Safeguarding Drosophila female germ cell identity depends on an H3K9me3 mini domain guided by a ZAD zinc finger protein.

Laura Shapiro-Kulnane and Helen K. Salz

Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH 44106-4955, USA.

Correspondence and requests for materials should be addressed to H.K.S. (email: hks@case.edu)

Abstract

H3K9me3-based gene silencing is a conserved and vital strategy for securing cell fate. But mechanisms controlling the tissue-specific installation of this epigenetic modification are largely unknown. To dissect the molecular mechanism underlying tissue-specific deposition of H3K9me3, we focus on the Drosophila *phf7* gene, whose repression in female germ cells is essential for maintaining female fate. Through our work focused on identifying essential cis-regulatory elements and trans-acting factors, we discovered that *phf7* repression depends on *CG4936*. This gene, which we name identity crisis (*idc*), encodes a zinc finger-associated domain (ZAD) containing C2H2 zinc finger protein. Loss of *idc* in germ cells interferes with H3K9me3 deposition at the *phf7* locus and leads to ectopic PHF7 protein expression. Furthermore, IDC localizes to the *phf7* gene. Collectively, our studies establish that IDC guides the installation of the H3K9me3 repressive marks onto *phf7*, thereby preventing accidental female-to-male programming.

Introduction

Gene silencing is critical to establishing and maintaining cell fates. Once made, the decision to silence a gene is fortified by the acquisition of repressive histone modifications. While cell-type specific epigenetic silencing is primarily associated with tri-methylation of histone H3 lysine 27 (H3K27me3)-marked chromatin, tissue-specific genes can also be repressed by H3K9 methylation (Becker et al., 2016; Nicetto and Zaret, 2019; Ninova et al., 2019; Padeken et al., 2022). For example, in *S. pombe,* discrete H3K9me3 islands silence meiotic genes in vegetative cells [e.g., (Martienssen and Moazed, 2015; Zofall et al., 2012)]. In *C. elegans,* H3K9 methylation silences inappropriate cell type-specific genes, including germline genes, in somatic cells [e.g., (Methot et al., 2021; Rechtsteiner et al., 2019)]. In the *D. melanogaster* female germline, H3K9me3 silences male germline genes (Smolko et al., 2018). In the mouse, H3K9me3 silences germline genes during early embryonic development [e.g., (Karimi et al.,

2011; Mochizuki et al., 2021)]. Studies carried out in mammalian tissue culture systems further identify H3K9me3-mediated gene silencing as a conserved and vital strategy for maintaining cell fates in a wide range of tissues [e.g., (Biferali et al., 2021; Bilodeau et al., 2009; Cheng et al., 2021; Du et al., 2018; Jiang et al., 2017; Koide et al., 2016; Lohmann et al., 2010; McCarthy et al., 2021; Nicetto et al., 2019; Schultz et al., 2002; Tan et al., 2012; Wu et al., 2020)]. But mechanisms controlling tissue-specific installation of this epigenetic modification onto protein-coding genes are largely unknown.

The repressive H3K9me3 histone modification has well-characterized roles in constitutive heterochromatin formation, repression of satellite repeats, and repression of transposable elements (TEs) (Janssen et al., 2018; Marsano and Dimitri, 2022; Wallrath et al., 2022). These studies identified two mechanisms of H3K9me3 recruitment. One mechanism involves small RNAs that guide localization through a complementary base pairing mechanism. In *Drosophila* and mammals, for example, the PIWI-associated small RNAs (piRNAs) guide the H3K9me3 silencing machinery to TEs (Aravin, 2020). A second mechanism involves sequence-specific DNA binding proteins. In mammals, for example, H3K9me3 can be guided to TEs by members of the vertebrate-specific Kruppel-associated box zinc finger (KRAB-ZNF) family of DNA binding proteins (Bruno et al., 2019; Ecco et al., 2017; Senft and Macfarlan, 2021; Yang et al., 2017). An analogous mechanism might exist in Drosophila, as two zinc finger proteins, Kipferl and Sov, were recently shown to have roles in TE silencing (Andreev et al., 2022; Benner et al., 2019; Jankovics et al., 2018).

To study how the H3K9me3 silencing machinery is installed at protein-coding genes, we focused on the Drosophila female germline where H3K9me3-mediated repression of one key testis gene, *PHD finger protein 7* (*phf7*), prevents accidental female-to-male programming (Smolko et al., 2018; Smolko et al., 2020). *phf7* encodes a predicted chromatin reader, first identified in a screen for genes expressed in male but not female embryonic germ cells (Yang,

2012, 2017). Curiously, *phf7* is not essential for male fertility, but it is critical that female germ cells not express the PHF7 protein. Forced expression of PHF7 activates a toxic gene expression program enriched for genes usually restricted to the male germline (Smolko et al., 2020). Even though PHF7 protein is limited to male germ cells, *phf7* is transcribed in both male and female germ cells. Sex-specificity is achieved by alternative transcription start sites (TSS; Figure 1A). In testes, transcription from the upstream TSS (TSS1) produces a long mRNA isoform that makes protein. In ovaries, an H3K9me3 mini-domain prevents the selection of TSS1. Instead, transcription initiates from the downstream TSS (TSS2) to produce a short mRNA isoform that is not translated and has no function. Of the three H3K9 methyltransferases in Drosophila, only SETDB1 has a specific and nonredundant role in repressing *phf7* (Smolko et al., 2018). Germ cell-specific loss of the H3K9 methyltransferase SETDB1, its binding partner ATF7IP, or the H3K9 reader HP1a produced ovarian germ cell tumors that inappropriately express the PHF7 protein. Notably, derepression results from losing the H3K9me3 mini domain. Thus, *phf7* regulation provides a biologically relevant platform for elucidating the regulatory circuitry controlling H3K9me3 installation onto a protein-coding gene.

Here, we address the mechanism of targeted H3K9me3 deposition onto *phf7* by identifying essential cis-regulatory elements and trans-acting factors. This analysis led us to discover that repression depends on *CG4936*. This gene, which we have name *identity crisis* (*idc),* encodes a zinc finger-associated domain (ZAD) containing C2H2 zinc finger protein. Notably, loss of *idc* in germ cells interferes with *phf7* repression by reducing H3K9me3 deposition. Together with the observation that IDC localizes to the *phf7* gene in ovarian extracts, our analysis supports a model in which IDC guides H3K9me3 installation, thereby preventing accidental female-to-male programming.

Results

Identification of cis-regulatory elements required for H3K9me3 deposition

H3K9me3 accumulates over a ~3 kb region within the *phf7* gene (Smolko et al., 2018). This discrete peak covers the male TSS1, the first exon, and most of the first intron. Interestingly, the intron contains seven copies of ~250 bp sequence not found anywhere else in the genome (element R; **Figure 1A**; **Figure S1**). Repetitive elements can regulate gene expression in *cis* by serving directly as H3K9me3 nucleation sites or in *trans* by encoding small RNAs that serve as guides via a base-pairing mechanism. We, therefore, hypothesized that the repeats play a role in H3K9me3 deposition and silencing of TSS1. To test this hypothesis, we used CRISPR to delete the repeats in the endogenous locus (*phf7*$^{\Delta R}$) (**Figure 1A**). We were surprised, however, to discover that this deletion has no mutant phenotype, suggesting that the repeats may not be essential for repression. To evaluate the impact of deleting the repeats in more detail, we measured H3K9me3 accumulation in wild-type and *phf7*$^{\Delta R}$ homozygous ovaries using chromatin immunoprecipitation followed by quantitative PCR (ChIP-qPCR). In wild-type (control) ovaries, sequences in the intron positioned 319 bp downstream of TSS1 (within intronic element A) are significantly enriched in the H3K9me3 modification when compared to the region within the coding sequence (CDS; $p<0.0001$) (**Figure 1B**). Significant H3K9me3 enrichment within element A compared to the CDS region ($p<0.0001$) was also observed in *phf7*$^{\Delta R}$ homozygous ovaries. However, the deletion had a mild impact on H3K9me3 accumulation compared to control ovaries. To determine if the *phf7*$^{\Delta R}$ deletion disrupted sex-specific transcriptional regulation, we used RT-qPCR to assay the testis-specific *phf7-RC* expression levels in gonads. As in wild-type gonads, we could detect *phf7-RC* expression in *phf7*$^{\Delta R}$ homozygous testis but not ovaries (**Figure 1C**). These results indicate that the repeats are not essential for H3K9me3-mediated repression.

To complement these studies, we used a transgenic approach to identify the sequences capable of promoting H3K9me3 to an ectopic location. For these assays, we created transgenic flies that carry portions of the *phf7* locus, beginning at TSS1 and ending at the beginning of the

open reading frame in exon 2, inserted into the same heterologous genomic location on the 3rd chromosome using site-specific integration. To assay for H3K9me3 accumulation on the transgene, but not at the endogenous locus, we crossed each line into a *phf7*$^{\Delta 13}$ background. We measured the level of H3K9me3 deposition by ChIP-qPCR at a region that is present in the transgenes but absent in *phf7*$^{\Delta 13}$ (intronic element B, **Figure 2A**). In agreement with our previous findings, deleting the seven tandem repeats (element R) from the first intron did not interfere with ectopic H3K9me3 accumulation. (**Figure 2A, B**). In contrast, removing elements A and R abolished H3K9me3 accumulation (**Figure 2A, B**). These data demonstrate that cis-regulatory determinants required for H3K9me3 recruitment are within element A.

