Genome Biology

RESEARCH

Open Access

A H3K27me3 reader complex couples H3K27me3 accumulation to nascent transcription of transposable elements in *Paramecium*

Thomas Balan[1], Ambre Petitalot[2], Julien Richard Albert[1], Louise Abbou[1], Guillaume Chevreux[1], Bassem Al-Sady[3], Bihter Özdemir Aygenli[4], Till Bartke[4], Raphaël Margueron[2] and Sandra Duharcourt[1*]

*Correspondence:
sandra.duharcourt@ijm.fr

[1] Université Paris Cité, CNRS, Institut Jacques Monod, 75013 Paris, France
[2] Institut Curie, Paris Sciences Et Lettres Research University, INSERM, CNRS, 75005 Paris, France
[3] Department of Microbiology & Immunology, George Williams Hooper Foundation, University of California San Francisco, San Francisco, CA, USA
[4] Institute of Functional Epigenetics, Helmholtz Zentrum München, 85764 Neuherberg, Germany

Abstract

Background: The ability to deposit histone H3K27-trimethyl (me3) marks is essential for transcriptional repression by Polycomb Repressive Complex 2 (PRC2). This is largely attributed to Polycomb repressive complex 1 (PRC1), whose recruitment is H3K27me3-dependent. Yet it is unclear how H3K27me3 contributes to transcription regulation independently of PRC1.

Results: To address this question, we identified H3K27me3-binding proteins in the unicellular eukaryote *Paramecium* where PRC2 targets transposable elements (TEs) and PRC1 is absent. We show that the chromodomain protein Firefly is a H3K27me3 reader in vitro and accumulates onto TEs in a H3K27me3-dependent manner. We also identify Firefly interactors: Sleepy, a coiled-coil containing protein and TfIIs4, a transcription elongation factor. We show that Firefly, Sleepy and TfIIs4 are jointly required in vivo for correct H3K27me3 accumulation and nascent transcription at TEs.

Conclusions: This positive feedback loop to enrich H3K27me3 at TEs is analogous to self-reinforcing loops leading to H3K9 methylation at repeats in fungi, plants and animals. Our work reveals a unique association between H3K27me3 readers and active transcription.

Highlights

- Firefly recognizes H3K27me3 through its chromodomain.
- Firefly associates with Sleepy and TfIIs4.
- The three proteins are required for H3K27me3 accumulation, nascent transcription of TEs and their elimination.

Keywords: Polycomb repressive complex, Transposable elements, Histone modifications, Chromodomain, Nascent transcription, Programmed DNA elimination, Ciliates

Background

Polycomb complexes are central components of chromatin-based gene transcription regulation. They play crucial developmental roles by controlling cell-type-specific transcription programs that determine cell identity in both animals and plants [1]. Polycomb repressive complex 1 (PRC1) mono-ubiquitylates histone H2A and Polycomb repressive complex 2 (PRC2) trimethylates histone H3 at lysine 27 (H3K27me3) [2–6]. H3K27me3 is essential for PRC2-mediated repression [7, 8]. Repression associated with H3K27me3 is largely attributed to PRC1, which binds to H3K27me3 via the chromodomain-containing CBX subunit, highlighting the coupling between PRC1 and PRC2 [9, 10]. To date, our knowledge of the function of Polycomb complexes stems mostly from studies of animals and plants, where the specific roles of PRC1 and PRC2 in gene transcription repression are difficult to disentangle because both complexes are required. Interestingly, however, PRC1 and PRC2 have evolved independently, and, in many organisms, PRC1 is absent [11]. Notably, no clear PRC1 ortholog has been identified in the unicellular ciliate *Paramecium tetraurelia* [11], which thus provides an ideal system to determine how H3K27me3 contributes to transcription regulation independently of PRC1.

In the ciliate *Paramecium tetraurelia*, deposition of H3K27me3 and H3K9me3 at TEs is catalyzed by the Enhancer-of-zeste Ezl1 enzyme, the catalytic subunit of PRC2 [12]. There is increasing evidence that TEs are associated with H3K27me3 in diverse eukaryotes, including ciliates [13], and that PRC2 silences TEs [12, 14–16]. In *P. tetraurelia*, the enrichment of H3K27me3 and H3K9me3 at TEs is necessary to trigger their physical elimination during development of the somatic nucleus [12, 17]. Ciliates have thus evolved a TE silencing mechanism that purges them from the somatic genome during the sexual cycle.

In *P. tetraurelia*, two types of nuclei coexist in the same cell: the germline micronucleus (MIC) and the somatic macronucleus (MAC). During the self-fertilization process of autogamy, the MIC undergoes meiosis and karyogamy to produce the zygotic nucleus. New MICs and new MACs develop from mitotic products of the zygotic nucleus [18]. The MAC contains a reduced genome (72 Mb) compared to the MIC genome (108 Mb) as a result of massive and reproducible DNA elimination that occurs at each sexual cycle. During the development of the new MAC, at least one third of the MIC genome is removed [19–22]. Eliminated sequences include 45,000 Internal Eliminated Sequences (IESs) that are remnants of TEs scattered throughout the genome [19, 21]. Other Eliminated Sequences (OES) [23] correspond to large regions comprising repeats such as transposable elements or satellites [20, 22].

The sites of DNA elimination do not seem to be specified by the DNA sequence in the MIC genome because, beyond an invariable TA at IES extremities, no strictly conserved sequence motif has been identified in or around IESs, or in TE-containing regions. Understanding how such diverse sequences are recognized and eliminated remains challenging. The DNA elimination process involves small RNA-guided histone marks and subsequent DNA cleavage and repair [24]. 25-nt long scnRNAs are produced from the entire germline genome, followed by the degradation of the subpopulation of scnRNAs matching MAC sequences, while germline-specific scnRNAs escape degradation and are thus selected [25–27]. Germline-specific scnRNAs bound to the Argonaute Ptiwi01/09 proteins guide the deposition of H3K9me3 and H3K27me3 onto TEs in the developing

MAC [17]. Loss of histone marks prevents the elimination of all TEs and of 70% of IESs [12, 17, 28, 29]. We and others recently discovered that the RNAi effector Ptiwi09 physically interacts with Ezl1 [17, 29] and that this interaction is essential for the correct targeting of histone marks and transcriptional repression of TEs in the developing MAC [17]. The physical association with the RNAi machinery is thought to tether PRC2 specifically to TEs in order to initiate silent chromatin formation. In this mechanism, scnRNA-Piwi complexes guide histone modification deposition, therefore providing specificity despite the lack of a conserved motif on eliminated sequences. The putative transcription elongation factor, TfIIs4, is involved in DNA elimination and required for the non-coding transcription of several IESs [30], suggesting that TfIIs4-mediated transcripts may serve as a recruitment platform for the scnRNA-Piwi complexes. However, the mechanism by which deposition of repressive histone modifications at TEs leads to their elimination is currently unknown.

The elimination machinery is composed of the domesticated transposase Pgm, which introduces DNA double-strand breaks at the boundaries of IESs [31], of its PgmL partners [32], and of the non-homologous end joining (NHEJ) components, which repair the breaks [33–35]. Given that H3K27me3 and H3K9me3 still accumulate at TEs in the absence of Pgm [12, 17], it is believed that PRC2-mediated histone marks specifically tether the elimination machinery at these genomic sites, but how this is achieved is not known [36].

Here, we identify proteins that physically interact with recombinant nucleosomes bearing H3K27me3 by pull-down and mass spectrometry. We assessed the ability of the chromodomain of Firefly to directly bind H3K27me3 peptides in vitro and its genome-wide occupancy by chromatin immunoprecipation followed by sequencing. We found that Firefly is a H3K27me3 reader that accumulates at TEs. We also identified proteins that interact with Firefly: an uncharacterized coiled-coil protein Sleepy, and the transcription elongation factor TfIIs4. We show that depletions of these three proteins phenocopy each other, that their localization is interdependent in vivo, and that they appear to form a complex in a heterologous expression system. Genome-wide analyses reveal that Firefly and its partners are required for correct targeting of H3K27me3 to TEs and for their elimination. We further show that, through the Firefly chromodomain, the three partners are required for nascent transcription of TEs, providing a positive feedback loop to couple H3K27me3 and nascent transcription of TEs.

Results

Modified recombinant nucleosome affinity purification identifies Firefly, a H3K27me3 reader that acumulates on transposable elements

To decipher the role of H3K27me3 and H3K9me3 in the process of DNA elimination, we sought to identify proteins that interact specifically with nucleosomes carrying these modifications [37]. We assembled recombinant biotinylated human dinucleosomes bearing *Paramecium* histone H3 tails unmethylated or trimethylated on lysine 9 or lysine 27 (see Methods). Of note, the *Paramecium* H3 tail diverges substantially from that of human H3 [38]. These nucleosomes were incubated with *Paramecium* nuclear extracts to perform pull-downs ($n = 3$ in each condition), followed by label-free quantitative proteomics. The *Paramecium* nuclear extracts were prepared

at T = 10 h after the onset of autogamy (i.e. 10 h following the time when 50% of the cells in the population have a fragmented maternal MAC) (Fig. 1A, Additional file 1: Figure S1A-B). This time point corresponds to DNA elimination events [39], when both H3K27me3 and H3K9me3 are present in the developing MACs (Fig. 1A).

Statistical analyses revealed that very few proteins (2 out of 1973) were differentially enriched with H3K9me3 compared with unmodified nucleosomes (fold change > 5; p-value < 0.01; unique peptides ≥ 2) (Fig. 1B). Besides, no chromodomain-containing proteins, which are known for their ability to bind methylated histones, were identified [41].

In contrast, the pull-down with the H3K27me3 nucleosome retrieved many proteins (44 proteins out of 1973; fold change > 5; p-value < 0.01; unique peptides ≥ 2) (Fig. 1C). Among the five proteins that are encoded by genes from the early and intermediate expression peak clusters [40], we found a chromodomain-containing protein, Firefly, which was recently reported as a putative H3K27me3 binder involved in the elimination of a subset of IESs [42](Fig. 1D). Firefly proteins are encoded by two genes (Firefly1 and Firefly2) [42] originating from a recent whole genome duplication (WGD) [43] and belong to the intermediate expression peak clusters [40](Additional file 1: Figure S2). We renamed them Firefly-a and Firefly-b, according to the *Paramecium* gene nomenclature guidelines [44, 45]. Firefly proteins are among the 37 chromodomain-containing proteins found in the *Paramecium tetraurelia* genome (Additional file 1: Figure S2A-B) [42]. Because the two proteins share 85% identity, and RNAi-mediated gene knockdown indicated redundant functions [42], we will refer to them collectively as Firefly. Apart from its conserved chromodomain, the Firefly protein does not show conservation at the sequence or AlphaFold3 predicted structure levels with canonical chromodomain proteins (Additional file 1: Figure S2C-D).

To determine whether Firefly binds directly to H3K27me3, we performed pull-down assays with biotinylated *Paramecium* H3 peptides that are trimethylated either on K9 or on K27, or unmethylated on either residue. We purified full-length recombinant Firefly protein fused to GST (Additional file 1: Figure S1C). As a control, we used a full-length recombinant mutant Firefly protein (Firefly^{WYAA}), in which two conserved amino acids (W28 and Y31) of the aromatic cage in the chromodomain were substituted by alanines (Fig. 1D; Additional file 1: Figure S1E). The recombinant chromodomains of mouse Cbx5 and Cbx2, known to bind H3K9me3 or H3K27me3, respectively, were also fused to GST to serve as positive controls (Additional file 1: Figure S1C) [2, 5, 46, 47] (Methods). As expected, no peptide binding was detected with GST alone, while the chromodomains of Cbx5 or of Cbx2 show direct binding to H3K9me3 and H3K27me3, respectively. We found that GST-Firefly binds to H3K27me3, but not to H3K9me3 (Fig. 1E). As expected, the GST-Firefly^{WYAA} protein does not bind H3K27me3, indicating that the chromodomain of Firefly is required for H3K27me3 binding (Fig. 1E). We conclude that Firefly is a direct reader of H3K27me3, as others have recently speculated [42].

Given that Firefly binds to H3K27me3 (Fig. 1E) and that H3K27me3 in *Paramecium* is associated with Other Eliminated Sequences (OES) that contain TEs [12, 17], we hypothesized that Firefly and H3K27me3 colocalize over MIC-limited OES and TEs. To test this possibility, we identified Firefly binding sites on chromatin genome-wide. To circumvent the absence of a Firefly antibody, we expressed a tagged FLAG-HA-Firefly

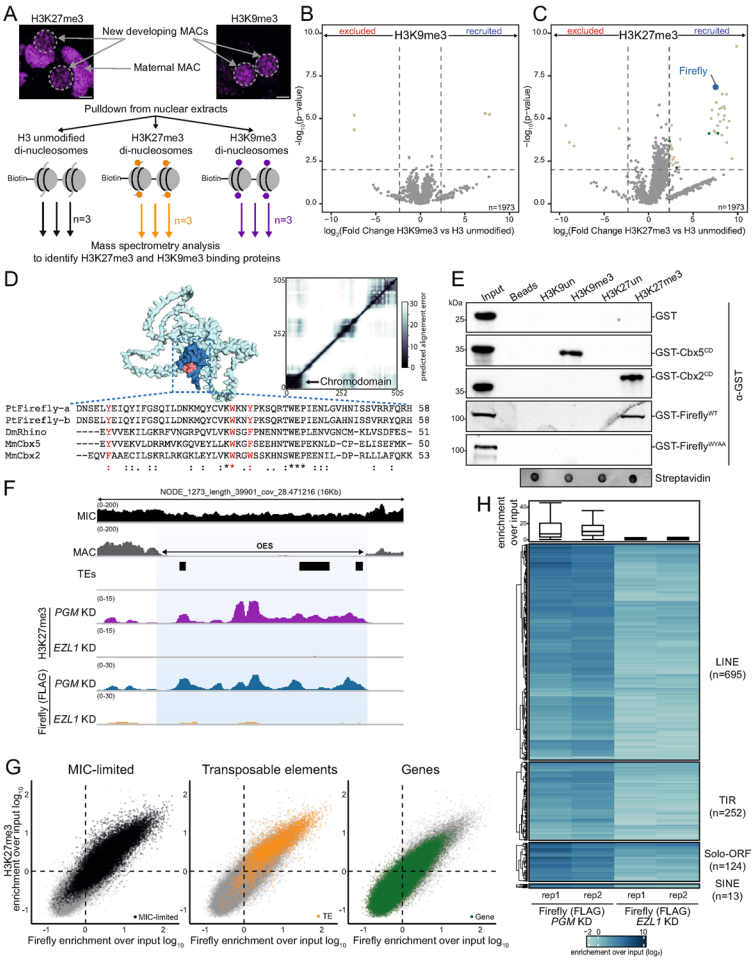

Fig. 1 Modified recombinant nucleosome affinity purification identifies Firefly, a H3K27me3-reader that accumulates on transposable elements. **A** Experimental design for the identification of H3K27me3 and H3K9me3 readers by nucleosome affinity purification followed by quantitative mass spectrometry. **B**, **C** Volcano plot of the quantitative label-free mass-spectrometry analysis of H3K9me3 (**B**) or H3K27me3 (**C**) nucleosome pull-downs. 3 biological replicates were analyzed for each condition. The total number of identified proteins is indicated at the bottom right. Significantly enriched proteins over unmodified nucleosome pull-downs (statistical t-test FC $\geq$ 5 and *p*-value $\leq$ 0.01, unique peptides $\geq$ 2) are colored. Enriched proteins expressed from the early and intermediate expression gene clusters [40] are highlighted in dark green, and Firefly is in blue. **D** Structure prediction of Firefly-a using AlphaFold3. The chromodomain is highlighted in dark blue and the amino acids of the aromatic cage are in red. Predicted aligned error (PAE) plot for the top ranked AlphaFold3 prediction is displayed on the right. Alignment of the chromodomain of Firefly-a and Firefly-b from *P. tetraurelia* (PTET.51.1.P0220149 and PTET.51.1.P0020347), Rhino of *D. melanogaster* and Cbx2 of *M. musculus*. (*) conserved sequence, (:) conservative substitution, (.) semi-conservative substitution. **E** Western blot of biotin-conjugated *Paramecium* histone H3 peptide pull-down assays using GST, GST-Cbx5^{CD} (H3K9me3 positive control), GST-Cbx2^{CD} (H3K27me3 positive control) and full-length GST-Firefly^{WT} and GST-Firefly^{WYAA} purified recombinant proteins (Additional file 1: Figure S1E). Dot blot with Streptavidin-HRP serves as a loading control for each pull-down. **F** Genome browser screenshot of an example of a MIC-limited region containing TEs displaying H3K27me3 and Firefly (FLAG) ChIPseq enrichments in *PGM* and *EZL1* KDs. Coordinates are indicated in the figure. **G** Scatter plot showing global H3K27me3 and Firefly (FLAG) enrichments over 1-kb bins. Data points ($n = 97{,}658$) are colored to highlight MIC-limited regions (black); transposable elements (orange) and genes (green). **H** Heatmap and boxplot of Firefly (FLAG) enrichment over input as determined by ChIP-seq for the considered TE copies upon *PGM* or *EZL1* KD (two biological replicates). The TE copies are ordered by hierarchical clustering of DNA coverage in each family. LINEs: class I Long Interspersed Nuclear Elements. TIRs: class II DNA transposons with Terminal Inverted Repeats. SINEs: class I non-autonomous Short Interspersed Nuclear Elements. SOLO: class I non-autonomous elements carrying ORF1 only. The boxplots show the coverage distribution (enrichment over input) for all TE copies. The box shows the second and third quartiles. The median is displayed as a horizontal line. The whiskers run from the minimum to the maximum value

fusion protein from an RNAi-resistant transgene (see Methods). To assess the functionality of the *FLAG-HA-FIREFLY* transgene, we performed in vivo genetic complementation assays. *FIREFLY* KD impairs the production of viable sexual progeny after autogamy (Additional file 1: Table S1 and [42]). We found that the expression of the *FLAG-HA-FIREFLY* transgene rescues the production of viable sexual progeny in cells subjected to *FIREFLY* KD (Additional file 1: Table S1), indicating that the *FLAG-HA-FIREFLY* transgene is functional. Taking advantage of the tagged protein, we performed chromatin immunoprecipitation (ChIP) of Firefly (FLAG) in duplicate on cells expressing the functional *FLAG-HA-FIREFLY* transgene. Since H3K27me3-marked regions are normally eliminated, we subjected these cells to *PGM* KD, which enables us to overcome both the absence of synchrony in *Paramecium* autogamous cell cultures and the loss of histone marks as development and DNA elimination proceed. Indeed, Pgm-depleted cells are blocked at a late developmental stage, after H3K27me3 deposition in the developing MAC [12, 17]. To determine whether genome occupancy of Firefly is dependent on H3K27me3, we also performed ChIP of Firefly (FLAG) in duplicate on cells subjected to KD of *EZL1*, the catalytic subunit of PRC2 [12, 17].

