as Trim24, which contains a PHD and a Bromo domain targeting unmodified H3K4 and acetylated H3K23, respectively, has been shown to bind a single histone tail bearing these two marks in vitro (Tsai et al. 2010). However recent work indicates that H3K23ac is absent from Trim24-binding sites in mESCs and does not appear to play a role in Trim24 localization in vivo (Isbel et al. 2023).

Here, our data suggest a direct role for a dual histone code in determining Rhi binding, likely in a dimerization-dependent manner. We showed that Rhi chromodomain dimers have a similar binding pattern as bivalent chromodomain fusion proteins expected to bind H3K9me3/H3K27me3 decorated domains, strongly suggesting that Rhi chromodomain dimers bind simultaneously to both H3K9me3/H3K27me3. This hypothesis is further supported by protein modelling that found Rhi chromodomain dimers sterically able to bind to H3K9me3/H3K27me3 dual-modified histone peptides. However, considering that previous work found that the Rhi

CD does not bind H3K27me3 in vitro (Yu 2015; Mohn 2014), we cannot rule out the possibility of a yet to be discovered co-factor that provides H3K27me3 specificity. In summary, our study provides an unexpected example highlighting *in vivo* the importance of a combinatorial histone code in the binding specificity of a chromatin binding protein.

MATERIALS AND METHODS

Data generation

Fly husbandry and stocks

All flies were kept at 18 °C or 25 °C on standard cornmeal or propionic food. Flies were obtained from the Vienna *Drosophila* Resource Center (VDRC), or from the Bloomington *Drosophila* Stock Center (BDSC). All used fly stocks are listed in **Table S3**. For germline-specific knockdowns, we used a fly line containing a UAS-Dcr2 transgene and a nos-GAL4 driver (Czech et al. 2013), a pTOsk-GAL4 driver (ElMaghraby et al. 2022), or a nos-GAL4 driver (Czech et al. 2013), each crossed to stocks expressing shRNAs or dsRNAs under the UAS promoter. After mating at 27 °C for five days, parental flies were removed from the vials. Hatching F1 offspring were collected and aged with yeast for 2-3 days prior to use for follow-up experiments.

RNA isolation and qRT-PCR

Samples were lysed in 1 ml TRIzol reagent (Thermo Fisher Scientific) and RNA was extracted according to manufacturer's instructions. One microgram of total RNA was treated with DNaseI (Thermo Fisher Scientific), and reverse transcribed with the Superscript III First Strand Synthesis Kit (Thermo Fisher Scientific), using oligo(dT)20 primers. Primer sequences are listed in **Table S4.**

TE screen

1-2 different fly lines per gene (1 replicate per line) expressing shRNAs or dsRNAs under the UAS promoter were crossed with a fly line containing a UAS-Dcr2 transgene and a nos-GAL4 driver (Czech et al. 2013) as previously described. RNA was isolated using 5-10 ovaries per cross and reverse transcribed as described above. Multiplexed qPCRs were carried out using TaqMan Universal Master Mix II, no UNG (Applied Biosystems) as previously described (Czech et al. 2013). Experiments were performed on a CFX96 Real-Time System C1000 Touch Thermal Cycler (BioRad). Z-scores for transposon expression were calculated on ΔCT values [CT(transposon) – CT(*rp49* control)] (Livak and Schmittgen 2001). The primers and probes used are listed in **Table S4**. Positive hits were confirmed by qPCRs using SYBR green Master mix (ThernoFisher Scientific). QPCRs were performed on a QuantStudio Real-Time PCR Light Cycler (Thermo Fisher Scientific).

RNA FISH

Single-molecule RNA fluorescence in situ hybridization (smRNA-FISH) for transcripts derived from piRNA clusters *42AB*, *80F* and *20A*, as well as the transposons *Burdock*, *Gypsy12*, and *copia* was performed using Stellaris probes (Biosearch Technologies). Probe sequences are listed in **Table S4**.

Ovaries from 3 to 6-day-old flies were dissected in Schneider's *Drosophila* Medium and fixed in Fixing Buffer (4% formaldehyde, 0.3% Triton X-100, 1x PBS) for 20 min at room temperature, rinsed three times in 0.3% Triton X-100, once in PBS, and permeabilized in 70% ethanol at 4°C overnight. Permeabilized ovaries were rehydrated in smRNA FISH wash buffer (10% formamide in 2x SSC) for 10 min. Ovaries were resuspended in 50µL hybridization buffer (10% dextran sulphate, 10% formamide in 2xSSC) supplemented with 1.5 µL of smRNA FISH probes. Hybridization was performed with rotation at 37°C overnight. Ovaries were then washed twice with smRNA FISH wash buffer at 37°C for 30 min and twice with 2xSSC solution. Then, DNA was stained with DAPI (1/500 dilution in 2x SSC) at room temperature for 20 min. Ovaries were mounted in 30 µL Vectashield mounting medium and imaged on a Zeiss Zeiss LSM-800 confocal microscope. The resulting images were processed using FIJI/ImageJ.

Immunofluorescence

Fly ovaries were dissected in ice-cold PBS, fixed for 14 min in 4% PFA (Alfa Aesar) at RT and permeabilized with 3x10 min washes in PBS with 0.3% Triton (PBS-Tr). Samples were blocked in PBS-Tr with 1% BSA for 2 hrs at RT and incubated overnight at 4 °C with primary antibodies in PBS-Tr and 1% BSA. After 3x10 min washes at RT in PBS-Tr, secondary antibodies were incubated overnight at 4 °C in PBS-Tr and 1% BSA. After 4x10min washes in PBS-Tr at RT with 4′,6-diamidino-2-phenylindole (DAPI; Invitrogen) added during the third wash, and 2x5 min washes in PBS, samples were mounted with ProLong Diamond Antifade Mountant (Thermo Fisher Scientific) and imaged on a Leica SP8 or Zeiss LSM-800 confocal microscope. Images were deconvoluted using Huygens Professional or using FIJI/ImageJ. All antibodies used are given in **Table S5**.

Molecular cloning and constructs

All constructs were cloned using the NEBuilder HiFi DNA Assembly kit (New England Biolabs E2621) according to manufacturer's instructions. Chromodomain (CD) sequences were amplified from cDNA prepared from ovaries or ordered as gBlock DNA fragments from Integrated DNA Technologies (IDT). The final constructs expressed the CDs of interest tagged amino-terminally with an NLS-3xFlag-EGFP cassette under the control of the *D. simulans ubiquitin* promoter. Tandem chromodomains were linked using the following linker GGGGSGGGGSGGGGSGGGGSTR (A list of all constructs used in this study are provided in **Table S6**).

mRNA-seq (polyA selected)

Total RNA was extracted from 30 ovaries from 3-6 day-old flies using TRIzol (Thermo Fisher Scientific) in three replicates. One microgram of total RNA was subjected to polyA selection and subsequent fragmentation, reverse transcription, and library preparation according to the manufacturer's instructions using the Illumina stranded mRNA Prep for sequencing. Sequencing was performed by Novogene on an Illumina Novaseq 6000 instrument.

ribo-depleted RNA-seq

Total RNA was extracted from 10 to 20 ovaries from 3-6 day-old flies using TRizol (Thermo Fisher Scientific) following the manufacturer's instructions.