Conservation of non-coding cis-regulatory elements

Cis-regulatory elements are often conserved. Thus, we might expect the essential sequences within element A might be retained in evolution. To test this prediction, we chose to examine the level of conservation between *D. melanogaster, D. simulans,* and *D. yakuba.* Although *D. simulans* and *D. yakuba* are separated from *D. melanogaster* by 5-15 million years, the *phf7* intron/exon structure is conserved (**Figure 3A**). Furthermore, RNA-seq data obtained from wild-type testis and ovaries (Rogers et al., 2014; VanKuren and Vibranovski, 2014) shows that *phf7* expression is also sex-specific in *D. simulans* and *D. yakuba* (**Figure 3A)**. These data suggest that regulation through alternative TSS selection is conserved between *D. melanogaster, D. simulans,* and *D. yakuba*. When we compared the level of sequence conservation in the first intron between *D. melanogaster, D. simulans,* and *D. yakuba,* we found that only the first 235 bp of intron element A was conserved (**Figure 3B; Figure S2**). Interestingly, the first non-coding exon also displays a high level of conservation, suggesting that additional cis-regulatory elements may be present within the first non-coding exon.

Core components of the piRNA pathway are not required for *phf7* sex-specific transcriptional control

Prior studies have shown that SETDB1 is required for both TE and *phf7* silencing in germ cells (Osumi et al., 2019; Rangan et al., 2011; Sienski et al., 2012; Smolko et al., 2018). Although there are no recognizable TE sequences at *phf7*, the piRNA pathway may nevertheless play a role in repressing *phf7.* To test this possibility, we analyzed published RNA-seq data sets from wild-type and ovaries carrying mutations in genes encoding the dedicated piRNA pathway components, *piwi, aubergine (aub), rhino, panoramix (panx,* also known as *silenceo),* and *nuclear RNA export factor 2 (nxf2)* (Fabry et al., 2019; Mohn et al., 2014; Teixeira et al., 2017). Loss of each component causes the piRNA pathway to collapse: Piwi, Aub, and Rhino are essential for piRNA biogenesis (Mohn et al., 2014; Sienski et al., 2012; Théron et al., 2018). Panx and Nxf2, while dispensable for piRNA biogenesis, are essential for SETDB1-mediated heterochromatin formation at PIWI target genes (Batki et al., 2019; Fabry et al., 2019; Murano et al., 2019; Sienski et al., 2015; Yu et al., 2015; Zhao et al., 2019). In agreement with our prior studies (Smolko et al., 2018), RNA-seq data shows that while only the shorter *phf7-RA* transcript in wild-type ovaries, the longer testis-specific *phf7-RC* transcript is present in *setdb1* mutant ovaries **(Figure 4)**. However, data from *piwi, aub, rhino, panx,* and *nxf2* mutant ovaries show only the shorter *phf7-RA* transcript. Finding that these five essential piRNA pathway-specific protein components are dispensable for *phf7* regulation indicates that the mechanisms controlling H3K9me3 deposition onto *phf7* and TEs are distinct.

The ZAD-ZNF protein IDC is required for *phf7* sex-specific transcriptional control

In mammalian cells, SETDB1 can be recruited to its targets by members of the large KRAB-zinc finger family of sequence-specific DNA binding proteins (Bruno et al., 2019; Ecco et al., 2017; Yang et al., 2017). The KRAB family is confined to mammals. It has been hypothesized that members of the insect-specific ZAD-ZNF family are KRAB family paralogues

(Chung et al., 2002; Chung et al., 2007; Kasinathan et al., 2020). We recently tested the function of 68 of the 93 ZAD-ZNF encoding genes in female germ cells by performing an RNAi screen (Shapiro-Kulnane et al., 2020). These data revealed that loss of CG4936 in germ cells phenocopies the ovarian defects caused by ectopic PHF7 expression. CG4936, which we name Identity crisis (IDC), is a 521 amino acid (aa) protein that contains an N-terminal ZAD (zinc finger-associated domain, 22-95 aa), an unstructured linker region and a C-terminal domain that includes an array of 5 C2H2 zinc fingers (386-491 aa). Studies of other ZAD-ZNF proteins suggest that the ZAD mediates protein-protein interactions and the C2H2 zinc fingers bind DNA (Bonchuk et al., 2021; Bonchuk et al., 2022; Jauch et al., 2003; Maksimenko et al., 2020; Zolotarev et al., 2016). We, therefore, hypothesize that IDC is required for *phf7* repression.

As the first step in understanding the role of IDC in the female germline, we determined its localization pattern in the adult ovary. We used a genomic fosmid transgene for these studies that encodes a C-terminally GFP-tagged IDC protein (IDC-GFP). We infer that the GFP tag does not interfere with *idc* function because the *idc-GFP* rescues the idc^{1} lethal phenotype (see materials and methods). Each ovary comprises 16-20 ovarioles, which contain germ cells spanning the range of maturity from germline stem cells (GSCs) at the anterior end to fully mature eggs at the posterior end (Hinnant et al., 2020). In the egg chambers, we find that IDC-GFP is detectable in the nucleus of several different cell types, including the nurse cells, the oocytes, and the somatic follicle cells of the mature egg chambers (**Figure 5A**). IDC-GFP is tightly associated with the polytene chromosomes in the nurse cells, consistent with its presumed DNA binding activity (**Figure 5B**).

To determine the identity of the idc-expressing germ cells, we compared expression of IDC-GFP with the cytoplasmic differentiation factor Bag-of Marbles (BAM). At the anterior end of the ovariole, in the germarium, we found that the IDC-GFP protein accumulates to high levels in the nucleus of only 3 to 6 germ cells (boxed region; **Figure 5A)**. This expression pattern is like that described for the female-specific sex determination protein SEX-LETHAL (SXL), which is

also required for *phf7* H3K9me3 silencing (Chau et al., 2009; Chau et al., 2012; Smolko et al., 2018). Indeed, co-staining experiments reveal that SXL accumulates in the cytoplasm of all IDC-GFP-expressing germ cells (**Figure 5C**). Studies have shown that SXL accumulates in GSCs and their immediate daughter cells. When GSCs divide, the daughter cells at the tip remain a GSC, while the more posterior daughter cells, called a cystoblast (CB), express BAM. In agreement with published studies, BAM expression, assessed with a transgene that encodes a C-terminally HA-tagged BAM protein expressed from a *bam* promoter, is readily detectable in the cytoplasm of just a few germ cells. Co-staining experiments reveal that IDC-GFP is expressed in the GSCs where HA-BAM is not expressed and in the adjacent two to three cells where HA-BAM is detectable (**Figure 5D**). Based on these observations, we conclude that the nuclear IDC protein is expressed in the GSCs and their immediate daughter cells.

To investigate the possibility that loss of *idc* in female germ cells disrupts *phf7* repression, we first used RT-qPCR to assay for the presence of the testis-specific *phf7-RC* transcript. Germ cell-specific depletion was achieved using an inducible RNA interference (RNAi) transgene and the germ cell-specific *nos-GAL4* driver. We demonstrated RNAi efficiency by showing that Germ Line specific Knock Down (GLDK) of *idc* abolished the intense germ cell staining pattern exhibited by the IDC-GFP rescue transgene in germ cells (**Figure 6A, B).** Using primer pairs capable of detecting *phf7-RC,* we found that in contrast to control ovaries, *phf7-RC* is detectable in *idc* GLKD mutant ovaries (**Figure 6C**). Next, we asked whether ectopic *phf7-RC* expression correlated with ectopic PHF7 protein expression. Previous work has shown that a BAC transgene that encodes an N-terminally HA-tagged PHF7 protein (HA-PHF7) is a faithful reporter of the sex-specific protein expression pattern of PHF7. We found that, in contrast to wild-type ovaries, the HA-PHF7 protein is detectable in the cytoplasm and the nucleus of the *idc GLKD* mutant ovaries (**Figure 6 D,E**). We, therefore, conclude that IDC is required for *phf7* regulation.

Lastly, we asked whether the loss of IDC affects H3K9me3 accumulation at the *phf7* locus. To this end, we compared the amount of H3K9me3 accumulation within element A using ChIP-qPCR. We found that H3K9me3 was significantly reduced in *idc* GLKD ovaries compared to control ovaries ($p<0.0001$; **Figure 6F**). Based on these studies, we conclude that IDC regulates *phf7* transcription by promoting H3K9me3 deposition.

IDC binds to sequences within the male-specific exon

IDC contains five zinc fingers suggesting that it has sequence-specific DNA binding activity. We speculated that IDC might promote H3K9me3 deposition by binding to the *phf7* locus. When we tested this prediction by ChIP-qPCR, we found that the IDC-GFP tagged protein associates with DNA within the first non-coding exon but not outside of it (**Figure 7A**). Collectively, our studies indicate that IDC regulates *phf7* directly by serving as an H3K9me3 guidance factor (**Figure 7B**).

