ADMA-histones define genomic loading of Rhino at the initial step of piRNA cluster formation

Raku Saito[1], Hirotsugu Ishizu[2], Ritsuko Harigai[1], Kensaku Murano[2], and Mikiko C. Siomi[1]*

[1]Department of Biological Sciences, Faculty of Science, The University of Tokyo, Tokyo 113-0032, Japan

[2]Department of Molecular Biology, Keio University School of Medicine, Tokyo 162-8582, Japan

*Corresponding author: M.C.S.: siomim@bs.s.u-tokyo.ac.jp

Keywords: piRNA cluster, Rhino, histone arginine methylation, ADMA, piRNA

Abstract

In *Drosophila* germ cells, piRNAs arise from dual-strand piRNA clusters. These clusters are occupied by H3K9me3, but are transcribed from internal sites in a manner dependent on the binding of HP1 homolog Rhino to H3K9me3 on the clusters. However, how initial loading of Rhino onto the clusters occurs remains unknown. Here, we used cultured ovarian somatic cells (OSCs), which lack endogenous Rhino and Rhino stabilizer Kipferl, the absence of which renders the dual-strand clusters inert, and found that exogenous Rhino tends to bind to the ends of dual-strand clusters with asymmetric dimethylarginine histones (ADMA-histones). Depletion of the arginine methyltransferases responsible for ADMA modification affected the genomic localization of Rhino in OSCs and in the ovary. We also identified genomic regions, termed INSECTs, where ADMA-dependent Rhino propagation begins. We propose that ADMA-histones define the initial genomic loading of Rhino and stabilize Rhino–genome association during cluster formation.

Introduction

PIWI-interacting RNAs (piRNAs) are germline-enriched small non-coding RNAs that repress transposons in collaboration with PIWI proteins to maintain fertility[1,2,3]. In *Drosophila* germ cells, piRNAs are primarily derived from dual-strand piRNA clusters, intergenic elements rich in transposon fragments[4,5]. These clusters are occupied with the repressive histone mark H3K9me3 and lack typical promoters, but are transcribed in both directions from the internal sites[6,7]. This unconventional transcription requires special factors, including heterochromatin protein 1 (HP1) homolog HP1d/Rhino (Rhino), and provides piRNA precursors with a variety of sequences, which are then processed to mature piRNAs in the cytoplasm[5,6,7,8,9,10,11,12,13,14]. The mechanism dedicated to this piRNA processing is known as the ping-pong cycle, where Aubergine and Argonaute 3 cleave the precursors in a manner dependent on the sequences of piRNAs already bound to them[1,2,3]. This pathway also consumes transposon RNA transcripts as the sources of mature piRNAs, resulting in transposon repression at the post-transcriptional level[2,3].

Like other HP1 family members, Rhino has a chromodomain through which it binds to H3K9me3 scattered on dual-strand piRNA clusters[6,8,9,12,13,14,15,16]. Deadlock and Cutoff then join Rhino, assembling the Rhino–Deadlock–Cutoff (RDC) complex, which further recruits the transcription factor IIA-L (TFIIA-L) paralogue Moonshiner, the transcription factor IIA-S (TFIIA-S), and the animal TATA-box-binding protein (TBP) paralog TRF2, initiating transcription of the clusters[6,7,10]. Upon transcription, Cutoff binds to piRNA precursors and protects them from splicing, contributing to the production of an abundance of piRNAs[17,18,19,20].

A recent study identified the zinc-finger protein Kipferl as a factor required for

Rhino to distinguish between H3K9me3 located on transposons and dual-strand piRNA clusters[12]. Kipferl stabilizes Rhino on guanine-rich sequence motifs present in dual-strand clusters in an H3K9me3-independent manner[12]. Maternally inherited piRNAs also play a role in the localization of Rhino to dual-strand clusters during early embryogenesis via Piwi-dependent H3K9me3 epigenetic marks [9,21].

Rhino has a propensity to spread across the genome. For example, depletion of the histone demethylase Kdm3 in the female germline causes Rhino to spread to adjacent chromatin[13]. Deletion of the region proximal to the transcription start site of the *phospholipase D* gene, which is in the region flanking piRNA cluster 42AB, also resulted in Rhino spreading toward the deleted genomic region[10]. In addition, Rhino tethering experiments with artificial constructs showed that Rhino spreads as the constructs are transcribed[7]. In each case, production of piRNAs from regions after Rhino spreading was detected. These findings suggest that piRNA clusters are defined by Rhino spreading over the genome and imply that the process of piRNA cluster formation is composed of three steps: Rhino spreading, Kipferl-dependent stabilization of Rhino–genome association, and epigenetic inheritance of the Rhino positional memory by maternally inherited piRNAs. It is imagined that the initial loading of Rhino on piRNA clusters precedes the spread of Rhino, but the mechanism of such Rhino behavior remains unclear.

In this Article, we describe the important role of asymmetric dimethylarginine histones (ADMA-histones) in the initial loading of Rhino onto piRNA clusters and their stabilization. We first established cultured ovarian somatic cells (OSCs)[22] expressing FLAG-Rhino by induction with doxycycline (Dox). Unlike ovarian germ cells, OSCs lack Rhino and other factors required for the activation of dual-strand piRNA clusters, such

as Deadlock and Cutoff[23]; therefore, dual-strand clusters are not in an active state for piRNA biogenesis[24]. We found that Rhino in OSCs assembles nuclear foci similar to Rhino foci in the ovary. We then performed chromatin immunoprecipitation sequencing (ChIP-seq) and showed that Rhino tends to bind to the ends of dual-strand clusters with ADMA-histones, such as H3R17me2a. Depletion of the arginine methyltransferases responsible for ADMA modification affected the genomic localization of Rhino in OSCs and in the ovary. This study also identified genomic regions where H3R17me2a-dependent Rhino propagation begins. piRNAs were mapped to these regions, albeit at low levels; therefore, these regions are considered cluster precursors.

Results

Rhino localizes to the ends of dual-strand clusters in the OSC genome

To observe how Rhino in OSCs behaves intracellularly, we first performed immunofluorescence to examine the subcellular localization of FLAG-Rhino upon Dox induction in FLAG-Rhino inducible OSCs. This showed that Rhino assembled nuclear speckles in OSCs similar to Rhino-positive foci in the ovary[5,10] (Fig. 1a). It is noted that at lower Dox concentrations (< 0.05 ng/mL) and lower levels of Rhino expression, no foci appeared, but at higher concentrations (e.g., 5 ng/mL), the speckles appeared (Fig. 1a, Extended Data Fig. 1a, b). This suggests that Rhino-positive foci form in a manner dependent on the level of Rhino expression.

We next conducted ChIP-seq of FLAG-Rhino under conditions where Rhino assembles nuclear foci as in Fig. 1a. Genome mapping of the ChIP-seq reads showed that Rhino can be found on dual-strand piRNA clusters, such as 38C1 and 38C2 on chromosome 2 and 102F on chromosome 4, which are in an active state in piRNA

production in the ovary, but the signals tend to be confined to the ends of the clusters (Fig. 1b). These results suggest that OSCs contain a factor(s) that promotes Rhino binding to the ends of the clusters, but unlike the ovary, they lack a factor(s) that promotes and stabilizes Rhino spreading throughout the clusters. The Rhino stabilizer Kipferl is not expressed in OSCs[12].

We next compared the mode of Rhino genome binding in OSCs with that in the ovary in a comprehensive manner. This showed that, in genomic regions where Rhino binds locally within the ovary, Rhino binds similarly in OSCs (Fig. 1b, c). However, in genomic regions where Rhino binds extensively within the ovary, such as 38C1 and 38C2 on chromosome 2, Rhino failed to exhibit similar behavior in OSCs (Fig. 1b, c). Further analysis showed that, in OSCs, Rhino preferentially binds to mononucleosomal units, but not to multiple units, unlike in the ovary (Fig.1d). This specific localization does not appear to be due to H3K9me2, because it tends to be rather widely localized (Fig. 1b, e, f). As well, H3K9me3 was not always found in Rhino peaks: While most H3K9me3 localizes broadly in the genome, Rhino co-localized only with sharp H3K9me3 peaks (Fig. 1b, e–g). Immunofluorescence also showed weak overlap between the prominent H3K9me3 speckles and FLAG-Rhino foci (Fig. 1a). Furthermore, annotation of FLAG-Rhino localization in OSCs showed that Rhino preferentially binds to euchromatic regions rather than heterochromatin and annotated piRNA clusters (Fig. 1h, i). These results suggest that chromatin scaffolds other than H3K9me3 are involved in Rhino recruitment in the OSC genome.

Rhino co-localizes with asymmetric dimethylarginine histones on the OSC genome

To identify genomic scaffolds that are potentially involved in determining the genomic localization of Rhino in OSCs, we performed ChIP-seq of key histone modifications and compared the mapping data with Rhino ChIP-seq data. Heatmap showed that ADMA-histones, such as H3R2me2a, H3R8me2a, H3R17me2a, H3R26me2a, and H4R3me2a, tend to co-localize with Rhino in the OSC genome (Fig. 2a). Consistent with this, ADMA-histones showed a tendency to localize to the ends of Rhino-bound dual-strand piRNA clusters (Fig. 2b). Examination of the correlation coefficients revealed that signal peaks of each ADMA-histone were significantly correlated with those of FLAG-Rhino and H3K9me3 (Fig. 2c). Involvement of ADMA-histones in the genomic localization of Rhino was implicated.

The overlaps of ChIP-seq peaks were examined. In the diagram shown in Fig. 1e, approximately 23% of Rhino ChIP peaks did not overlap with H3K9me3 (187 in 823 sites), but these Rhino peaks were almost encapsulated by the read sets of ADMA-histones (Fig. 2d). This suggests that ADMA-histones are conjugated with H3K9me3 to define the initial genomic loading of Rhino in OSCs. Each ADMA-histone showed a tendency to localize to the same genomic annotations as Rhino localization (Fig. 1h, i, Extended Data Fig. 2a, b). However, the degree of overlap of those sites is not very high in each case (Extended Data Fig. 2c). It appears that each ADMA-histone contributes differently to Rhino localization.

Unlike in the case of Rhino, genomic localization of HP1a, another member of the HP1 family essential for constitutive heterochromatin, does not appear to be determined by ADMA-histones and H3K9me3 (Extended Data Fig. 2d). These results suggest that ADMA-histones, together with H3K9me3, contribute significantly and specifically to the recruitment of Rhino to the ends of dual-strand clusters in OSCs.

Arginine methyltransferases DART1 and DART4 differently regulate Rhino genomic binding

Arginine methylation of histones is catalyzed by protein arginine methyltransferases (PRMTs), and the *Drosophila* genome encodes 11 of these PRMTs, also known as *Drosophila* arginine methyltransferases (DARTs)[25,26]. Of those, DART1, a homolog of mammalian PRMT1, and DART4/CARMER (DART4), a homolog of mammalian PRMT4/CARM1 (CARM1), were reported to be type I PRMTs that catalyze the methylation of monomethylarginine to asymmetric dimethylarginine[27,28,29,30,31,32,33,34]. DART8 is involved in asymmetric dimethylation of H3R2[35]. CG17726 is considered to be a DART member because of its sequence similarity and is expressed in the ovary (Flybase 2024). However, its function remains unknown. If ADMA-histones play an important role in the local recruitment of Rhino to the OSC genome, it is highly likely that the loss of DARTs responsible for ADMA modification would influence the subcellular localization of Rhino.