ChIP-seq reads were mapped onto the MIC reference genome [20] and enrichment over input was measured in 1-kb bins for Firefly and for H3K27me3 in *PGM* and *EZL1* KD conditions. The genome browser screenshot illustrates the co-localization of Firefly and H3K27me3 on an OES that comprises TEs in *PGM* KD (Fig. 1F; Additional file 1: Figure S3A). Indeed, both Firefly and H3K27me3 are enriched over the vast majority (73%) of 1-kb bins overlapping MIC-limited OES regions, including all TE families in *PGM* KD (Fig. 1G; Additional file 1: Figure S3A-C). In contrast, 24% of 1-kb bins overlapping protein-coding genes are enriched for both (Fig. 1G; Additional file 1: Figure S3D-E). To examine Firefly enrichment on TEs, we selected TE copies that are well covered (>10 RPKM) in the inputs (695 LINEs, 252 TIRs, 124 SOLO-ORFs and 13 SINEs). These TE copies are enriched in Firefly in *PGM* KD, and this enrichment is greatly reduced in *EZL1* KD (Fig. 1F and H; Additional file 1: Figure S3A-C). We conclude that Firefly is enriched in H3K27me3-associated regions that comprise TEs, and not in protein-coding genes, and that Firefly occupancy depends on H3K27me3.

Altogether, the results from our nucleosome affinity purification, peptide pull-down and genomics assays all show that Firefly is a *bona fide* H3K27me3 reader that accumulates on TEs.

Firefly, Sleepy and TfIIs4 interact in vivo, co-elute on glycerol gradient and their localization is interdependent

To investigate Firefly's mode of action, we identified its interacting partners through immunoprecipitation followed by MS/MS analysis. Nuclear extracts were prepared from *Paramecium* cells expressing the functional FLAG-HA-Firefly fusion protein at T = 10 h after the onset of autogamy, since immunofluorescence experiments indicated that the Firefly fusion protein localizes in the developing new MAC at this time (Fig. 2A). We then tested nuclear extractions with increasing saline concentrations to determine the conditions where the most Firefly is recovered in the soluble nuclear fraction (Fig. 2B). Western blot analysis indicated that Firefly is mostly detected in the nuclear fractions at 15 and 150 mM NaCl, while little is extracted at 300 mM NaCl.

This extraction profile mirrors that of histone H3, suggesting a tight interaction of Firefly with chromatin (Fig. 2B).

Immunoprecipitation of the FLAG-tagged Firefly protein in triplicate from nuclear extracts prepared at 150 mM NaCl was followed by quantitative label-free mass spectrometry analysis (Additional file 1: Figure S4A). Stringent statistical analyses (fold change > 5; p-value < 0.01; unique peptides > 5) identified 11 enriched proteins, among which 3 displayed similar, high abundance over all replicates (Fig. 2C): Firefly, the putative transcription elongation factor TfIIs4 involved in IES noncoding transcription and elimination [30] and an uncharacterized coiled-coil-containing protein that we named Sleepy (Fig. 2C). Like Firefly and TfIIs4, Sleepy is encoded by a gene expressed during the same developmental window (intermediate peak cluster) (Additional file 1: Figure S4B) [40]. RNAi-mediated *SLEEPY* gene knock-down during autogamy led to lack of viable sexual progeny (Additional file 1: Table S1), as previously reported upon *FIREFLY* or *TFIIS4* KD [30, 42] (Additional file 1: Table S1). We found that upon *SLEEPY* KD, similarly to *FIREFLY* or *TFIIS4* KDs, new MACs start to develop during autogamy but no viable sexual progeny are recovered, indicating that the new MACs are not functional under these conditions.

We speculated that the three proteins form a complex together. To test this hypothesis, we co-infected Sf9 insect cells with baculoviruses expressing FLAG-Firefly, HA-Sleepy and Strep-TfIIs4 to perform co-immunoprecipitation experiments and assess direct interactions. The three proteins were well expressed, however, interaction studies

(See figure on next page.)

Fig. 2 Firefly, Sleepy and TfIIs4 interact in vivo, co-elute on glycerol gradient and their localization is interdependent. **A** Schematic representation of Firefly immunoprecipitation (IP) experiments. Anti-FLAG immunostaining of cells transformed with a *FLAG-HA-FIREFLY*^{WT} transgene, at the same time point as the preparation of nuclear extracts for immunoprecipitation (T = 10 h). Developing new MACs are circled with a dotted line. Scale bar, 5 µm. **B** Western blot analysis of nuclear extracts prepared with increasing salt concentrations (15 mM; 150 mM; 300 mM NaCl), and the insoluble fraction (ins) from cells expressing a *FLAG-HA-FIREFLY*^{WT} transgene and from non-injected cells (CTL) at T = 10 h. Anti-HA antibodies are used to detect Firefly and histone H3 antibodies for normalization. Predicted MW for FLAG-HA-Firefly: 64.7 kDa. **C** Volcano plot of the quantitative label-free mass-spectrometry analysis of FLAG-HA-Firefly affinity purification. 3 biological replicates were analyzed for each condition. (Left) Significantly enriched proteins in Firefly IP over control IP (statistical t-test, fold change (FC) ≥ 2 and p-value ≤ 0.05) are colored. The total number of identified proteins is indicated at the bottom right. Enriched proteins expressed from the early and intermediate expression gene clusters [40] are highlighted in dark green, and include Firefly, TfIIs4 and Sleepy. (Right) Abundance plot of the quantitative label-free mass-spectrometry analysis of FLAG-HA-Firefly affinity purification. Only proteins with a FC ≥ 5, a p-value ≤ 0.01 and a number of quantified unique peptides greater than 5 were considered. Enriched proteins expressed from the early and intermediate expression gene clusters [40] are highlighted in dark green, and include Firefly, TfIIs4 and Sleepy. **D** Western blot analysis of density sedimentation assay performed on Sf9 lysate co-expressing FLAG-Firefly, HA-Sleepy and Strep-TfIIs4. Firefly, Sleepy and TfIIs4 exhibit similar elution profiles. Anti-His antibodies are used to visualize Firefly, Sleepy and TfIIs4. The two membranes (see Zenodo [48]) are separated by a dotted line. **E**, **F**, **G** (Left) Immunostaining of cells expressing a *FLAG-HA-FIREFLY*^{WT} (**E**), *FLAG-HA-SLEEPY*^{WT} (**F**), *GFP-TFIIS4*^{WT} (**G**) transgenes at T = 10 h after the onset of autogamy upon *ICL7* (CTL) *FIREFLY*, *SLEEPY* or *TFIIS4* KD. The developing new MACs (indicated by the dotted line) are round and display a faint and smooth Hoechst signal. The other Hoechst-stained nuclei are the fragments of the maternal MAC and the MICs. Scale bar, 5 µm. (Middle) Quantification of FLAG (Firefly or Sleepy) or GFP (TfIIs4) fluorescence signal in the new MAC (see Methods). Bars correspond to mean ± SD. Mann–Whitney statistical tests. (Right) Western blot analysis of Firefly, Sleepy, and TfIIs4 protein levels from whole cell extracts (T = 10 h after the onset of autogamy) in wild-type cells (WT), and cells expressing *FIREFLY*^{WT}, *SLEEPY*^{WT} or *TFIIS4*^{WT} transgenes upon *ICL7* (CTL) *FIREFLY*, *SLEEPY* or *TFIIS4* KD. *: contaminant band. Anti-HA antibodies are used to detect Firefly and Sleepy, and anti-GFP antibodies to detect TfIIs4. Anti-α-tubulin antibodies are for normalization

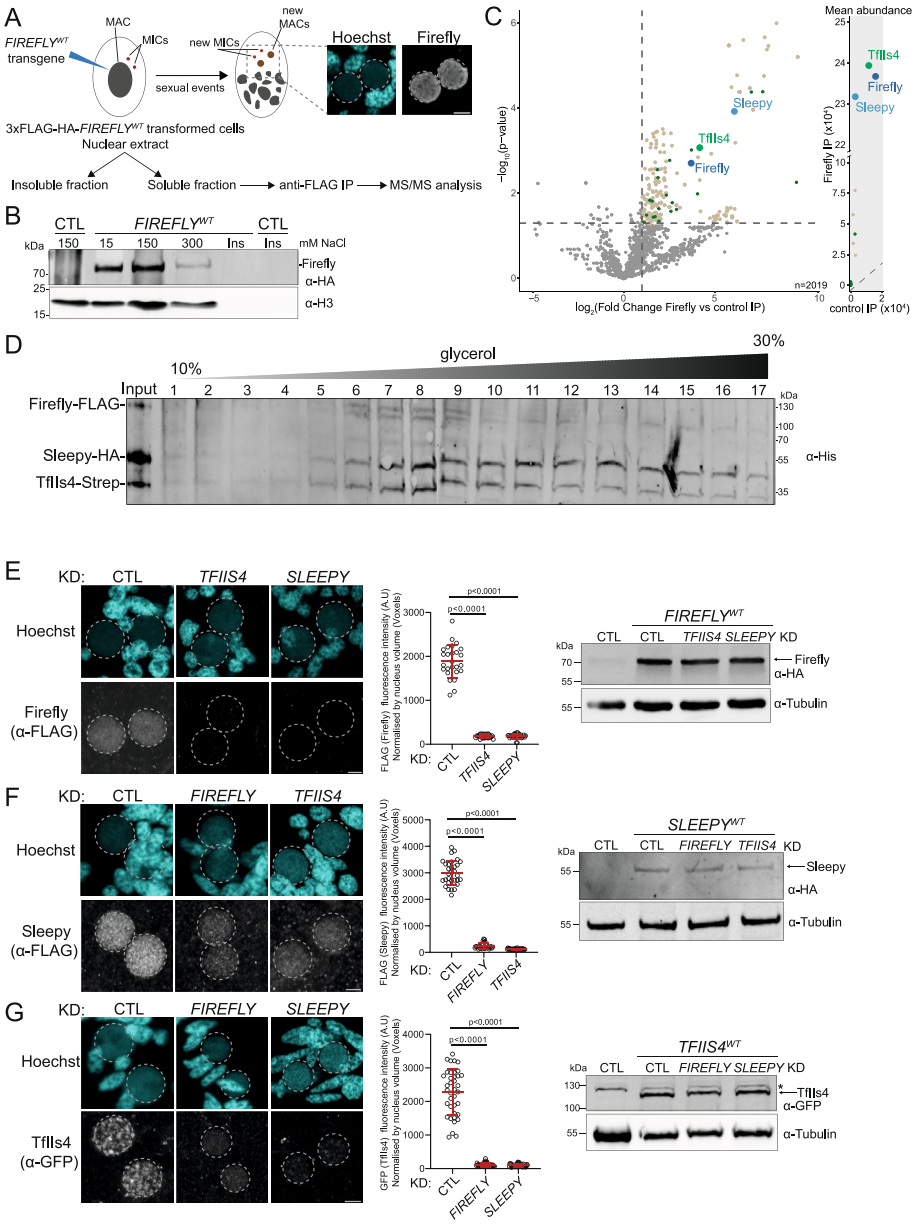

Fig. 2 (See legend on previous page.)

were limited by the tendency of these proteins, especially Firefly, to be retained non-specifically by the beads used for the co-immunoprecipitation assays. We therefore moved to an alternative strategy by co-expressing the three proteins and analysing their elution pattern after density sedimentation assay as described in [49] (Fig. 2D). We loaded the lysate of Sf9 cells expressing all three proteins on a 10%−30% glycerol gradient and monitored their elution by western blot (Fig. 2D). This approach again revealed an unusual behavior of the three proteins, which display a very broad elution pattern stretching from the 6th fraction almost until the end of the gradient. Such a pattern is reminiscent of what is observed with proteins that tend to oligomerize as reported for Polyhomeotic [50]. Interestingly, all three proteins display the same pattern of elution, which supports the existence of a complex. The relative intensity of the western blot signals

further suggests that Firefly might be a substochiometric subunit as compared to Sleepy and TfIIs4 (Fig. 2D). Given the presence of a coiled-coil domain, Sleepy might have a structural or scaffolding role in the assembly of the Firefly-Sleepy-TfIIs4 complex. Interestingly, coiled-coil proteins that interact with PRC2 have been shown to be important for the association of PRC2 to chromatin in rice [51] and in fungi [52].

Given the likely existence of a complex formed around Firefly, we investigated the interdependence of the cellular localization of the three subunits. *Paramecium* cells expressing either *FLAG-HA-FIREFLY*, *FLAG-HA-SLEEPY* or *GFP-TFIIS4* [30] transgenes were used to determine the localization of each fusion protein upon depletion of the other two proteins. Of note, we validated the functionality of the RNAi-resistant *FLAG-HA-SLEEPY* transgene by in vivo genetic complementation (Additional file 1: Table S1). Immunofluorescence experiments showed that each fusion protein localizes in the developing new MACs in control conditions (Fig. 2E-G), consistent with their role in MAC development and in agreement with previous work [30, 42]. However, a significant decrease in the nuclear immunofluorescence signal of all three proteins is observed upon depletion of either of the other two members of the complex (e.g. TfIIs4 and Sleepy depletion for FLAG-HA-Firefly in Fig. 2E), as shown by the quantification of the signal in the new MAC (Fig. 2E-G; Additional file 1: Additional file 1: 4C). However, this does not reflect an effect on steady-state protein levels as shown by western blot analysis of whole cell extracts (Fig. 2E-G). Thus, the localization of Firefly, Sleepy and TfIIs4 is strongly interdependent.

To assess whether Sleepy and TfIIs4, like Firefly (Fig. 2B), are bound to chromatin, we isolated nuclear extracts with increasing salt concentrations from cells expressing *FLAG-HA-SLEEPY* or *GFP-TFIIS4* transgenes. Western blot analysis indicated that the tagged Sleepy protein is found in the soluble nuclear extracts at 15 mM and even more at 150 mM NaCl (Additional file 1: Figure S4D), very much like what is observed for Firefly (Fig. 2B), suggesting an important fraction of these proteins is chromatin bound. In contrast, GFP-TfIIs4 is mostly detected in the soluble nuclear extracts at 15 mM NaCl (Additional file 1: Figure S4E), suggesting an important fraction of TfIIs4 is not tightly bound to chromatin. Given its different pattern, it is very likely that TfIIs4 exerts functions outside its interaction with Firefly and Sleepy.

Altogether, the results from our IP/MS, co-sedimentation, fractionation and localization experiments show that Firefly, Sleepy and TfIIs4 are strongly interdependent for their localization in the new MAC and appear to work within the same complex.