Discussion

SETDB1-controlled H3K9 methylation plays an essential role in securing female germ cell fate by silencing lineage-inappropriate *phf7* transcription (Smolko et al., 2018; Smolko et al., 2020). SETDB1 is also required for TE silencing, where a piRNA-guided mechanism guides H3K9me3 deposition (Osumi et al., 2019; Rangan et al., 2011; Sienski et al., 2012). In this work, we establish that *phf7* is silenced by a piRNA-independent mechanism and discover that regulation depends on IDC, a previously uncharacterized member of the ZAD-ZNF family of DNA binding proteins. Regulation appears direct, as IDC localizes to the *phf7* locus and is required for H3K9me3 deposition. Collectively, our data establish that the sequence-specific DNA binding protein IDC guides the installation of the H3K9me3 mini-domain onto the *phf7* locus (Figure 5H), thereby preventing accidental female-to-male reprogramming. In addition to extending our understanding of how female germ cell fate is maintained, our studies provide one of the first examples of a ZAD-ZNF protein guiding H3K9me3-mediated gene silencing.

Our work is consistent with a simple model in which IDC recruits the SETDB1 H3K9me3 methyltransferase to *phf7*. However, we could not detect an interaction between IDC and SETDB1 in co-immunoprecipitation experiments. It is, therefore, possible that another, as yet undiscovered, protein recruits SETDB1 to *phf7* and that IDC acts at a later step in the pathway to ensure the sequence-specific establishment of the H3K9me3 mini domain.

Considering that IDC protein is also expressed in male germ cells, we hypothesize that another protein confers female germ cell specificity to *phf7* silencing. One attractive contender is *stonewall (stwl)*. STWL is a female-specific heterochromatin-associated protein that acts as a transcriptional repressor *in vitro* (Yi et al., 2009), associates with SETDB1 in yeast (Shokri et al., 2019), and localizes to the *phf7* locus in S2 cells (Zinshteyn and Barbash, 2022). Importantly, loss of *stwl* in ovaries leads to the inappropriate expression of the testes *phf7* transcript (Zinshteyn and Barbash, 2022). Together, these studies raise the possibility that *stwl* collaborates with IDC to silence *phf7*.

The H3K9me3 reader protein, HP1a, is essential for *phf7* silencing (Smolko et al., 2018). A requirement for HP1a is not surprising, as HP1a drives chromatin compaction and transcriptional silencing (Allshire and Madhani, 2018). HP1a also recruits H3K9 methyltransferases, enabling the spreading of the H3K9me3 repressive domain through a positive feedback mechanism. Therefore, the spread of H3K9me3 over the ~3 kb region within the *phf7* gene likely occurs by an analogous mechanism. But, at the *phf7* locus, the H3K9me3 domain does not extend into the open reading frame or the neighboring genes. Therefore, more work is needed to understand what stops the spreading of this repressive chromatin domain.

The SETDB1-controlled H3K9 methylation silencing pathway in female germ cells is not restricted to *phf7.* Genome-wide H3K9me3 profiling in wild-type and mutant ovaries has shown that this mark silences two classes of protein-encoding genes. One type includes genes usually expressed in other tissues, such as the testis or the central nervous system (Smolko et al., 2018). A second class includes genes such as RpS19b that are typically expressed in

undifferentiated female germ cells but whose transcriptional silencing is required for entry into meiosis (Sarkar et al., 2022). How might SETDB1 be recruited to these genes in a stage and tissue-specific manner? We think it is unlikely that the piRNA pathway is involved, as a collapse of the piRNA pathway does not cause defects in germ cell differentiation (Batki et al., 2019; Fabry et al., 2019; Murano et al., 2019; Sienski et al., 2015; Théron et al., 2018; Yu et al., 2015; Zhao et al., 2019). We, therefore, favor the idea that SETDB1 is recruited to these target genes by a dedicated DNA binding protein.

In summary, by focusing on a biologically relevant gene, we discovered that a putative DNA binding protein recruits H3K9me3. These findings are reminiscent of how TEs can be silenced in mammals, where members of the KRAB-ZNF protein family recruit SETDB1 to establish epigenetic repression (Bruno et al., 2019; Ecco et al., 2017; Senft and Macfarlan, 2021; Yang et al., 2017). Interestingly, the KRAB-ZNFs are vertebrate-specific, and the ZAD-ZNFs are insect specific. Yet, both gene families exhibit similar patterns of species-specific gene expansion and diversification. These observations have fueled the speculation that ZAD-ZNFs and KRAB-ZNFs perform similar functions (Chung et al., 2002; Chung et al., 2007; Kasinathan et al., 2020). In fact, IDC's closest human relative is ZNF75D, a KRAB-ZNF protein of unknown function. Whether this or other members of the KRAB-ZNF protein family are required for tissue-specific H3K9me3 gene silencing will be an exciting area of further exploration.

Materials and Methods

Drosophila stocks and culture conditions

Fly strains were kept on standard media at 25 °C unless otherwise noted. Wild-type control flies were from the lab y^1 w^1 stock. Stocks used to report on protein gene activity include the HA-tagged *phf7* transgene, *PBac{3XHA-PHF7}* (Yang et al., 2012); the GFP-tagged *idc* transgene, *FlyFos{fTRG00376.sfGFP-TVPTBF}* (VDRC #318556) described in (Sarov et al.,

2016)and the HA-tagged *bam* transgene, *P{Bam::HA}* (Li et al., 2009). *phf7*$^{\Delta 13}$, *phf7*$^{\Delta R}$, the *idc*^{1} null allele, and the ΔR and ΔAR chromatin reporters were generated for this study, as described below.

Crosses for knock-down studies were set up at 29°C, and adults were aged 3-5 days before gonad dissection. Knock-down studies were carried out with the following germline-optimized lines generated by the *Drosophila* Transgenic RNAi Project (Hu et al., 2020): *idc-P{TRiP.HMC05569}* (BDSC #64550), and the control *w-P{TRiP.GL00094}* (BDSC #35573). Expression of the UAS-RNAi transgenic lines was driven by the third chromosome *P{GAL4::VP16-nos.UTR}* insertion (BDSC #4937), which is strongly expressed in germ cells.

Generation of the *phf7*$^{\Delta R}$ allele

We generated the *phf7*$^{\Delta R}$ allele in two steps. First, CRISPR was used to replace the repeats within the first intron with a 3XP3::DsRed cassette by Rainbow Transgenic Flies, Inc. The deletion was generated with the following guide RNAs: agttaaaaaaaatcaatcgatgg and cgcagcgattgaatgttaatggg. 1 kb homology arms were generated through PCR and cloned into the pHD-dsRed-attP (Addgene #51019) (Gratz et al., 2014). Guide RNAs and the donor vector were co-injected into *nos-Cas9* embryos by Rainbow Transgenic Flies. Flies were screened for DsRed expression in the eyes, and sequence verified for accuracy. *phf7*$^{\Delta R}$ was generated by removing the DsRed cassette. Homozygous *phf7*$^{\Delta R+dsRED}$ females were crossed to a Cre-expressing fly line (BDSC #1092), and the male progeny screened for the loss of dsRED expression in the eyes, followed by sequence confirmation of precise tag excision.

Generation of the *phf7*$^{\Delta 13}$ allele

The *phf7*$^{\Delta 13}$ allele was generated by P-element mediated imprecise excision as follows: females homozygous for the P-element insertion P{EPgy2}Phf7^{EY03023} (BDSC #15894) were

crossed with the Δ2-3 transposase expressing line (BDSC #3664). Potential excision lines were established from the male progeny of the F1 females who had lost the *white*$^{+}$ marker carried by the P element. Stocks that carried deletions were identified by PCR. The exact breakpoints were determined by DNA sequencing the PCR amplified DNA fragments.

Generation of the ΔR and ΔAR chromatin transgenic reporter lines

Constructs were generated by VectorBuilder's custom cloning services. The transgenes were constructed in their "user-defined promoter" modification of the pUASTattB expression vector. The transgenic constructs were sent to Rainbow Transgenic Flies Inc. for *phi-C31* catalyzed integration into the 65B2 PBac{y[+]-attP-3B}VK00033 site. Transgenic flies were sequence verified for accuracy.

Generation of the *idc*^{1} null allele and rescue by the GFP-tagged *idc* transgene

The *idc*^{1} allele was generated by *in vivo* CRISPR mutagenesis (Port et al., 2020). Females expressing two sgRNAs under control of the GAL4/UAS system, P{HD_CFD00598} (VDRC #341525), were crossed at 25°C to M{vas-Cas9}; P{GAL4::VP16-nos.UTR} (BDSC #55821 + #4937) males. The *vas-Cas9; nos>gRNA* male offspring were crossed to a third chromosome balancer line to isolate and balance each putative *idc* allele in the next generation. *idc* alleles were identified by the failure to complement Df(3R)ED6027 (BDSC #9479) and checked for the presence of the predicted indel by PCR.

We found that *idc*^{1}*/Df(3R)ED6027* animals do not survive to adulthood (n>100). To verify that the GFP-tagged *idc* transgene, PBac{fTRG00376.sfGFP-TVPTBF} (VDRC #318556), rescues *idc*^{1}*,* we first made the double mutant *idc*^{1}*, idc-GFP.* To test for rescue, we crossed *idc*^{1}*, idc-GFP/TM3* to *Df(3R)ED6027/TM3* males. This cross yielded 84% (31/37) of the expected *idc*^{1}*, idc-GFP/ Df(3R)ED6027* progeny, all of which were fertile.

qRT-PCR and data analysis

RNA was extracted from dissected gonads using TRIzol (Thermo Fisher, cat# 15596026). Quantity and quality were measured using a NanoDrop ND-1000 spectrophotometer. The RNA was treated with DNase RQ1 (Promega, cat# M6101). cDNA was generated by reverse transcription using a SuperScript First-Strand Synthesis System for RT-PCR kit (Thermo Fisher, cat# 11904018) using random hexamers. qPCR was performed using *Power* SYBR Green PCR Master Mix (Thermo Fisher, cat# 4368706) with an Applied Biosystems QuantStudio 3 Real-Time PCR system. PCR steps were as follows: 95°C for 10 minutes followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. Melt curves were generated with the following parameters: 95°C for 15 seconds, 60°C for 1 minute, 95°C for 15 seconds, and 60°C for 15 seconds. Measurements were taken in biological triplicate with two technical replicates each. Relative transcript levels were calculated using the 2-ΔΔCt method (Livak and Schmittgen, 2001). Using GraphPad Prism software, P values were calculated using unpaired two-tailed Student's t-tests. The primers used for measuring RNA levels are listed in Table S1.