A previous study showed that HP1a germline knockdown (GLKD) causes Rhino foci to be more aggregated in the ovary and reduces Rhino occupancy at the clusters[36]. Because a similar phenotype was observed for Rhino foci in HP1a-knockdown (KD) OSCs (Fig. 3a), we used this phenotype to examine the loss-of-function effects of DARTs. Similar phenotypes were observed in DART1- and DART4-KD OSCs as in HP1a-KD OSCs, suggesting that DART1/4-dependent histone ADMA modification is involved in Rhino recruitment (Fig. 3a, b, Extended Data Fig. 3a, b). In contrast, this phenotype was not observed for HP1a, suggesting that the effects of DART1/4 are specific to Rhino (Extended Data Fig. 3c).

Fluctuation in the localization of Rhino in the DART1- and DART4-KD OSCs was also observed in ChIP experiments (Fig. 3c–g). In ChIP-seq, Rhino tended to be decreased in DART1 KD and DART4 KD (Fig. 3c, d), which was confirmed by ChIP-qPCR (Fig. 3e). Among the ADMA-histones that co-localized with Rhino, H3R17me2a decreased in a DART4-dependent manner and H4R3me2a decreased in a DART1-dependent manner (Fig. 3c), consistent with previous findings for human PRMT1 and CARM1[37,38,39,40]. Detailed analysis revealed that the genomic localizations of H4R3me2a lost in DART1 KD (575 sites) and H3R17me2a lost in DART4 KD (653 sites) overlapped at 48 sites (Fig. 3f). Correspondingly, the genomic localizations of Rhino that decreased in DART1 KD (115 sites) and those that decreased in DART4 KD (207 sites) overlapped at 25 sites (Fig. 3g). These results suggest that DART1 and DART4 are involved in Rhino recruitment at distinct genomic sites through the decreases in ADAM-histones in each of their KD conditions (H4R3me2a and H3R17me2a, respectively).

Among ADMA-histones, only H3R8me2a tended to decrease in both DART1 and DART4 KD (Fig. 3c). To examine the specificity of the regulation of H3R8me2a by DART1 and DART4, the degree of overlap was visualized in Venn diagrams for H3R8me2a, as well as H3R17me2a and H4R3me2a. The genomic localizations of H3R8me2a decreased by DART1 KD (235 sites) and DART4 KD (58 sites) weakly overlapped (8 sites) (Extended Data Fig. 3d). The genomic localizations of H3R8me2a (58 sites) and H3R17me2a (652 sites) that decreased in DART4 KD overlapped only at 7 sites (Extended Data Fig. 3e), as was the case for H3R8me2a (235 sites) and H4R3me2a (575 sites), which were decreased in DART1 KD (Extended Data Fig. 3f). These results suggest that DART1 and DART4 regulate Rhino localization through H3R8me2a at different sites, as well as through H3R17me2a and H4R3me2a.

FLAG-Rhino expression was slightly decreased in DART1/4-KD cells, but only to the extent that it did not cause Rhino foci destabilization (Extended Data Fig. 1a, b, Extended Data Fig. 3g). Western blotting with anti-ADMA antibodies showed that Rhino itself is not arginine-methylated, suggesting that the effect on Rhino foci of the lack of DART1/4 is mediated through histone arginine methylation (Extended Data Fig. 4).

ADMA-histones co-localize with Rhino in a manner dependent of DART4 in the ovary

We next examined the localization of ADMA-histones in the Oregon-R ovary. In the wild-type (WT) ovary, we detected weak Rhino peaks in FLAG-Rhino localization sites in OSCs and ADMA-histones co-localized with them (Fig. 1b, Fig. 4a, b, Extended Data Fig. 5a). Localization of ADMA-histones in the ovary was restricted to the narrow Rhino-positive regions, as observed in OSCs, and did not co-localize with broad Rhino-containing piRNA clusters (Fig. 4b).

In Kdm3 GLKD, Rhino spreading occurred, and ectopic piRNA clusters appeared as H3K9me2 increased[13]. To confirm that ADMA-histones function as the spreading origins of the piRNA clusters, we examined their localization at the ends of these ectopic piRNA clusters. The ADMA-histones were localized near the 5′ and 3′ ends of Kdm3 GLKD-dependent ectopic piRNA clusters, as well as in the general piRNA clusters (Fig. 4c, d, Extended Data Fig. 5b). In addition, ADMA-histones were not localized inside the clusters, even in the Kdm3 GLKD-dependent ectopic piRNA clusters (Fig. 4c, d). These results support the notion that ADMA-histones define initial Rhino loading prior to the spread of Rhino.

To investigate whether the destabilization of initial Rhino loading in DART4-KD

OSCs can also be observed in the ovary, we performed GLKD of DART4 using the GAL4-UAS system. Rhino expression was not reduced under the DART4 GLKD conditions (Extended Data Fig. 6a). However, Rhino foci were destabilized (Fig. 5a, b, Extended Data Fig. 6b). ChIP-seq experiments showed that H3R17me2a was lost in the ovary and that the initial loading of Rhino was significantly decreased (Fig. 5c–f, Extended Data Fig. 6c, d). These results indicate that the relationship between ADMA-histones and Rhino found in OSCs is conserved in the ovary.

Interestingly, Rhino spreading across clusters annotated in the proTRAC database[41], such as 42AB, 38C, 80F, and 102F, was little affected by DART4 GLKD (Fig. 5d, f, Extended Data Fig. 6e). These clusters may have reached the stable phase by Kipferl or the memory phase by maternally inherited piRNAs.

Initial-stage piRNA clusters (INSECTs) are the genomic sites from which Rhino spreads

Counting the genomic sites where H3R17me2a was no longer localized in DART4-GLKD ovaries, 3,082 sites were identified. The genomic sites where Rhino was decreased under the same circumstance were also counted (2,078 sites). Of these, 497 sites overlapped in the two groups (Fig. 6a). This indicates that 76.1% of the genomic localization of Rhino decreased in DART4 GLKD are H3R17me2a-independent. Genomic snapshot shows representative regions where loss of broad Rhino localization does not depend on the loss of H3R17me2a (Fig. 6b, Extended Data Fig. 7a). These regions had H3R17me2a at their ends and Rhino localization in their internal regions. However, in DART4 GLKD, loss of H3R17me2a at their ends resulted in concomitant loss of Rhino localization in their internal regions (Fig. 6c, d). Forty-three such regions

were found in the *Drosophila* genome (Fig. 6e). We named these regions initial-stage piRNA clusters (hereinafter INSECTs).

INSECTs tend to be longer than the initial loading regions of FLAG-Rhino in OSCs and shorter than the regular piRNA clusters (Fig. 6f). In addition, piRNAs were mapped on the INSECTs, but in small amounts (Fig. 6g). The proportion of piRNAs mapped to transposons was also small (Extended Data Fig. 7b); therefore, INSECTs are considered to be “precursory” clusters from a functional perspective. Not only H3R17me2a, but also other ADMA-histones were found to localize to the ends of INSECTs (Extended Data Fig. 8a). H3K9me2 accumulated in the internal regions (Extended Data Fig. 8a), consistent with previous studies implying H3K9me2 involvement in Rhino spread[13]. These results indicate that ADMA-histones function as Rhino spread origins and that INSECTs are precursory piRNA clusters from which Rhino spreads, but stabilization has not yet been reached. FLAG-Rhino also localized at the ends of INSECTs (Fig. 6d). This again indicates that OSCs are suitable materials for characterizing the initial loading step.

Kipferl stabilizes the Rhino–genome binding after the spread of Rhino[12]. ChIP-seq using anti-Kipferl antibodies showed that Kipferl co-localizes with Rhino in INSECTs and decreases in accordance with Rhino loss in DART4 GLKD (Extended Data Fig. 8b, c). This suggests that Kipferl is already localized even in DART4-dependent INSECTs but not sufficient to stabilize Rhino spread. Thus, the existence of an unknown factor(s) that stabilizes Rhino in “mature” piRNA clusters, but not in INSECTs, is suggested. These results also raise the possibility that Kipferl may also be involved in the spread of Rhino. However, low levels of Kipferl binding were observed in some INSECTs (Extended Data Fig. 8d); thus, this possibility does not apply to all INSECTs, even if it is likely to be.

Based on these results, we propose that the process of piRNA cluster formation begins with the initial loading of Rhino onto bivalent nucleosomes containing H3K9me3 and ADMA-histones (Fig. 7). In OSCs, the absence of Kipferl and other necessary factors means that Rhino loading into the genome does not proceed to the next step. In the ovaries, however, Kipferl is present, and it stabilizes Rhino within the cluster as it spreads out into the genomic regions. In INSECTs, Rhino spreads but does not stabilize. In the ovary, the positional information of H3K9me3 and Rhino is conserved by the transgenerational inheritance of piRNAs in most of the piRNA clusters, reinforcing the idea that Rhino finally becomes the “mature” piRNA clusters.

Discussion

In this study, we took advantage of cultured OSCs for our analysis and found that chromatin marks (i.e., ADMA-histones) define the loading of Rhino onto the dual-stranded clusters. Comparative analysis of OSC and ovarian ChIP-seq data also allowed us to identify precursory piRNA clusters INSECTs. These findings led us to propose a model (Fig. 7) in which ADMA-histone-dependent Rhino loading onto the genome is the first step in the piRNA cluster formation process, followed by Rhino spreading. ADMA-histones were predominantly found in euchromatin rather than pre-existing piRNA clusters (Extended Fig. 2a, b), suggesting that they serve as a preparatory step for the *de novo* formation of piRNA clusters when a novel transposon insertion event occurs in a neighboring genomic region. In general, ADMA-histones are known to function as transcriptional coactivators[42]. In mammalian cells, H3R8me2a does not co-localize with H3K9me3, but with enhancer modifications such as H3K4me1 and H3K27ac[43]. This may indicate that the function of ADMA-histones has evolved in a *Drosophila*-specific manner,

which is not unlikely given that Rhino itself is *Drosophila*-specific.

This study showed that multiple ADMA-histones, such as H3R17me2a and H4R3me2a, play a role in the initial Rhino loading. This suggests that a co-factor(s) with an affinity for ADMAs might be involved. Tudor domains are known to bind partner proteins through methylation of arginine and lysine residues[44], and Tudor domain-containing protein 3 (TDRD3) was identified as a factor that specifically binds to H3R17me2a and H4R3me2a in humans[45]. Therefore, TDRD protein family members may be involved in Rhino loading via ADMAs. ADMAs of other histone arginine residues such as H3R42 may also be involved in the regulation of Rhino behavior[42]. In this study, ChIP-seq was not performed for these modifications due to the lack of commercially available antibodies.