Firefly and its partners are required for H3K27me3 accumulation at TEs

Given that Firefly binds to H3K27me3 (Fig. 1) and impacts H3K27me3 accumulation in vivo [42], we examined whether depletion of Firefly, Sleepy, and TfIIs4 has an effect on H3K27me3 during autogamy using immunofluorescence. In control conditions, H3K27me3 is detected in the fragments of the maternal MAC and in the developing new MACs, where it accumulates in nuclear foci, as previously described [17, 28]. In cells depleted for Ezl1, the catalytic subunit of PRC2, the H3K27me3 signal disappears in both nuclei [12, 28]. Similarly, upon *SLEEPY*, *TFIIS4*, or *FIREFLY* KDs, we could not detect any accumulation of H3K27me3 in the developing new MACs (Fig. 3A). Quantification of H3K27me3 fluorescence confirmed the apparent reduction of H3K27me3

accumulation in the developing MACs upon *FIREFLY, SLEEPY* or *TFIIS4* KDs, although the effect is less pronounced than upon *EZL1* KD (Fig. 3A; Additional file 1: Figure S5A). In contrast, H3K27me3 accumulation was still observed in the fragments of the maternal MAC in these KDs, unlike the *EZL1* KD (Fig. 3A). The selective loss of H3K27me3 signal in the developing new MACs made us think of a sleepy face, with closed eyes, which prompted us to name the previously uncharacterized Firefly partner Sleepy. We also examined the effect of depleting Firefly and its partners on H3K9me3 by immunofluorescence. While H3K9me3 is detected in the developing new MACs in control conditions, as previously reported [28], we found that the H3K9me3 signal is significantly

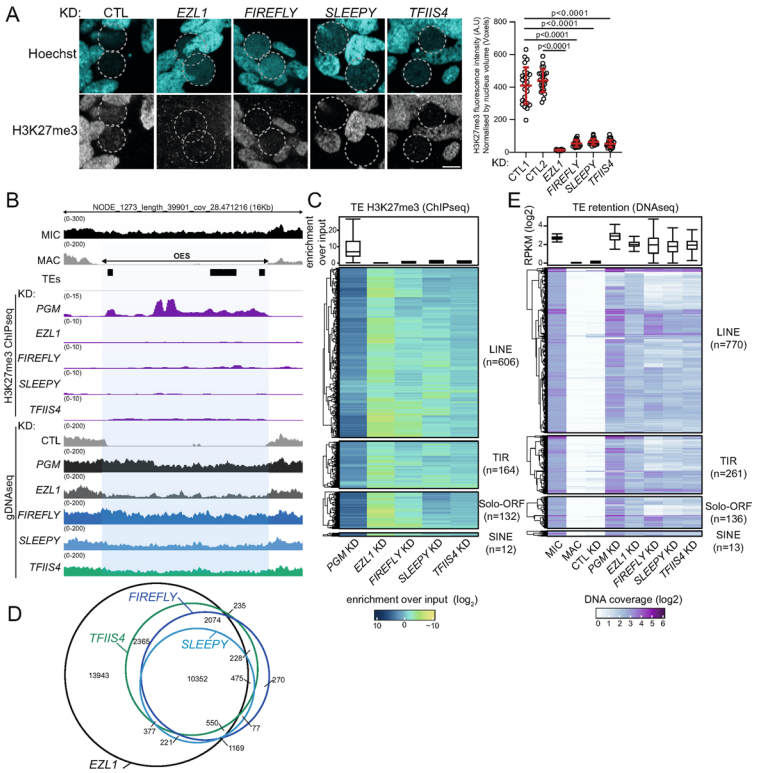

Fig. 3 Firefly and its partners are required for H3K27me3 accumulation and DNA elimination. **A** Immunostaining of H3K27me3 of wild-type cells at T = 10 h after the onset of autogamy upon *ICL7* (CTL1 and CTL2), *EZL1*, *FIREFLY*, *SLEEPY* or *TFIIS4* KD. Developing new MACs are circled with a dotted line. Scale bar, 5 µm. Quantification of H3K27me3 fluorescence signal in the new MAC (see Methods). Bars correspond to mean ± SD. Mann–Whitney statistical tests. **B** Genome browser screenshot of an example of a MIC-limited region containing TEs displaying spike-in normalized H3K27me3 ChIPseq enrichment upon *PGM*, *EZL1*, *FIREFLY*, *SLEEPY* or *TFIIS4* KD. Normalized DNA coverage from whole-genome resequencing of flow cytometry sorted developing new MACs upon *ICL7* (CTL) *PGM*, *EZL1*, *FIREFLY*, SLEEPY or *TFIIS4* KD is also shown. Coordinates are indicated in the figure. **C** Heatmap and boxplot of H3K27me3 enrichment over input as determined by ChIP-seq for the considered TE copies upon *PGM*, *EZL1*, *FIREFLY*, *SLEEPY* or *TFIIS4* KD. The TE copies are ordered by hierarchical clustering of DNA coverage in each family. The boxplots show the coverage distribution (enrichment over input) for all TE copies. The box shows the second and third quartiles. The median is displayed as a horizontal line. The whiskers run from the minimum to the maximum value. **D** Venn diagram of significantly retained IESs from whole-genome resequencing of flow cytometry sorted developing new MACs upon *ICL7* (CTL) *PGM*, *EZL1*, *FIREFLY*, SLEEPY or *TFIIS4* KD. **E** Heatmap of normalized TE DNA coverage on the MAC [28] and MIC (DNA from flow-cytometry sorted MIC) [20] genomes and upon *PGM*, *EZL1*, *FIREFLY*, *SLEEPY* and *TFIIS4* KD sorted new MACs as in B and D. The TE copies are ordered by hierarchical clustering of DNA coverage in each family. The coverage distribution (RPKM: Reads Per Kilobase per Million mapped reads) for all TE copies is shown as a boxplot on top (log_2 scale). The box shows the second and third quartiles. The median is displayed as a horizontal line. The whiskers run from the minimum to the maximum value

reduced upon *SLEEPY* or *TFIIS4* KDs, as also seen for the *FIREFLY* KD (Additional file 1: Figure S5B-C).

To determine H3K27me3 genome occupancy, we devised a spike-in ChIP procedure (Methods) to measure H3K27me3 enrichment upon *FIREFLY*, *SLEEPY* and *TFIIS4* KDs. *PGM* and *EZL1* KDs were used as positive and negative controls, respectively. H3K27me3 enrichment over input was calculated on the MIC reference genome [20] in these different conditions and normalized with *Drosophila* spike-in chromatin (Methods). Upon *SLEEPY* or *TFIIS4* KDs, similarly to *FIREFLY* KD (Fig. 1), we found that H3K27me3 was no longer enriched on OES (Fig. 3B). To examine H3K27me3 enrichment on TEs, we selected TE copies that are well covered (>10 RPKM) in the inputs (606 LINEs, 164 TIRs, 132 SOLO-ORFs and 12 SINEs). These TE copies are enriched in H3K27me3 upon *PGM* KD, and this enrichment is greatly reduced upon *EZL1* KD [12, 17] (Fig. 3C; Additional file 1: Figure S5D). Again, H3K27me3 enrichment is decreased upon *FIREFLY*, *SLEEPY* or *TFIIS4* KDs, although the effect is milder than upon *EZL1* KD (Fig. 3C; Additional file 1: Figure S5D). Altogether, Firefly, Sleepy and TfIIs4 appear to be required for proper H3K27me3 accumulation at TEs.

Firefly and its partners are required for DNA elimination

Given the involvement of Firefly and TfIIs4 in IES elimination [30, 42] and the important role of H3K27me3 in TE elimination [12, 17], we examined whether Sleepy is important for IES and TE elimination. For this, we performed sequencing of genomic DNA extracted from flow-cytometry-sorted developing new MACs (see Methods) upon *SLEEPY* KD, and *ICL7*, *EZL1* and *PGM* KDs as controls (Additional file 1: Figure S6A), when DNA elimination is normally completed (T = 50 h after the onset of autogamy). We also included *FIREFLY* and *TFIIS4* KDs to examine TE elimination. Consistent with previous studies done on genomic DNA extracted from enriched—but not sorted—new MACs, we found that 15,195 IESs (out of 44,928) are significantly retained upon *FIREFLY* KD, similarly to *SLEEPY* and *TFIIS4* KDs (12,098 and 16,129 IESs, respectively (p-value < 0.05)) (Fig. 3D; Additional file 1: Figure S6B-D). The retained IESs belong to the subset of IESs that are retained upon *EZL1* KD (Fig. 3D).

To analyze the effects on the elimination of MIC-limited sequences other than IESs, we calculated the sequencing read coverage of MIC-limited regions in the different KDs and found that these regions are covered upon *FIREFLY, SLEEPY* and *TFIIS4* KDs, as is the case upon *EZL1* and *PGM* KDs (Additional file 1: Figure S6E). All four major TE families are retained upon *FIREFLY, SLEEPY* and *TFIIS4* KDs, although the effect is milder and more heterogeneous among TE copies than upon *PGM* and *EZL1* KDs (Fig. 3E; Additional file 1: Figure S6F).

Given that TE transcript levels are increased upon depletion of PRC2 components [12, 17], we examined whether this is also the case upon depletion of Firefly, Sleepy and TfIIs4. We performed RNA-seq at different developmental stages upon *FIREFLY, SLEEPY* and *TFIIS4* KDs (T = 0, T = 10, T = 35 and T = 50 h after the onset of autogamy) (Additional file 1: Figure S7A-D). Strikingly, while a total of 25% of all annotated TE copies become expressed (>1 RPKM) during MAC development (T = 50 h) upon *EZL1* KD [12], we could not detect any TE expression in cells depleted of Firefly, Sleepy and TfIIs4,

similarly to Pgm-depleted cells (% of TE copies > 1 RPKM: *FIREFLY* 1.3%, *TFIIS4* 0.1%, *SLEEPY* 0.5%, *PGM* 2.5%) (Additional file 1: Figure S7D).

The expression of a few thousand developmental protein-coding genes is deregulated upon *EZL1* KD [12], raising the question of whether depletion of Firefly, Sleepy and TfIIs4 also impacts their expression. We therefore used our RNAseq datasets to evaluate the effect of *FIREFLY*, *SLEEPY* and *TFIIS4* KDs on the expression of genes whose expression is upregulated ($N = 1505$) upon *EZL1* KD. We found that these genes are indeed deregulated upon *FIREFLY*, *SLEEPY* and *TFIIS4* KDs (Additional file 1: Figure S7E). Focusing on the 628 developmental genes that are upregulated upon KD of the elimination machinery (Pgm, Ku80 and Xrcc4) [53], we found that depletion of Firefly, Sleepy and TfIIs4 also affects the expression of these genes (Additional file 1: Figure S7E). Thus, Firefly, Sleepy and TfIIs4 depletion, similarly to depletion of other proteins essential for DNA elimination, results in the deregulation of a distinctive subset of developmental genes. This is likely a response to defective IES elimination in the new MAC [53].

Thus, knockdowns of *FIREFLY*, *SLEEPY* and *TFIIS4* phenocopy each other, consistent with the hypothesis of a functional complex, and indicate that these proteins are required for H3K27me3 accumulation on TEs, DNA elimination and the production of viable sexual progeny.

Firefly, Sleepy, TfIIs4 and Ezl1 are required for nascent transcription of TEs

Given that TfIIs4 is required for the accumulation of noncoding transcripts of some IESs [30], we analyzed nascent transcription by performing a 45-min in vivo pulse labelling during DNA elimination (at T = 9 h after the onset of autogamy) with the cell-permeable uridine analog 5-ethynyluridine (EU). Nascent RNA molecules that contain EU were conjugated to biotin by click chemistry. In order to implement this technique in *Paramecium*, the incorporation of EU in RNA was validated with streptavidin antibodies by dot blot (Additional file 1: Figure S8A). The biotin EU RNAs were then enriched by streptavidin pull-down, and used for RT-qPCR to measure relative nascent RNA levels in control (*ICL7*), *FIREFLY*, *SLEEPY*, *TFIIS4* and *EZL1* KDs. No obvious differences in the relative nascent RNA levels were detected for five protein-coding genes tested in the different KD conditions, indicating that nascent transcription of genes is not dependent on Firefly and its partners, and Ezl1, the PRC2 catalytic subunit Ezl1 (Fig. 4A). In contrast, relative nascent RNA levels for five individual TE copies were reduced by approximately fivefold in *FIREFLY*, *SLEEPY*, *TFIIS4* and *EZL1* KDs compared to control KD (Fig. 4A). We conclude that Firefly, Sleepy, TfIIs4 and PRC2-Ezl1 are all required for nascent transcription of TEs, but not of protein-coding genes.

Because TfIIs4-dependent transcripts are thought to recruit the scnRNA-Ptiwi01/09 complexes, which in turn guide histone modifications via the interaction with PRC2 and its cofactor Rf4 [17], we examined Ptiwi09 localization by immunofluorescence in cells depleted for TfIIs4, Firefly, and Sleepy. Using an immunofluorescence protocol where cells expressing a FLAG-tagged version of Ptiwi09 [17] are subjected to a permeabilization step before fixation (pre-extraction conditions, see Methods), we found that Ptiwi09 accumulates in the developing new MACs in control conditions (Fig. 4B). In contrast,

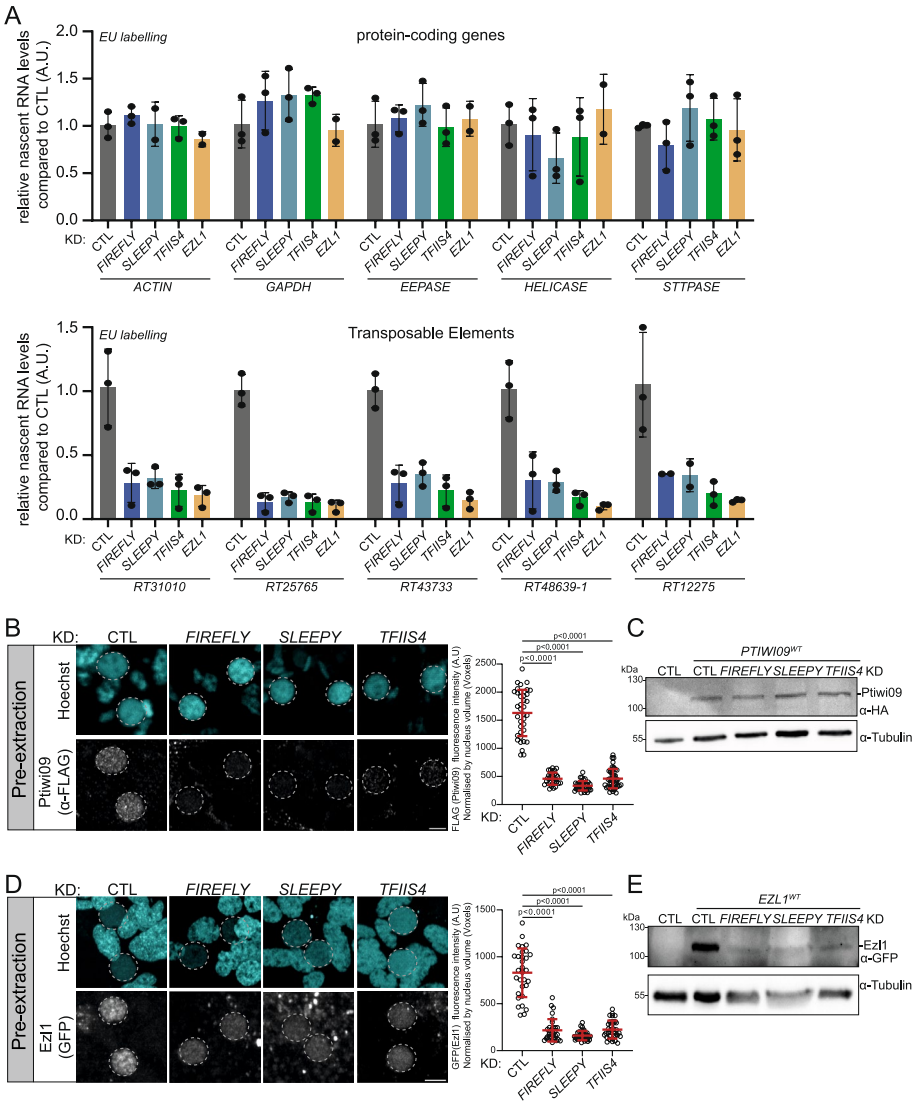

Fig. 4 Firefly, Sleepy, TfIIs4 and Ezl1 are required for nascent transcription of TEs. **A** Relative nascent RNA levels for individual genes or TE copies at T = 10 h after the onset of autogamy upon *ICL7* (CTL), *FIREFLY*, *SLEEPY*, *TFIIS4* or *EZL1* KD as measured by RT-qPCR. Nascent RNA levels of genes and TEs are normalized with nascent RNA levels of the housekeeping *T1b* gene. The relative nascent RNA levels are obtained by dividing these normalized values by the mean of the *ICL7* (CTL) KD replicates. **B** Immunostaining of Ptiwi09 (FLAG) of cells expressing a *FLAG-HA-PTIWI09* transgene at T = 10 h after the onset of autogamy upon *ICL7* (CTL), *FIREFLY*, *SLEEPY* or *TFIIS4* KD. Cells were permeabilized then fixed (Methods). Developing new MACs are circled with a dotted line. Scale bar, 5 µm. Quantification of Ptiwi09 (FLAG) fluorescence signal in the new MAC (see Methods). Bars correspond to mean ± SD. Mann–Whitney statistical tests. **C** Western blot analysis of Ptiwi09 protein levels from whole cell extracts (T = 10 h after the onset of autogamy) in wild-type cells (WT), and cells expressing a *FLAG-HA-PTIWI09* transgene (*PTIWI09*^{WT}) upon *ICL7* (CTL) *FIREFLY*, *SLEEPY* or *TFIIS4* KD. Anti-HA antibodies are used to detect Ptiwi09 and anti-α-tubulin antibodies for normalization. **D** Immunostaining of Ezl1 (GFP) on cells expressing GFP-Ezl1 at T = 10 h after the onset of autogamy upon *ICL7* (CTL), *FIREFLY*, *SLEEPY* or *TFIIS4* KD. Cells were permeabilized then fixed (see Methods). Developing new MACs are circled with a dotted line. Scale bar is 5 µm. Quantification of Ezl1 (GFP) fluorescence signal in the new MAC (see Methods). Bars correspond to mean ± SD. Mann–Whitney statistical tests. **E** Western blot analysis of Ezl1 protein levels from whole cell extracts (T = 10 h after the onset of autogamy) in wild-type cells (WT), and cells expressing a *GFP-EZL1*^{WT} transgene (*EZL1*^{WT}) upon *ICL7* (CTL) *FIREFLY*, *SLEEPY* or *TFIIS4* KD. GFP antibodies are used to detect anti-GFP and anti-α-tubulin antibodies for normalization

the nuclear immunofluorescence signal of Ptiwi09 is lower upon *FIREFLY*, *SLEEPY* or *TFIIS4* KDs (Fig. 4B; Additional file 1: Figure S8B), while no change in total cellular Ptiwi09 levels is observed, indicating that neither Ptiwi09 expression nor its stability are affected in these conditions (Fig. 4C). In conditions where cells are fixed before permeabilization, Ptiwi09 localization remains unchanged in all KD conditions (Additional file 1: Figure S8C), likely because the pre-extraction protocol favors the extraction of proteins freely moving in the nuclear fraction. We hypothesize that the loss of the Ptiwi09 immunofluorescence signal in pre-extraction conditions results from a weaker association of Ptiwi09 with chromatin in the absence of Firefly and its partners, consistent with the idea that nascent transcription mediated by Firefly and its partners is required for Ptiwi09 recruitment to its target sites.