For all the experiments conducted with the pTOsk-Gal4 driver, ribosomal RNA was depleted using RiboPOOL (siTOOLs, Biotech) following the manufacturer's protocol. RNA-seq libraries were produced using NEBNext Ultra Directional Library Prep Kit for Illumina, following the manufacturer's instructions for rRNA depleted RNA. Library size distribution was analysed on a TapeStation instrument (Agilent Technologies) using a High Sensitivity D1000 ScreenTape. Libraries were pooled in equal molar ratio and quantified with KAPA Library Quantification Kit for Illumina (Kapa Biosystems) and sequenced paired-end 50 on an Illumina NovaSeq 6000 instrument.

For the experiments conducted using the nos-Gal4d driver, rRNA depletion was performed from 1µg total RNA using the RNA Depletion stranded Library Prep kit (BGI). The samples

were then sequenced as PE100 reads on the DNBSEQ G400 sequencer, and adapter-clipped reads were provided by BGI.

Small RNA-seq

Regulatory small RNA extraction was performed as described in (Grentzinger et al. 2020). Argonaute-small RNA complexes were isolated from 10 to 30 pairs of ovaries from 3-5 day-old flies using TraPR ion exchange spin columns (Lexogen, Catalog Nr.128.08). Libraries generations were performed according to the manufacturer's instructions.

For all experiments conducted with the pTOsk-Gal4 driver and ovaries of the five *Drosophila* species, sequencing was performed at the CRUK CI Genomics core on an Illumina NovaSeq 6000 instrument. Sequencing for the experiments conducted using the nos-Gal4d driver was performed by Fasteris SA (Geneva, Switzerland) on an Illumina NextSeq550 instrument.

ChIP-seq

Chromatin immunoprecipitation (ChIP) was performed as previously described in (Lee, Johnstone and Young 2006) with minor modifications.

Briefly, 100 ovary pairs were manually dissected into Schneider media and cross-linked in 1% formaldehyde/PBS for 10 min at room temperature with agitation. The cross-linking reaction was quenched by STOP buffer (PBS 1X, Triton 0.1%, Glycine 1 M) and ovaries were washed in PBS and homogenized in a glass douncer: first slightly dounced in PBST 0.1% and centrifugated 1 min 400g, followed by strong douncing in cell lysis buffer (KCL 85 mM, HEPES 5 mM, NP-40 0.5% , Sodium butyrate 10 mM, EDTA free protease inhibitor cocktail Sigma) following by 5 min centrifugation at 2000g. We performed 2 washes with cell lysis buffer. The homogenates were then lysed on ice for 30 min in nucleus lysis buffer (HEPES 50 mM, EDTA 10 mM, N lauryl sarkosyl 0.5%, sodium butyrate 10 mM, EDTA free protease inhibitor cocktail Sigma). DNA was sheared using a Bioruptor pico from Diagenode for 10 cycles (30 sec on, 30 sec off). The sonicated lysates were cleared by centrifugation and then incubated overnight at 4°C with 5µl of specific antibodies (**Table S5**). Then 40 µL of Protein A Dynabeads was then added and allowed to bind antibody complexes by incubation for 1h at 4°C. Following four washing steps with high salt buffer (Tris pH 7.5 50 mM, NaCl 500 mM, Triton 0.25%, NP-40 0.5%, BSA 0.5%, EDTA pH 7.5 5 mM), DNA-protein complexes were eluted and de-cross-linked 10 h at 65°C. RNA and protein was digested by RNase A and Proteinase K treatments, respectively, before purification using Phenol:Chloroform:Isoamyl Alcohol 25:24:1 (Sigma) according to the manufacturer's instructions.

Barcoded libraries were prepared using Illumina technology, and subsequently sequenced on a NextSeq High (Illumina) by Novogene (Rhi and H3K9me3 ChIP-seq), or by the Jean Perrin facility (H3K27me3 ChIP-seq).

CUT&RUN

1 million cells per sample were harvested and washed three times with Wash buffer (20 mM HEPES, pH7.5, 150 mM NaCl, 0.5 mM spermidine supplemented with protease inhibitors) and resuspended in 1 ml of Wash buffer. 10-20 fly ovaries were dissected in ice cold PBS per sample. Ovaries were digested in 250 µl of dissociation buffer (0.5% Trypsin and 2.5 mg/ml Collagenase A in PBS) for an hour with shaking at 800 rpm at 30 °C. The digestion was stopped with 250 µl of Schneider medium containing 10% FBS. Cell suspensions were filtered through 40 µm strainers, pelleted and washed three times with Wash buffer and resuspended in 1 ml of Wash buffer.

Following sample preparation, CUT&RUN was performed according to instructions in CUT&RUN protocol V.3 with some modifications (Meers et al. 2019). 10 µl of activated Concanavalin A-coated magnetic beads (Bangs Laboratories) were added to each sample and rotated for 10 min at RT. Bead-bound cells were incubated with 5 µl of antibody (**Table S5**) and 95 µl of antibody buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 0.05% digitonin) overnight at 4 °C. Bead-bound cells were washed twice with Dig-wash buffer (0.05% digitonin in Wash buffer). Bead-bound cells were resuspended in Dig-wash buffer containing 1x CUTANA™ pAG-MNase (Epicypher) and rotated for 1 hr at RT. Following pAG-MNase binding, bead-bound cells were washed twice with Dig-wash buffer and resuspended in 100 µl Dig-wash buffer. Chromatin digestion was performed on ice for 30 min by adding 2 µL of 100 mM $CaCl_2$. Digestion was stopped by the addition of 2x STOP Buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.05% digitonin, 100 µg/ml RNase A (Thermo Fisher Scientific), 50 µg/ml glycogen) and samples were incubated at 37 °C for 30 min to release DNA fragments into the solution. After centrifugation, 0.1% SDS and 0.2 µg/µl Proteinase K were added to the supernatant and samples were incubated for 1 hour at 50 °C. DNA was extracted using Phenol:Chloroform:Isoamyl Alcohol 25:24:1 (Sigma) according to the manufacturer's instructions.

Libraries were prepared following the manufacturer's instructions with NEBNext Ultra II DNA Library Prep Kit for Illumina. Sequencing was performed on a NovaSeq 6000 instrument (Illumina).