The mechanism by which ADMA-histones is involved in Rhino spreading remains unclear. Because Rhino spreading has been shown to be correlated with transcription[7,10], one interpretation is that ADMA-histones may affect Rhino spreading via transcriptional activity. Insertion of insulators into piRNA clusters inhibits piRNA production[46], indicating that the three-dimensional (3D) structure of the genome within the piRNA cluster also affects the extent of Rhino spreading. ADMA-histones were found at the ends of piRNA clusters, reminiscent of the boundaries of topologically associated domains. An unknown stabilizing factor(s) suggested by the absence of Kipferl in some DART4-dependent INSECTs (Extended Date Fig. 8d) may be a relevant factor(s) determining such genomic 3D structures.

Maternally inherited siRNAs initiate piRNA biogenesis from a transgene inserted into the uni-strand piRNA cluster 20A[14]. In this previous report, evidence was presented that siRNAs induce "*de novo*" Rhino loading onto this transgene (20A-X) via H3K9me3.

This led us to hypothesize that ADMA-histones should be present in the vicinity of 20A-X. Indeed, we detected ADMA-histones in the region and Rhino, albeit weakly (Extended Data Fig. 5a). Thus, this study provided experimental evidence in support of the "*de novo* model" by reinforcing the findings of siRNA induced piRNA biogenesis. This model has been becoming supported by evolutionary analyses[47,48,49], but experimental results had been limited.

In *Drosophila*, major piRNA clusters such as 42AB and 38C are located at the boundary between heterochromatin and euchromatin[4,50]. It was unclear why these "mature" clusters, which are capable of producing substantial piRNAs, tend to appear in regions where two different types of chromatins collide. Our model can answer this question by showing that cluster formation begins with initial Rhino loading by ADMA-histones onto euchromatic regions and then spreads to adjacent regions of H3K9me2-rich heterochromatin (Fig. 7). Rhino shows a tendency to spread from an INSECT on chromosome 2 to heterochromatin/euchromatin boundary regions where 42AB and 38C are present (Fig. 6e), supporting our hypothesis.

Methods

Establishment and cloning of FLAG-Rhino inducible OSCs

To construct Tet-ON 3G inducible FLAG-Rhino plasmid, cDNA encoding FLAG-Rhino was cloned into pPB-2× Ty1-Tjen-EGFP-P2A-BlastR[51] using In-Fusion cloning (TAKARA), bearing pPB-FLAG-Rhino-Tjen-EGFP-P2A-BlastR. The Actin promoter and *EGFP* cDNA were replaced respectively with the TRE3G promoter and *rtTA* cDNA from inducible Caspex expression plasmid (Addgene). PCR was performed using PrimeSTAR MAX (TAKARA) with primers (Supplementary Table 1). FLAG-Rhino inducible OSCs

were established using the piggyBac system as described previously[51]. In brief, 2 × 10^6 OSCs were transfected with 4 μg of pPB-TRE3G-FLAG-Rhino-Tjen-trTA-P2A-BlastR and 1 μg of pHsp70-Myc-hyPBase using Nucleofector Kit V (Lonza). After incubation in media containing blasticidin (50 μg/mL), colonies were isolated. One of the colonies contained a single FLAG-Rhino copy in the intron of the *jing interacting gene regulatory 1* gene, and this was used in this study.

Cell culture and RNAi

FLAG-Rhino inducible OSCs were cultured in Shields and Sang M3 Insect Medium (Sigma) supplemented with 10% fly extract, 10% fetal bovine serum, 0.6 mg/mL glutathione, and 10 mU/mL insulin[26,52]. To perform RNAi, 400 pmol of siRNA duplex was introduced into 3.0 × 10^6 cells using the Cell Line 96-well Nucleofector Kit SF (Lonza) with program DG150 of the 96-well Shuttle Device (Lonza). After 48 h incubation, cells were transfected with another 400 pmol of siRNA and incubated further for 48 h. For HP1a KD, RNAi was performed once to avoid cell damage. All siRNAs used are shown in Supplementary Table 2.

Fly stock

Flies were raised in standard Bloomington medium at 26 °C. Oregon-R was employed as a WT strain. For DART4 GLKD, BDSC 4937 (nosGAL4, w[1118]; P{w[+mC]=GAL4::VP16-nanos.UTR}CG6325[MVD1]) and BDSC 36833 (y[1] sc[*] v[1] sev[21]; P{y[+t7.7] v[+t1.8]=TRiP.GL01073}attP2) were used as the maternal GAL4 driver and DART4 UAS-RNAi lines, respectively. BDSC 36303 (y[1] v[1]; P{y[+t7.7]=CaryP}attP2) was used as a control (ctrl).

Western blotting

Western blotting was performed as described previously[53]. Antibodies are summarized in Supplementary Table 3.

Immunofluorescence

Immunofluorescence of OSCs was performed essentially as described previously[54]. OSCs were fixed using 4% formaldehyde (Wako) and permeabilized in PBS containing 0.1% Triton X-100 for 24 h at RT. For DAPI staining, VECTASHIELD Mounting Medium with DAPI (Vector Laboratories) was used. Immunofluorescence of ovaries was performed as described previously[10]. All images were obtained using a confocal laser scanning microscope (Olympus FV3000 or Carl Zeiss LSM 710). For image processing, ZEN software (Carl Zeiss) was used. Antibodies used are listed in Supplementary Table 3.

Generation of mouse anti-Rhino monoclonal antibodies

Anti-Rhino monoclonal antibodies were raised against N-terminal 200 residues of Rhino using a protocol published previously[55]. The N-terminal 200 residues of Rhino were fused to glutathione S-transferase (GST) and the purified fusion protein was used as antigen to immunize mice. The expression vector was produced by inserting the cDNA encoding the Rhino fragment into pGEX-5X-3 (Cytiva) by Infusion (Takara). The primers used are listed in Supplementary Table 1.

RNA extraction and RT-qPCR

Total RNAs were extracted from OSCs or ovaries with the ISOGEN II reagent (Nippon Gene) and treated with DNase (Life Technologies) to eliminate DNA contaminations. The cDNA libraries were prepared using the Transcriptor First Strand cDNA Synthesis Kit (Roche). For RT-qPCR, cDNAs were amplified using the StepOnePlus Real-Time PCR System (Applied Biosystems) with PowerUp SYBR Green Master Mix (Thermo Fisher Scientific). The expression level of *Rhino, HP1a, Kipferl, DART1, DART4, DART8*, and *CG17726* in OSCs and ovary were calculated by ΔΔCt method using *ribosomal protein 32* (*RP49/RpL32*) as an internal control. Data are presented as mean values ± SE. The primers used are listed in Supplementary Table 1.

OSC ChIP

OSCs were fixed in PBS containing 1% formaldehyde for 10 min at RT and glycine was added to 150 mM to stop the fixation. Cells were collected and suspended in swelling buffer [25 mM HEPES-KOH (pH 7.3), 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM DTT, 0.1% NP-40, supplemented with cOmplete ULTRA Tablets, EDTA-free (Roche)]. The nuclei were collected by centrifugation at 1,000 g for 5 min at 4°C, suspended in sonication buffer [50 mM HEPES-KOH (pH 7.4), 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, supplemented with cOmplete ULTRA Tablets, EDTA-free], and sonicated using a Covaris S220 Focused ultrasonicator for 10 min at 4°C with setting of peak power of 200, duty factor of 10.0, and 200 cycles/burst. The chromatin fraction was obtained by centrifugation at 20,000 g for 15 min at 4°C and diluted with sonication buffer to make the DNA concentration to 0.2 mg/mL. Immunoprecipitation was performed O/N at 4°C using 1mL of chromatin fraction per ChIP. Antibodies used are listed in Supplementary Table 3. The reaction mixture was mixed with 25 μL of

Dynabeads ProteinG (Invitrogen) and incubated for 1 h at 4°C. Beads were washed one time each with low-salt buffer [20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM EDTA, 0.1% Triton X-100, 0.1% SDS, and 1 mM phenylmethylsulfonyl fluoride (PMSF)], high-salt buffer [20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 2 mM EDTA, 0.1% Triton X-100, 0.1% SDS, and 1 mM PMSF], and LiCl buffer [20 mM Tris-HCl (pH 8.0), 250 mM LiCl, 2 mM EDTA, 1% NP-40, 1% sodium deoxycholate, and 1 mM PMSF], and then twice with TE buffer [20 mM Tris-HCl (pH 8.0), 1 mM EDTA, and 1 mM PMSF]. The beads were suspended in elution buffer [50 mM Tris-HCl (pH 8.0), 10 mM EDTA, and 1% SDS] and incubated at 65°C for 30 min. For de-crosslinking, the supernatant was incubated O/N at 65°C after supplement of NaCl to 200mM. RNAs and proteins were eliminated from the sample by incubating at 37°C for 30 min with 10 mg/mL RNase A and 20 mg/mL proteinase K at 55°C for 1 h. The DNA in the supernatant was purified by EconoSpin column IIa (Funakoshi). 50 μL of nuclease-free water was added to it and the sample was centrifuged at 13,000 rpm for 1 min to obtain ChIP product.

Ovary ChIP

150 ovary pairs were crosslinked in PBS containing 1% formaldehyde for 10 min at RT, quenched with 180 mM glycine, and washed twice in PBS. Ovaries were flash frozen and suspended with swelling buffer (see “OSC ChIP”). Ovaries were disrupted using a Dounce homogenizer and incubated at 4°C for 10 min. The nuclei were collected by centrifugation at 1,000 g for 5 min at 4°C, suspended in sonication buffer, and sonicated (see “OSC ChIP”). The process after this is as described above (see “OSC ChIP”). Antibodies used are listed in Supplementary Table 3.

ChIP-seq library preparation

ChIP-seq libraries were constructed using NEBNext Ultra II FS DNA Library Prep Kit for Illumina (New England BioLabs) using ChIP products from “OSC ChIP” and “Ovary ChIP”. In the adaptor ligation step, the original adaptor reagent was diluted 10-fold. Adapter ligated DNAs were PCR-amplified (10–12 cycles) with Illumina primers. The PCR products of 100–1,000 bp (including adaptor and primer sequences) were purified using AMPure XP beads (Beckman Coulter) and captured on an Illumina flow cell. The libraries were sequenced using the Illumina MiSeq.

Computational analysis of ChIP-seq

Prior to mapping, low-quality reads were removed and adapters were trimmed using Fastp[56]. The filtered reads were mapped to the genome of *Drosophila melanogaster* (dm6) using Bowtie 2[57]. Only unique mapped reads were used in the analysis using samtools view command option -q 42[58]. For genome browser view of piRNA cluster, reads were extracted that map to multiple locations without this option applied because piRNA cluster contains multiple transposons. We performed all ChIP-seq experiment using two biological replicates. The replicates were highly correlated, so we merged them together and used the merged libraries for analyses. Because of the low mapping score, four replicates were prepared for Rhino and three for Kipferl in the GLKD condition. Peakcall was completed with MACS2 with q = 0.01 settings using input ChIP-seq data as control unless otherwise noted in figure legends[59]. For FLAG-Rhino peakcall in OSCs, anti-FLAG ChIP-seq data in Dox = 0 ng/mL condition was used as control. For Rhino peakcall in Oregon-R ovary in Fig. 1c and Fig. 4b, and H3K9me3 peakcall in OSCs in Fig. 1g, options --broad --broad-cutoff 0.01 were used. For heatmaps and profile plots,

ChIP scores were calculated by deeptools[60]. Bigwig files of ChIP were produced by the bamCoverage / bamCompare program of deeptools and matrix data were calculated by the computeMatrix program of deeptools using bigwig files and plotted. For bamCompare program, anti-FLAG ChIP data with Dox = 0 (ng/mL) was subtracted as control for FLAG-Rhino in OSCs and input for the other data. Mapping file correlation and hierarchical clustering were computed using multibamSummary programs of deeptools. To draw the Venn diagram, the narrowPeak file obtained with Macs2 callpeak was stored in the findOverlapsOfPeaks function in R's ChIPpeakanno library and visualized with makeVennDiagram program[61]. The manorm was used for quantitative comparison of called peaks[62].