Following the same logic, we investigated the localization of Ezl1, the catalytic subunit of PRC2, using cells expressing a GFP-Ezl1 fusion protein [28]. Similarly to what we observed for Ptiwi09, Ezl1 accumulates in the developing new MACs in control conditions, while the nuclear immunofluorescence signal is greatly reduced upon *FIREFLY*, *SLEEPY* and *TFIIS4* KD in pre-extraction conditions (Fig. 4D; Additional file 1: Figure S8D). When cells are fixed before permeabilization, we observed that Ezl1 localization is similar in all KD conditions (Additional file 1: Figure S8E). As for Ptiwi09, we conclude that the strength of the association of Ezl1 with chromatin is decreased in the absence of Firefly, Sleepy and TfIIs4, a condition where H3K27me3 no longer accumulates in the new MAC (Fig. 3A). We note, however, that steady-state Ezl1 protein levels detected by western blot analysis of whole cell extracts are lower upon *FIREFLY*, *SLEEPY* and *TFIIS4* KD compared to control conditions (Fig. 4E). Given that RNA expression levels of *EZL1* are not diminished upon *FIREFLY*, *SLEEPY* and *TFIIS4* KD compared to control (Additional file 1: Figure S7C), this suggests a reduction in the stability but not the expression of Ezl1. It likely reflects that the PRC2-Ezl1 complex collapses when it is not chromatin-bound.

Altogether, our data indicate that Firefly, Sleepy, and TfIIs4 are required for nascent transcription of TEs and for the nuclear anchoring of Ptiwi09 and of the PRC2 catalytic subunit, Ezl1.

The Firefly chromodomain is required for H3K27me3 accumulation in vivo and TE nascent transcription

Our data indicate that Firefly, Sleepy and TfIIs4 recognize H3K27me3 (Fig. 1), promote nascent transcription of TEs (Fig. 4) and are essential for proper deposition of H3K27me3 (Fig. 3). We hypothesized that a positive feedback loop, mediated by these three proteins, couples nascent transcription of TEs with accumulation of H3K27me3 at TEs. A prediction of our model is that disrupting the ability of Firefly to bind H3K27me3 should prevent nascent transcription of TEs. To test this hypothesis, we used two transgenes expressing Firefly mutant proteins in the conserved chromodomain (Fig. 5A). In addition to the Firefly^{WYAA} mutant that prevents H3K27me3 binding in vitro (Fig. 1E), we generated a mutant transgene (ΔCD), in which the entire chromodomain (between positions 2 and 66) was deleted (Fig. 5A). We performed in vivo genetic complementation assays and showed that the wild-type RNAi-resistant *FLAG-HA-FIREFLY*^{WT} transgene is functional and capable of

rescuing the lethality caused by *FIREFLY* KD and IES elimination, while the mutant transgenes failed to complement (Fig. 5A; Additional file 1: Figure S8F; Additional file 1: Table S1). Furthermore, immunostaining indicated accumulation of H3K27me3 in the developing new MACs in cells expressing the *FLAG-HA-FIREFLY*^{WT} transgene, while this was not the case in cells expressing the mutant transgenes (Fig. 5B; Additional file 1: Figure S8G). Yet similar amounts of the wild-type and mutant fusion proteins were detected by immunostaining or by western blot analysis of whole cell extracts (Fig. 5B-C). Thus, the chromodomain-mutant transgenes are unable to rescue the phenotypes caused by the depletion of endogenous Firefly, indicating that its chromodomain is essential for its function, including H3K27me3 accumulation in the new MAC.

To examine Firefly association to chromatin when it no longer binds H3K27me3, we isolated nuclear extracts at increasing saline concentrations from cells expressing Firefly^{WYAA} or Firefly^{WT} upon *FIREFLY* KD. Western blot analysis revealed that the extraction profile of the chromodomain mutant Firefly^{WYAA} is shifted toward lower saline concentrations compared to the wild-type protein (Fig. 5D). We conclude that

(See figure on next page.)

Fig. 5 The chromodomain of Firefly is required for H3K27me3 accumulation and nascent transcription of TEs. **A** Production of post-autogamous sexual progeny of wild-type cells (CTL), or cells expressing a *FLAG-HA-FIREFLY* transgene (*FIREFLY*^{WT}), cells expressing a *FLAG-HA-FIREFLY* transgene with a W28AY31A mutation (*FIREFLY*^{WYAA}) and cells expressing a *FLAG-HA-FIREFLY* transgene deleted for the sequence corresponding to the chromodomain (*FIREFLY*$^{\Delta CD}$), upon *ICL7* (CTL) or *FIREFLY* KD. The total number of cells analyzed for each KD is indicated in parentheses, as well as the number of independent experiments. **B** Co-immunostaining with FLAG and H3K27me3 antibodies of cells expressing *FIREFLY*^{WT}, *FIREFLY*^{WYAA} or *FIREFLY*$^{\Delta CD}$ upon *FIREFLY* KD at T = 10 h after the onset of autogamy. (Left) Representative confocal images are displayed. Overlay of Z-projections of magnified views of Hoechst staining (top), FLAG- (middle) and H3K27me3- (bottom) specific antibodies are presented. Developing new MACs are circled with a dotted line. Scale bar, 5 µm. (Right) Quantification of Hoechst, FLAG and H3K27me3 fluorescence signal in the new MAC (Methods). Bars correspond to mean ± SD. Mann–Whitney statistical tests. **C** Western blot analysis of Firefly^{WT}, Firefly^{WYAA} or Firefly$^{\Delta CD}$ protein levels from whole cell extracts (T = 10 h after the onset of autogamy) in cells expressing *FLAG-HA-FIREFLY*^{WT}, *FLAG-HA-FIREFLY*^{WYAA} or *FLAG-HA-FIREFLY*$^{\Delta CD}$, respectively, upon *FIREFLY* KD. Anti-HA antibodies are used to detect Firefly^{WT}, Firefly^{WYAA} and Firefly$^{\Delta CD}$ and anti-α-tubulin antibodies for normalization. **D** Western blot analysis of nuclear extracts prepared with increasing salt concentrations (15 mM; 150 mM; 300 mM NaCl) from cells expressing *FLAG-HA-FIREFLY*^{WT} or *FLAG-HA-FIREFLY*^{WYAA} at T = 10 h upon *FIREFLY* KD. Nuclear extracts prepared at 150 mM NaCl from non-injected cells (CTL) are used as a control. Anti-HA antibodies are used to detect Firefly^{WT} and Firefly^{WYAA} proteins and histone H3 antibodies are used for normalization. **E** Immunostaining of cells expressing *FLAG-HA-FIREFLY*^{WT} or *FLAG-HA-FIREFLY*^{WYAA} at T = 10 h after the onset of autogamy upon *FIREFLY* KD. Cells were permeabilized then fixed (pre-extraction) (Methods). Developing new MACs are circled with a dotted line. Scale bar, 5 µm. Quantification of Firefly^{WT} and Firefly^{WYAA} (FLAG) fluorescence signal in the new MAC (Methods). Bars correspond to mean ± SD. Mann–Whitney statistical tests. **F** Western blot analysis of Firefly^{WT} and Firefly^{WYAA} protein levels from whole cell extracts (T = 10 h after the onset of autogamy) in wild-type cells (CTL), and cells expressing *FLAG-HA-FIREFLY*^{WT}or *FLAG-HA-FIREFLY*^{WYAA} upon *FIREFLY* KD, after EU labelling. Anti-HA antibodies are used to detect Firefly^{WT} and Firefly^{WYAA} and anti-α-tubulin antibodies for normalization. **G** Relative nascent RNA levels at individual genes and TE copies at T = 10 h after the onset of autogamy in cells expressing *FLAG-HA-FIREFLY*^{WT}or *FLAG-HA-FIREFLY*^{WYAA} upon *FIREFLY* KD, as measured by RT-qPCR. Nascent RNA levels of genes and TEs are normalized with nascent RNA levels of the housekeeping *T1b* gene. The relative nascent RNA levels are obtained by dividing these normalized values by the mean of the Firefly^{WT} (WT) replicates. **H** Model for the role of Firefly and its partners. A complex composed of Firefly, Sleepy and Tflls4 is required for nascent transcription of TEs. TE transcripts pair with complementary scnRNA-Ptiwi1/9 complexes, which guide PRC2 to TEs where H3K27me3 and H3K9me3 are deposited. The binding of the chromodomain of Firefly to H3K27me3 engages Firefly and its partners in a positive feedback loop that couples H3K27me3 accumulation at TEs and nascent transcription of TEs

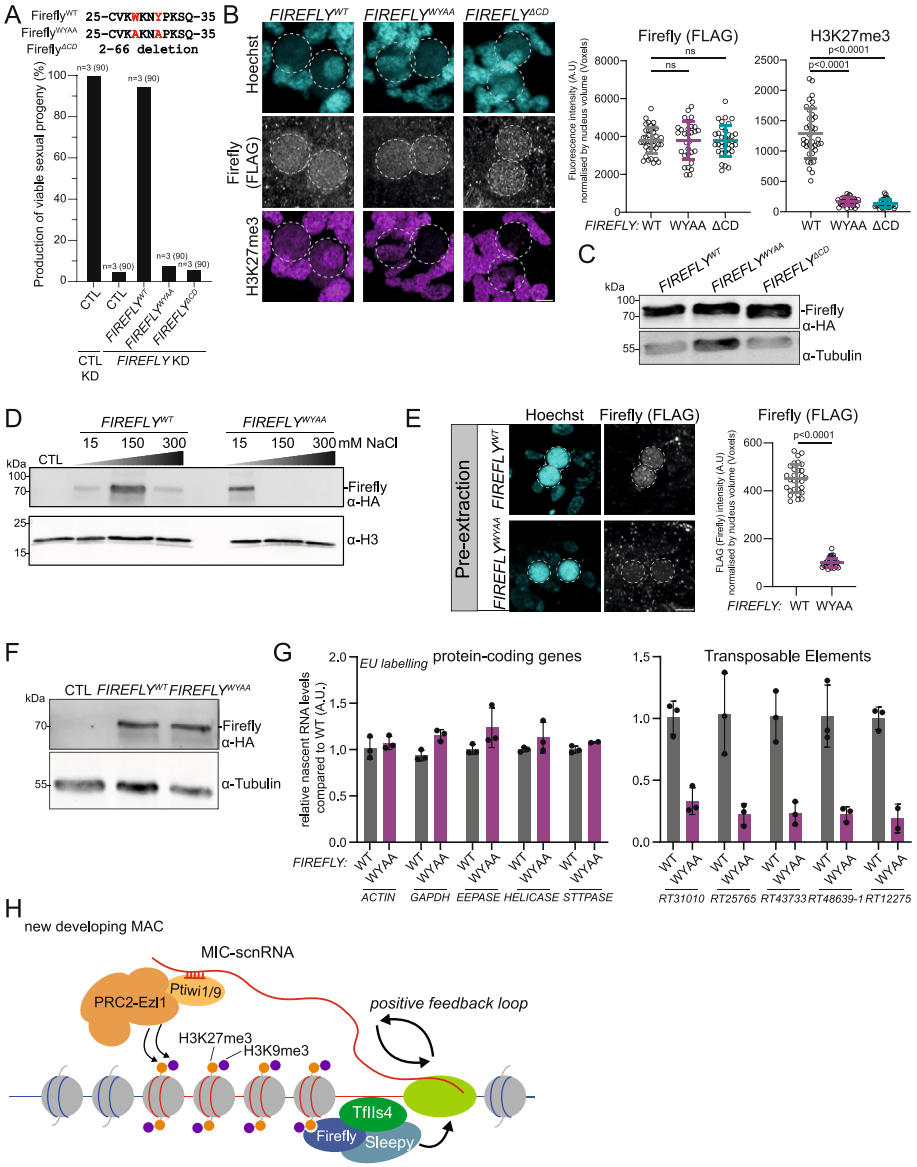

Fig. 5 (See legend on previous page.)

the strength of the association of Firefly with chromatin is decreased in the absence of a functional chromodomain. Furthermore, the nuclear immunofluorescence signal of the chromodomain mutant is greatly reduced compared to the wild-type protein in pre-extraction conditions (Fig. 5E; Additional file 1: Figure S8H). When cells are fixed before permeabilization, we observed that the localization of the mutant Firefly is similar to the wild-type protein (Additional file 1: Figure S8I). Altogether, our data indicate that Firefly binding to chromatin requires a functional chromodomain.

To evaluate nascent transcription in conditions where Firefly can no longer bind to H3K27me3, we performed EU labelling at T = 9 h after the onset of autogamy, biotinylation and streptavidin pull-down in cells expressing *FIREFLY*^{WT} or *FIREFLY*^{WYAA} upon *FIREFLY* KD. Similar levels of the wild-type and mutant proteins were detected

by western blot analysis (Fig. 5F), and genetic complementation was observed with the expression of the wild-type protein, but not with the mutant protein (Additional file 1: Figure S8J; Additional file 1: Table S1), as expected (Fig. 5A). Comparable relative nascent RNA levels were detected for the five protein-coding genes tested in presence of the wild-type or of the mutant Firefly proteins (Fig. 5G). Thus, the mutation in the chromodomain of Firefly does not affect nascent transcription of coding genes, consistent with Firefly, Sleepy and TfIIs4 being dispensable (Fig. 4A). In contrast, the relative nascent RNA levels of five individual TEs were reduced ~ fivefold in cells expressing the Firefly^{WYAA} mutant protein compared to the wild-type control condition (Fig. 5G). We therefore conclude that Firefly binding to H3K27me3 is required for nascent transcription of TEs.

Discussion

Previous studies have shown that H3K27me3 and H3K9me3, which are both catalyzed by PRC2-Ezl1 [12], are required for the elimination of TEs during development in *P. tetraurelia.* Yet, the underlying molecular mechanisms considering that PRC1 appears to be absent in this organism remain elusive [11, 17]. To address this question, we identify proteins that act downstream of H3K27me3. Our affinity purification of recombinant H3K27me3-modified nucleosomes uncovered the chromodomain-containing protein Firefly, which we show is a H3K27me3 reader (Fig. 1). Co-immunoprecipitation experiments and co-sedimentation assays with recombinant proteins show that Firefly appears to form a stable complex with the coiled-coil protein Sleepy and the transcription elongation factor TfIIs4 [30] (Fig. 2). The three partners are required for the accumulation of H3K27me3 at TEs (Figs. 1 and 3), and for their elimination during development of the somatic genome (Fig. 3). We show that nascent transcription of TEs requires the three proteins, and PRC2-Ezl1, and that point mutations of the chromodomain of Firefly mitigate this nascent transcription (Figs. 4 and 5). We thus demonstrate that the binding of Firefly, and by extension its partners, to H3K27me3 couples nascent transcription of TEs and accumulation of H3K27me3 at TEs. This contrasts with the reported opposition between transcription and H3K27me3 described in mammals [54, 55] and suggests that PRC2 might obey different regulatory rules in ciliates and mammals.

Our results also contrast with the prevailing hypothesis that histone modifications tether the elimination machinery to TEs [18, 24, 56]. Indeed, no protein belonging to the elimination machinery (Pgm, its partners PgmL or NHEJ components) was found enriched in the H3K9me3- or H3K27me3- nucleosome pull-downs. This raises the question of the functional relationship between the histone modifications and the elimination machinery, a question that will require further investigation.

Intriguingly, whereas readers of H3K9me3 have been identified in most of the eukaryotic models tested to our knowledge (see for instance Bartke et al., [57]), here very few proteins were significantly enriched with H3K9me3 nucleosomes compared to control unmodified nucleosomes. In particular, no chromodomain-containing proteins were identified, even though the expression of most of the 37 chromodomain-containing proteins encoded in the somatic genome is upregulated during DNA elimination and one of them was recently reported to bind to H3K9me3 human peptides in vitro (Additional file 1: Figure S2) [42]. One possible explanation is that, in *P. tetraurelia,* H3K9me3

readers have weak affinity for the marks, making them difficult to capture in our assay. Alternatively, in *P. tetraurelia,* H3K9me3 may be recognized in combination with other post-translational modifications that are not present on the recombinant nucleosomes used for affinity purification. H3K9me3 may also be recognized in combination with the nucleosome backbone. The use of recombinant hybrid (human/*Paramecium)* nucleosomes in our experiments may mitigate the affinity of readers for H3K9me3. It is also possible that the chromodomain proteins bind to other post-translational modifications of H3 or to nucleic acids [58–61]. However, while we cannot rule out technical limitations in identifying H3K9me3 readers, our data suggests that H3K27me3 plays a predominant role in TE regulation and elimination in *P. tetraurelia.*

According to the current model to explain how TEs are targeted for histone mark deposition, nascent TfIIs4-mediated transcripts serve as a recruitment platform for scnRNA-Piwi complexes. These complexes interact with PRC2, which, in turn, guide the deposition of H3K27me3 and H3K9me3 at TEs, eventually leading to the removal of the marked chromatin by the elimination machinery. Our demonstration of a coupling between H3K27me3 accumulation and nascent transcription of TEs (Figs. 3, 4 and 5) suggests that the establishment of the histone marks at TEs by PRC2-Ezl1 is followed by an additional amplification step allowing the accumulation of the marks at the same loci through the action of a positive feedback loop depending on Firefly and its partners (Fig. 5). Indeed, our data show that accumulation of H3K27me3 involves the three partners, Sleepy, Firefly and TfIIs4 (Fig. 3) and the binding of Firefly via its chromodomain to H3K27me3 (Figs. 1 and 5).

In support of the proposed model (Fig. 5H), the reduced nuclear anchoring of Ptiwi09 and Ezl1, the catalytic subunit of PRC2, observed when interfering with Firefly, Sleepy and TfIIs4 (Fig. 4B-E), is consistent with the idea that nascent transcription mediated by these three proteins is required for tethering Ptiwi09 and PRC2 to their target sites on chromatin. We imagine that the Firefly-Sleepy-TfIIs4 complex allows the RNA-mediated tethering of Ptiwi09 and PRC2. While previous work reported that TfIIs4 is required for the production of transcripts corresponding to several IESs [30], we demonstrate here that this protein is necessary for nascent transcription of TEs and their elimination, challenging previous speculation of distinct transcriptional status of IESs and TEs [42]. In addition, we note that the chromatin remodeler ISWI1 is required for the elimination of TEs (Additional file 1: Figure S6F), and not only for a subset of IESs as previously reported [62].