Tissue culture, transfection and knockdowns

Drosophila Schneider 2 (S2) cells were purchased from Thermo Fisher Scientific and grown at 26 °C in Schneider media supplemented with 10% FBS. S2 cells were transfected using the TransIT®-Insect Transfection Reagent, using 2 million cells per transfection and 1 µg of plasmid DNA. Cells were collected after 48hrs. For knockdown experiments, two rounds of electroporation (48hrs apart) were performed using the Cell Line Nucleofector kit V (Amaxa Biosystems) and the program G-030, as described (Batki et al. 2019). Chromodomain constructs were co-transfected with 200 pmol of siRNA duplexes (oligo sequences are listed in **Table S4**) during the second nucleofection round or transfected using the TransIT®-Insect Transfection Reagent 24hrs after the second nucleofection round. Cells were collected 5 days after the first nucleofection round.

<u>Data analysis</u>

Processing of mRNA-seq and ribo-depleted RNA-seq

Adapters were removed using Trim Galore! (v0.6.4, --stringency 6) and with additional parameters "-a CTGTCTCTTATA --clip_R1 1 --clip_R2 1 --three_prime_clip_R1 1 --three_prime_clip_R2 1" if appropriate. The resulting reads were mapped to *dm6* using STAR (v2.7.3a, --outMultimapperOrder Random --outSAMmultNmax 1 --outFilterMultimapNmax 1000 --winAnchorMultimapNmax 2000 --alignSJDBoverhandMin 1 --sjdbScore 3) and a genome index built using NCBI RefSeq. Gene expression was quantified using featureCounts (v1.5.3, -s 2 -O --largestOverlap -Q 50) with gene models from Ensembl (release 97) or annotations corresponding to each transposon consensus sequences (RepBase, downloaded 2022-04-20).

Processing of sRNA-seq

The analyses included nos-driven *E(z)* or *w* knockdown (2 replicates each), pTOsk-driven *E(z)*, *rhi*, or *w* knockdown (2 replicates each), and previously published (Baumgartner et al. 2022) Gal4-driven *kipf*, *rhi* or *w* knockdown (1 replicate each) and Kipf-KO, rhi-KO and w1118 flies (1 replicate each). Adapters were removed using Trim Galore! (v0.6.4, --stringency 6 --length 18 --max_length 29 -q 0). For previously published sRNA-seq, we

further used '-a' to specify the adapter, and '--clip_R1' and '--three_prime_clip_R1' to remove non-standard adapters and random nucleotides at the read ends (see **Table S7** for details). The resulting reads were mapped to *dm6* using Bowtie (v2.3.1, -S -n 2 -M 1 --best --strata --nomaqround --chunkmbs 1024 --no-unal). Gene expression was quantified using featureCounts (v1.5.3, -s 1 [or -s 2] -O --largestOverlap -Q 40) with gene models from Ensembl (release 97) or annotations corresponding to each transposon consensus sequences (RepBase, downloaded 2022-04-20).

Differential gene expression and piRNA abundance analysis

RNA-seq differential expression analyses were performed using the DESeq2 package from R/Bioconductor. For gene expression, we first applied the DESeq2 function with default parameters followed by fold change shrinkage using the 'ashr' method. To analyse transposon expression, we performed a similar analysis, but used the size factors previously derived based on gene expression, while dispersion and fold change shrinkage was estimated based on both gene and transposon expression to increase robustness in the estimates. RPKM values were calculated following the robust median implementation in DESeq2's 'fpm' function. To consider a gene or transposon to be differentially expressed, we required a 4-fold change and padj<0.05 unless otherwise specified.

sRNA-seq was analysed in a similar way, except that the analysis was restricted to reads of length 23-30 nt mapping sense or antisense, respectively, to transposon consensus sequences. Size factors were estimated separately using the estimateSizeFactors function with siRNAs mapping antisense to annotated genes as input.

Processing of CUT&RUN data

CUT&RUN data from S2 cells and *D melanogaster* ovaries were generated in 2-4 replicates per condition as specified in **Table S7**. Sequencing adapters were removed using Trim Galore! (v0.6.4, --paired --stringency 6). The resulting reads were mapped to the *dm6* reference genome using Bowtie (v1.2.3, -S -y -M 1 --best --strata --fr --minins 10 --maxins 600 --chunkmbs 2000 --nomaqround), reporting at most one hit for each read and considering insert sizes between 10 and 600nt. PCR duplicates were removed using MarkDuplicates from Picard tools (v2.21.2). All the following analyses used the deduplicated data, except for the genome browser visualization.

Processing of ChIP-seq data

ChIP-seq reads were aligned to the dm6 genome using Bowtie2 (v2.4.2), with the alignment process set to report at most one hit for each read. In case of alignments with the same MAPQ score, the best alignment was randomly selected from among those equally scored alignments.

Peaks were called using MACS2 (v2.2.7.1) to capture narrow (-q 0.05 -g dm) and broad peaks (-q 0.05 -g dm –broad –broad-cutoff 0.1). Only uniquely mapped reads were used for the peak calling. As a control we used Input for each condition.

Processing of publicly available ChIP-seq data

Some publicly available ChIP-seq libraries used for comparison to CUT&RUN were analysed slightly differently. Specifically, HP1a ChIP-seq data for from ovaries was downloaded from GEO (accession GSE140539) (Zenk et al. 2021). The libraries were paired-end 2x101nt and performed in two replicates. Rhi ChIP-seq samples from control, Rhi, and Moon knockout ovaries and two ChIP-seq input samples were downloaded from GEO (accession GSE97719) (Andersen et al. 2017). The Rhi libraries were paired-end 2x50nt and performed in one replicate per condition. The input samples were 2x100nt and performed in two replicates. Rhi and Kipferl ChiP-seq samples from controls and Rhi and Kipf MTD-Gal4-mediated knockdown ovaries and corresponding input samples were downloaded from GEO (accession GSE202468) (Baumgartner et al. 2022). This data was had variable read length (50, 74, or 100 nt) and was processed as single-end 50 nt. An overview of all ChIP-seq samples is available in **Table S7**.

Sequencing adapters were removed using Trim Galore! (v0.6.4 or v0.6.6, --paired --stringency 6), or alternatively '--hardtrim5 50'. The resulting reads were mapped to the *dm6* reference genome using either Bowtie (v1.2.3, -S -y -M 1 --best --strata --fr --maxins 500 --chunkmbs 2000 --nomaqround), reporting at most one hit for each read. PCR duplicates were removed using MarkDuplicates from Picard tools (v2.21.2). All the following analyses used the deduplicated data, except for the genome browser visualization.