ChIP-qPCR

Real-time qPCR was performed using PowerUp SYBR Green Master Mix and StepOnePlus real-time PCR system. Data are presented as mean values ± SE. The primers used are listed in Supplementary Table 2.

Definition of DART4-dependent INSECTs

DART4-dependent INSECTs were defined using manorm results based on the macs2 callpeak with option p = 0.01 and --broad --broad cutoff 0.1 output files that met the following seven criteria. First, M value > 0.5 is required for Rhino to be predominantly enriched against DART4 GLKD. Second, the sharp peaks of the mononucleosome units were excluded because of their non-spreading initial Rhino localization. Third, peaks overlapped with DART4-dependent H3R17me2a were excluded because these peaks cannot distinguish spread Rhino and initial Rhino loading. Fourth, each Rhino peak was

checked in the genome browser, and if there were consecutive Rhino peaks around it that met the first, second and third criteria, they were grouped together as one peak. Fifth, peaks that overlap with major piRNA clusters such as 42AB, 38C, 80F, and 102F were excluded because they are “mature” piRNA clusters. Sixth, peaks on subtelomeres and pericentromeres were excluded because ChIP-seq reads cannot distinguish piRNA clusters or repeats. Finally, peaks whose localization spanned the entire length of a single transposon were excluded because they may reflect genomic differences between ctrl GLKD and DART4 GLKD, such as copy number variation of transposons. The combined peaks that met these seven criteria were defined as single INSECT together with the surrounding contiguous ctrl GLKD-specific Rhino localizations if they were more than 1 kb. The genomic location of INSECTs in Fig. 6d was visualized using Phenogram software[63]. The regions of INSECTs are listed in Supplementary Table 4.

Calculation of piRNA RPKM in piRNA clusters

GSM1580790 and GSM1580793 datasets[64] were used as piRNA-seq data in the ovaries of ctrl GLKD and Rhino GLKD, respectively. Raw reads were adapter trimmed and the random 4 bases of each side were removed using cutadapt[65]. Trimmed reads were mapped to the dm6 genome using bowtie[66]. For Fig. 6g, only unique mapped reads were used in the analysis using samtools view command option -q 42[58]. For read count in Extended Data Fig. 7b, this option was excluded. Read counts for each piRNA cluster were performed using bedtools[67].

Data availability

Sequencing data generated in this study have been deposited in the Sequence Read

Archive (SRA) under accession number PRJNAXXXXXXX and the Gene Expression Omnibus (GSE) under accession number GSEXXXXXX. All other data supporting the findings of this study are available from the corresponding authors upon request.

Acknowledgments

We are grateful to all members of the Siomi laboratory for discussions and comments on this work, especially K. Sato. We also thank S. Yamanaka, T. Kakutani, and S. Inagaki for critical reading of the manuscript, Y. Koga for technical assistance with *Drosophila* genetics, and A. Taguchi for support with the Illumina sequencing. We are also grateful to J. Brennecke and L. Tirian for antibodies and Harvard TRiP and Bloomington stock centers for *Drosophila* flies. This study was supported by MEXT KAKENHI Grant Number JP19H05466 (to M.C.S.), JSPS KAKENHI Grant Number 21K20629 and 23K14232 (to R.S.), and JSPS KAKENHI Grant Numbers 17K08644 and 20H03439 (to K.M.).

Author contributions

R.S. and R.H. conducted biochemical experiments. R.S. performed bioinformatic analysis. H.I. produced FLAG-Rhino inducible OSCs. K.M. produced anti-Rhino antibodies. R.S. and M.C.S. conceived the project and designed the experiments. R.S. and M.C.S. wrote the manuscript with input from all authors.

Disclosure and competing interest statement

The authors declare that they have no conflicts of interest.

References

1. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. *Nat Rev Genet* 10,94-108 (2009)
2. Iwasaki YW, Siomi MC, Siomi H. PIWI-Interacting RNA: Its Biogenesis and Functions. *Annu Rev Biochem* 84,405-33 (2015).
3. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet* 20,89-108 (2019).
4. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128,1089-103 (2007).
5. Klattenhoff C, Xi H, Li C, Lee S, Xu J, Khurana JS, Zhang F, Schultz N, Koppetsch BS, Nowosielska A, Seitz H, Zamore PD, Weng Z, Theurkauf WE. The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* 138,1137-49 (2009).
6. Mohn F, Sienski G, Handler D, Brennecke J. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell* 157,1364-1379 (2014).
7. Zhang Z, Wang J, Schultz N, Zhang F, Parhad SS, Tu S, Vreven T, Zamore PD, Weng Z, Theurkauf WE. The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell* 157,1353-1363 (2014).
8. Yu B, Cassani M, Wang M, Liu M, Ma J, Li G, Zhang Z, Huang Y. Structural insights into Rhino-mediated germline piRNA cluster formation. *Cell Res* 25,525-8 (2015).
9. Akkouche A, Mugat B, Barckmann B, Varela-Chavez C, Li B, Raffel R, Pélisson A, Chambeyron S. Piwi Is Required during *Drosophila* Embryogenesis to License Dual-Strand piRNA Clusters for Transposon Repression in Adult Ovaries. *Mol Cell* 66,411-419 (2017).
10. Andersen PR, Tirian L, Vunjak M, Brennecke J. A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature* 549,54-59 (2017).
11. Yu B, Lin YA, Parhad SS, Jin Z, Ma J, Theurkauf WE, Zhang ZZ, Huang Y. Structural insights into Rhino-Deadlock complex for germline piRNA cluster specification. *EMBO Rep* 19,e45418 (2018).
12. Baumgartner L, Handler D, Platzer SW, Yu C, Duchek P, Brennecke J. The *Drosophila* ZAD zinc finger protein Kipferl guides Rhino to piRNA clusters. *Elife* 11,e80067 (2022).
13. Casier K, Autaa J, Gueguen N, Delmarre V, Marie PP, Ronsseray S, Carré C,

Brasset E, Teysset L, Boivin A. The histone demethylase Kdm3 prevents auto-immune piRNAs production in *Drosophila*. *Sci Adv* 9, eade3872 (2023).
14. Luo Y, He P, Kanrar N, Fejes Toth K, Aravin AA. Maternally inherited siRNAs initiate piRNA cluster formation. *Mol Cell* 83,3835-3851 (2023).
15. Volpe AM, Horowitz H, Grafer CM, Jackson SM, Berg CA. *Drosophila* rhino encodes a female-specific chromo-domain protein that affects chromosome structure and egg polarity. *Genetics* 159,1117-34 (2001).
16. Vermaak D, Henikoff S, Malik HS. Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in *Drosophila*. *PLoS Genet* 1,196-108 (2005).
17. Pane A, Jiang P, Zhao DY, Singh M, Schüpbach T. The Cutoff protein regulates piRNA cluster expression and piRNA production in the *Drosophila* germline. *EMBO J* 30,4601-15 (2011).
18. Czech B, Preall JB, McGinn J, Hannon GJ. A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway. *Mol Cell* 50,749-61 (2013).
19. Chen YA, Stuwe E, Luo Y, Ninova M, Le Thomas A, Rozhavskaya E, Li S, Vempati S, Laver JD, Patel DJ, Smibert CA, Lipshitz HD, Toth KF, Aravin AA. Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. *Mol Cell* 63,97-109 (2016).
20. Pritykin Y, Brito T, Schupbach T, Singh M, Pane A. Integrative analysis unveils new functions for the *Drosophila* Cutoff protein in noncoding RNA biogenesis and gene regulation. *RNA* 23,1097-1109 (2017).
21. Le Thomas A, Stuwe E, Li S, Du J, Marinov G, Rozhkov N, Chen YC, Luo Y, Sachidanandam R, Toth KF, Patel D, Aravin AA. Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev* 28,1667-80 (2014).
22. Saito K, Inagaki S, Mituyama T, Kawamura Y, Ono Y, Sakota E, Kotani H, Asai K, Siomi H, Siomi MC. A regulatory circuit for piwi by the large Maf gene traffic jam in *Drosophila*. *Nature* 461,1296-9 (2009).
23. Sumiyoshi T, Sato K, Yamamoto H, Iwasaki YW, Siomi H, Siomi MC. Loss of l(3)mbt leads to acquisition of the ping-pong cycle in *Drosophila* ovarian somatic cells. *Genes Dev* 30,1617-22 (2016).
24. Sato K, Siomi MC. The piRNA pathway in *Drosophila* ovarian germ and somatic cells. *Proc Jpn Acad Ser B Phys Biol Sci* 96,32-42 (2020)
25. Wang YC, Li C. Evolutionarily conserved protein arginine methyltransferases in

non-mammalian animal systems. *FEBS J* 279,932-45 (2012).
26. Wang YC, Wang JD, Chen CH, Chen YW, Li C. A novel BLAST-Based Relative Distance (BBRD) method can effectively group members of protein arginine methyltransferases and suggest their evolutionary relationship. *Mol Phylogenet Evol* 84,101-11 (2015).
27. Boulanger MC, Miranda TB, Clarke S, Di Fruscio M, Suter B, Lasko P, Richard S. Characterization of the *Drosophila* protein arginine methyltransferases DART1 and DART4. *Biochem J* 379,283-9 (2004).
28. Scaramuzzino C, Monaghan J, Milioto C, Lanson NA Jr, Maltare A, Aggarwal T, Casci I, Fackelmayer FO, Pennuto M, Pandey UB. Protein arginine methyltransferase 1 and 8 interact with FUS to modify its sub-cellular distribution and toxicity *in vitro* and *in vivo*. *PLoS One* 8,e61576 (2013).
29. Xu J, Wang AH, Oses-Prieto J, Makhijani K, Katsuno Y, Pei M, Yan L, Zheng YG, Burlingame A, Brückner K, Derynck R. Arginine Methylation Initiates BMP-Induced Smad Signaling. *Mol Cell* 51,5-19 (2013).
30. Jäckel S, Summerer AK, Thömmes CM, Pan X, Voigt A, Schulz JB, Rasse TM, Dormann D, Haass C, Kahle PJ. Nuclear import factor transportin and arginine methyltransferase 1 modify FUS neurotoxicity in *Drosophila*. *Neurobiol Dis* 74,76-88 (2015).
31. Yuan Q, Tian R, Zhao H, Li L, Bi X. Multiple Arginine Residues Are Methylated in *Drosophila* Mre11 and Required for Survival Following Ionizing Radiation. *G3* 8,2099-2106 (2018).
32. Chen D, Ma H, Hong H, Koh SS, Huang SM, Schurter BT, Aswad DW, Stallcup MR. Regulation of transcription by a protein methyltransferase. *Science* 284,2174-7 (1999).
33. Cakouros D, Daish TJ, Mills K, Kumar S. An arginine-histone methyltransferase, CARMER, coordinates ecdysone-mediated apoptosis in *Drosophila* cells. *J Biol Chem* 279,18467-71 (2004).
34. Urwyler O, Zhang L, Li X, Imboden H, Suter B. Tissue-dependent subcellular localization of *Drosophila* arginine methyl-transferase 4 (DART4), a coactivator whose overexpression affects neither viability nor differentiation. *Differentiation* 75,757-65 (2007).
35. Villar-Garea A, Forne I, Vetter I, Kremmer E, Thomae A, Imhof A. Developmental regulation of N-terminal H2B methylation in *Drosophila* melanogaster. *Nucleic Acids Res* 40,1536-49 (2012).
36. Teo RYW, Anand A, Sridhar V, Okamura K, Kai T. Heterochromatin protein 1a

functions for piRNA biogenesis predominantly from pericentric and telomeric regions in *Drosophila*. *Nat Commun* 9,1735 (2018).