Depletion of Firefly, Sleepy or TfIIs4 has a milder and more heterogeneous effect on TE elimination than Ezl1 depletion (Fig. 3E; Additional file 1: Figure S6F). Furthermore, while Ezl1 depletion results in complete loss of H3K27me3 in immunofluorescence experiments (Fig. 3A), we found that residual H3K27me3 was detected in the new MAC upon *FIREFLY*, *SLEEPY* or *TFIIS4* KD (Fig. 3A). Similarly, H3K27me3 enrichment on TEs is reduced but not completely abolished as in Ezl1 depletion, as measured by spike-in normalized ChIP-seq (Fig. 3C). The phenotypes observed in Ezl1-depleted cells are also not the same as those in cells depleted for the three partners when looking at IESs. Indeed, PRC2 depletion results in the retention of 69% of IESs (Additional file 1: Figure S6B; and [17]), while only approximately 30% of IESs are retained upon the depletion of Firefly, Sleepy or TfIIs4 (Additional file 1: Figure S6B). Development of the new MAC

involves successive DNA endoreplication cycles to reach a final ploidy of around 1,600. Most IESs are eliminated within one endoreplication cycle (32C to 64C) and display a temporal program of elimination with short IESs excised earlier than longer ones [39]. We note that IESs that are correctly eliminated independently of Firefly/Sleepy/TfIIs4 belong to the "very early" and "early" excision timing (Additional file 1: Figure S6D) [39], suggesting that sequences that are eliminated early may depend less on these proteins than on PRC2.

As presented in the model (Fig. 5H), we propose that the initial deposition of H3K27me3 onto TE chromatin solely relies on PRC2 histone methyltransferase activity, with no need for the mark to be recognized by Firefly. In contrast, the amplification loop that leads to H3K27me3 accumulation at TEs seems to depend both on PRC2 and on the chromodomain of Firefly. The weaker phenotypes observed in conditions of depletion of Firefly/Sleepy/TfIIs4 compared to Ezl1 depletion are entirely consistent with the differential requirements of PRC2 and H3K27me3 reading ability in the initiation and amplification steps of our model (Fig. 5H). In particular, the residual methylation upon Firefly/Sleepy/TfIIs4 depletion likely reflects a basal PRC2 activity independent of the feedback loop. The amplification loop revealed in our study is possibly linked to the endoreplication cycles that occur during MAC development. As H3K27me3 is deposited on newly replicated DNA, the amplification loop mediated by the H3K27me3 reader would allow H3K27me3 accumulation to keep pace as endoreplication proceeds. Elimination of TEs begins after 64C [39]. Consistent with this model, the TEs less affected upon depletion of Firefly/Sleepy/TfIIs4 compared with depletion of Ezl1 appear to be those that are eliminated earlier, and undergo fewer endoreplication cycles (Additional file 1: Figure S6F). Firefly/Sleepy/TfIIs4 may therefore contribute to the maintenance of H3K27me3 domains across multiple rounds of endoreplication, a process that is essential to establish the correct chromatin environment to ensure TE elimination from the somatic nucleus and cell survival.

Such self-reinforcing regulatory loops leading to accumulation of histone H3 lysine methylation at TEs and other repeats have been described in fungi, plants and animals (reviewed in [24, 63, 64]). In these organisms, small RNA-guided chromatin silencing involves the binding of small RNA/Piwi complexes to complementary nascent transcripts, which triggers the recruitment of histone methyltransferases to the targeted locus. In one of the best-studied cases, the yeast *Schizosaccharomyces pombe*, constitutive heterochromatin is established at the pericentromeric DNA repeats and positive feedbacks mediate the amplification of small RNA populations and H3K9 methylation and ensure maintenance of heterochromatin. In particular, the chromodomain of Chp1 binds to H3K9 methylation and Chp1 itself is required for methylation of histone H3 on lysine 9 [65]. Similarly, in animal gonads, piRNAs, which originate from transposon-rich genomic loci termed piRNA clusters, repress the transcription of TEs in the nucleus. In *Drosophila* germ cells, H3K9me3 accumulates at piRNA clusters, and H3K9me3 is bound by the chromodomain HP1 variant Rhino [66]. Rhino is recruited at many sites by the zinc finger domain protein, Kipferl [67], while at Kipferl-independent sites it requires dual recognition of H3K27me3 and H3K9me3 for binding [14]. Rhino associated with Deadlock and Cutoff drive the synthesis of piRNA precursors [68–70]. Through positive feedback loops,

piRNA-mediated targeting of histone methylation leads to more piRNA production and reinforces TE repression. The mechanisms appear very similar in *Paramecium*, except that H3K27me3 deposited by PRC2, instead of H3K9me, is guided by small RNAs to TEs [12, 17, 25]. We also note that in contrast to the other feed-forward loops where small RNAs and histone methylation are amplified, the positive feedback loop that amplifies H3K27me3 at TEs does not lead to scnRNA amplification, since KDs of Firefly or TfIIs4 have no impact on scnRNA biogenesis [30, 42].

This work is, to our knowledge, the first robust evidence of a functional association of H3K27me3 with nascent transcription. While H3K9 methylation readers, such as SHH1 in plants and Rhino in *Drosophila*, recruit the transcription machinery to heterochromatin to transcribe small RNA precursor loci [68–72], our work demonstrates that Firefly recognizes H3K27me3 and together with Sleepy and TfIs4 promotes the nascent transcription of TEs (Figs. 4 and 5). While SHH1 recruits the plant-specific RNA polymerase IV [72], and the Rhino-associated protein Deadlock interacts with Moonshiner, a germline-specific paralog of a basal transcription factor IIA (TFIIA) subunit [73], we show that Firefly and Sleepy interact with the transcription elongation factor TfIIs4 (Fig. 2). Thus, coupling of histone mark readers—of either H3K9 or H3K27 methylation—to basal transcription factors appears as a recurring theme to mediate transcription of heterochromatin, as suggested by Andersen et al. 2017 [73].

PRC2 itself is involved in positive feedback loops that mediate the propagation of H3K27me3 on newly assembled nucleosomes. Binding of the mammalian PRC2 subunit EED to H3K27me3 via its aromatic cage stimulates the methyltransferase activity of PRC2 in vitro [74, 75], providing a mechanism for the propagation of H3K27me3. It is believed that, when PRC2 is recruited to appropriate chromatin domains, the presence of pre-existing H3K27me3 marks on neighbouring nucleosomes activates the complex to carry out further methylation of unmodified H3K27. In plants, a chromodomain protein called LHP1 binds to both the PRC2 component RbAp46/48/Nurf55, and H3K27me2/3, thereby linking PRC2 to H3K27me3 in a positive feedback loop to recruit PRC2 to chromatin that carries H3K27me3 [76]. Similarly, in the pathogenic yeast *Cryptococcus neoformans*, the PRC2-associated chromodomain protein that contains a coiled-coil domain Ccc1 binds to H3K27me3 and this binding prevents ectopic deposition of H3K27me3 [52]. Thus, PRC2 typically harbors modules, either encoded within the enzyme complex (EED), or on an associated protein (Ccc1 or LHP1), that recognize H3K27me3, the product of PRC2. This can be important for the local spread of the modification, or the anchoring of the complex to specific sites. Our findings demonstrate a distinct mechanism whereby recognition of the PRC2 product is mediated by a protein that does not belong to nor associates with PRC2. Indeed, there is no evidence that PRC2 and Firefly-Sleepy-TfIIs4 directly interact.

Furthermore, our work unravels the existence of a positive feedback loop involving PRC2 and its H3K27me3 product, in a context where PRC1 is absent. This is distinct from another instance of PRC1-independent function of PRC2 described in mammals, where BAH-containing proteins bind to H3K27me3 and promote gene repression through HDAC recruitment [77, 78]. Also conceptually different, the role of PRC2 is here to promote active nascent transcription, instead of repression.

Conclusions

We discovered a novel protein complex, formed around the chromodomain containing protein Firefly, that reads the histone mark H3K27me3 and promotes active transcription. The complex comprises a transcription elongation factor TfIIs4 and a coiled-coil-containing protein, Sleepy. The complex is required in vivo both for H3K27me3 accumulation and for nascent transcription of TEs.

In addition to this unprecedented interaction between H3K27me3 readers and transcription factors, our work identifies a positive feedback loop that enriches H3K27me3 at TEs. While PRC2 is well known to act through positive feedback loops wherein H3K27me3 allosterically stimulates PRC2 to mediate its own propagation, our results demonstrate a distinct mechanism involving recognition of H3K27me3 by another reader protein as well as transcription of the underlying DNA.

We unravel an unexpected mode of transcriptional regulation by PRC2 that acts independently of PRC1. Our discovery of a positive feedback loop coupling H3K27me3 enrichment to nascent transcription of TEs demonstrates that proper silencing of TEs requires that they first be transcribed. Remarkably, it is PRC2 that promotes this transcriptional activity through a novel H3K27me3 reader complex.

Methods

Paramecium strains cultivation

All experiments were carried out with the entirely homozygous strain 51 of *P. tetraurelia*. Cells were grown in wheat grass powder (WGP) (Pines International) infusion medium bacterized the day before use with *Klebsiella pneumoniae*, unless otherwise stated, and supplemented with 0.8 mg/mL β-sitosterol (Merck). Cultivation and autogamy were carried out at 27 °C as described [79, 80].

Cytological stages monitoring and description

Progression of autogamy was followed by cytology with Hoechst staining. The progression through autogamy is not synchronous in the cell population [81]. The time points refer to hours after T = 0 h (the onset of autogamy) that is defined as 50% of cells are autogamous (approximately 50% have a fragmented maternal MAC), as evaluated by cytological observation.

Gene silencing experiments

Plasmids used for T7Pol-driven dsRNA production in silencing experiments were obtained by cloning PCR products from each gene using plasmid L4440 and *Escherichia coli* strain HT115 DE3, as previously described [82]. Sequences used for silencing of *ICL7a* (pLA29), *EZL1* (pL4440_EZL1_1), *FIREFLYa* + *b* (pTB22), *SLEEPY* (pMG12), *TFIIS4* (pTB11), and *PGM* (pL4440_PGM) were segments 1–580 of PTET.51.1.G0700039 (*ICL7a*); 989–1501 of PTET.51.1.G1740049 (*EZL1*) [28]; 774–1275 of PTET.51.1.G0220149 (*FIREFLYa*); 457–950 of PTET.51.1.G0020347 (*FIREFLYb*); 94–528 of PTET.51.1.G0310084 (*SLEEPY*); 18–1219 of PTET.51.1.G0900102 (*TFIIS4*) [30]; 873–1440 of PTET.51.1.G0490162 (*PGM*) [31]. Preparation of silencing

medium and RNAi during autogamy were performed as described in [31]. Lethality of post-autogamous cells after RNAi was assessed by transferring 30–60 individual post-autogamous cells to standard growth medium. Cells with a functional new MAC were identified as normally growing survivors unable to undergo a novel round of autogamy if starved after ~8 divisions. Cells usually divided one to three times before dying upon *FIREFLYa* + *b* KD as for *SLEEPY* and *TFIIS4* KD, and unlike *EZL1* or *PGM* KD cells, which usually did not divide or only did so once before dying. See Additional file 1: Table S1 for all survival assays.

Transformation with tagged transgenes

For the construction of in-frame *3xFLAG-HA-FIREFLY*^{WT} (pAH38), *3xFLAG-HA-SLEEPY* (pTB29) fusion plasmids, 3xFLAG-HA tags that were codon-optimized for the *P. tetraurelia* genetic code were added to the 3' of the gene for *3xFLAG-HA-FIREFLY*^{WT} or 5' for *3xFLAG-HA-SLEEPY*. As a result, the tag is fused to the C-terminus of Firefly^{WT} or the N-terminus of Sleepy. The fusion proteins are expressed under the control of their endogenous regulatory regions (promoter and 3'UTR). *FIREFLY* contains 152-bp upstream and 29-bp downstream of its open reading frame, and *SLEEPY* 515-bp upstream and 221-bp downstream. The *3xFLAG-HA-FIREFLY*^{WT} and *3xFLAG-HA-SLEEPY* fusion transgenes are RNAi-resistant. The 774–1275 DNA fragments of *FIREFLY* and 94–528 DNA fragments of *SLEEPY* coding sequences were replaced with synthetic DNA sequences (Eurofins Genomics) designed to maximize nucleotide sequence divergence with the endogenous genomic loci without modifying the amino acid sequences of the encoded proteins. *3xFLAG-HA-FIREFLY*^{WYAA} (pTB21) and *3xFLAG-HA-FIREFLY*$^{\Delta CD}$ (pTB18) were derived from *3xFLAG-HA-FIREFLY*^{WT} to either obtain the W28A Y31A mutant (*3xFLAG-HA-FIREFLY*^{WYAA}) and a deletion of the chromodomain (amino acids 2 to 66, nucleotides 3 to 197). *3xFLAG-HA-FIREFLY*^{WT}, *3xFLAG-HA-FIREFLY*^{WYAA}, *3xFLAG-HA-FIREFLY*$^{\Delta CD}$, *3xFLAG-HA- SLEEPY*, *GFP-TFIIS4* (pMG2) [30] *and GFP-EZL1* (pCM10) [28] were linearized by XmnI, *3xFLAG-HA-PTIWI09* (pAH30) [17] by SpeI, for microinjection into the MAC of vegetative cells. No lethality was observed in the post-autogamous progeny of injected cells, indicating that none of the fusion constructs interfered with the normal progression of autogamy.

Serial nuclear extraction

Paramecium nuclear protein extracts were performed as previously described [12]. Autogamous cells (T = ~10 h) were lysed with a Potter–Elvehjem homogenizer in 3 volumes of lysis buffer (10 mM Tris pH 6.8, 10 mM MgCl2, 0.2% Nonidet P-40, 1 mM PMSF, 4 mM benzamidine, 1 × Complete EDTA-free Protease Inhibitor Cocktail tablets (Roche)). The nuclei-containing pellet was collected by centrifugation and washed with the addition of 2.5 volumes of washing solution (0.25 M sucrose, 10 mM MgCl2, 10 mM Tris pH 7.4, 1 mM PMSF, 4 mM benzamidine, 1 × Complete EDTA-free Protease Inhibitor Cocktail tablets (Roche)). The pellet was incubated in 1 volume of nuclear extraction buffer 2X (100 mM Hepes pH 7.8, 100 mM KCl, 30 mM NaCl, 0.2 mM EDTA, 20% Glycerol, 2 mM DTT, 0.02% Nonidet P-40, 2 mM PMSF, 2 × Complete EDTA-free Protease Inhibitor Cocktail tablets (Roche)) for 1 h at 4 °C. The salt-extractable fraction at 15 mM NaCl was recovered following centrifugation for 3 min at 10,000 g at 4 °C. The pellet was

resuspended in nuclear extraction buffer 2X (same as above, 300 mM NaCl) and incubated for 1 h at 4 °C. The salt-extractable fraction at 150 mM NaCl was recovered following centrifugation for 3 min at 10,000 g at 4 °C. The pellet was resuspended in nuclear extraction buffer 2X (same as above, 600 mM NaCl) and incubated for 1 h at 4 °C. The salt-extractable fraction at 300 mM NaCl was recovered following centrifugation for 3 min at 10,000 g at 4 °C. The remaining pellet corresponds to the insoluble fraction.

Nucleosome pull-downs

Nucleosome pull-downs were performed as described in [37]. Briefly, C-terminal thioester peptides corresponding to *Paramecium tetraurelia* histone H3 (H3P1, PTET.51.1.P1520006) tail that were either unmodified (ARTKQTARKSTAGNKKPTKHLATKAARKTAPAVGA-S-Bzl), K9 trimethylated (ARTKQTARK(me3)STAGNKKPTKHLATKAARKTAPAVGA-S-Bzl) or K27 trimethylated (ARTKQTARKSTAGN KKPTKHLATKAARK(me3)TAPAVGA-S-Bzl) (Cambridge Peptides) were fused to *H. sapiens* histone H3.1 Δ1–31 T32C by native chemical ligation. The full length H3 was then used with recombinant *H. sapiens* histones H4 (P62805), H2A (Q6FI13) and H2B (P33778) to form octamers. Biotinylated 601 dinucleosomal DNA was used for di-nucleosome formation. The di-nucleosomes were immobilized on streptavidin-coated beads (Cytiva, GE17-5113–01) and incubated for 4 h at 4 °C with *Paramecium* nuclear extracts prepared at 300 mM NaCl at T = 10 h after the onset of autogamy, as described in [12]. Beads were then washed 4 times in pull-down buffer (20 mM HEPES–KOH pH7.9, 150 mM NaCl, 0.2 mM EDTA, 10% glycerol).

Recombinant protein production and purification

Codon optimized GST (pGEX-2T), GST-Cbx2^{CD} (*M. musculus*, P30658, Cbx2 2–63, pTB7), and GST-Cbx5^{CD} (*M. musculus*, Q61686, Cbx5 2–66, pTB6), were cloned into pGEX-2T expression vector. After transformation into BL21 (DE3), bacteria were plated on LB + ampicillin (100 µg/mL) + chloramphenicol (25 µg/mL) for incubation overnight at 37 °C. The next day, 1–2 colonies were picked into 1 mL LB + AMP + CAM and incubated at 37 °C with shaking for 2 h, then diluted 1:1000 into 1L LB + AMP + CAM at 37 °C with shaking. After reaching OD = 0.5, cells were put on ice to cool down before 30 min incubation at 25 °C with shaking. Then IPTG (0.2 mM final) was added to induce protein expression for 2 h at 25 °C with shaking. Cells were collected by centrifugation at 5000 g for 10 min at 4 °C, washed three times in Phosphate-buffered saline (PBS) before freezing in liquid nitrogen for storage at −80 °C. Cells were lysed by sonication in lysis buffer (300 mM NaCl, 20 mM Tris pH8, 0.5 mM EDTA, 0.5% NP-40) and the supernatant was recovered after centrifugation at 12,000 g for 10 min at 4 °C. Codon optimized full length GST-Firefly^{WT} (pLA9) and GST-Firefly^{WYAA} (pLA32) were cloned in pFastBac (pLA9). Baculovirus infected Sf9 cells were grown in Insect XPRESS (Lonza) supplemented with 2%FBS, Penn/Strep and fungizone. Infected cells were lysed in benzonase lysis buffer (20 mM Tris pH8, 300 mM NaCl, 0.1% NP40, 10% Glycerol, 1 mM DTT, 1 mM PMSF, 1 × Complete EDTA-free Protease Inhibitor Cocktail tablets (Roche), 10 mM MgCl2, 100 kU benzonase (E1014, Millipore) for 45 min at 4 °C on a wheel and the supernatant was recovered after centrifugation at 12,000 g for 10 min at 4 °C.