ChIP-qPCR and data analysis

ChIP was performed as described in (Smolko et al., 2018) with some modifications. Buffers are listed in Table S2. Briefly, ovaries from 200 1-2 day old adults were dissected in PBSP, homogenized with a pellet pestle, and crosslinked with 1.8% methanol-free formaldehyde (Thermo Fisher, cat#28906) for 5 minutes. Fixation was quenched for 5 minutes by adding glycine to a final concentration of 225mM. The solution was removed, and the pellet was washed in PBSP 3 times. Samples were resuspended in 500μl PBSP, frozen, and stored at -80°C. Samples were thawed, centrifuged, and resuspended in 1ml lysis buffer 1. Samples were

homogenized using sterile homogenizing beads in a Bullet Blender Homogenizer at 4°C for six cycles of 30 seconds on and 1 minute off. Following centrifugation, cell pellets were washed at 4°C for 10 minutes in lysis buffer 1, centrifuged, washed for 10 min at 4°C in lysis buffer 2, and then resuspended in 750μl lysis buffer 3. Chromatin was sheared to 100–500 base pairs at 4°C with a Diagenode BioRuptor Pico for 15 cycles of 30 seconds on, and 30 seconds off.

For H3K9me3 ChIP, the chromatin lysates were precleared with a 1:1 mix of protein A/G Dynabead magnetic beads (Thermo Fisher, cat# 10001D and 10003D) for 1 hour at 4°C and then incubated overnight at 4°C with 1:1 mix of protein A/G Dynabead magnetic beads conjugated to H3K9me3 antibody (Abcam, cat# 8898 RRID: AB_306848).

For GFP ChIP, the chromatin lysates were precleared with Chromotek Magnetic Binding Control Agarose Beads (Proteintech, cat# bmab) for 1 hour at 4°C, and then incubated overnight at 4°C with Chromotek GFP-Trap Magnetic Agarose beads (Proteintech, cat# gtma).

Following immunoprecipitation, the samples were washed six times with ChIP-RIPA buffer, two times with ChIP-RIPA/500 buffer, two times with ChIP-LiCl buffer, and twice with TE buffer. DNA was eluted and reverse crosslinked from beads in 200μl elution buffer overnight at 65°C. Following RNaseA and proteinase K treatment, DNA was recovered by phenol-chloroform extraction and used for qPCR.

qPCR was performed using *Power* SYBR Green PCR Master Mix (Thermo Fisher, cat# 4367659) with an Applied Biosystems QuantStudio 3 Real-Time PCR system. PCR steps were as follows: 95°C for 10 minutes followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. Melt curves were generated with the following parameters: 95°C for 15 seconds, 60°C for 1 minute, 95°C for 15 seconds, and 60°C for 15 seconds. Primers as listed in Table S1. ChIP experiments were performed on three independent biological samples. qPCR measurements on each sample were performed on at least two technical replicates. For H3K9me3 ChIP-qPCR, the ChIP to input enrichment (presented as percent input) is normalized to a negative control

genomic region in the *rp49* locus. Statistical analysis was carried out by the GraphPad Prism software. The P values were calculated using unpaired two-tailed Student's t tests.

Immunofluorescence and image analysis

Females were aged 3-5 days before gonad dissection. Ovaries were fixed and stained according to standard procedures with the following primary antibodies: mouse anti-drosophila sex-lethal (1:100, Developmental Studies Hybridoma Bank, cat# M18, RRID: AB_528464), rabbit anti-GFP (1:2500, Thermo Fisher, cat# A-11122, RRID: AB_221569) or FITC conjugated goat anti-GFP (1:750, Abcam, cat# ab6662, RRID: AB_305635), and rat anti-HA high affinity (1:500, Sigma, cat# 11867423001, RRID: AB_390919). The following secondary antibodies were used at 1:200: Alexa Fluor 555 goat anti-rat (Thermo Fisher, cat# A-21434, RRID: AB_2535855), FITC goat anti-mouse (Jackson Immunoresearch Laboratories, cat# 115-095-003, RRID: AB_2338589), or FITC goat anti-rabbit (Jackson Immunoresearch Laboratories, cat# 111-095-003, RRID: AB_2337972). TO-PRO-3 Iodide carbocyanine monomer nucleic acid stain (1:1000, Thermo Fisher, cat# T3605) was used to stain DNA. Images were taken on a Leica SP8 confocal with 1024x1024 pixel dimensions, a scan speed of 600 Hz, and a frame average of 3. Sequential scanning was done for each channel, and three Z-stacks were combined for each image. Processed images were compiled with Microsoft PowerPoint.

Analysis of publicly available RNA-seq data sets

Table S3 lists the sources of the publicly available RNA-seq data generated from wild-type and mutant Drosophila tissues. The data were downloaded from the SRA database at NCBI and analyzed using the tools available on the Galaxy web platform (https://usegalaxy.org/). FastQC assessed read quality and STAR was used to align the reads to the Drosophila reference genome (UCSC dm6). Screenshots of the expression data displayed on The Integrated Genome Viewer (IGV) are shown.

D. simulans and *D. yakuba phf7* orthologs were identified using the ortholog database at www.flybase.org. RNA-seq data from wild-type *D. simulans* and *D. yakuba* gonads was downloaded from the SRA database at NCBI (Table S4). We uploaded the sequencing data to the Galaxy web platform (https://usegalaxy.org/), assessed read quality with FastQC and used STAR to align the reads to their respective reference genomes (listed in Table S4). Browser screenshots of the expression data are displayed on the Integrated Genome Viewer (IGV).

DNA sequence alignments

Multiple sequence alignments were performed in SnapGene (https://www.snapgene.com) using Muscle with default options. Pairwise sequence alignments were performed in BLASTn (https://blast.ncbi.nlm.nih.gov).

Acknowledgments

We would like to thank the members of the Cleveland fly community for helpful discussions, Jane Heatwole for fly food, and FlyBase as an essential resource. Stocks obtained from the Bloomington (NIH P40D018537) and The Vienna Drosophila Stock Centers were used in this study. Imaging was performed using equipment purchased through NIH Grant S10-OD016164, housed in the CWRU School of Medicine Imaging Core.

Author contributions

HS conceived this study. LSK and HS generated reagents, designed and executed the experiments, and interpreted the data. MS performed the evolutionary analysis. HS wrote the manuscript with input from LSK and MS.

Declaration of interests

The authors declare no competing interests

Availability of data and material

All genomic sequence and expression data were obtained from public databases and are described in Table S3 and S4. *Drosophila* strains are available upon request.

References

Allshire, R. C. and Madhani, H. D. (2018). Ten principles of heterochromatin formation and function. *Nature reviews. Molecular cell biology* 19, 229–244.

Andreev, V. I., Yu, C., Wang, J., Schnabl, J., Tirian, L., Gehre, M., Handler, D., Duchek, P., Novatchkova, M., Baumgartner, L., et al. (2022). Panoramix SUMOylation on chromatin connects the piRNA pathway to the cellular heterochromatin machinery. *Nat Struct Mol Biol* 29, 130–142.

Aravin, A. A. (2020). Pachytene piRNAs as beneficial regulators or a defense system gone rogue. *Nat Genet* 52, 644–645.

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

Becker, J. S., Nicetto, D. and Zaret, K. S. (2016). H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends in genetics : TIG* 32, 29–41.

Benner, L., Castro, E. A., Whitworth, C., Venken, K. J. T., Yang, H., Fang, J., Oliver, B., Cook, K. R. and Lerit, D. A. (2019). Drosophila Heterochromatin Stabilization Requires the Zinc-Finger Protein Small Ovary. *Genetics* 213, 877–895.

Biferali, B., Bianconi, V., Perez, D. F., Kronawitter, S. P., Marullo, F., Maggio, R., Santini, T., Polverino, F., Biagioni, S., Summa, V., et al. (2021). Prdm16-mediated H3K9 methylation controls fibro-adipogenic progenitors identity during skeletal muscle repair. *Sci Adv* 7, eabd9371.

Bilodeau, S., Kagey, M. H., Frampton, G. M., Rahl, P. B. and Young, R. A. (2009). SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes & Development* 23, 2484–2489.

Bonchuk, A., Boyko, K., Fedotova, A., Nikolaeva, A., Lushchekina, S., Khrustaleva, A., Popov, V. and Georgiev, P. (2021). Structural basis of diversity and homodimerization specificity of zinc-finger-associated domains in Drosophila. *Nucleic Acids Res* 49, 2375–2389.

Bonchuk, A. N., Boyko, K. M., Nikolaeva, A. Y., Burtseva, A. D., Popov, V. O. and Georgiev, P. G. (2022). Structural insights into highly similar spatial organization of zinc-finger associated domains with a very low sequence similarity. *Structure* 30, 1004-1015.e4.