37. Schurter BT, Koh SS, Chen D, Bunick GJ, Harp JM, Hanson BL, Henschen-Edman A, Mackay DR, Stallcup MR, Aswad DW. Methylation of histone H3 by coactivator-associated arginine methyltransferase 1. *Biochemistry* 40,5747-56 (2001).
38. Miao F, Li S, Chavez V, Lanting L, Natarajan R. Coactivator-associated arginine methyltransferase-1 enhances nuclear factor-kappaB-mediated gene transcription through methylation of histone H3 at arginine 17. *Mol Endocrinol* 20,1562-73 (2006).
39. Wang H, Huang ZQ, Xia L, Feng Q, Erdjument-Bromage H, Strahl BD, Briggs SD, Allis CD, Wong J, Tempst P, Zhang Y. Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor. *Science* 293,853-7 (2001).
40. Huang S, Litt M, Felsenfeld G. Methylation of histone H4 by arginine methyltransferase PRMT1 is essential *in vivo* for many subsequent histone modifications. *Genes Dev* 19,1885-93 (2005).
41. Rosenkranz D, Zischler H. proTRAC--a software for probabilistic piRNA cluster detection, visualization and analysis. *BMC Bioinformatics* 13,5 (2012).
42. Blanc RS, Richard S. Arginine Methylation: The Coming of Age. *Mol Cell* 65,8-24 (2017).
43. Dong F, Li Q, Yang C, Huo D, Wang X, Ai C, Kong Y, Sun X, Wang W, Zhou Y, Liu X, Li W, Gao W, Liu W, Kang C, Wu X. PRMT2 links histone H3R8 asymmetric dimethylation to oncogenic activation and tumorigenesis of glioblastoma. *Nat Commun* 9,4552 (2018).
44. Šimčíková D, Gelles-Watnick S, Neugebauer KM. Tudor-dimethylarginine interactions: the condensed version. *Trends Biochem Sci* 48,689-698 (2023).
45. Yang Y, Lu Y, Espejo A, Wu J, Xu W, Liang S, Bedford MT. TDRD3 is an effector molecule for arginine-methylated histone marks. *Mol Cell* 40,1016-23 (2010).
46. Radion E, Sokolova O, Ryazansky S, Komarov PA, Abramov Y, Kalmykova A. The Integrity of piRNA Clusters is Abolished by Insulators in the *Drosophila* Germline. *Genes (Basel)* 10,209 (2019).
47. Luo S, Zhang H, Duan Y, Yao X, Clark AG, Lu J. The evolutionary arms race between transposable elements and piRNAs in *Drosophila* melanogaster. *BMC Evol Biol* 20,14 (2020).
48. Wierzbicki F, Kofler R. The composition of piRNA clusters in *Drosophila*

melanogaster deviates from expectations under the trap model. *BMC Biol* 21,224 (2023).

49. Srivastav S, Feschotte C, Clark AG. Rapid evolution of piRNA clusters in the *Drosophila* melanogaster ovary. *bioRxiv* Preprint at https://www.biorxiv.org/content/10.1101/2023.05.08.539910v2 (2023).
50. Yamanaka S, Siomi MC, Siomi H. piRNA clusters and open chromatin structure. *Mob DNA* 5,22 (2014).
51. Takeuchi C., Murano K., Ishikawa M., Okano H., Iwasaki W.Y. Generation of stable *Drosophila* ovarian somatic cell lines using piggyBac system. *Methods Mol Biol* 2509,143–153 (2022).
52. Niki Y, Yamaguchi T, Mahowald AP. Establishment of stable cell lines of *Drosophila* germ-line stem cells. *Proc Natl Acad Sci U S A* 103,16325-30 (2006).
53. Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev* 20,2214-22 (2006).
54. Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, Siomi H, Siomi MC. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes Dev* 24,2493-2498 (2010).
55. Murano K, Iwasaki YW, Ishizu H, Mashiko A, Shibuya A, Kondo S, Adachi S, Suzuki S, Saito K, Natsume T, Siomi MC, Siomi H. Nuclear RNA export factor variant initiates Piwi-piRNA-guided co-transcriptional silencing. *EMBO J* 38,e102870 (2019).
56. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34,i884-i890 (2018).
57. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9,357-9 (2012).
58. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25,2078-9 (2009).
59. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9,R137 (2008).
60. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44,W160-5 (2016).

61. Zhu L, Gazin C, Lawson N, Pagès H, Lin S, Lapointe D, Green M (2010). "ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data."*BMC Bioinformatics*, **11**(1), 237. ISSN 1471-2105
62. Shao Z, Zhang Y, Yuan GC, Orkin SH, Waxman DJ. MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biol* 13,R16 (2012).
63. Wolfe D, Dudek S, Ritchie MD, Pendergrass SA. Visualizing genomic information across chromosomes with PhenoGram. BioData Min 6,18 (2013).
64. Mohn F, Handler D, Brennecke J. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science* 348,812-817 (2015).
65. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* 17, 10–12 (2011).
66. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10,R25 (2009).
67. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26,841-2 (2010)

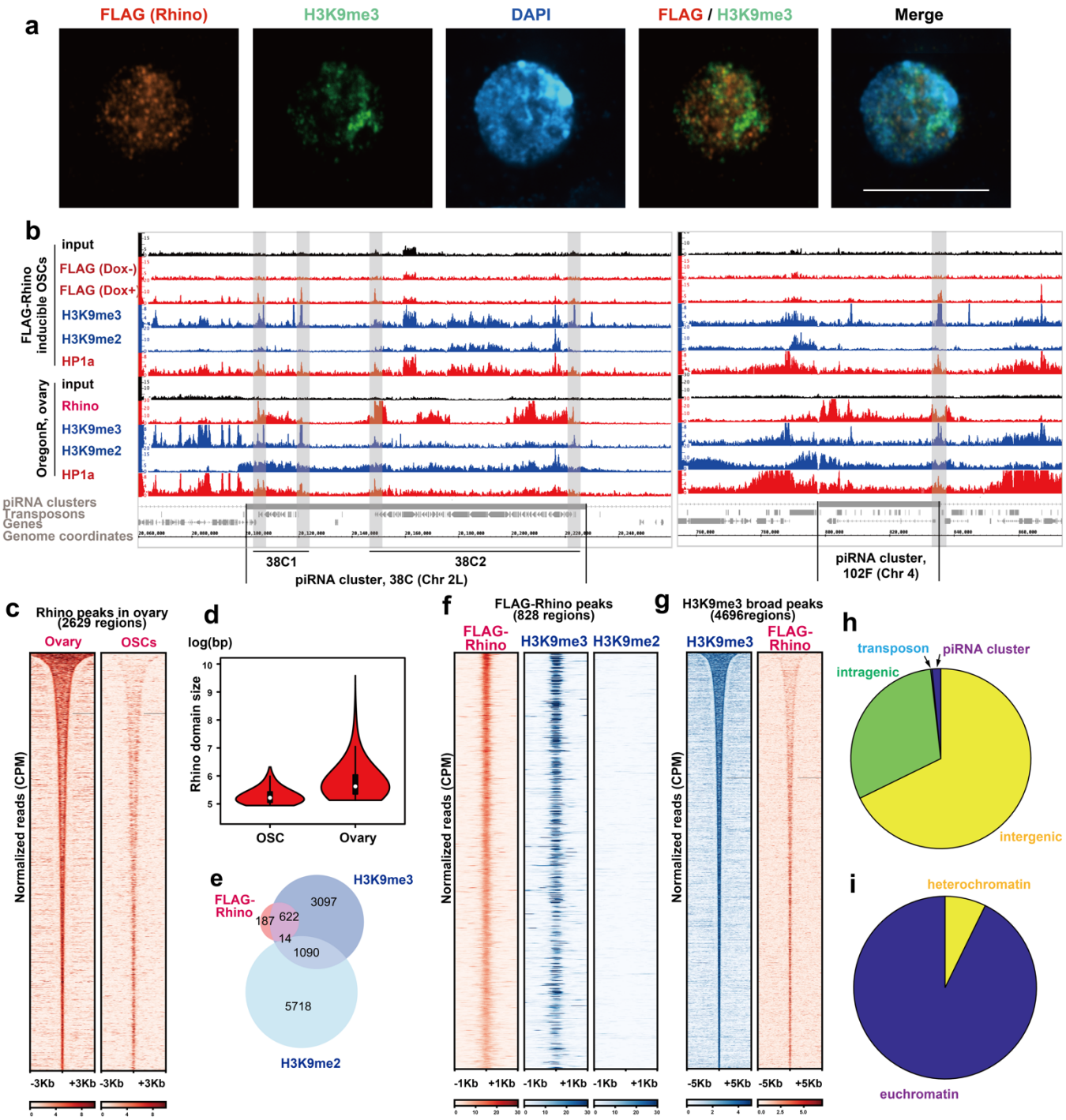
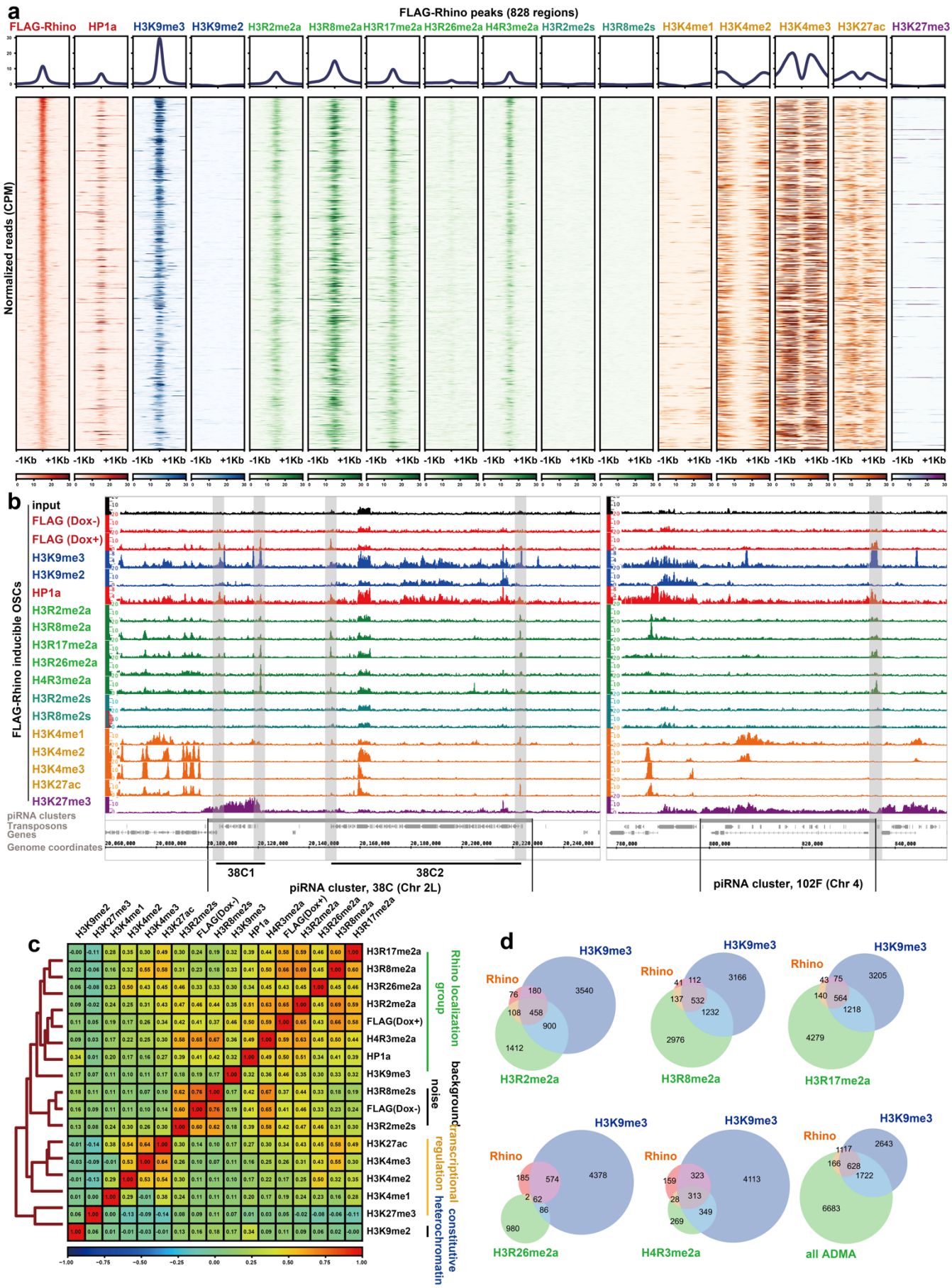