Peak calling using CUT&RUN or publicly available ChIP-seq data

Peaks were called using MACS2 to capture either narrow (callpeak -f BAMPE -g dm -q 0.01) or broad (callpeak -f BAMPE -g dm -q 0.05 --broad --broad-cutoff 0.1) peaks. Only uniquely mapped reads were used for the peak calling. As controls we used the corresponding IgG libraries for CUT&RUN, or input libraries for ChIP-seq, except for when Rhi ChIP-seq were compared to CUT&RUN and we used the Rhi knockout ChIP-seq as control. For samples with two or more replicates, we derived a consensus peak set by first merging all peaks and then excluding peaks that were not supported at least two replicates.

Binning analysis in *Drosophila melanogaster*

The binning analysis was done using either 1kb or 10kb bins. For 1kb bins, the genome was divided into 144,916 non-overlapping 1kb bins. To avoid regions with low mappability, we excluded bins with less than 20% mappability (fewer than 200 mappable positions) resulting in 125,519 mappable 1kb bins. For 10kb bins, the genome was divided into 29,918 bins using a 10kb sliding window that moves 5kb at a time. Bins at the chromosome ends with size less than 5kb were removed. To avoid regions with low mappability, we excluded bins with 20% mappability (fewer than 2,000 mappable positions) resulting in 25,865 mappable 10kb bins. Bins derived from mitochondria were further removed, producing a final set of 125,499 bins of size 1kb and 25,862 of size 10kb. Selection of mappable bins were done with respect for the CUT&RUN libraries, and the same bins were used for RNA-seq and sRNA-seq analyses.

CUT&RUN and publicly available ChIP-seq signal was quantified per 50nt window using the bamCompare module from deepTools (v3.3.2, --binSize 50 --ignoreForNormalization chrM -p 4 --scaleFactorsMethod SES --extendReads --centerReads --exactScaling --minMappingQuality 255 -of bedgraph --operation subtract --pseudocount 0) using pooled IgG CUT&RUN samples as background, except for Rhi ChIP-seq where we used the Rhi knockout ChIP-seq as background. Next, values below zero were set to zero, before the average normalized signal per window was calculated using bedtools map.

RNA-seq and sRNA-seq signal was quantified in a stand-specific manner by converting uniquely mapped reads to BED format and counting the number of reads from each strand falling into a bin with at least half of their length (bedtools intersect, -c -F 0.5).

Genome browser visualization

For CUT&RUN, RNA-seq and sRNA-seq, we first counted the number of all and uniquely mapped reads, respectively, using samtools, per strand if applicable. Uniquely mapped reads were converted into bigWig files using the deepTools bamCoverage

module (v3.3.2, --binSize 1 --ignoreForNormalization chrM --normalizeUsing CPM --exactScaling --scaleFactor *s* --skipNonCoveredRegions --minMappingQuality 255), where the scale factor, *s*, was set to the number of uniquely mapped reads divided by all mapped reads. Additionally, we used '--extendReads --centerReads' for CUT&RUN samples to centre the reads at fragments midpoints, '--filterRNAstrand' for RNA-seq and sRNA-seq to separate the two strands, and '--minFragmentLength 23 --maxFragmentLength 30' to select piRNAs for sRNA-seq.

Uniquely mapped reads were converted into bigwig files using deepTools bamCoverage function (v3.5.0 –binSize 1 –ignoreForNormalisation ChrM –normalizeUzing CPM –extendReads --centerReads –skipNonCoveredRegions).

Assessing classification performance using AUC

Classification performance for individual histone marks or combinations was assessed using a receiver operating characteristic (ROC) curve displaying true positive rate against false positive rate. Positive instances were defined as bins overlapping the cluster(s) of interest, with the remaining bins considered negative. The cumulative distribution was calculated in R using the 'cumsum' function with area under the curve calculated using the trapezoidal method.

Assessing Kipf-dependency of Rhi peaks

Rhi ChIP-seq peaks were derived using MACS2 as described above. To focus on the most reliable binding sites, we merged peaks with at least 3-fold enrichment over background in individual control knockdown replicates into a high-confidence peak set. Any peak that was located to an unplaced contig or that was also present in Rhi-KD was excluded from the analysis. Finally, the remaining peaks were subdivided into Kipf-independent (n=463) or Kipf-dependent (n=1166) peaks, depending on whether they overlapped a Rhi peak in Kipf-KD or not.

Coverage plots at Rhi peaks

To visualize read coverage over Rhi peaks (**Fig. 3g**) we used the deepTools computeMatrix module (v3.3.2, reference-point --bin-size 50 -b 5000 -a 5000 –missingDataAsZero –reference-point center). The signal in the peak region was shown as difference between the knock-down and the control conditions.

Contribution of E(z) and Kipf, respectively, to piRNA production

Analysis was performed per 1kb bin across the genome. First, we identified 632 bins where piRNA production depends on Rhi (>2-fold reduction in cpm, across pTOsk-Gal4- or MTD-Gal4-driven knockdown, and *rhi* knockout; 1-2 replicates each). Next, we calculated the change in piRNA abundance in E(z)- or Kipf-depleted ovaries (nos-Gal4- or pTOsk-Gal4-driven knockdown of *E(z)*, 2 replicates each; MTD-Gal4-driven knockdown of *kipf* or *kipf* knockout, 1 replicate each). Bins with >4-fold reduction in ovaries depleted for Kipf, E(z), or both were considered to be Kipf-dependent, E(z)-dependent or dependent on both.

Correlation heatmaps

Heatmaps were produced using the R package pheatmap (v1.0.12) with Euclidean distance.

Analysis of CUT&RUN in E(z) knockdown samples

To study the genome-wide effect of *E(z)* knockdown in S2 cells, we represented the genome as 117,300 1kb bins, and divided it into six equally sized groups based on the change in H3K27me3 signal upon E(z) loss. Four groups represented variable levels of H3K27me3 loss, one group represented no change and one group displayed a relative gain in H3K27me3 signal. Next, across each group, we calculated the change in 2xPc, 2xCBX1, 2xRhi, and Pc-CBX1 binding affinity. Although 2xPc was strongly responsive to H3K27me3 loss, the other chromatin binders were largely unaffected. Hence, we next restricted the analysis to 46,583 bins across the six groups with H3K9me3 (defined as CD^{2xCBX1} above 90th percentile of euchromatic regions).

Binding affinity analysis using CUT&RUN data

Differential binding affinity analysis was performed using DiffBind (v2.14.0) applied on the MACS2 narrow or broad peaks from *siRen* and *siEz* samples, using the corresponding IgG libraries as controls and without specifying any blacklist regions. Two samples (siRen.2xRhi.3 and siEx.2xPc.5) were excluded due to library preparation failures characterized by low library complexity (61% duplication rate with estimated complexity an order of magnitude less than other libraries) and missing library (only 3,224 reads), respectively.