Bruno, M., Mahgoub, M. and Macfarlan, T. S. (2019). The Arms Race Between KRAB–Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. *Annu Rev Genet* 53, 1–24.

Chau, J., Kulnane, L. S. and Salz, H. K. (2009). Sex-lethal facilitates the transition from germline stem cell to committed daughter cell in the Drosophila ovary. *Genetics* 182, 121–132.

Chau, J., Kulnane, L. S. and Salz, H. K. (2012). Sex-lethal enables germline stem cell differentiation by down-regulating Nanos protein levels during Drosophila oogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 109, 9465–9470.

Cheng, E.-C., Hsieh, C.-L., Liu, N., Wang, J., Zhong, M., Chen, T., Li, E. and Lin, H. (2021). The Essential Function of SETDB1 in Homologous Chromosome Pairing and Synapsis during Meiosis. *Cell Reports* 34, 108575.

Chung, H.-R., Schäfer, U., Jäckle, H. and Böhm, S. (2002). Genomic expansion and clustering of ZAD-containing C2H2 zinc-finger genes in Drosophila. *EMBO reports* 3, 1158–1162.

Chung, H.-R., Löhr, U. and Jäckle, H. (2007). Lineage-specific expansion of the zinc finger associated domain ZAD. *Molecular biology and evolution* 24, 1934–1943.

Du, D., Katsuno, Y., Meyer, D., Budi, E. H., Chen, S.-H., Koeppen, H., Wang, H., Akhurst, R. J. and Derynck, R. (2018). Smad3-mediated recruitment of the methyltransferase SETDB1/ESET controls Snail1 expression and epithelial-mesenchymal transition. *EMBO reports* 19, 135–155.

Ecco, G., Imbeault, M. and Trono, D. (2017). KRAB zinc finger proteins. *Development (Cambridge, England)* 144, 2719–2729.

Fabry, M. H., Ciabrelli, F., Munafò, M., Eastwood, E. L., Kneuss, E., Falciatori, I., Falconio, F. A., Hannon, G. J. and Czech, B. (2019). piRNA-guided co-transcriptional silencing coopts nuclear export factors. *Elife* 8, e47999.

Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M. and O'Connor-Giles, K. M. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics* 196, 961–971.

Hinnant, T. D., Merkle, J. A. and Ables, E. T. (2020). Coordinating Proliferation, Polarity, and Cell Fate in the Drosophila Female Germline. *Frontiers in cell and developmental biology* 8, 19.

Hu, Y., Comjean, A., Rodiger, J., Liu, Y., Gao, Y., Chung, V., Zirin, J., Perrimon, N. and Mohr, S. E. (2020). FlyRNAi.org—the database of the Drosophila RNAi screening center and transgenic RNAi project: 2021 update. *Nucleic Acids Res* 49, gkaa936-.

Jankovics, F., Bence, M., Sinka, R., Faragó, A., Bodai, L., Pettkó-Szandtner, A., Ibrahim, K., Takács, Z., Szarka-Kovács, A. B. and Erdelyi, M. (2018). Drosophila small ovary gene is required for transposon silencing and heterochromatin organization, and ensures germline stem cell maintenance and differentiation. *Development (Cambridge, England)* 145, dev170639.

Janssen, A., Colmenares, S. U. and Karpen, G. H. (2018). Heterochromatin: Guardian of the Genome. *Annual review of cell and developmental biology* 34, 265–288.

Jauch, R., Bourenkov, G. P., Chung, H.-R., Urlaub, H., Reidt, U., Jäckle, H. and Wahl, M. C. (2003). The zinc finger-associated domain of the Drosophila transcription factor grauzone is a novel zinc-coordinating protein-protein interaction module. *Structure (London, England : 1993)* 11, 1393–1402.

Jiang, Y., Loh, Y.-H. E., Rajarajan, P., Hirayama, T., Liao, W., Kassim, B. S., Javidfar, B., Hartley, B. J., Kleofas, L., Park, R. B., et al. (2017). The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain. *Nature genetics* 49, 1239–1250.

Karimi, M. M., Goyal, P., Maksakova, I. A., Bilenky, M., Leung, D., Tang, J. X., Shinkai, Y., Mager, D. L., Jones, S., Hirst, M., et al. (2011). DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. *Cell Stem Cell* 8, 676–687.

Kasinathan, B., Colmenares, S. U., McConnell, H., Young, J. M., Karpen, G. H. and Malik, H. S. (2020). Innovation of heterochromatin functions drives rapid evolution of essential ZAD-ZNF genes in Drosophila. *eLife* 9,.

Koide, S., Oshima, M., Takubo, K., Yamazaki, S., Nitta, E., Saraya, A., Aoyama, K., Kato, Y., Miyagi, S., Nakajima-Takagi, Y., et al. (2016). Setdb1 maintains hematopoietic stem and progenitor cells by restricting the ectopic activation of nonhematopoietic genes. *Blood* 128, 638–649.

Li, Y., Minor, N. T., Park, J. K., Mckearin, D. M. and Maines, J. Z. (2009). Bam and Bgcn antagonize Nanos-dependent germline stem cell maintenance. *Proceedings of the National Academy of Sciences* 106, 9304–9309.

Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif.)* 25, 402–408.

Lohmann, F., Loureiro, J., Su, H., Fang, Q., Lei, H., Lewis, T., Yang, Y., Labow, M., Li, E., Chen, T., et al. (2010). KMT1E mediated H3K9 methylation is required for the maintenance of embryonic stem cells by repressing trophectoderm differentiation. *Stem cells (Dayton, Ohio)* 28, 201–212.

Maksimenko, O., Kyrchanova, O., Klimenko, N., Zolotarev, N., Elizarova, A., Bonchuk, A. and Georgiev, P. (2020). Small Drosophila zinc finger C2H2 protein with an N-terminal zinc finger-associated domain demonstrates the architecture functions. *Biochimica Et Biophysica Acta Bba - Gene Regul Mech* 1863, 194446.

Marsano, R. M. and Dimitri, P. (2022). Constitutive Heterochromatin in Eukaryotic Genomes: A Mine of Transposable Elements. *Cells* 11, 761.

Martienssen, R. and Moazed, D. (2015). RNAi and Heterochromatin Assembly. *Csh Perspect Biol* 7, a019323.

McCarthy, R. L., Kaeding, K. E., Keller, S. H., Zhong, Y., Xu, L., Hsieh, A., Hou, Y., Donahue, G., Becker, J. S., Alberto, O., et al. (2021). Diverse heterochromatin-associated proteins repress distinct classes of genes and repetitive elements. *Nat Cell Biol* 23, 905–914.

Methot, S. P., Padeken, J., Brancati, G., Zeller, P., Delaney, C. E., Gaidatzis, D., Kohler, H., Oudenaarden, A. van, Großhans, H. and Gasser, S. M. (2021). H3K9me selectively blocks transcription factor activity and ensures differentiated tissue integrity. *Nat Cell Biol* 23, 1163–1175.

Mochizuki, K., Sharif, J., Shirane, K., Uranishi, K., Bogutz, A. B., Janssen, S. M., Suzuki, A., Okuda, A., Koseki, H. and Lorincz, M. C. (2021). Repression of germline genes by PRC1.6 and SETDB1 in the early embryo precedes DNA methylation-mediated silencing. *Nat Commun* 12, 7020.

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

Murano, K., Iwasaki, Y. W., Ishizu, H., Mashiko, A., Shibuya, A., Kondo, S., Adachi, S., Suzuki, S., Saito, K., Natsume, T., et al. (2019). Nuclear RNA export factor variant initiates piRNA-guided co-transcriptional silencing. *Embo J* 38, e102870.

Nicetto, D. and Zaret, K. S. (2019). Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. *Current Opinion in Genetics & Development* 55, 1–10.

Nicetto, D., Donahue, G., Jain, T., Peng, T., Sidoli, S., Sheng, L., Montavon, T., Becker, J. S., Grindheim, J. M., Blahnik, K., et al. (2019). H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. *Science* 363, 294–297.

Ninova, M., Tóth, K. F. and Aravin, A. A. (2019). The control of gene expression and cell identity by H3K9 trimethylation. *Development (Cambridge, England)* 146,.

Osumi, K., Sato, K., Murano, K., Siomi, H. and Siomi, M. C. (2019). Essential roles of Windei and nuclear monoubiquitination of Eggless/SETDB1 in transposon silencing. *Embo Rep* 20, e48296.

Padeken, J., Methot, S. P. and Gasser, S. M. (2022). Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. *Nat Rev Mol Cell Bio* 1–18.

Port, F., Strein, C., Stricker, M., Rauscher, B., Heigwer, F., Zhou, J., Beyersdörffer, C., Frei, J., Hess, A., Kern, K., et al. (2020). A large-scale resource for tissue-specific CRISPR mutagenesis in Drosophila. *Elife* 9, e53865.

Rangan, P., Malone, C. D., Navarro, C., Newbold, S. P., Hayes, P. S., Sachidanandam, R., Hannon, G. J. and Lehmann, R. (2011). piRNA production requires heterochromatin formation in Drosophila. *Current biology : CB* 21, 1373–1379.

Rechtsteiner, A., Costello, M. E., Egelhofer, T. A., Garrigues, J. M., Strome, S. and Petrella, L. N. (2019). Repression of Germline Genes in Caenorhabditis elegans Somatic Tissues by H3K9 Dimethylation of Their Promoters. *Genetics* 212, 125–140.