Figure 1. Rhino forms nuclear foci and binds the genome in OSCs

(a) Confocal immunofluorescent images of FLAG-Rhino (red) in FLAG-Rhino inducible OSCs upon Dox induction (5 ng/mL). H3K9me3 and OSC nucleus (DAPI staining) are shown in green and blue. Scale bar represents 10 μm. (b) Genome snapshot showing ChIP-seq in major dual-strand piRNA clusters (38C and 102F). The gray-shaded areas indicate the sites at which FLAG-Rhino is localized in each piRNA cluster in OSCs. piRNA cluster 38C has two piRNA-producing regions, labeled 38C1 and 38C2. (c) Heatmap showing FLAG-Rhino localization of OSCs in Rhino peaks of *Drosophila* wild-

type Oregon-R ovary. The regions of 6 kb around the Rhino localization site were plotted based on macs2 callpeak results. (d) Violin plot comparing Rhino localization units in OSCs and Oregon-R ovary. The box indicates the lower (25%) and upper (75%) quartiles, and the open circle indicates the median. Whiskers extend to the most extreme data points. (e) Venn diagram showing the overlap of Rhino, H3K9me3, and H3K9me2 peaks in OSCs. (f) Heatmap showing the localization of H3K9me3 and H3K9me2 in FLAG-Rhino peaks of OSCs. The regions of 2 kb around the FLAG-Rhino localization sites were plotted based on macs2 callpeak results. (g) Heatmap showing the localization of FLAG-Rhino and H3K9me3 in H3K9me3 broad peaks of OSCs. The regions of 10 kb around the FLAG-Rhino localization sites were plotted based on macs2 callpeak results. (h) Annotation of FLAG-Rhino localization in OSCs, when annotations overlapped, the annotation was assigned based on the following priority: piRNA cluster > transposon > gene. (i) Distribution of euchromatin/heterochromatin at Rhino localization sites in OSCs. Euchromatin/heterochromatin classification follows a previous study[12].



1 **Figure 2. ADMA-histones tend to co-localize with FLAG-Rhino in OSCs**

2 (a) Heatmap showing key histone modifications and HP1a ChIP-seq results at FLAG-

Rhino localization sites (828 regions) in OSCs. The regions of 2 kb around the FLAG-Rhino localization sites were plotted based on macs2 callpeak results. (b) Genome snapshot showing all histone modifications' ChIP-seq data in major dual-strand piRNA clusters (38C and 102F). The gray-shaded areas indicate the localization sites of FLAG-Rhino in each piRNA cluster in OSCs. (c) Pearson correlation coefficient comparing the coverage per 100 bp bin size of all ChIP-seq mapping results to the dm6 genome and its hierarchical clustering. (d) Venn diagram showing sites of overlapping localization of FLAG-Rhino, H3K9me3, H3R2me2a, H3R8me2a, H3R17me2a, H3R26me2a, H4R3me2a, and all ADMA-histones in OSCs.

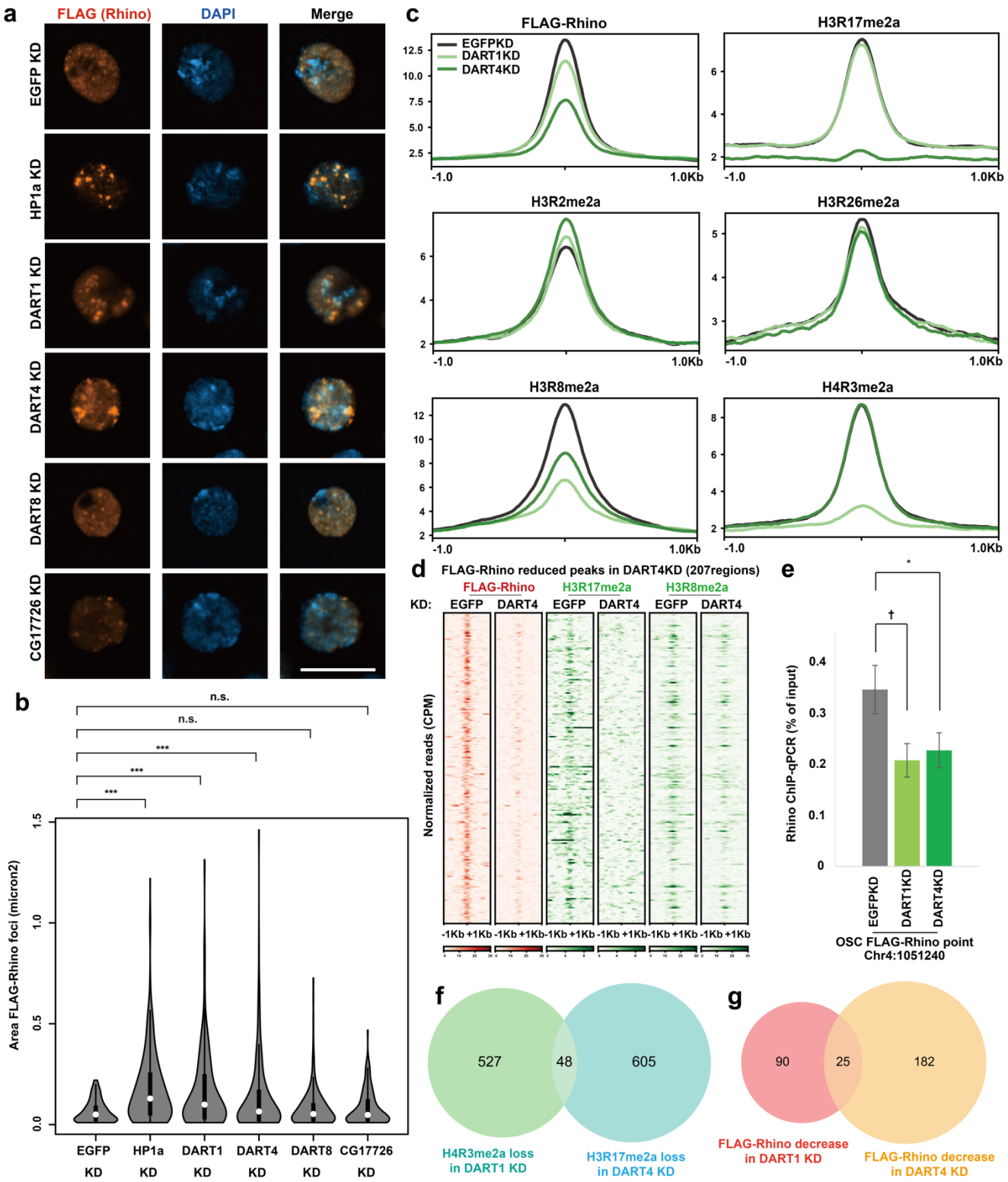


Figure 3. Arginine methyltransferases DART1 and DART4 play a role in Rhino genomic localization and foci formation in OSCs

(a) Confocal immunofluorescent images of FLAG-Rhino (red) and nuclei (blue; DAPI stained) in the EGFP-, HP1a-, DART1-, DART4-, DART8-, and CG17726-KD OSCs. Scale bar represents 10 μm. (b) Comparison of FLAG-Rhino foci sizes among the EGFP-, HP1a-, DART1-, DART4-, DART8-, and CG17726-KD OSCs in (a). Thirty OSCs from

three biological replicates were analyzed. Statistical significance in differences of Rhino foci between the genotypes was examined using Kolmogorov–Smirnov test. *P < 0.05, **P < 0.01, and ***P < 0.001. HP1a KD: ***P = 8.49×10^{-13}, DART1 KD: ***P = 1.83×10^{-7}, DART4 KD: ***P = 2.83×10^{-4}, DART8 KD: n.s. P = 0.655, CG17726 KD: n.s. P = 0.454. The box indicates the lower (25%) and upper (75%) quartiles, and the open circle indicates the median. Whiskers extend to the most extreme data points. (c) ChIP-seq profile showing the localization of FLAG-Rhino, H3R2me2a, H3R8me2a, H3R17me2a, H3R26me2a, and H4R3me2a in EGFP-, DART1-, and DART4-KD OSCs at FLAG-Rhino localization sites. The genomic regions used to draw the profile are the same as in Fig. 1f. (d) Heatmap showing the localization of FLAG-Rhino, H3R17me2a, and H3R8me2a in EGFP- and DART4-KD OSCs at the sites of decreased FLAG-Rhino localization in DART4 KD. The regions with decreased FLAG-Rhino were defined as those with an M value greater than 0.5 by manorm. macs2 callpeak p = 0.05 was used for manorm with input as a control. (e) FLAG-Rhino ChIP-qPCR results of one site at which Rhino initially localized. DART1 KD: †P = 5.88×10^{-2}, DART4 KD: * P = 3.17×10^{-2}. (f) Venn diagram showing the overlap of H3R17me2a localization decreased in DART4 KD and H4R3me2a localization decreased in DART1 KD. The regions with decreased ADMA-histones were defined as those with an M value greater than 1 in manorm. macs2 callpeak p = 0.01 was used for manorm. (g) Venn diagram showing the overlap of Rhino localizations that were decreased by DART1 KD and DART4 KD. The decreased genomic sites are identified as in (f).