The supernatant was incubated with Glutathione Sepharose ™ 4B (Cytiva, 17–0756-01) for 2 h at 4 °C under agitation before being washed 5 times in washing buffer (300 mM NaCl, 20 m M Tris pH8, 0.5 mM EDTA) and 10 elution fractions were collected using the elution buffer (50 mM Tris pH 8, 300 mM NaCl, 10 mM L-Glutathione reduced (Sigma, G4251)). The fractions containing the recombinant protein were determined by coomassie staining then combined before dialysis 3 times against dialysis buffer (1xPBS, 5% glycerol).

Firefly (pLA8), Sleepy (pLA7) and TfIIs4 (pLA6) were cloned in pFastBac through regular cloning strategy. Sf9 cells were grown in Insect XPRESS (Lonza) supplemented with 2%FBS, Penn/Strep and fungizone. Infected cells were lysed in BC300 (25 mM Tris pH 7.9, 1 mM EDTA, 10% glycerol, 300 mM KCl, 0.1 mM PMSF + Protease inhibitors, 1 mM DTT) and sonicated and supernatants were incubated on anti-FLAG M2 agarose beads or Strep-TactinXT 4Flow High Capacity Resin overnight at 4 °C. Beads were washed 3 times with BC300 and eluted in BC300 + 0.2 mg/ml FLAG-peptide or BC300 + 50 mM Biotin.

Density sedimentation assay

Density sedimentation assay was performed as described in [49]. Total cell lysate (500 µg) of Sf9 cells co-expressing Firefly, Sleepy and TfIIs4 was resuspended in 10% glycerol HEMG buffer (20 mM Tris pH8, 300 mM NaCl, 1 mM EDTA, 1 mM DTT) (supplemented with protease inhibitors) and deposited atop a 11 mL 10–30% glycerol (in HEMG buffer) gradient prepared in a 14 × 89 mm polyallomer centrifuge tube (Beckman Coulter, No. 331327). Tubes were centrifuged in an SW40 rotor at 4 °C for 16 h at 40,000 rpm. Fractions (0.550 mL) were collected and used for western blot analyses.

Peptide pull-down

Peptide pull-downs assays were performed by incubating 1 µg of purified GST, GST-Cbx2, GST-Cbx5, GST-Firefly or GST-Firefly^{WYAA} with 1.25 µg of *Paramecium* histone H3 biotinylated peptides (H3K9un: ARTKQTARKSTAGNKKPTK – biotin, H3K9me3: ARTKQTARK(me3)STAGNKKPTK – biotin, H3K27un: HLATKAARKTAPAVGATGG K –biotin, H3K27me3: HLATKAARK(me3)TAPAVGATGG K –biotin, Eurogentec) immobilized on streptavidin beads (65,001, Invitrogen) overnight at 4 °C in binding buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.1% NP-40, 1 mM DTT, 1 mM PMSF, 1 × Complete EDTA-free Protease Inhibitor Cocktail tablets (Roche)). Beads were washed 4 times with binding buffer before elution in 1 × Sample buffer at 95 °C for 5 min.

Immunoprecipitation

For Firefly immunoprecipitation, *Paramecium* nuclear protein extracts were performed as previously described [12]. Nuclear extracts obtained at 150 mM NaCl were incubated overnight at 4 °C with 150 µl anti-FLAG M2 magnetic beads (M8823, Sigma) that were pre-washed with 1 mL TEGN buffer (20 mM Tris pH 8, 0.1 mM EDTA, 10% Glycerol, 150 mM NaCl, 0.01% Nonidet P-40). Beads were washed five times with TEGN buffer before elution with 3xFLAG peptide (F4799, Sigma-Aldrich) at 4 °C for 5 h.

Western blot and silver staining

For western blot, electrophoresis and blotting were carried out according to standard procedures. Samples were run on 10% Tris–Glycine Biorad gels. Blotting was performed overnight using a nitrocellulose (GE10600002, Merck). FLAG (1:1,000) (MAI-91878, Thermo Fisher Scientific), *Paramecium* H3 (1:5,000) [12], (1:4,000), α-tubulin (1:5,000) (05–829, Merck), α-GFP (1:20,000) (ab290, Abcam) or HA (1:4,000) (H6908, Merck) were used for primary antibodies. Secondary horseradish peroxidase-conjugated anti-mouse (W4021, Promega) or anti-rabbit IgG antibodies (W4011, Promega) were used at 1:2,500 or 1:8,000 dilution followed by detection by ECL, or fluorescent secondary antibodies anti-Rabbit at 1:5000 dilution (12,004,161, Biorad).

Silver staining was carried out with SilverQuest (LC6070, Invitrogen) according to the manufacturer's instructions.

Mass spectrometry

Sample preparation

Beads from dinucleosome pull-down experiments were incubated overnight at 37 °C with 40 µL of 50 mM NH4HCO3 buffer containing 1 µg of sequencing-grade trypsin/Lys C mix (Promega). For Firefly IP, gel bands were washed with a destaining solution composed of 50/50 ACN/NH4HCO3 50 mM, then reduced with TCEP-HCl 10 mM and alkylated with MMTS 20 mM. Proteins were then in gel-digested overnight at 37 °C with 20 µL of 50 mM NH4HCO3 buffer containing 1 µg of sequencing-grade trypsin/Lys C mix (Promega). The digested peptides were loaded and desalted on evotips provided by Evosep (Odense, Denmark) according to manufacturer's procedure before LC–MS/MS analysis. Samples were analyzed on a timsTOF Pro 2 mass spectrometer (Bruker Daltonics, Bremen, Germany) coupled to an Evosep one system (Evosep, Odense, Denmark) operating with the 30SPD method developed by the manufacturer. Briefly, the method is based on a 44-min gradient and a total cycle time of 48 min with a C18 analytical column (0.15 × 150 mm, 1.9 µm beads, ref EV-1106) equilibrated at 40 °C and operated at a flow rate of 500 nL/min. H2O/0.1% FA was used as solvent A and ACN/0.1% FA as solvent B. The timsTOF Pro 2 was operated in DDA-PASEF mode1 over a 1.3 s cycle time. Mass spectra for MS and MS/MS scans were recorded between 100 and 1700 m/z.

Data analysis

MS raw files were processed using PEAKS Online 11 (build 1.9, Bioinformatics Solutions Inc.). Data were searched against the *Paramecium tetraurelia* database downloaded from ParameciumDB website (parameciumtetraurelia_mac_51_annotation_v2.0.protein.fasta, 40,460 entries). For dinucleosome pull-downs, the sequences of the proteins that make up the human nucleosome used for pull-down have also been added to the protein database. Parent mass tolerance was set to 20 ppm, with fragment mass tolerance to 0.05 Da. Specific tryptic cleavages were selected and a maximum of 2 missed cleavages were allowed. The following post-translational modifications were considered for identification: Oxidation (M), Deamidation (NQ), Acetylation (Protein N-term) as variable and Beta-methylthiolation (C) as fixed. Identifications were filtered based on a 1% FDR (False Discovery Rate) threshold at both peptide and protein group levels. Label free

quantification was performed using the PEAKS Online 11 quantification module, allowing a mass tolerance of 10 ppm, a CCS error tolerance of 0.02 and a 0.5-min retention time shift tolerance for match between runs for dinucleosome pull-downs. For Firefly IP, a mass tolerance of 20 ppm, a CCS error tolerance of 0.05 and a 1-min retention time shift tolerance for match between runs were allowed. Protein abundance was inferred using the top N peptide method and TIC was used for normalization. Multivariate statistics on protein measurements were performed using Qlucore Omics Explorer 3.9 (Qlucore AB, Lund, SWEDEN). A positive threshold value of 1 was set to allow a log2 transformation of abundance data for normalization i.e. all abundance data values below the threshold are replaced by 1 before transformation. The transformed data were finally used for statistical analysis i.e. the evaluation of differentially present proteins between two groups using a bilateral Student's t-test.

Chromodomain phylogenetics

To generate a phylogenetic tree of chromodomain containing proteins in *Paramecium*, *Paramecium tetraurelia* (strain 51) protein sequences (ptetraurelia_mac_51_annotation_v2.0.protein.fa), ohnolog annotations and gene expression data were downloaded from ParameciumDB [44]. Additionally, the protein sequence of key heterochromatin proteins were downloaded from the Uniprot database (release 2024_05) [83]: *Homo sapiens*: P83916/Cbx1, Q14781/Cbx2, Q13185/Cbx3, P45973/Cbx5; *Mus musculus*: P83917/Cbx1, P30658/Cbx2, P23198/Cbx3, Q61686/Cbx5; *Drosophila melanogaster*: P05205/HP1a, Q9W396/HP1b, Q9VCU6/HP1c, Q7JXA8/Hp1d/Rhino, P26017/Pc; *Ciona intestinalis*: H2XK44/Cbx1-l; *Arabidopsis thaliana*: Q946J8/LHP1; *Tetrahymena tetrahymena*: I7M1A7/Pdd1; *Schizosaccharomyces pombe*: O60016/CLR4, P40381/SWI6. Finally, hidden markov profiles of the Chromo (PF00385) and Chromoshadow (PF01393) domains were downloaded (November 24 2024) from Pfam [84]. *Paramecium* chromodomain proteins were identified using PF00385 and hmmsearch (HMMER v3.4, hmmer.org) using default parameters and an E-value threshold of 1e-4. Chromodomain sequences were then extracted using custom python scripts (adapted from https://github.com/nickatirwin/phylogenomic_scripts), aligned using MAFFT (v7.520) [85], and processed using iqtree (v2.0.3) [86] and substitution model LG4M and statistical support was calculated using SH-aLRT testing ($n = 1{,}000$ replicates) and visualized in the ITOL web server [87]. To annotate additional domains within chromodomain proteins, a *Paramecium*-tailored chromoshadow domain was first defined by performing an iterative search for PF01393 using HMMER and a database of 14 *Paramecium* proteomes downloaded from ParameciumDB. Following each search, chromoshadow domain sequences were extracted and highly similar sequences were removed by clustering using cd-hit (v4.8.1) [88] and aligned using linsi (MAFFT) to define a new HMM. The number of chromoshadow domain proteins converged at the fifth iteration and the fourth HMM was combined with Pfam-A models [89] and used to identify protein domains using hmmscan (E-value threshold 1e-5).

AlphaFold3 modelling and chromodomain comparison

AlphaFold3-Multimer [90] was used to predict Firefly protein structure on the AlphaFold server (https://alphafoldserver.com/). Protein structure was analyzed

using PyMOL (v2.5.2 http://www.pymol.org/pymol). *Mus musculus* CBX2 (P30658), CBX5 (Q61686) and *Drosophila melanogaster* Pc (P26017) protein sequences were downloaded from Uniprot (release 2025_04). Chromodomains were defined using hmmscan against the Pfam-A database using an E-value threshold of 1e-5 and extracted. The 3D conformation of extracted chromodomains was predicted using the AlphaFold3 webserver on 24 October 2025 [90]. Pairwise alignment and superposition were conducted using Reseek and 3D structure similarity and sequence similarity were reported (AQ and ID scores, respectively) [91]. ChimeraX was used to visualise structure superposition [92].

Immunofluorescence and quantification

Cells were fixed for 30 min in solution I (10 mM EGTA, 25 mM HEPES, 2 mM MgCl2, 60 mM PIPES pH 6.9, PHEM 1X; 1% formaldehyde, 2.5% Triton X-100, 4% sucrose), and for 10 min in solution II (PHEM 1X, 4% formaldehyde, 1.2% Triton X-100, 4% sucrose). Following blocking in 3% bovine serum albumin-supplemented Tris buffered saline-Tween 20 0.1% (TBST) for 10 min, fixed cells were incubated overnight at room temperature under agitation with primary antibodies as follows: rabbit anti-H3K9me3 (1:200) [12], rabbit anti-H3K27me3 (1:1,000) [12], rabbit anti-GFP (1:2,000) (ab290, Abcam) and mouse anti-FLAG (1:200) (MAI-91878, Thermo Fisher Scientific). Cells were labelled with Alexa Fluor 568-conjugated goat anti-rabbit IgG, Alexa Fluor 488-conjugated goat anti-rabbit IgG or Alexa Fluor 568-conjugated goat anti-mouse IgG at 1:500 for 1 h, stained with 1 µg/mL Hoechst for 5–10 min and finally mounted in Citifluor AF2 glycerol solution.

The pre-extraction procedure was performed as follows: cells were permeabilized for 4 min in PHEM 1X with 1% Triton X100 then fixed in PHEM 1X with 2% formaldehyde for 10 min. This procedure was compared with cells fixed for 10 min in PHEM 1X with 2% formaldehyde then permeabilized for 15 min in PHEM 1X with 1% Triton X100. Cells were then incubated with anti-FLAG or anti-GFP antibodies and processed as described above.

Images were acquired using a Zeiss LSM 980 laser-scanning confocal microscope with an Airyscan 2 post–processing treatment and a Plan-Apochromat 63 ×/1.40 oil DIC (420,782–9900) or a Plan-Apochromat 40 ×/1.3 oil DIC (420,762–9800) objective. Z-series were performed with Z-steps of 0.25 µm. Quantification was performed as previously described using ImageJ [93]. The volume of the nucleus (in voxels) was estimated as follows: using the Hoechst channel, the top and bottom Z stacks of the developing MAC were defined to estimate nucleus height in pixels. The equatorial Z stack of the developing MAC was defined, and the corresponding developing MAC surface was measured in pixels. The estimated volume of the developing MAC was then calculated as the product of the obtained nucleus height by the median surface. For each Z stack of the developing MAC, the H3K9me3, H3K27me3, FLAG or GFP fluorescence intensity was measured and corrected using the ImageJ "subtract background" tool. The sum of the corrected fluorescence intensities for all the Z stacks, which corresponds to the total fluorescence intensity, was divided by the estimated volume to obtain the fluorescence intensity per voxel in each nucleus. For each condition, at least 25 nuclei were quantified. Mann–Whitney statistical tests were performed with GraphPad Prism.

Chromatin immunoprecipitation and sequencing

Firefly ChIP experiments were performed using FLAG M2 antibody (F1804, Sigma) with a modified version of the protocol described in [12]: a fixation step with Di(N-succinimidyl) glutarate (DSG) for 45 min in the protein–protein crosslinking buffer (2 mM DSG, 1 mM MgCl2, PBS 1X pH8) under gentle agitation was added before the formaldehyde fixation. H3K27me3 ChIP experiments were performed as previously described [12] with the addition of a spike-in method as described in [94]. *Drosophila melanogaster* chromatin (53,083, Active-Motif) was added to each sonicated *Paramecium* chromatin sample before immunoprecipitation with H3K27me3 and spike-in antibodies (61,686, Active-Motif), according to the manufacturer's instructions. From ChIP-enriched samples and inputs, DNA was extracted with phenol, precipitated with glycogen in sodium acetate and ethanol and resuspended in deionized distilled water.

For H3K27me3 ChIP experiments, 5 ng of immunoprecipitated DNA was used for library construction with the Microplex Library Preparation kit V3 (Diagenode) following the manufacturer's recommendations. Final libraries quality was assessed on an Agilent Bioanalyzer 2100, using an Agilent High Sensitivity DNA Kit. Libraries were pooled and sequenced on an Illumina NextSeq2000 instrument, according to the manufacturer recommendations, in a Paired-End 2×101 pb run at I2BC high-throughput sequencing facility.

For Firefly ChIP-seq, 4 ng of immunoprecipitated DNA was used for library construction with the XGen DNA Lib Prep MC kit (IDT) following the manufacturer's recommendations. Final libraries quality was assessed on an Agilent 4200 TapeStation using a High Sensitivity D5000 DNA ScreenTapes. Paired-End sequencing was performed by Novogene.

DNA extraction and sequencing

DNA for deep sequencing was isolated from post-autogamous cells ($T = \sim 50$ h). Cells were lysed with a Potter–Elvehjem homogenizer in lysis buffer (0.25 M sucrose, 10 mM MgCl2, 10 mM Tris pH 6.8, 0.2% Nonidet P-40). The nuclei-containing pellet was fixed with the fixation buffer (0.25 M sucrose, 10 mM MgCl2, 10 mM Tris pH 7.4, 4% PFA) for 30 min under gentle agitation. The nuclei-containing pellet was washed twice with washing buffer (0.25 M sucrose, 10 mM MgCl2, 10 mM Tris pH 7.4). The nuclear pellet was subjected to flow cytometry sorting. 60,000 to 100,000 new MACs were sorted by a FACSAria Fusion flow cytometer based upon their propidium iodide (DNA content) and SSC characteristics [20] and used for subsequent genomic extraction using three volumes of proteinase K buffer (0.5 M EDTA pH 9, 1% N-lauryl sarcosine sodium, 1% SDS, 1 mg/mL proteinase K). Following overnight incubation at 55 °C, genomic DNA was purified. Genomic DNA quality was assessed on an Agilent Bioanalyzer 2100, using Agilent genomic DNA kit. Genomic libraries preparation and Paired-End sequencing was performed by Novogene.