Rogers, R. L., Shao, L., Sanjak, J. S., Andolfatto, P. and Thornton, K. R. (2014). Revised Annotations, Sex-Biased Expression, and Lineage-Specific Genes in the Drosophila melanogaster Group. *G3 Genes Genomes Genetics* 4, 2345–2351.

Sarov, M., Barz, C., Jambor, H., Hein, M. Y., Schmied, C., Suchold, D., Stender, B., Janosch, S., KJ, V. V., Krishnan, R., et al. (2016). A genome-wide resource for the analysis of protein localisation in Drosophila. *Elife* 5, e12068.

Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. and Rauscher, F. J. (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes & Development* 16, 919–932.

Senft, A. D. and Macfarlan, T. S. (2021). Transposable elements shape the evolution of mammalian development. *Nat Rev Genet* 22, 691–711.

Shapiro-Kulnane, L., Bautista, O. and Salz, H. K. (2020). An RNA-interference screen in Drosophila to identify ZAD-containing C2H2 zinc finger genes that function in female germ cells. *G3 Genes Genomes Genetics* 11, jkaa016.

Shokri, L., Inukai, S., Hafner, A., Weinand, K., Hens, K., Vedenko, A., Gisselbrecht, S. S., Dainese, R., Bischof, J., Furger, E., et al. (2019). A Comprehensive Drosophila melanogaster Transcription Factor Interactome. *Cell Reports* 27, 955-970.e7.

Sienski, G., Dönertas, D. and Brennecke, J. (2012). Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* 151, 964–980.

Sienski, G., Batki, J., Senti, K.-A., Dönertas, D., Tirian, L., Meixner, K. and Brennecke, J. (2015). Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. *Genes & Development* 29, 2258–2271.

Smolko, A. E., Shapiro-Kulnane, L. and Salz, H. K. (2018). The H3K9 methyltransferase SETDB1 maintains female identity in Drosophila germ cells. *Nature communications* 9, 4155.

Smolko, A. E., Shapiro-Kulnane, L. and Salz, H. K. (2020). An autoregulatory switch in sex-specific phf7 transcription causes loss of sexual identity and tumors in the Drosophila female germline. *Development (Cambridge, England)*.

Tan, S.-L., Nishi, M., Ohtsuka, T., Matsui, T., Takemoto, K., Kamio-Miura, A., Aburatani, H., Shinkai, Y. and Kageyama, R. (2012). Essential roles of the histone methyltransferase ESET in the epigenetic control of neural progenitor cells during development. *Development (Cambridge, England)* 139, 3806–3816.

Teixeira, F. K., Okuniewska, M., Malone, C. D., Coux, R.-X., Rio, D. C. and Lehmann, R. (2017). piRNA-mediated regulation of transposon alternative splicing in the soma and germ line. *Nature* 8, 272.

Théron, E., Maupetit-Mehouas, S., Pouchin, P., Baudet, L., Brasset, E. and Vaury, C. (2018). The interplay between the Argonaute proteins Piwi and Aub within Drosophila germarium is critical for oogenesis, piRNA biogenesis and TE silencing. *Nucleic Acids Res* 46, gky695-.

VanKuren, N. W. and Vibranovski, M. D. (2014). A Novel Dataset for Identifying Sex-Biased Genes in Drosophila. *J Genom* 2, 64–67.

Wallrath, L. L., Rodriguez-Tirado, F. and Geyer, P. K. (2022). Shining Light on the Dark Side of the Genome. *Cells* 11, 330.

Wu, K., Liu, H., Wang, Y., He, J., Xu, S., Chen, Y., Kuang, J., Liu, J., Guo, L., Li, D., et al. (2020). SETDB1-Mediated Cell Fate Transition between 2C-Like and Pluripotent States. *Cell Reports* 30, 25-36.e6.

Yang, S. Y., Baxter, E. M. and Doren, M. van (2012). Phf7 controls male sex determination in the Drosophila germline. *Developmental Cell* 22, 1041–1051.

Yang, P., Wang, Y. and Macfarlan, T. S. (2017). The Role of KRAB-ZFPs in Transposable Element Repression and Mammalian Evolution. *Trends in genetics : TIG* 33, 871–881.

Yi, X., Vries, H. I. de, Siudeja, K., Rana, A., Lemstra, W., Brunsting, J. F., Kok, R. M., Smulders, Y. M., Schaefer, M., Dijk, F., et al. (2009). Stwl modifies chromatin compaction and is required to maintain DNA integrity in the presence of perturbed DNA replication. *Molecular biology of the cell* 20, 983–994.

Yu, Y., Gu, J., Jin, Y., Luo, Y., Preall, J. B., Ma, J., Czech, B. and Hannon, G. J. (2015). Panoramix enforces piRNA-dependent cotranscriptional silencing. *Science* 350, 339–342.

Zhao, K., Cheng, S., Miao, N., Xu, P., Lu, X., Zhang, Y., Wang, M., Ouyang, X., Yuan, X., Liu, W., et al. (2019). A Pandas complex adapted for piRNA-guided transcriptional silencing and heterochromatin formation. *Nat Cell Biol* 21, 1261–1272.

Zinshteyn, D. and Barbash, D. A. (2022). Stonewall prevents expression of ectopic genes in the ovary and accumulates at insulator elements in D. melanogaster. *Plos Genet* 18, e1010110.

Zofall, M., Yamanaka, S., Reyes-Turcu, F. E., Zhang, K., Rubin, C. and Grewal, S. I. S. (2012). RNA elimination machinery targeting meiotic mRNAs promotes facultative heterochromatin formation. *Science* 335, 96–100.

Zolotarev, N., Fedotova, A., Kyrchanova, O., Bonchuk, A., Penin, A. A., Lando, A. S., Eliseeva, I. A., Kulakovskiy, I. V., Maksimenko, O. and Georgiev, P. (2016). Architectural proteins Pita, Zw5, and ZIPIC contain homodimerization domains and support specific long-range interactions in Drosophila. *Nucleic Acids Res* 44, 7228–7241.

Figure legends

Figure 1. H3K9me3 deposition and sex-specific transcriptional regulation are maintained in *phf7*$^{\Delta R}$ ovaries.

(A) Diagram of the *phf7*$^{+}$ and *phf7*$^{\Delta R}$ alleles. Exons are represented by boxes, flanking DNA, and introns by lines, black boxes are coding sequences, and grey boxes are untranslated regions (UTRs). The sequence elements A, R, and B location within the first intron are shaded in yellow, blue, and pink. In ovaries, transcription initiates from TSS2 (pink arrow), whereas an H3K9me3-mediated mechanism represses transcription from TSS1 (black arrow). Locations of primers for ChiP-qPCR and RT-qPCR are indicated by brackets & blue arrowheads. **(B)** H3K9me3 accumulation at the endogenous *phf7* locus in control (*y*^{1} *w*^{1}*)* and mutant ovaries. ChIP-qPCR measured H3K9me3 signal at the *phf7* locus. ChIP to input enrichment is normalized to *rp49.* **(C)** Expression of the testis-specific *phf7* transcript (*phf7-RC)* in control ((*y*^{1} *w*^{1}*)* and mutant ovaries. Expression is measured by RT-qPCR and is shown as the fold change relative to the testis. Expression is normalized to the total level of *phf7.* For both RT-qPCR and ChIP-qPCR experiments. Error bars represent the standard error of the mean (SD) of three biological replicates. A two-tailed Student's t-test estimates statistical significance. where ****p<0.0001.

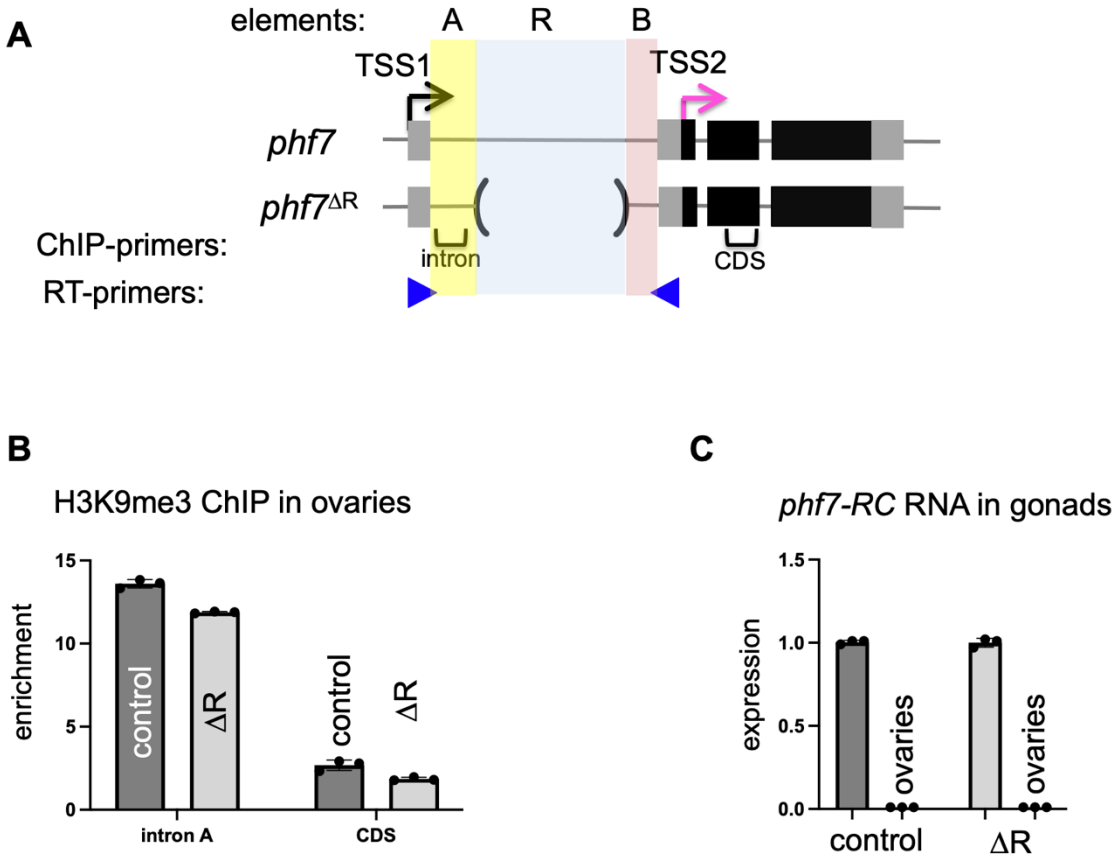


Figure 2. Non-coding sequences within the first intron are sufficient for H3K9me3 deposition.