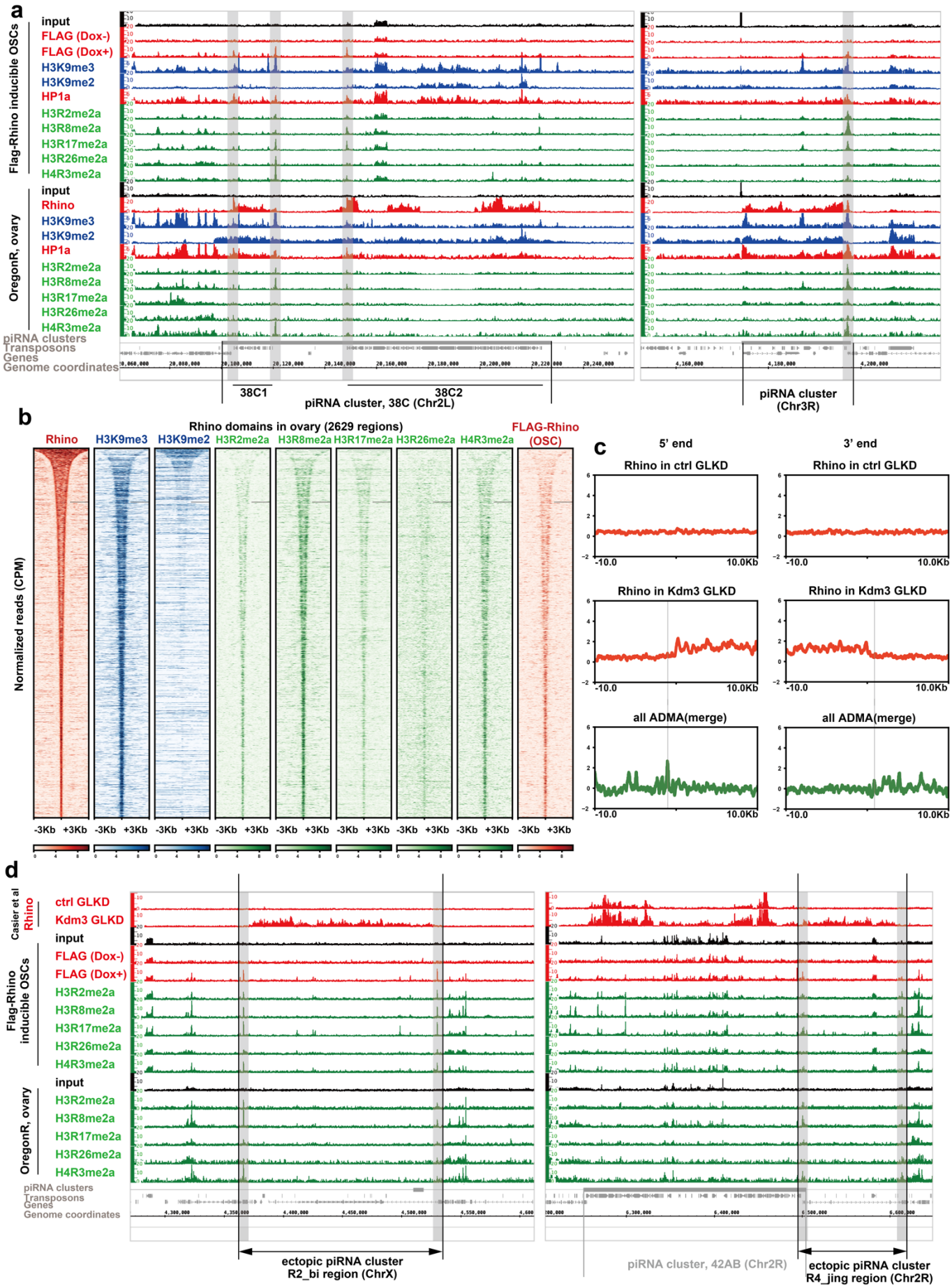


Figure 4. ADMA-histones co-localize with Rhino prior to Rhino spread in the ovary

(a) Genome browser snapshot showing the localization of H3R2me2a, H3R8me2a, H3R17me2a, H3R26me2a, and H4R3me2a in piRNA clusters in OSCs and Oregon-R ovary. The gray-shaded areas indicate the localization sites of FLAG-Rhino in each piRNA cluster in OSCs. (b) Heatmap showing the localization of H3R2me2a, H3R8me2a, H3R17me2a, H3R26me2a, and H4R3me2a at Rhino peaks in Oregon-R ovary. The Rhino peak regions shown in Fig. 1c were used. (c) Heatmap showing ADMA-histones at the 5′ and 3′ ends of the major ectopic piRNA clusters R1–R12 (R5 was excluded because of low Rhino localization in Kdm3 GLKD) that appeared in Kdm3 GLKD. Kdm3 GLKD-dependent ectopic piRNA cluster ends were extracted using the following method. Among Rhino peaks listed in the MACS2 callpeak files "GSE203279_MACS2_narrowpeak_Rhino_Ctrl.txt" and "GSE203279_MACS2_narrowpeak_Rhino_Kdm3GLKD.txt" in the GSE203279 dataset, we first extracted the Rhino peaks that appeared only in the latter[13]. The peaks closest to the 5′ and 3′ ends of the R1–R12 regions in Table S4 of the Casier et al. (2023) were designated as the ends of ectopic piRNA clusters[13]. (d) Genome browser snapshot of the indicated ChIP-seq data in representative ectopic piRNA clusters appearing in Kdm3 GLKD (R2_bi region and R4_jing region). The gray-shaded areas indicate the localization sites of potential FLAG-Rhino considered to be the ends of the ectopic piRNA cluster in OSCs.

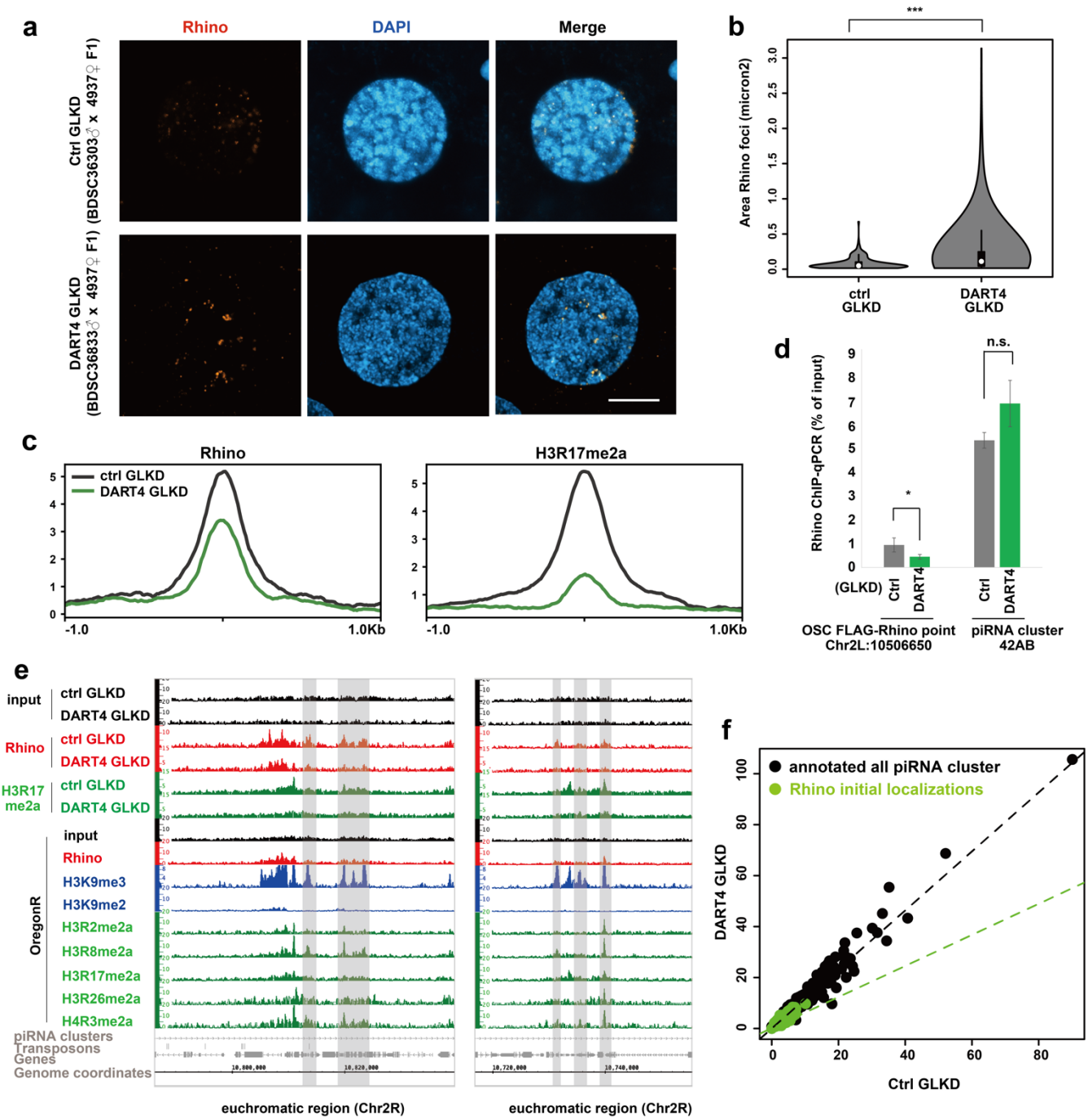
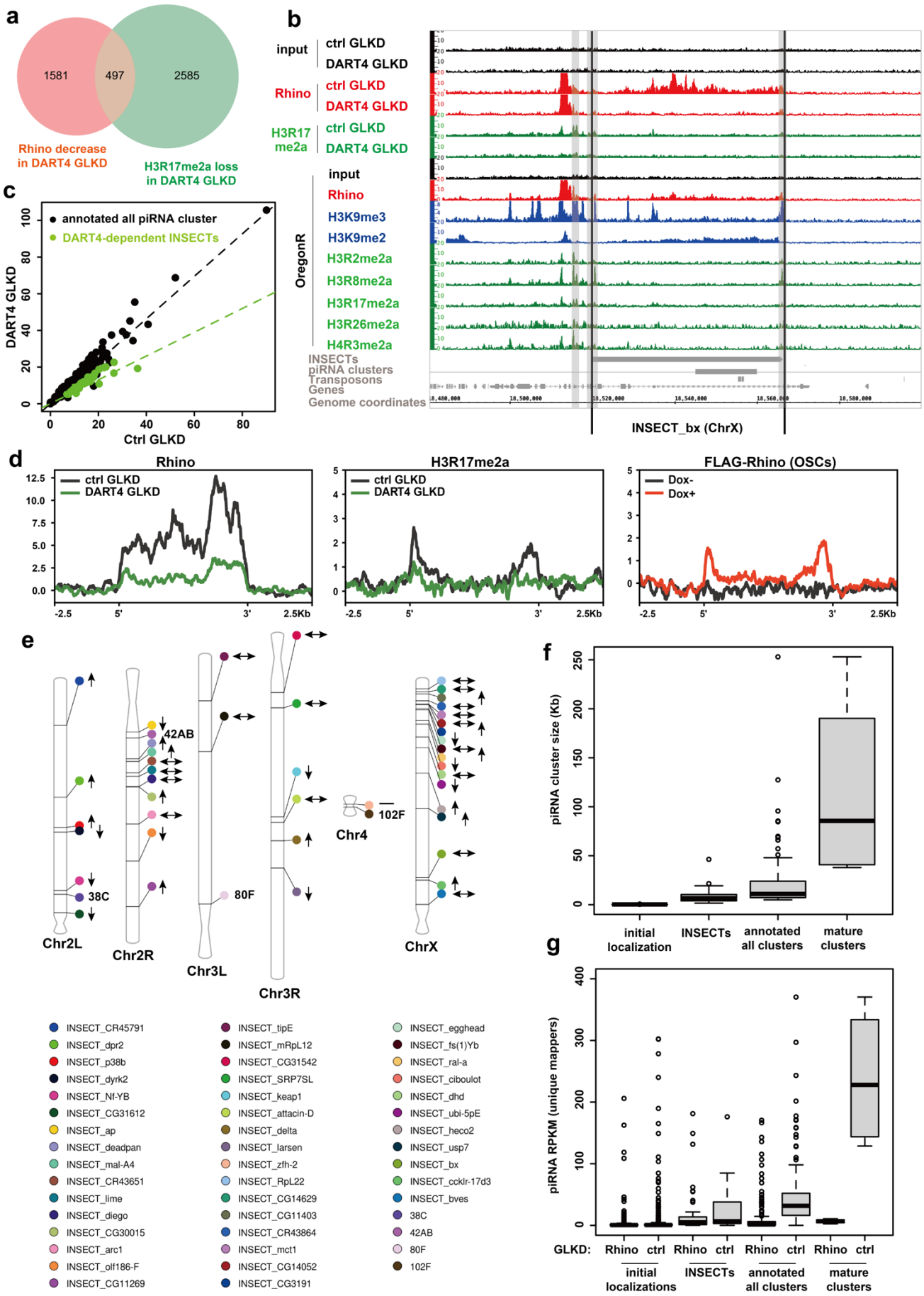