RNA extraction and sequencing

Total RNA samples were extracted as previously described [12] from 200–400 mL of culture at 500 cells/mL for vegetative cells or at 2,000–4,000 cells/mL at different

time-points during autogamy. Briefly, cells were centrifuged and flash-frozen in liquid nitrogen prior to TRIzol treatment, modified by the addition of glass beads for the initial lysis step.

RNA-seq libraries were prepared with the Illumina mRNA stranded ligation kit according to the manufacturer recommendations starting from 400 ng of total RNA. Libraries were pooled in equimolar proportion and sequenced on a NextSeq2000 instrument (Illumina) on a P2-Paired-end 2×58 pb run at I2BC high-throughput sequencing facility.

Nascent RNA extraction

Nascent RNA labelling was performed as follow: 200 mL of *Paramecium* cells at ~3000C/mL at T = 9 h after the onset of autogamy were labelled in vivo with 0.5 mM 5-ethynyl uridine (EU) (Invitrogen, E10345) for 45 min. Total RNA extraction was performed using TRIzol (Invitrogen). 10 µg of RNA was biotinylated and used for streptavidin pull-down with the Click-iT Nascent RNA Capture Kit (ThermoFisher, C10365) following the manufacturer's instructions. Control of EU incorporation into *Paramecium* RNA was verified by dot blot of biotinylated RNA using HRP-conjugated streptavidin (Millipore, 18–152) as in [95]. First-strand cDNA synthesis of the captured nascent RNA was done using SuperScript VILO cDNA synthesis kit (Invitrogen). cDNA was used for qPCR using standard protocols (Additional file 1: Table S2).

Analysis of sequencing data

Reference genomes and datasets

The sequencing data were mapped on *Paramecium tetraurelia* strain 51 MAC v1 (ptetraurelia_mac_51.fa), MAC+IES v1 (ptetraurelia_mac_51_with_ies.fa) (Arnaiz et al., 2012) and the MIC (ptetraurelia_mic2.fa) reference genomes [20]. The gene annotations v2.0 (ptetraurelia_mac_51_annotation_v2.0.bed), IES annotation v1 (internal_eliminated_sequence_PGM_ParTIES.pt_51.bed) and TE v1.0 annotation (ptetraurelia_mic2_TE_annotation_v1.0.bed) [20] were used in this study. All these files can be downloaded from ParameciumDB [44] or can be found in the associated Zenodo repository [48].

Bioinformatics analyses

Sequencing data were demultiplexed using bcl-converter (v4.1.5) and adapters were removed using cutadapt (v3.2). Reads were mapped using Bowtie2 (v2.2.9) [96] on known contaminants (mitochondrial genomes, ribosomal DNA, and bacterial genomes) to remove them. DNA-seq and ChIP-seq data were mapped on reference genomes using Bowtie2 (v2.4.5 –local -X 500), RNA-seq data using Hisat2 [97] (v2.2.1, –rna-strandness FR –minintronlen 20 –max-intronlen 100). Sequencing metrics are available in Additional file 1: Table S3. Only uniquely mapping reads were considered for all types of sequencing data used in this study (samtools v1.15.1 -q 30).

DNA-seq

IES retention was evaluated using ParTIES (MIRET module v1.05 default parameters) [98].

The RPKM coverage on TEs was calculated using the reads counts, determined by bedtools (v2.30.0 multicov –q 10) then normalized by the number of mapped reads on the MIC genome. Only TE copies of a length superior to 500 nucleotides on contigs bigger than 2 kb were considered ($N = 1180$; 770 LINEs, 261 TIRs, 136 Solo-ORF and 13 SINEs). The heatmaps of RPKM (log2) were generated with ComplexHeatmap R package (https://doi.org/10.1093/bioinformatics/btw313).

Using the MIC genome coverage analysis as in [93], the read counts were calculated with bedtools (v2.30.0 multicov -q 30) on MIC non overlapping 1-kb windows. Only the windows corresponding to MIC-limited or MAC-destined (22,521 and 75,106 respectively) regions were considered for this analysis [20]. The counts were normalized (RPKM) using the number of mapped reads on the MIC genome. After removing the outliers, the distribution of the normalized coverage was displayed using violin plot function (ggplot2 R package).

RNA-seq

The RPKM coverage on TEs or genes was calculated using the reads counts, determined by bedtools (v2.30.0 multicov –q 10) then normalized by the number of mapped reads on the MIC genome. Only TE copies of a length superior to 500 nucleotides on contigs bigger than 2 kb were considered ($N = 1180$; 770 LINEs 261, TIRs, 136 Solo-ORF and 13 SINEs). The heatmaps of RPKM (log2) were generated with ComplexHeatmap R package (https://doi.org/10.1093/bioinformatics/btw313).

ChIP-seq

PCR duplicates (samtools rmdup v1.15.1) were removed from the analysis. For Firefly ChIP, each sample was normalized by down-sampling to have equivalent numbers of unique mapping reads to the mic2 genome before enrichment over input was calculated. For H3K27me3 ChIP, reads were mapped to a mic2 + dm6 genome to calculate a spike-in normalization factor for down-sampling for all samples as: ((mic2 reads/dm6 reads)$^{\text{lowest ratio}}$/(mic2 reads/dm6 reads)$^{\text{sample}}$) x (mic2 reads/dm6 reads)$^{\text{Input}}$.

The bedtools multicov (-q 10) software was used to calculate the read coverage of TE copies (only TE copies > 500 nt on MIC contigs > 2 kb and covered by at least 10 RPKM in all input samples were considered). ChIP enrichments were calculated over the corresponding Input. Enrichment heatmaps (log2) were generated with ComplexHeatmap R package.

Alignment files were converted to BigWig using deepTools bamCoverage (v3.5.1 –binSize 1 –smoothLength 100) and enrichment over input was calculated using bigwigCompare (v3.5.1, –operation ratio). Counts on MIC non overlapping 1-kb windows were computed after conversion to bedgraph file using bigWigToBedGraph (v 0.1.0) by bedtools (v2.30.0 map –o mean). Scatter plots were generated using ggplot2 R package.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-026-04045-7.

Additional file 1: Figures S1-S8 and Tables S1-S3. All Supplementary Figures and Tables are cited in the main text.

Acknowledgements
We wish to thank the members of the Duharcourt lab for fruitful discussions, and Daniel Holoch for critical reading of the manuscript. We thank Ludovic Colombeau from Raphaël Rodriguez laboratory (Institut Curie) for his help for the native ligation experiments, and Olivier Arnaiz (I2BC) for his help using ParTIES. We acknowledge the ImagoSeine core facility of the Institut Jacques Monod, member of the France BioImaging infrastructure (https://ror.org/01y7vt929) supported by the French National Research Agency (ANR-24-INBS-0005 FBI BIOGEN) and GIS-IBiSA, Nicolas Valentin from the flow cytometry facility for assistance, and the sequencing and bioinformatics expertise of the I2BC High-throughput sequencing facility, supported by France Génomique (funded by the French National Program "Investissement d'Avenir" ANR-10-INBS-09). We thank Véronique Legros from ProteoSeine facility of Institut Jacques Monod for mass spectromery, Joël Marchand and the IT team for installing and maintaining computational resources, Takayuki Kawaguchi for his advice for the peptide pull down assay, Vincent Maupu-Massamba for medium preparation, and Adeline Humbert for technical support.

Peer review information
Hamish King and Veronique van den Berghe were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Authors' contributions
TB conducted most experiments, with the help of LA for cloning and silencing experiments, and performed the bioinformatic analyses of NGS data. AP generated the baculoviruses, produced the recombinant proteins and analyzed their interaction and elution pattern on glycerol gradients. JRA performed the phylogenetic analyses. BOA and T. Bartke provided the truncated H3 for native ligation, reagents and advices for nucleosome pull-down. BAS provided a full-length Firefly construct for heterologous expression. GC performed the mass spectrometry analyses. TB, RM, SD designed the experiments. RM supervised the recombinant proteins production and interaction. TB prepared the figures and SD together with TB wrote the paper with input from all co-authors. SD supervised the project. All authors read and approved the final manuscript.

Funding
This work was funded by Centre National de la Recherche Scientifique, the Agence Nationale pour la Recherche (ANR) [project "SELECTION" ANR-23-CE12-0027 to SD and GC] and [projects "POLYCHROME" ANR-19-CE12-0015 and "MODIFICATION" ANR-25-CE12-7757 to SD and RM]; the Fondation de la Recherche Médicale "Equipe FRM EQU202203014643" to SD. TB was recipient of doctoral fellowships from BioSPC Doctoral School from Université Paris Cité and Fondation de la Recherche Médicale (FDT202404018139), and a postdoctoral transition fellowship from the EUR G.E.N.E. Graduate School (reference #ANR 17 EURE 0013) that is part of the Université Paris Cité IdEx #ANR 18 IDEX 0001 funded by the French Government through its 'Investments for the Future' program. BOA and T. Bartke were supported by funding from the Deutsche Forschungsgemeinschaft (DFG Project ID 431163844) and the Helmholtz Gesellschaft.

Data availability
All datasets described in this study are publicly available. The sequencing data generated during the current study are available in the European Nucleotide Archive (ENA) at https://www.ebi.ac.uk/ena/browser/view/PRJEB106600 [99]. Previously published datasets used in this study were retrieved from public repositories as described in the original publications [17, 20, 28, 39, 53, 62]. Additional file 1: Table S3 provides information regarding the sequencing data, accession numbers, and mapping statistics for all datasets, either generated in this study or previously published. All bioinformatics scripts used to analyse data in this study have been deposited on GitHub (https://www.github.com/ThomasBalan/Sleepy_paper/tree/v1) [100] under GNU General Public License v3.0 and archived in Zenodo (https://doi.org/10.5281/zenodo.18860375) [101]. Mass spectrometry raw data generated during the current study were deposited in the PRoteomics IDEntifcations database (PRIDE) at https://www.ebi.ac.uk/pride/ under accession number PXD065449 [102]. A table with the statistical analysis of the quantitative proteomics data can be found in Zenodo (https://doi.org/10.5281/zenodo.18267557) [48]. Plasmid maps, uncropped blots, statistical data, and numerical data used in this article can be found in Zenodo (https://doi.org/10.5281/zenodo.18267557) [48]. Raw microscopy images displayed in the figures or used for quantification have been deposited on BioImage Archive (https://doi.org/10.6019/S-BIAD3020) [103].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 28 August 2025 Accepted: 16 March 2026
Published online: 30 March 2026

References

1. Blackledge NP, Klose RJ. The molecular principles of gene regulation by polycomb repressive complexes. Nat Rev Mol Cell Biol. 2021;22:815–33. https://doi.org/10.1038/s41580-021-00398-y.
2. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, et al. Role of histone H3 lysine 27 methylation in polycomb-group silencing. Science. 2002;298:1039–43. https://doi.org/10.1126/science.1076997.
3. Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V. *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal polycomb sites. Cell. 2002;111:185–96. https://doi.org/10.1016/s0092-8674(02)00975-3.
4. Fursova NA, Blackledge NP, Nakayama M, Ito S, Koseki Y, Farcas AM, et al. Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. Mol Cell. 2019;74:1020-1036.e8. https://doi.org/10.1016/j.molcel.2019.03.024.
5. Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev. 2002;16:2893–905. https://doi.org/10.1101/gad.1035902.
6. Müller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, et al. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell. 2002;111:197–208. https://doi.org/10.1016/s0092-8674(02)00976-5.
7. Pengelly AR, Copur Ö, Jäckle H, Herzig A, Müller J. A histone mutant reproduces the phenotype caused by loss of histone-modifying factor polycomb. Science. 2013;339:698–9. https://doi.org/10.1126/science.1231382.
8. Sankar A, Mohammad F, Sundaramurthy AK, Wang H, Lerdrup M, Tatar T, et al. Histone editing elucidates the functional roles of H3K27 methylation and acetylation in mammals. Nat Genet. 2022;54:754–60. https://doi.org/10.1038/s41588-022-01091-2.
9. Min J, Zhang Y, Xu R-M. Structural basis for specific binding of polycomb chromodomain to histone H3 methylated at Lys 27. Genes Dev. 2003;17:1823–8. https://doi.org/10.1101/gad.269603.
10. Wang L, Brown JL, Cao R, Zhang Y, Kassis JA, Jones RS. Hierarchical recruitment of polycomb group silencing complexes. Mol Cell. 2004;14:637–46. https://doi.org/10.1016/j.molcel.2004.05.009.
11. de Potter B, Raas MWD, Seidl MF, Verrijzer CP, Snel B. Uncoupled evolution of the polycomb system and deep origin of non-canonical PRC1. Commun Biol. 2023;6:1–12. https://doi.org/10.1038/s42003-023-05501-x.
12. Frapporti A, Miró Pina C, Arnaiz O, Holoch D, Kawaguchi T, Humbert A, et al. The Polycomb protein Ezl1 mediates H3K9 and H3K27 methylation to repress transposable elements in Paramecium. Nat Commun. 2019;10:2710. https://doi.org/10.1038/s41467-019-10648-5.
13. Déléris A, Berger F, Duharcourt S. Role of polycomb in the control of transposable elements. Trends Genet. 2021. https://doi.org/10.1016/j.tig.2021.06.003.
14. Akkouche A, Kneuss E, Bornelöv S, Renaud Y, Eastwood EL, van Lopik J, et al. Binding of heterochromatin protein Rhino to a subset of piRNA clusters depends on a combination of two histone marks. Nat Struct Mol Biol. 2025. https://doi.org/10.1038/s41594-025-01584-8.
15. Hisanaga T, Romani F, Wu S, Kowar T, Wu Y, Lintermann R, et al. The polycomb repressive complex 2 deposits H3K27me3 and represses transposable elements in a broad range of eukaryotes. Curr Biol. 2023. https://doi.org/10.1016/j.cub.2023.08.073.
16. Hure V, Piron-Prunier F, Yehouessi T, Vitte C, Kornienko AE, Adam G, et al. Alternative silencing states of transposable elements in Arabidopsis associated with H3K27me3. Genome Biol. 2025;26:11. https://doi.org/10.1186/s13059-024-03466-6.
17. Miró-Pina C, Charmant O, Kawaguchi T, Holoch D, Michaud A, Cohen I, et al. Paramecium Polycomb repressive complex 2 physically interacts with the small RNA-binding PIWI protein to repress transposable elements. Dev Cell. 2022;57:1037-1052.e8. https://doi.org/10.1016/j.devcel.2022.03.014.
18. Betermier M, Duharcourt S. Programmed rearrangement in ciliates: *Paramecium*. Microbiol Spectr. 2014. https://doi.org/10.1128/microbiolspec.MDNA3-0035-2014.
19. Arnaiz O, Mathy N, Baudry C, Malinsky S, Aury J-M, Denby Wilkes C, et al. The Paramecium germline genome provides a niche for intragenic parasitic DNA: evolutionary dynamics of internal eliminated sequences. PLoS Genet. 2012;8:e1002984. https://doi.org/10.1371/journal.pgen.1002984.
20. Guérin F, Arnaiz O, Boggetto N, Denby Wilkes C, Meyer E, Sperling L, et al. Flow cytometry sorting of nuclei enables the first global characterization of *Paramecium* germline DNA and transposable elements. BMC Genomics. 2017;18(1):327. https://doi.org/10.1186/s12864-017-3713-7.
21. Sellis D, Guérin F, Arnaiz O, Pett W, Lerat E, Boggetto N, et al. Massive colonization of protein-coding exons by selfish genetic elements in Paramecium germline genomes. PLoS Biol. 2021:19;e3001309. https://doi.org/10.1371/journal.pbio.3001309.
22. Arnaiz O, Guérin F, Couloux A, Miró-Pina C, Pellerin G, Nekrasova I, et al. The tiny germline chromosomes of Paramecium aurelia have an exceptionally high recombination rate and are capped by a new class of Helitrons. bioRxiv; 2025. p. 2025.11.06.686955. https://doi.org/10.1101/2025.11.06.686955. Cited 2026 Jan 8.
23. Swart EC, Wilkes CD, Sandoval PY, Hoehener C, Singh A, Furrer DI, et al. Identification and analysis of functional associations among natural eukaryotic genome editing components. F1000Res. 2017. https://doi.org/10.12688/f1000research.12121.1.
24. Balan T, Lerner LK, Holoch D, Duharcourt S. Small-RNA-guided histone modifications and somatic genome elimination in ciliates. Wiley Interdiscip Rev RNA. 2024;15:e1848. https://doi.org/10.1002/wrna.1848.