(A) Diagram of *phf7*$^{+}$, *phf7*$^{\Delta 13,}$ and the genomic transgenes. **(B)** H3K9me3 accumulation on the transgenes. ChIP-qPCR measured signal, and the ChIP to input enrichment normalized to *rp49.* Error bars represent the standard error of the mean (SD) of three biological replicates.

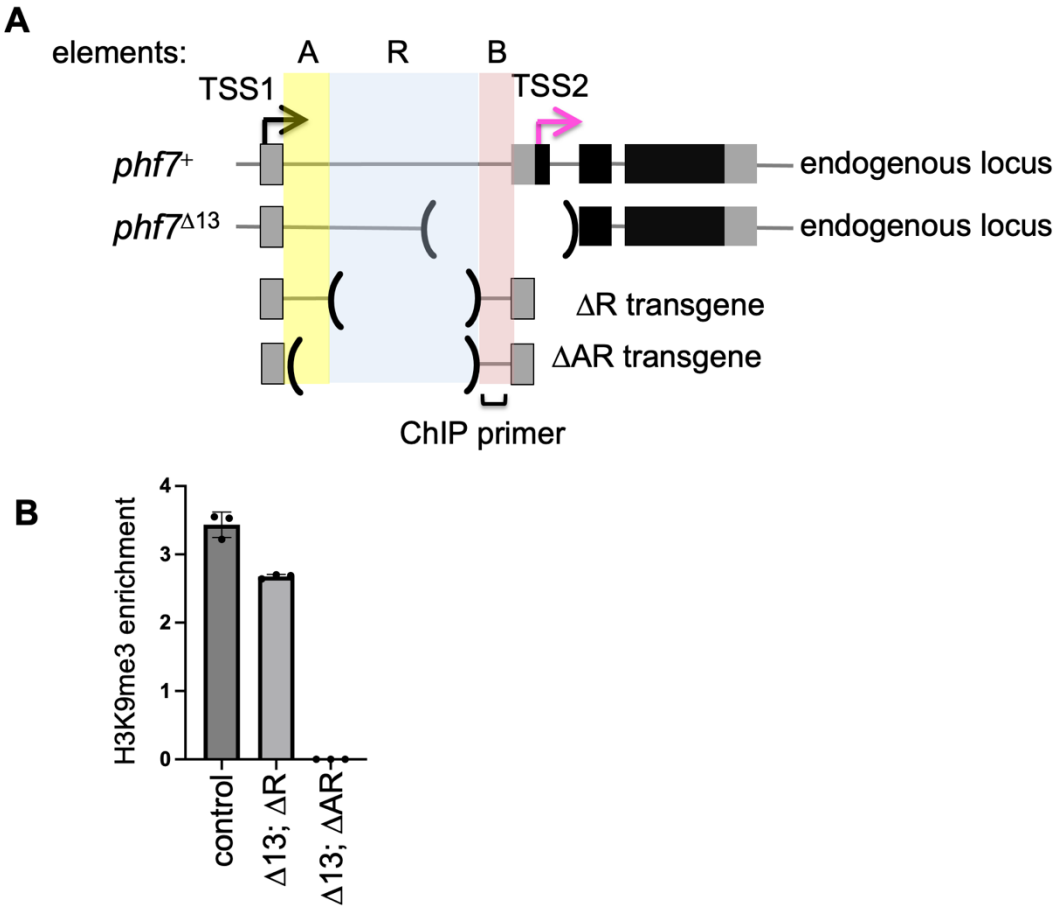
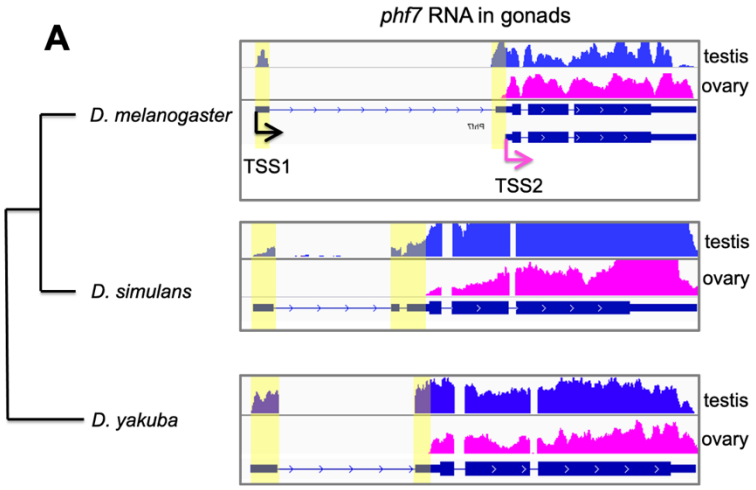


Figure 3. Sex-specific transcriptional regulation is conserved
(A) Genome browser views of RNA-seq data aligned to the *phf7* locus from *D. melanogaster, D. simulans,* and *D. yakuba* ovaries and testis. **(B)** The first exon and a portion of the first intron are conserved. Pairwise alignments between the first exon and the adjacent intron were identified using the Basic Local Alignment Search Tool (BLASTn) available at (https://blast.ncbi.nlm.nih.gov). % identity to the Drosophila melanogaster *phf7* gene is shown.



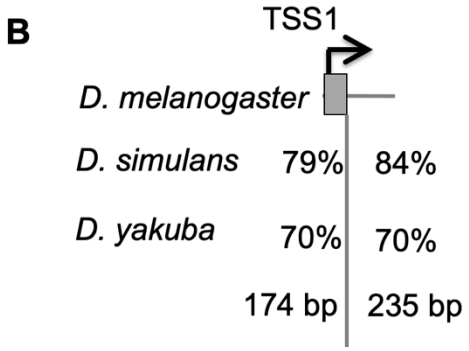


Figure 4. Loss of Setdb1, but not dedicated piRNA-pathway components, leads to female-to-male reprogramming at the *phf7* locus.
Genome browser views of the *phf7* locus. Tracks show RNA-seq reads from control and mutant ovaries aligned to the Drosophila genome (UCSC dm6). The screenshot is reversed so that the 5' end of the gene is on the left. The RNA-seq reads that are unique to *setdb1* mutant ovaries are highlighted. In wild-type control ovaries, only the *phf7*-RA transcript is detected (TSS2, pink arrow), whereas loss of *setdb1* leads to ectopic expression of the testis-specific *phf7-RC* isoform (TSS1, black arrow). In contrast, loss of the essential piRNA components, *rhino, piwi, aub, nxf2* and *panx,* does not disrupt sex-specific *phf7* transcription. The complete mutant genotypes and accession numbers of the RNA-seq data sets are listed in Table S3.

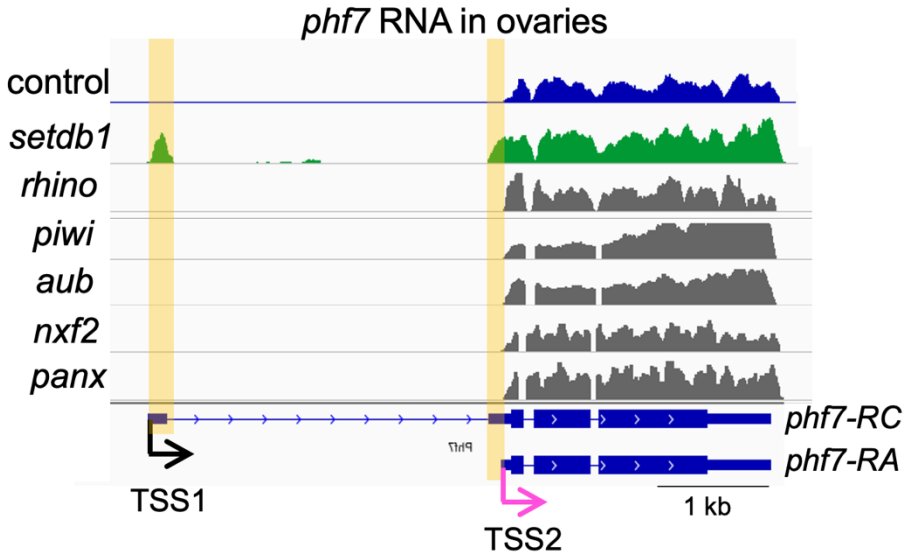


Figure 5. IDC is expressed in both somatic and germline cells.
(A) Confocal image of an ovariole dissected from a female carrying the IDC-GFP rescue transgene stained for GFP (green) and DNA (red). The germarium at the anterior end of the ovariole is boxed in. An oocyte, a nurse cell, and a somatic follicle cell are indicated. Scale bar 25 μm. **(B)** Confocal image of an early-stage egg chamber dissected from a female carrying the IDC-GFP transgene stained for GFP (green) and DNA (red). Scale bar 25 μm. **(C)** Confocal image of the anterior end of a germarium dissected from a female expressing the IDC-GFP fusion protein and co-stained for GFP (green), DNA (white), and SXL (red). Scale bar 25 μm. **(D)** Confocal image of the anterior end of a germarium dissected from a female expressing the IDC-GFP and the HA-BAM fusion proteins co-stained for GFP (green), DNA (white), and HA (red). Scale bar 25 μm.