Figure 5. DART4 GLKD has impacts on Rhino foci and Rhino genomic localization in the ovary

(a) Confocal immunostaining images of Rhino (red) in ctrl- and DART4-GLKD ovaries. DAPI shows nuclei (blue). Scale bar represents 10 μm. (b) Comparison of sizes of Rhino foci between the ctrl- and DART4-GLKD ovaries in (a). Ten ovaries from two biological replicates were analyzed. Statistical significance in differences of Rhino foci between the two genotypes was examined as in Fig. 3b. DART4 GLKD: ***$P = 1.26 \times 10^{-10}$. (c) ChIP-seq profile showing the localization of Rhino and H3R17me2a in ctrl- and DART4-GLKD ovaries at the FLAG-Rhino localization sites in OSCs. The FLAG-Rhino peak regions,

as in Fig. 1f, are used. (d) Rhino ChIP-qPCR results of one site at which Rhino initially localized and piRNA cluster 42AB. t-test was used for statistical hypothesis testing. Rhino initial localization point: * $P = 4.03 \times 10^{-2}$, piRNA cluster 42AB: n.s. $P = 0.11$. (e) Genome snapshot of ChIP-seq data in euchromatic regions. The gray-shaded areas indicate genomic locations where Rhino localization is decreased in a H3R17me2a-dependent manner. (f) Scatter plot comparing Rhino enrichment in ChIP-seq between ctrl GLKD and DART4 GLKD. Each dot represents the square root of the read count per million mapped reads (RPM) within the annotated piRNA clusters in proTRAC database (colored in black, 184 clusters, without clusters on chromosome Y and uni-strand piRNA clusters flamenco/20A) and Rhino initial localizations as in Fig. 5c (colored in green, 828 regions). The dotted lines represent an approximation by a linear regression model.



1 **Figure 6. Initial-stage piRNA clusters (INSECTs) are lost in DART4 GLKD**

(a) Venn diagram showing the overlap of the sites with decreased Rhino localization and the sites with lost H3R17me2a localization in DART4 GLKD. The regions where Rhino was decreased were defined as those with an M value greater than 0.5 by manorm, and the regions where H3R17me2a was lost were defined as those with an M value greater than 0. For manorm, the output results of macs2 callpeak p = 0.01 were used with input as a control. (b) Genome snapshot showing ChIP-seq results of DART4-dependent INSECTs found around the *bx* gene. Black lines indicate INSECT boundaries. The gray-shaded areas indicate the localization sites of potential ADMA-histones considered to be the ends of DART4-dependent INSECTs. (c) Scatter plot comparing Rhino enrichment in ChIP-seq between ctrl GLKD and DART4 GLKD. Each dot represents the square root of the read count per million mapped reads (RPM) within the annotated piRNA clusters in proTRAC database (colored in black, 184 clusters, without clusters on chromosome Y and uni-strand piRNA clusters flamenco/20A) and DART4-dependent INSECTs (colored in green, 43 clusters). The dotted lines represent an approximation by a linear regression model. (d) ChIP-seq profiles of Rhino, H3R17me2a in ctrl GLKD and DART4 GLKD in all 43 INSECTs and Rhino in OSCs. (e) Chromosomal location of the identified DART4-dependent INSECTs and the direction of Rhino spread estimated based on the initial Rhino localization in OSCs. Depressed areas indicate pericentromeres. Arrows indicate the direction of Rhino spread estimated based on the genomic location of ADMA-histones. Bidirectional arrows indicate that ADMA-histones are present at both ends and the direction of spread cannot be determined. Horizontal bars indicate that there are no ADMA-histones at both ends. (f) Box plot showing the lengths of Rhino localization sites in OSCs, DART4-dependent INSECTs, all piRNA clusters annotated in proTRAC (as in Fig. 6c), and mature piRNA clusters (42AB, 38C, 80F, and 102F). (g) Box plot showing

read count per kilobase of cluster per million mapped reads (RPKM) of piRNAs uniquely mapped to Rhino localization sites in OSCs, DART4-dependent INSECTs, all piRNA clusters annotated in proTRAC, and mature piRNA clusters (42AB, 38C, 80F, and 102F). Clusters where RPKMs were under 400 are shown.

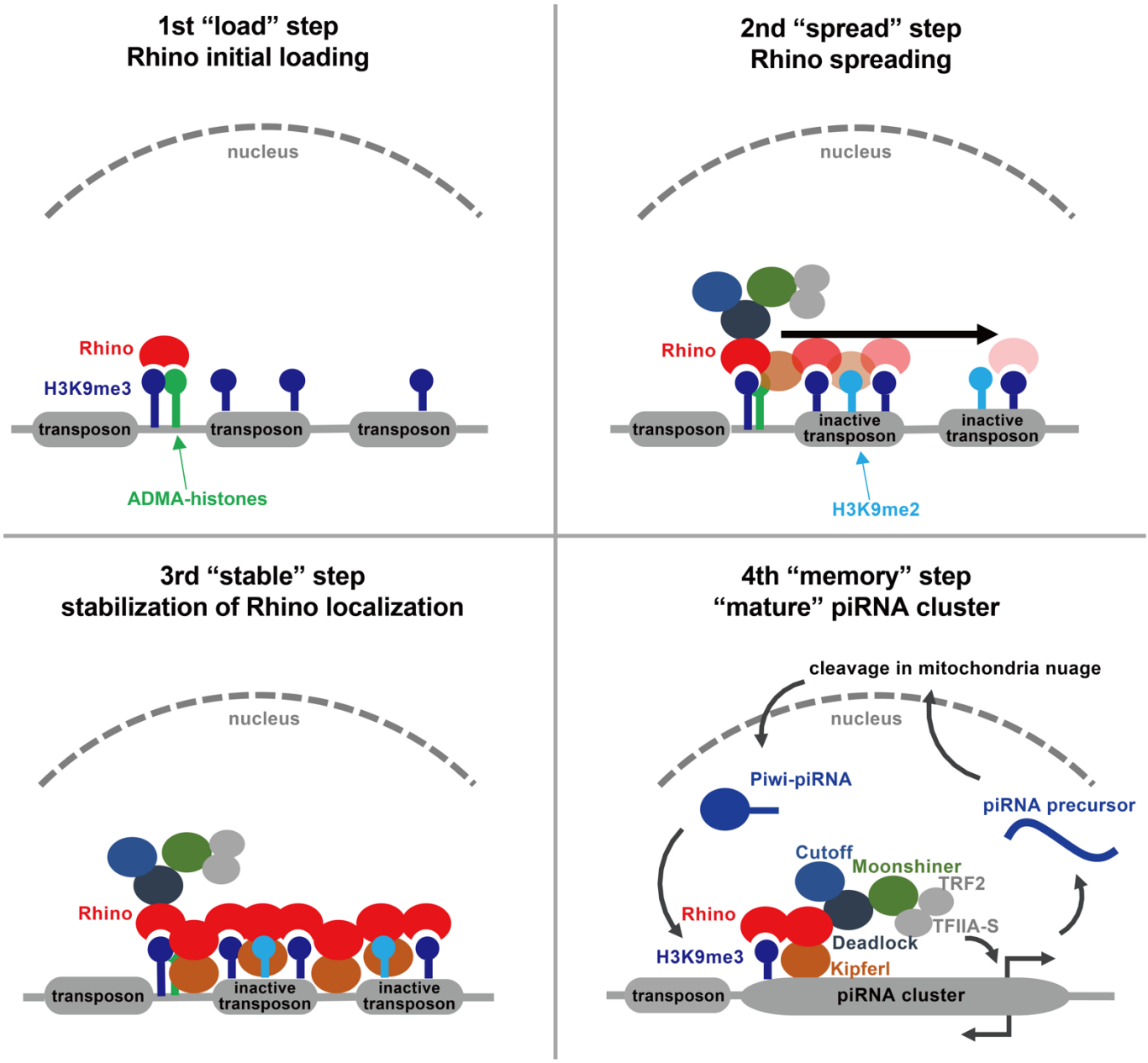


Figure 7. Model for the piRNA cluster formation

Rhino first localizes to bivalent nucleosomes containing H3K9me3 and ADMA-histones in the euchromatin regions. In the second step, Rhino spreading occurs. Transcription of adjacent regions and/or H3K9me2 enrichment may be involved in this step. In the third step, Rhino spreading is stabilized by an unknown factor(s). Finally, after completion of piRNA cluster formation, Rhino positional information is maintained by epigenetic inheritance of Piwi-piRNAs.