For the analysis, we used the dba, dba.count, dba.normalize, dba.contrast, and dba.analyze modules with default options. This will create a consensus set of peaks present in at least two samples, and then re-quantify the signal intensity at each consensus peak, and re-center and re-size the peaks to a 401nt region around their maxima. This resulted in 30,459 narrow consensus or 8,598 broad consensus peaks. The contrasts were specified as *siE(z)* against *siRen* for each target (H3K9me3, H3K27me3, 2xPc, 2xCBX1, 1xPc-1xCBX1, 2xRhi).

Euchromatin and heterochromatin coordinates

We used the following coordinates from Fabry and colleagues (Fabry et al. 2021) to define euchromatin (chr2R:6460000-25286936, chr2L:1-22160000, chr3L :1-23030000, chr3R:4200000-32079331, chrX:250000-21500000) and heterochromatin (chr2R:1-6460000, chr2L:22160000-23513712, chr3L:23030000-28110227, chr3R:1-4200000) in *D melanogaster*.

Reference genomes

We used the *dm6* assembly for *D melanogaster* downloaded from the UCSC genome browser and the following assemblies downloaded from NCBI: GCF_003285975 for *D annanassae*, GCF_003286155 for *D erecta*, GCF_016746395 for *D simulans*, and GCF_016746365 for *D yakuba*.

Genome-wide mappability

To estimate genome mappability, we divided the genome into all possible *n*-mers, where *n* is either 26 (for sRNA) or 100 (for CUT&RUN). Those sequences were then mapped back onto the genome and mappability for each position was estimated as the number of reads overlapping a position, divided by *n*.

Analysis of CUT&RUN in five *Drosophila* species

CUT&RUN was performed in *D annanassae*, *D melanogaster*, *D erecta*, *D simulans*, and *D yakuba* for H3K27me3 and H3K9me3 with two replicates per condition, followed by 2x50nt paired-end sequencing. Sequencing adapters were removed using Trim Galore! (v0.6.4, --paired --stringency 6 -a GATCGGAAGAGCACACGTCTGAACTCCAGTCAC). The resulting reads were mapped to their respective reference genome using Bowtie (v1.2.3, -S -y -M 1 --best --strata --fr --minins 10 --maxins 600 --chunkmbs 2000 --nomaqround), reporting at most one hit for each read and considering insert sizes between 10 and 600nt. One of the H3K9me3 replicates for *D yakuba* displayed very low complexity (1.3 million uniquely mapped reads with estimated library size 1.7 million fragments) and was therefore excluded from the analysis.

Analysis of sRNA-seq in five *Drosophila* species

Small RNA-seq was performed in *D annanassae*, *D melanogaster*, *D erecta*, *D simulans*, and *D yakuba* with three replicates per species. The libraries were sequenced as 2x50nt paired-end sequencing, but only the first read was used for the analysis. Trim Galore! (v0.6.4) was used to remove an abundant rRNA (--stringency 30 -a TGCTTGGACTACATATGGTTGAGGGTTGTA --length 18 -q 0) and sequencing adapters (--stringency 5 -a TGGAATTCTCGG --length 18 --max_length 35 -q 0). Next, Bowtie was used to exclude reads mapping to known *Drosophila* viruses (v.2.3.1, -S -M 1 --best --strata --nomaqround --chunkmbs 1024) using the '--max and --un' options to extract unmapped reads. These reads were mapped to their respective reference genome (-S -M 1 --best --strata --nomaqround --chunkmbs 1024), reporting at most one hit for each read.

Binning analysis in five *Drosophila* species

For the binning analysis, the genome was divided into either 1kb non-overlapping bins or 10kb bins with 5kb overlap. To estimate the mappability within each bin, we divided the genome into all possible 100-mers and mapped those sequences back onto the genome. Mappability was calculated as the number of 100-mers mapping uniquely within each bin divided by the bin size. Bins with size <5kb (for 10kb bins) or with <20% mappability were excluded, leaving between 121,081 and 178,773 1kb or between 24,687 and 38,112 10kb bins per species. We next quantified the number of uniquely mapped CUT&RUN or sRNA-seq reads per library mapping ≥50% within each bin, considering the forward and reverse strand separately for the sRNA-seq data. To correct for differences in sequencing depth, mappability, and bin size, the resulting bin counts were normalized by their mappability and converted to fpkm values. Bins from mitochondria were excluded from all subsequent analyses.

Scatterplots and violin plots in five *Drosophila* species

The normalized rpkm values were capped at 300 (or 200 for 10kb bins) and log10-transformed using a pseudo-count of 1. To focus on bins with piRNAs derived from both strands, we represented the sRNA-seq data as minimum signal on the forward and reverse strands. For H3K9me3 and H3K27me3 we used the mean of two replicates and for sRNA-seq we used the mean of three replicates. For the violin plots, we grouped the bins based on their piRNA level and displayed the piRNA, H3K27me3 or H3K9me3 level within each group. Briefly, we first constructed one group for bins with no piRNAs. Next, we sorted all remaining bins by piRNA level and extracted 15 equidistant breakpoints, including the lowest and highest piRNA level. Each pair of consecutive non-identical breakpoints were used to construct a piRNA level interval, resulting in between 11 or 14 additional groups per species (1kb) or between 14 and 15 additional groups (10kb). All scatterplots and violin plots were constructed in R (v3.6.2) using base graphics, imageScale from sinkr (v0.6), gridExtra, and ggplot2.

Structure modelling of Rhi dimer using AlphaFold multimer

AlphaFold v2.3.2 Colab was used to generate a prediction model (Jumper et al. 2021). The full-length Rhino chromodomain protein (amino acids 20-90) reported in the crystal structure of **4U68** was used as input twice to generate the multimer (Yu et al. 2015). The structure of the Rhino chromodomain dimer was found to resemble the original structure of **4U68** well (Yu et al. 2015).

Structure modelling of histone 3 peptide using Zdock

Zdock was used to evaluate potential binding modes of the peptide to the dimer. To generate the structure of the input histone 3 peptide, the existing crystal structure of the K9me3- and K27me3-containing peptides in **4U68** and **1PFB** were used (Min, Zhang and Xu 2003; Yu et al. 2015). Within the structure of **4U68**, information for the structure of KQTARK(Me3)S, of the K9me3 section of the histone peptide was available. The structure of LATKAAR(Me3)SAP, of the K27me3 section of the histone peptide was available via **1PFB**. A structure alignment was done using **4U68** and **1PFB** via PyMOL, and the connecting chain of 'TGGKAPRKQ' was added using Modeller tool in UCSF Chimera 1.16 (Fiser and Sali 2003; Pettersen et al. 2004). This peptide was then docked using Zdock server, with ZDOCK 3.0.2 (Pierce et al. 2014) to yield a predictive model of the binding motif.