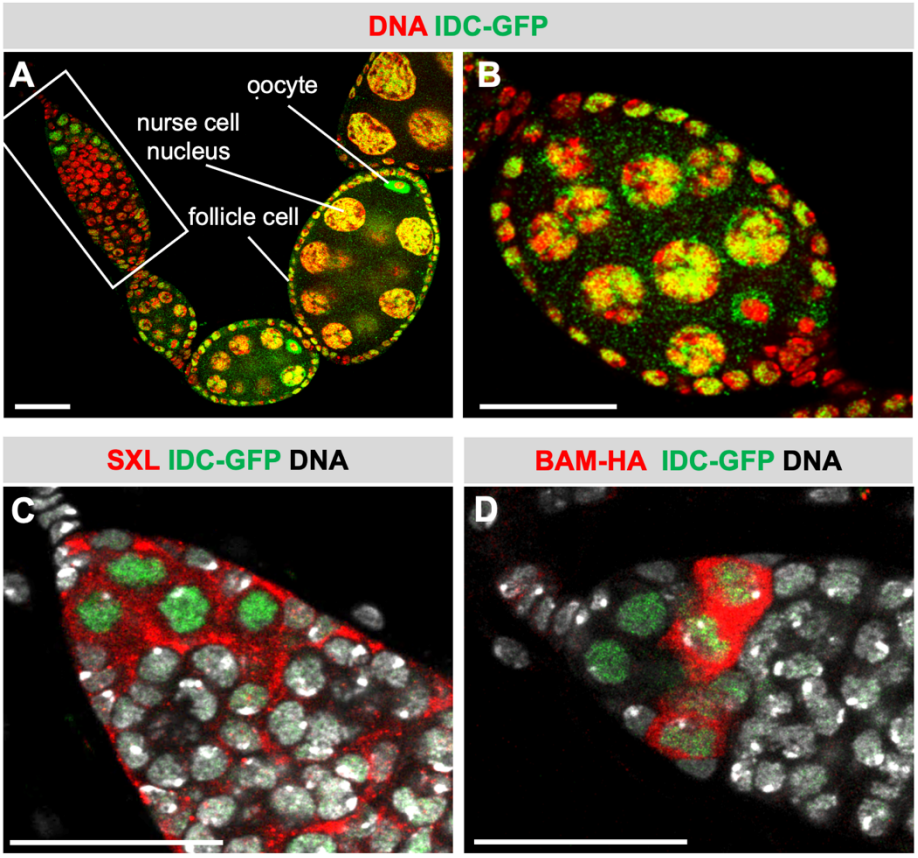


Figure 6. *idc* germ cell-specific knock-down disrupts *phf7* regulation.
(A, B) The intense germ cell IDC-GFP staining pattern observed in the wild-type anterior end of the germarium is abolished in *idc* GLKD germ cells. Confocal images of ovaries were dissected from control (*nos>white-RNAi*) and mutant (*nos>idc-RNAi*) females carrying an IDC-GFP transgene and stained for GFP (white). Scale bar 50 μm. **(C)** *idc* GLKD germ cells express the testis *phf7-RC* transcript. RT-qPCR measures expression normalized to the total level of *phf7* in RNA extracted from control *(nos>white-RNAi*) and *idc* GLKD (*nos>idc-RNAi*) ovaries. The histograms show the mean ± SD of three biological replicates. **(D, E)** *idc* GLKD germ cells inappropriately express the PHF7 protein. Confocal images of ovaries were dissected from control (*nos>white-RNAi*) and mutant (*nos>idc-RNAi*) females carrying an HA-PHF7 transgene and stained for HA (green) and DNA (red). Scale bar 50 μm. **(F)** H3K9me3 accumulation at the endogenous *phf7* locus is reduced in *idc* GLKD ovaries. ChIP-qPCR measured H3K9me3 signal at the *phf7* locus (primer in region A) in control (*nos>white-RNAi)* and mutant (*nos>idc-RNAi*) ovaries. ChIP to input enrichment is normalized to *rp49.* The histogram shows the mean ± SD of three biological replicates. A two-tailed Student's t-test estimates statistical significance. where ****p<0.0001.

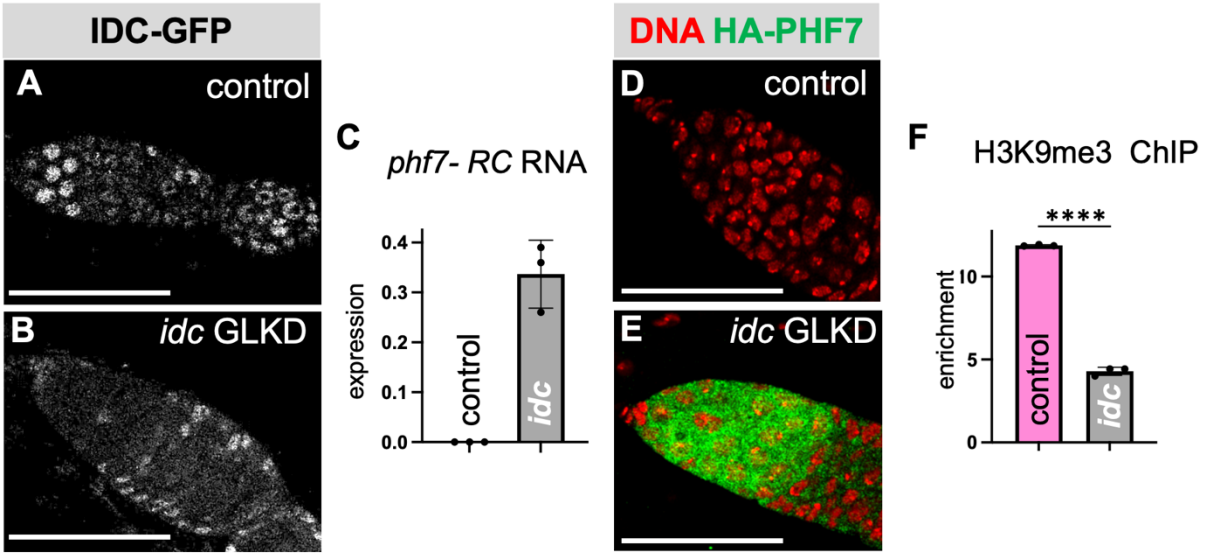


Figure 7. IDC binds to *phf7.*
(A) ChIP-qPCR shows that IDC binds to the *phf7* first exon. Left: The histogram shows the mean ± SD of three biological replicates. Right: Cartoon showing the location of the primers used for ChIP-qPCR. **(B)** Model of H3K9me3-mediated silencing of *phf7.* IDC binds to *phf7,* directing H3K9me3 deposition by the SETDB1 H3K9 methytransferase (HMT) complex. HP1 binds to H3K9me3, resulting in transcriptional silencing.

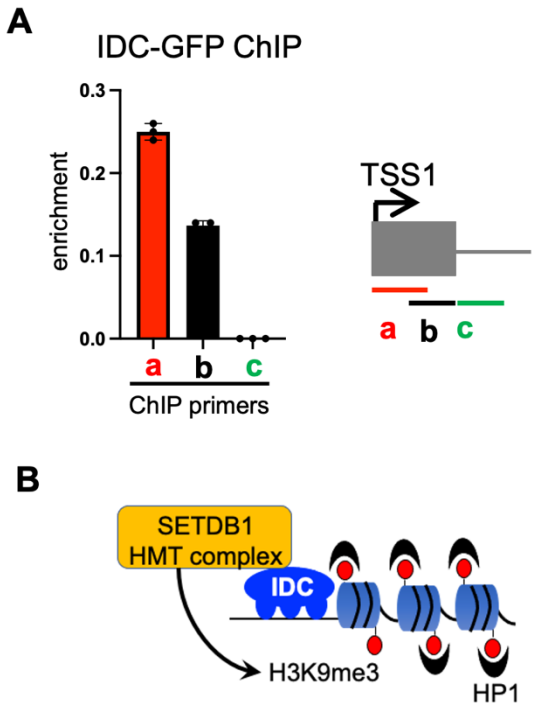


Figure S1. The first intron of *phf7* contains seven copies of an ~250 bp DNA sequence
Sequence alignment of the 7 ~250 bp sequences generated with the multiple sequence alignment tool MUSCLE embedded in the SnapGene software (snapgene.com).