25. Charmant O, Gruchota J, Arnaiz O, Nowak KP, Moisan N, Zangarelli C, et al. The PIWI-interacting protein Gtsf1 controls the selective degradation of small RNAs in *Paramecium*. Nucleic Acids Res. 2024. https://doi.org/10.1093/nar/gkae1055.
26. Wang C, Lyv L, Solberg T, Zhang H, Wen Z, Gao F. GTSF1 is required for transposon silencing in the unicellular eukaryote *Paramecium tetraurelia*. Nucleic Acids Res. 2024;52:13206–23. https://doi.org/10.1093/nar/gkae925.
27. Miró-Pina C, Charmant O, Giovannetti M, de Vanssay A, Frapporti A, Humbert A, et al. A histone methyltransferase-independent function of PRC2 controls small RNA dynamics during programmed DNA elimination in *Paramecium*. Nucleic Acids Res. 2025;53:gkaf1048. https://doi.org/10.1093/nar/gkaf1048.
28. Lhuillier-Akakpo M, Frapporti A, Denby Wilkes C, Matelot M, Vervoort M, Sperling L, et al. Local effect of enhancer of zeste-like reveals cooperation of epigenetic and cis-acting determinants for zygotic genome rearrangements. PLoS Genet. 2014;10:e1004665. https://doi.org/10.1371/journal.pgen.1004665.
29. Wang C, Solberg T, Maurer-Alcalá XX, Swart EC, Gao F, Nowacki M. A small RNA-guided PRC2 complex eliminates DNA as an extreme form of transposon silencing. Cell Rep. 2022;40:111263. https://doi.org/10.1016/j.celrep.2022.111263.
30. Maliszewska-Olejniczak K, Gruchota J, Gromadka R, Denby Wilkes C, Arnaiz O, Mathy N, et al. TFIIS-Dependent Non-coding Transcription Regulates Developmental Genome Rearrangements. PLoS Genet. 2015;11:e1005383. https://doi.org/10.1371/journal.pgen.1005383.
31. Baudry C, Malinsky S, Restituito M, Kapusta A, Rosa S, Meyer E, et al. PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia. Genes Dev. 2009;23:2478–83. https://doi.org/10.1101/gad.547309.
32. Bischerour J, Bhullar S, Denby Wilkes C, Régnier V, Mathy N, Dubois E, et al. Six domesticated PiggyBac transposases together carry out programmed DNA elimination in Paramecium. Elife. 2018;7:e37927. https://doi.org/10.7554/eLife.37927.
33. Abello A, Régnier V, Arnaiz O, Le Bars R, Bétermier M, Bischerour J. Functional diversification of Paramecium Ku80 paralogs safeguards genome integrity during precise programmed DNA elimination. PLoS Genet. 2020;16:e1008723. https://doi.org/10.1371/journal.pgen.1008723.
34. Kapusta A, Matsuda A, Marmignon A, Ku M, Silve A, Meyer E, et al. Highly precise and developmentally programmed genome assembly in Paramecium requires ligase IV-dependent end joining. PLoS Genet. 2011;7:e1002049. https://doi.org/10.1371/journal.pgen.1002049.
35. Verron B, Arnaiz O, Zangarelli C, Mathy N, Bétermier M, Bischerour J. The linker region of a development-specific DNA polymerase X ensures efficient repair of programmed DNA double-strand breaks in *Paramecium tetraurelia*. Nucleic Acids Res. 2025;53:gkaf286. https://doi.org/10.1093/nar/gkaf286.
36. Guérineau M, Bessa L, Moriau S, Lescop E, Bontems F, Mathy N, et al. The unusual structure of the PiggyMac cysteine-rich domain reveals zinc finger diversity in PiggyBac-related transposases. Mob DNA. 2021;12:12. https://doi.org/10.1186/s13100-021-00240-4.
37. Tvardovskiy A, Nguyen N, Bartke T. Identifying Specific Protein Interactors of Nucleosomes Carrying Methylated Histones Using Quantitative Mass Spectrometry. In: Margueron R, Holoch D, editors. Histone Methyltransferases: Methods and Protocols. New York, NY: Springer US; 2022. p. 327–403. https://doi.org/10.1007/978-1-0716-2481-4_16. Cited 2023 Apr 26.
38. Lhuillier-Akakpo M, Guérin F, Frapporti A, Duharcourt S. DNA deletion as a mechanism for developmentally programmed centromere loss. Nucleic Acids Res. 2016;44:1553–65. https://doi.org/10.1093/nar/gkv1110.
39. Zangarelli C, Arnaiz O, Bourge M, Gorrichon K, Jaszczyszyn Y, Mathy N, et al. Developmental timing of programmed DNA elimination in *Paramecium tetraurelia* recapitulates germline transposon evolutionary dynamics. Genome Res. 2022. https://doi.org/10.1101/gr.277027.122.
40. Arnaiz O, Van Dijk E, Bétermier M, Lhuillier-Akakpo M, de Vanssay A, Duharcourt S, et al. Improved methods and resources for paramecium genomics: transcription units, gene annotation and gene expression. BMC Genomics. 2017;18:483. https://doi.org/10.1186/s12864-017-3887-z.
41. Yap KL, Zhou M-M. Structure and mechanisms of lysine methylation recognition by the chromodomain in gene transcription. Biochemistry. 2011;50:1966–80. https://doi.org/10.1021/bi101885m.
42. Solberg T, Wang C, Matsubara R, Wen Z, Nowacki M, John Wiley & Sons, Ltd. Heterochromatin-dependent transcription links the PRC2 complex to small RNA-mediated DNA elimination. EMBO Rep. 2025;26:273–96. https://doi.org/10.1038/s44319-024-00332-1.
43. Aury J-M, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, et al. Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. Nature. 2006;444:171–8. https://doi.org/10.1038/nature05230.
44. Arnaiz O, Meyer E, Sperling L. ParameciumDB 2019: integrating genomic data across the genus for functional and evolutionary biology. Nucleic Acids Res. 2020;48:D599-605. https://doi.org/10.1093/nar/gkz948.
45. The Seventh International Meeting on Ciliate Molecular Biology Genetics Nomenclature Committee, Allen SL, Altschuler MI, Bruns PJ, Cohen J, Doerder FP, et al. Proposed genetic nomenclature rules for *Tetrahymena thermophila*, *Paramecium primaurelia* and *Paramecium tetraurelia*. Genetics. 1998;149:459–62. https://doi.org/10.1093/genetics/149.1.459.
46. Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature. 2001;410:120–124. https://doi.org/10.1038/35065138.
47. Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature. 2001;410:116–120. https://doi.org/10.1038/35065132.
48. Balan T, al. Raw data for "A H3K27me3 reader complex couples H3K27me3 accumulation to nascent transcription of transposable elements in Paramecium". Zenodo. [Internet]. 2026. https://doi.org/10.5281/zenodo.18267557
49. Pan J, McKenzie ZM, D'Avino AR, Mashtalir N, Lareau CA, St. Pierre R, et al. The ATPase module of mammalian SWI/SNF family complexes mediates subcomplex identity and catalytic activity–independent genomic targeting. Nat Genet. 2019;51:618–626. https://doi.org/10.1038/s41588-019-0363-5.
50. Gambetta MC, Müller J. O-GlcNAcylation prevents aggregation of the Polycomb group repressor polyhomeotic. Dev Cell. 2014;31:629–39. https://doi.org/10.1016/j.devcel.2014.10.020.

51. Tan F-Q, Wang W, Li J, Lu Y, Zhu B, Hu F, et al. A coiled-coil protein associates Polycomb Repressive Complex 2 with KNOX/BELL transcription factors to maintain silencing of cell differentiation-promoting genes in the shoot apex. Plant Cell. 2022;34:2969–88. https://doi.org/10.1093/plcell/koac133.
52. Dumesic PA, Homer CM, Moresco JJ, Pack LR, Shanle EK, Coyle SM, et al. Product binding enforces the genomic specificity of a Yeast Polycomb Repressive Complex. Cell. 2015;160:204–18. https://doi.org/10.1016/j.cell.2014.11.039.
53. Bazin-Gélis M, Eleftheriou E, Zangarelli C, Lelandais G, Sperling L, Arnaiz O, et al. Inter-generational nuclear cross-talk links the control of gene expression to programmed genome rearrangement during the Paramecium sexual cycle. Nucleic Acids Res. 2023;51:12337–51. https://doi.org/10.1093/nar/gkad1006.
54. Holoch D, Wassef M, Lövkvist C, Zielinski D, Aflaki S, Lombard B, et al. A cis-acting mechanism mediates transcriptional memory at Polycomb target genes in mammals. Nat Genet. 2021;53:1686–97. https://doi.org/10.1038/s41588-021-00964-2.
55. Riising EM, Comet I, Leblanc B, Wu X, Johansen JV, Helin K. Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. Mol Cell. 2014;55:347–60. https://doi.org/10.1016/j.molcel.2014.06.005.
56. Coyne RS, Lhuillier-Akakpo M, Duharcourt S. RNA-guided DNA rearrangements in ciliates: is the best genome defence a good offence? Biol Cell. 2012;104:1–17. https://doi.org/10.1111/boc.201100057.
57. Bartke T, Vermeulen M, Xhemalce B, Robson SC, Mann M, Kouzarides T. Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell. 2010;143:470–84. https://doi.org/10.1016/j.cell.2010.10.012.
58. Akhtar A, Zink D, Becker PB. Chromodomains are protein–RNA interaction modules. Nature. 2000;407:405–9. https://doi.org/10.1038/35030169.
59. Bouazoune K, Mitterweger A, Längst G, Imhof A, Akhtar A, Becker PB, et al. The dMi-2 chromodomains are DNA binding modules important for ATP-dependent nucleosome mobilization. EMBO J. 2002;21:2430–40. https://doi.org/10.1093/emboj/21.10.2430.
60. Ishida M, Shimojo H, Hayashi A, Kawaguchi R, Ohtani Y, Uegaki K, et al. Intrinsic Nucleic Acid-Binding Activity of Chp1 Chromodomain Is Required for Heterochromatic Gene Silencing. Molecular Cell Elsevier. 2012;47:228–41. https://doi.org/10.1016/j.molcel.2012.05.017.
61. Shirai A, Kawaguchi T, Shimojo H, Muramatsu D, Ishida-Yonetani M, Nishimura Y, et al. Impact of nucleic acid and methylated H3K9 binding activities of Suv39h1 on its heterochromatin assembly. eLife Sciences. 2017;6:e25317. https://doi.org/10.7554/eLife.25317.
62. Singh A, Maurer-Alcalá XX, Solberg T, Häußermann L, Gisler S, Ignarski M, et al. Chromatin remodeling is required for sRNA-guided DNA elimination in Paramecium EMBO J. 2022;41:e111839. https://doi.org/10.15252/embj.202211839.
63. Castel SE, Martienssen RA. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat Rev Genet. 2013;14:100–12. https://doi.org/10.1038/nrg3355.
64. Holoch D, Moazed D. RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet. 2015;16:71–84. https://doi.org/10.1038/nrg3863.
65. Partridge JF, Scott KSC, Bannister AJ, Kouzarides T, Allshire RC. Cis-acting DNA from fission yeast centromeres mediates histone H3 methylation and recruitment of silencing factors and cohesin to an ectopic site. Curr Biol. 2002;12:1652–60. https://doi.org/10.1016/s0960-9822(02)01177-6.
66. Klattenhoff C, Xi H, Li C, Lee S, Xu J, Khurana JS, et al. The Drosophila HP1 Homolog Rhino Is Required for Transposon Silencing and piRNA Production by Dual-Strand Clusters. Cell Elsevier. 2009;138:1137–49. https://doi.org/10.1016/j.cell.2009.07.014.
67. Baumgartner L, Handler D, Platzer SW, Yu C, Duchek P, Brennecke J, et al. The *Drosophila* ZAD zinc finger protein Kipferl guides Rhino to piRNA clusters. Elife. 2022;11:e80067. https://doi.org/10.7554/eLife.80067.
68. Le Thomas A, Stuwe E, Li S, Du J, Marinov G, Rozhkov N, et al. Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. Genes Dev. 2014;28:1667–80. https://doi.org/10.1101/gad.245514.114.
69. Mohn F, Sienski G, Handler D, Brennecke J. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. Cell. 2014;157:1364–79. https://doi.org/10.1016/j.cell.2014.04.031.
70. Zhang Z, Wang J, Schultz N, Zhang F, Parhad SS, Tu S, et al. The HP1 homolog Rhino anchors a nuclear complex that suppresses piRNA precursor splicing. Cell. 2014;157:1353–63. https://doi.org/10.1016/j.cell.2014.04.030.
71. Law JA, Vashisht AA, Wohlschlegel JA, Jacobsen SE, Public Library of Science. SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. PLoS Genet. 2011;7:e1002195. https://doi.org/10.1371/journal.pgen.1002195.
72. Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AMS, et al. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature. 2013;498:385–9. https://doi.org/10.1038/nature12178.
73. Andersen PR, Tirian L, Vunjak M, Brennecke J. A heterochromatin-dependent transcription machinery drives piRNA expression. Nature. 2017;549:54–9. https://doi.org/10.1038/nature23482.
74. Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury WJ, et al. Role of the polycomb protein Eed in the propagation of repressive histone marks. Nature. 2009;461:762–7. https://doi.org/10.1038/nature08398.
75. Poepsel S, Kasinath V, Nogales E. Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. Nat Struct Mol Biol. 2018;25:154–62. https://doi.org/10.1038/s41594-018-0023-y.
76. Derkacheva M, Steinbach Y, Wildhaber T, Mozgová I, Mahrez W, Nanni P, et al. Arabidopsis MSI1 connects LHP1 to PRC2 complexes. EMBO J. 2013;32:2073–85. https://doi.org/10.1038/emboj.2013.145.
77. Fan H, Guo Y, Tsai Y-H, Storey AJ, Kim A, Gong W, et al. A conserved BAH module within mammalian BAHD1 connects H3K27me3 to Polycomb gene silencing. Nucleic Acids Res. 2021;49:4441–55. https://doi.org/10.1093/nar/gkab210.
78. Zhao D, Zhang X, Guan H, Xiong X, Shi X, Deng H, et al. The BAH domain of BAHD1 is a histone H3K27me3 reader. Protein Cell. 2016;7:222–6. https://doi.org/10.1007/s13238-016-0243-z.

79. Beisson J, Bétermier M, Bré M-H, Cohen J, Duharcourt S, Duret L, et al. Maintaining clonal *Paramecium tetraurelia* cell lines of controlled age through daily reisolation. Cold Spring Harb Protoc. 2010;2010:pdb.prot5361. https://doi.org/10.1101/pdb.prot5361.
80. Beisson J, Bétermier M, Bré M-H, Cohen J, Duharcourt S, Duret L, et al. Mass culture of Paramecium tetraurelia. Cold Spring Harb Protoc. 2010;2010:pdb.prot5362. https://doi.org/10.1101/pdb.prot5362
81. Berger JD. Autogamy in *Paramecium* cell cycle stage-specific commitment to meiosis. Exp Cell Res. 1986;166:475–85. https://doi.org/10.1016/0014-4827(86)90492-1.
82. Galvani A, Sperling L. Rna interference by feeding in *Paramecium*. Trends Genet. 2002;18:11–2. https://doi.org/10.1016/s0168-9525(01)02548-3.
83. UniProt Consortium. Uniprot: the universal protein knowledgebase in 2023. Nucleic Acids Res. 2023;51:D523–31. https://doi.org/10.1093/nar/gkac1052.
84. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49:D412–9. https://doi.org/10.1093/nar/gkaa913.
85. Nakamura T, Yamada KD, Tomii K, Katoh K. Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics. 2018;34:2490–2. https://doi.org/10.1093/bioinformatics/bty121.
86. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37:1530–4. https://doi.org/10.1093/molbev/msaa015.
87. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52:W78-82. https://doi.org/10.1093/nar/gkae268.
88. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2. https://doi.org/10.1093/bioinformatics/bts565.
89. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, et al. InterPro in 2022. Nucleic Acids Res. 2023;51:D418–27. https://doi.org/10.1093/nar/gkac993.
90. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature Nature Publishing Group. 2024;630:493–500. https://doi.org/10.1038/s41586-024-07487-w.
91. Edgar RC. Protein structure alignment by Reseek improves sensitivity to remote homologs. Bioinformatics. 2024;40:btae687. https://doi.org/10.1093/bioinformatics/btae687.
92. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, et al. UCSF chimerax: tools for structure building and analysis. Protein Sci. 2023;32:e4792. https://doi.org/10.1002/pro.4792.
93. de Vanssay A, Touzeau A, Arnaiz O, Frapporti A, Phipps J, Duharcourt S. The Paramecium histone chaperone Spt16-1 is required for Pgm endonuclease function in programmed genome rearrangements. PLoS Genet. 2020. https://doi.org/10.1371/journal.pgen.1008949.
94. Egan B, Yuan C-C, Craske ML, Labhart P, Guler GD, Arnott D, et al. An alternative approach to ChIP-Seq normalization enables detection of genome-wide changes in histone H3 lysine 27 trimethylation upon EZH2 inhibition. PLoS ONE. 2016;11:e0166438. https://doi.org/10.1371/journal.pone.0166438.
95. Gregersen LH, Mitter R, Svejstrup JQ, Nature Publishing Group. Using ttchem-seq for profiling nascent transcription and measuring transcript elongation. Nat Protoc. 2020;15:604–27. https://doi.org/10.1038/s41596-019-0262-3.
96. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9. https://doi.org/10.1038/nmeth.1923.
97. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907–15. https://doi.org/10.1038/s41587-019-0201-4.
98. Denby Wilkes C, Arnaiz O, Sperling L. ParTIES: a toolbox for*Paramecium* interspersed DNA elimination studies. Bioinformatics. 2016;32:599–601. https://doi.org/10.1093/bioinformatics/btv691.
99. Balan T, al. Sequencing raw data for "A H3K27me3 reader complex couples H3K27me3 accumulation to nascent transcription of transposable elements in Paramecium". ENA. 2026.https://www.ebi.ac.uk/ena/browser/view/PRJEB106600
100. Balan T, al. Custom scripts for sequencing analysis for "A H3K27me3 reader complex couples H3K27me3 accumulation to nascent transcription of transposable elements in Paramecium". Github. 2026. https://github.com/ThomasBalan/Sleepy_paper/tree/v1
101. Balan T, al. Custom scripts for sequencing analysis for "A H3K27me3 reader complex couples H3K27me3 accumulation to nascent transcription of transposable elements in Paramecium". Zenodo. 2026. https://doi.org/10.5281/zenodo.18860375
102. Balan T, al. Proteomics raw data for "A H3K27me3 reader complex couples H3K27me3 accumulation to nascent transcription of transposable elements in Paramecium". PRIDE. 2026. https://www.ebi.ac.uk/pride/archive/projects/PXD065449
103. Balan T, al. Confocal images for for "A H3K27me3 reader complex couples H3K27me3 accumulation to nascent transcription of transposable elements in Paramecium". BioImage Archive. 2026. https://doi.org/10.6019/S-BIAD3020.

Publisher's Note