Molecular dynamics simulations

The structural prediction from AlphaFold multimer was used as a basis for molecular dynamic (MD) simulations. MD simulations in explicit water were performed using the graphics processing unit accelerated code (PMEMD) of the Amber 16 and AmberTools 20 packages (Case et al. 2016; Case et al. 2020). Protonation states were calculated using the PDB2PQR server. For the protein scaffold, an evolved version of the Stony Brook modification of the Amber 99 force field (ff14SB) (Maier et al. 2015) parameters were applied (see below for additional non-standard parameters). TIP3P parameters were assigned to water molecules (Jorgensen et al. 1983). To incorporate the non-canonical amino acid K(Me3), into the AlphaFold model, the structure of the K(Me3) residue was used from **4U68** and the partial charges of the K(Me3) residue were set to fit the electrostatic potential generated at the HF/6-31G(d) level by the restrained electrostatic potential model (Bayly et al. 1993). The charges were calculated according to the Merz–Singh–Kollman scheme using Gaussian (Besler, Merz Jr. and Kollman 1990). Each protein complex was immersed in a pre-equilibrated cubic box with a 12Å buffer of TIP3P water molecules using the Leap module. The systems were neutralized by the addition of explicit counterions (Na^+ or Cl^-). Long-range electrostatic effects were modelled using the particle mesh Ewald method with periodic boundary conditions (Darden, York and Pedersen 1993). An 8Å cut-off was applied to Lennard-Jones and electrostatic interactions.

MD simulations were performed according to the following steps: (1) Minimization was performed with a maximum cycle of 5000 and with the steepest descent algorithm for the first 2500 cycles, with a periodic boundary for constant volume (canonical ensemble, NVT) and without the SHAKE (an algorithm for constrained molecular dynamics) algorithm activated. Positional restraints of 2 $kcalmol^{-1}Å^{-2}$ were applied on heavy atoms of the protein backbone and heavy atoms of the ligand. (2) A 1 ns heating process was performed with a periodic boundary for constant volume (NVT) with the SHAKE algorithm turned on such that the angle between the hydrogen atoms was kept fixed. Temperature increased from 0 K to 300 K in a time period of 1 ns with heat bath coupling with time constant of 2 ps. Positional restraints of 2 $kcalmol^{-1}Å^{-2}$ were applied on heavy atoms of the protein backbone. (3) A 2 ns equilibrium process was performed with a periodic boundary for constant volume (NVT) with the SHAKE algorithm turned on such that the angle between the hydrogen atoms was kept fixed. An Andersen-like temperature coupling scheme is used to maintain the temperature at 300 K. (4) A 2 ns equilibrium process was performed with a periodic boundary for constant pressure (isothermal-isobaric ensemble, NPT) and with a constant temperature of 300 K, maintained using Langevin dynamics with the collision frequency of 5 ps^{-1}. (5) A 100 ns equilibrium process was performed with a periodic boundary for constant pressure (NPT) and with a constant temperature of 300 K. (6) A 1000 ns production was performed with a periodic boundary for constant pressure (NPT) and with a constant

temperature of 300 K. A representative frame was obtained from 1 µS MD simulations using Chimera's cluster analysis tool (Kelley, Gardner and Sutcliffe 1996).

Data availability

Sequencing data generated in this study has been deposited to the GEO under accession (GSExxxxx).

ACKNOWLEDGEMENTS

We thank members of the Brasset and Hannon groups for fruitful discussions. We thank Lisa Baumgartner and Julius Brennecke for sharing unpublished data and scientific discussion. We thank Marie Bao from the Life Science Editors for feedback and comments on the manuscript. We thank the Scientific Computing, Genomics, RICS and microscopy core facilities at CRUK Cambridge Institute and the CLIC facility (Clermont Imagerie Confocale) at iGReD for support. GJH is a Royal Society Wolfson Research Professor (RSRP\R\200001). Work in the Hannon lab was funded in whole, or in part, by Cancer Research UK (G101107) and the Wellcome Trust (110161/Z/15/Z). PC and MJ were supported by Cancer Research UK grants C9685/A26398, C9685/A27415 and C9545/A29580. Work in the Brasset lab was funded by the Agence Nationale pour la Recherche (CHApiTRE) ANR-20-CE12-0005, BiopiC ANR-21-CE12-0022, and EpiTET 699 ANR-17-CE12-0030-03), the La Fondation ARC pour la recherche sur le cancer (PJA20171206129). This research was financed also by the French government IDEX-ISITE initiative 16-IDEX-0001 (CAP 20-25).

REFERENCES

Andersen PR, Tirian L, Vunjak M, Brennecke J. 2017. A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature* **549**: 54-59.

Azuara V, Perry P, Sauer S, Spivakov M, Jorgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M, Fisher AG. 2006. Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* **8**: 532-538.

Barral A, Pozo G, Ducrot L, Papadopoulos GL, Sauzet S, Oldfield AJ, Cavalli G, Dejardin J. 2022. SETDB1/NSD-dependent H3K9me3/H3K36me3 dual heterochromatin maintains gene expression profiles by bookmarking poised enhancers. *Mol Cell* **82**: 816-832 e812.

Batki J, Schnabl J, Wang J, Handler D, Andreev VI, Stieger CE, Novatchkova M, Lampersberger L, Kauneckaite K, Xie W et al. 2019. The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation. *Nat Struct Mol Biol* **26**: 720-731.

Baumgartner L, Handler D, Platzer SW, Yu C, Duchek P, Brennecke J. 2022. The Drosophila ZAD zinc finger protein Kipferl guides Rhino to piRNA clusters. *Elife* **11**.

Bayly CI, Cieplak P, Cornell W, Kollman PA. 1993. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* **97**: 10269-10280.

Beltran T, Barroso C, Birkle TY, Stevens L, Schwartz HT, Sternberg PW, Fradin H, Gunsalus K, Piano F, Sharma G et al. 2019. Comparative Epigenomics Reveals that RNA Polymerase II Pausing and Chromatin Domain Organization Control Nematode piRNA Biogenesis. *Dev Cell* **48**: 793-810 e796.

Bernstein E, Duncan EM, Masui O, Gil J, Heard E, Allis CD. 2006. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol Cell Biol* **26**: 2560-2569.

Besler BH, Merz Jr. KM, Kollman PA. 1990. Atomic charges derived from semiempirical methods. *Journal of Computational Chemistry* **11**: 431-439.

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

Case DA, Belfon K, Ben-Shalom I, Brozell SR, Cerutti DS, Cheatham TE, Cruzeiro V, Darden T, Duke R, Giambasu G. 2020. Amber 2020: University of california. *San Francisco*.

Case DA, Betz R, Cerutti D, Cheatham T, Darden T, Duke R, Giese T, Gohlke H, Goetz A, Homeyer N. 2016. AMBER 2016 reference manual. *University of California: San Francisco, CA, USA*: 1-923.

Chen YA, Stuwe E, Luo Y, Ninova M, Le Thomas A, Rozhavskaya E, Li S, Vempati S, Laver JD, Patel DJ et al. 2016. Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. *Mol Cell* **63**: 97-109.

Czech B, Munafo M, Ciabrelli F, Eastwood EL, Fabry MH, Kneuss E, Hannon GJ. 2018. piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annu Rev Genet* **52**: 131-157.

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

Darden T, York D, Pedersen L. 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**: 10089-10092.

Deleris A, Berger F, Duharcourt S. 2021. Role of Polycomb in the control of transposable elements. *Trends Genet* **37**: 882-889.

Deleris A, Stroud H, Bernatavichute Y, Johnson E, Klein G, Schubert D, Jacobsen SE. 2012. Loss of the DNA methyltransferase MET1 Induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in Arabidopsis thaliana. *PLoS Genet* **8**: e1003062.

DeLuca SZ, Ghildiyal M, Pang LY, Spradling AC. 2020. Differentiating Drosophila female germ cells initiate Polycomb silencing by regulating PRC2-interacting proteins. *Elife* **9**.

Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E, Vashisht AA, Terragni J, Chin HG, Tu A et al. 2012. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell* **151**: 167-180.

Eastwood EL, Jara KA, Bornelov S, Munafo M, Frantzis V, Kneuss E, Barbar EJ, Czech B, Hannon GJ. 2021. Dimerisation of the PICTS complex via LC8/Cut-up drives co-transcriptional transposon silencing in Drosophila. *Elife* **10**.

ElMaghraby MF, Andersen PR, Puhringer F, Hohmann U, Meixner K, Lendl T, Tirian L, Brennecke J. 2019. A Heterochromatin-Specific RNA Export Pathway Facilitates piRNA Production. *Cell* **178**: 964-979 e920.

ElMaghraby MF, Tirian L, Senti KA, Meixner K, Brennecke J. 2022. A genetic toolkit for studying transposon control in the Drosophila melanogaster ovary. *Genetics* **220**.

Evans KJ, Huang N, Stempor P, Chesney MA, Down TA, Ahringer J. 2016. Stable Caenorhabditis elegans chromatin domains separate broadly expressed and

developmentally regulated genes. *Proc Natl Acad Sci U S A* **113**: E7020-E7029.

Fabry MH, Falconio FA, Joud F, Lythgoe EK, Czech B, Hannon GJ. 2021. Maternally inherited piRNAs direct transient heterochromatin formation at active transposons during early Drosophila embryogenesis. *Elife* **10**.

Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD, Khorasanizadeh S. 2003. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* **17**: 1870-1881.

Fiser A, Sali A. 2003. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol* **374**: 461-491.

Franke A, DeCamillis M, Zink D, Cheng N, Brock HW, Paro R. 1992. Polycomb and polyhomeotic are constituents of a multimeric protein complex in chromatin of Drosophila melanogaster. *EMBO J* **11**: 2941-2950.

Frapporti A, Miro Pina C, Arnaiz O, Holoch D, Kawaguchi T, Humbert A, Eleftheriou E, Lombard B, Loew D, Sperling L et al. 2019. The Polycomb protein Ezl1 mediates H3K9 and H3K27 methylation to repress transposable elements in Paramecium. *Nat Commun* **10**: 2710.

Gaertner B, Johnston J, Chen K, Wallaschek N, Paulson A, Garruss AS, Gaudenz K, De Kumar B, Krumlauf R, Zeitlinger J. 2012. Poised RNA polymerase II changes over developmental time and prepares genes for future expression. *Cell Rep* **2**: 1670-1683.

Goriaux C, Desset S, Renaud Y, Vaury C, Brasset E. 2014. Transcriptional properties and splicing of the flamenco piRNA cluster. *EMBO Rep* **15**: 411-418.

Grentzinger T, Oberlin S, Schott G, Handler D, Svozil J, Barragan-Borrero V, Humbert A, Duharcourt S, Brennecke J, Voinnet O. 2020. A universal method for the rapid isolation of all known classes of functional silencing small RNAs. *Nucleic Acids Res* **48**: e79.

Hisanaga T, Romani F, Wu S, Kowar T, Wu Y, Lintermann R, Fridrich A, Cho CH, Chaumier T, Jamge B et al. 2023. The Polycomb repressive complex 2 deposits H3K27me3 and represses transposable elements in a broad range of eukaryotes. *Curr Biol*.

Ho JW, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, Sohn KA, Minoda A, Tolstorukov MY, Appert A et al. 2014. Comparative analysis of metazoan chromatin organization. *Nature* **512**: 449-452.

Huang X, Cheng P, Weng C, Xu Z, Zeng C, Xu Z, Chen X, Zhu C, Guang S, Feng X. 2021. A chromodomain protein mediates heterochromatin-directed piRNA expression. *Proc Natl Acad Sci U S A* **118**.

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

Ichihara S, Nagao K, Sakaguchi T, Obuse C, Sado T. 2022. SmcHD1 underlies the formation of H3K9me3 blocks on the inactive X chromosome in mice. *Development* **149**.

Isbel L, Iskar M, Durdu S, Weiss J, Grand RS, Hietter-Pfeiffer E, Kozicka Z, Michael AK, Burger L, Thoma NH, Schubeler D. 2023. Readout of histone methylation by Trim24 locally restricts chromatin opening by p53. *Nat Struct Mol Biol* **30**: 948-957.

Iyer SS, Sun Y, Seyfferth J, Manjunath V, Samata M, Alexiadis A, Kulkarni T, Gutierrez N, Georgiev P, Shvedunova M, Akhtar A. 2023. The NSL complex is required for piRNA production from telomeric clusters. *Life Sci Alliance* **6**.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**: 926-935.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**: 583-589.

Kelley LA, Gardner SP, Sutcliffe MJ. 1996. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Engineering, Design and Selection* **9**: 1063-1065.

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

Kneuss E, Munafo M, Eastwood EL, Deumer US, Preall JB, Hannon GJ, Czech B. 2019. Specialization of the Drosophila nuclear export family protein Nxf3 for piRNA precursor export. *Genes Dev* **33**: 1208-1220.

Kumar D, Cinghu S, Oldfield AJ, Yang P, Jothi R. 2021. Decoding the function of bivalent chromatin in development and cancer. *Genome Res* **31**: 2170-2184.

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

Lee TI, Johnstone SE, Young RA. 2006. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc* **1**: 729-748.

Liu T, Rechtsteiner A, Egelhofer TA, Vielle A, Latorre I, Cheung MS, Ercan S, Ikegami K, Jensen M, Kolasinska-Zwierz P et al. 2011. Broad chromosomal domains of histone modification patterns in C. elegans. *Genome Res* **21**: 227-236.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402-408.

Lopik Jv, Trapotsi M-A, Hannon GJ, Bornelöv S, Nicholson BC. 2023. Unistrand piRNA clusters are an evolutionary conserved mechanism to suppress endogenous retroviruses across the Drosophila genus. *bioRxiv*: 2023.2002.2027.530199.

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**: 3696-3713.

Margueron R, Reinberg D. 2011. The Polycomb complex PRC2 and its mark in life. *Nature* **469**: 343-349.

Mashtalir N, Dao HT, Sankar A, Liu H, Corin AJ, Bagert JD, Ge EJ, D'Avino AR, Filipovski M, Michel BC et al. 2021. Chromatin landscape signals differentially dictate the activities of mSWI/SNF family complexes. *Science* **373**: 306-315.

Mathieu O, Probst AV, Paszkowski J. 2005. Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. *EMBO J* **24**: 2783-2791.

Meers MP, Bryson TD, Henikoff JG, Henikoff S. 2019. Improved CUT&RUN chromatin profiling tools. *Elife* **8**.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK,

Koche RP et al. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**: 553-560.

Min J, Zhang Y, Xu RM. 2003. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev* **17**: 1823-1828.

Miro-Pina C, Charmant O, Kawaguchi T, Holoch D, Michaud A, Cohen I, Humbert A, Jaszczyszyn Y, Chevreux G, Del Maestro L et al. 2022. Paramecium Polycomb repressive complex 2 physically interacts with the small RNA-binding PIWI protein to repress transposable elements. *Dev Cell* **57**: 1037-1052 e1038.

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

Montgomery SA, Tanizawa Y, Galik B, Wang N, Ito T, Mochizuki T, Akimcheva S, Bowman JL, Cognat V, Marechal-Drouard L et al. 2020. Chromatin Organization in Early Land Plants Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin. *Curr Biol* **30**: 573-588 e577.

Mozzetta C, Pontis J, Fritsch L, Robin P, Portoso M, Proux C, Margueron R, Ait-Si-Ali S. 2014. The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. *Mol Cell* **53**: 277-289.

Musselman CA, Lalonde ME, Cote J, Kutateladze TG. 2012. Perceiving the epigenetic landscape through histone readers. *Nat Struct Mol Biol* **19**: 1218-1227.

Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. 2019. PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet* **20**: 89-108.

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

Parhad SS, Tu S, Weng Z, Theurkauf WE. 2017. Adaptive Evolution Leads to Cross-Species Incompatibility in the piRNA Transposon Silencing Machinery. *Dev Cell* **43**: 60-70 e65.

Peng JC, Valouev A, Liu N, Lin H. 2016. Piwi maintains germline stem cells and oogenesis in Drosophila through negative regulation of Polycomb group proteins. *Nat Genet* **48**: 283-291.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605-1612.

Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, Weng Z. 2014. ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* **30**: 1771-1773.

Rangan P, Malone CD, Navarro C, Newbold SP, Hayes PS, Sachidanandam R, Hannon GJ, Lehmann R. 2011. piRNA production requires heterochromatin formation in Drosophila. *Curr Biol* **21**: 1373-1379.

Rougee M, Quadrana L, Zervudacki J, Hure V, Colot V, Navarro L, Deleris A. 2021. Polycomb mutant partially suppresses DNA hypomethylation-associated phenotypes in Arabidopsis. *Life Sci Alliance* **4**.

Saksouk N, Barth TK, Ziegler-Birling C, Olova N, Nowak A, Rey E, Mateos-Langerak J, Urbach S, Reik W, Torres-Padilla ME et al. 2014. Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. *Mol Cell* **56**: 580-594.

Tsai WW, Wang Z, Yiu TT, Akdemir KC, Xia W, Winter S, Tsai CY, Shi X, Schwarzer D, Plunkett W et al. 2010. TRIM24 links a non-canonical histone signature to breast cancer. *Nature* **468**: 927-932.

Van Doren M, Williamson AL, Lehmann R. 1998. Regulation of zygotic gene expression in Drosophila primordial germ cells. *Curr Biol* **8**: 243-246.

Veluchamy A, Rastogi A, Lin X, Lombard B, Murik O, Thomas Y, Dingli F, Rivarola M, Ott S, Liu X et al. 2015. An integrative analysis of post-translational histone modifications in the marine diatom Phaeodactylum tricornutum. *Genome Biol* **16**: 102.

Vermaak D, Henikoff S, Malik HS. 2005. Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in Drosophila. *PLoS Genet* **1**: 96-108.

Villasenor R, Pfaendler R, Ambrosi C, Butz S, Giuliani S, Bryan E, Sheahan TW, Gable AL, Schmolka N, Manzo M et al. 2020. ChromID identifies the protein interactome at chromatin marks. *Nat Biotechnol* **38**: 728-736.

Walter M, Teissandier A, Perez-Palacios R, Bourc'his D. 2016. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife* **5**.

Weinhofer I, Hehenberger E, Roszak P, Hennig L, Kohler C. 2010. H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. *PLoS Genet* **6**.

Xu J, Zhao X, Mao F, Basrur V, Ueberheide B, Chait BT, Allis CD, Taverna SD, Gao S, Wang W, Liu Y. 2021. A Polycomb repressive complex is required for RNAi-mediated heterochromatin formation and dynamic distribution of nuclear bodies. *Nucleic Acids Res* **49**: 5407-5425.

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

Yu B, Huang Y. 2015. Rhino defines H3K9me3-marked piRNA clusters. *Oncotarget* **6**: 20740-20741.

Zenk F, Loeser E, Schiavo R, Kilpert F, Bogdanovic O, Iovino N. 2017. Germ line-inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition. *Science* **357**: 212-216.

Zenk F, Zhan Y, Kos P, Loser E, Atinbayeva N, Schachtle M, Tiana G, Giorgetti L, Iovino N. 2021. HP1 drives de novo 3D genome reorganization in early Drosophila embryos. *Nature* **593**: 289-293.

Zhang G, Tu S, Yu T, Zhang XO, Parhad SS, Weng Z, Theurkauf WE. 2018. Co-dependent Assembly of Drosophila piRNA Precursor Complexes and piRNA Cluster Heterochromatin. *Cell Rep* **24**: 3413-3422 e3414.

Zhang G, Yu T, Parhad SS, Ho S, Weng Z, Theurkauf WE. 2021. piRNA-independent transposon silencing by the Drosophila THO complex. *Dev Cell* **56**: 2623-2635 e2625.

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

Zhao X, Xiong J, Mao F, Sheng Y, Chen X, Feng L, Dui W, Yang W, Kapusta A, Feschotte C et al. 2019. RNAi-dependent Polycomb repression controls transposable elements in Tetrahymena. *Genes Dev* **33**: 348-364.