The H3K9 methyltransferase SETDB1 maintains female identity in *Drosophila* germ cells

Anne E. Smolko\(^1\), Laura Shapiro-Kulnane\(^1\) & Helen K. Salz\(^1\)

The preservation of germ cell sexual identity is essential for gametogenesis. Here we show that H3K9me3-mediated gene silencing is integral to female fate maintenance in *Drosophila* germ cells. Germ cell specific loss of the H3K9me3 pathway members, the H3K9 methyltransferase SETDB1, WDE, and HP1a, leads to ectopic expression of genes, many of which are normally expressed in testis. SETDB1 controls the accumulation of H3K9me3 over a subset of these genes without spreading into neighboring loci. At *phf7*, a regulator of male germ cell sexual fate, the H3K9me3 peak falls over the silenced testis-specific transcription start site. Furthermore, H3K9me3 recruitment to *phf7* and repression of testis-specific transcription is dependent on the female sex determination gene *Sxl*. Thus, female identity is secured by an H3K9me3 epigenetic pathway in which *Sxl* is the upstream female-specific regulator, SETDB1 is the required chromatin writer, and *phf7* is one of the critical SETDB1 target genes.
In metazoans, germ cell development begins early in embryogenesis when the primordial germ cells are specified as distinct from somatic cells. Specified primordial germ cells then migrate into the embryonic gonad, where they begin to exhibit sex-specific division rates and gene expression programs, ultimately leading to meiosis and differentiation into either eggs or sperm. Defects in sex-specific programming interfere with germ cell differentiation leading to infertility and germ cell tumors. Successful reproduction, therefore, depends on the capacity of germ cells to maintain their sexual identity in the form of sex-specific regulation of gene expression.

In *Drosophila melanogaster*, germ cell sexual identity is specified in embryogenesis by the sex of the developing somatic gonad. However, extrinsic control is lost after embryogenesis and sexual identity is preserved by a cell-intrinsic mechanism. The SEX-LETHAL (SXL) female-specific RNA binding protein is an integral component of the cell-intrinsic mechanism, as loss of SXL specifically in germ cells leads to a global upregulation of spermatogenesis genes and a germ cell tumor phenotype. Remarkably, sex-inappropriate transcription of a single gene, PHD finger protein 7 (phf7), a key regulator of male identity, is largely responsible for the tumor phenotype. Depletion of phf7 in mutants lacking germline SXL suppresses the tumor phenotype and restores oogenesis. Moreover, forcing PHF7 protein expression in ovarian germ cells is sufficient to disrupt female fate and give rise to a germ cell tumor. Interestingly, sex-specific regulation of phf7 is achieved by a mechanism that relies primarily on alternative promoter choice and transcription start site (TSS) selection. Sex-specific transcription produces mRNA isoforms with different 5′ untranslated regions that affect translation efficiency, such that PHF7 protein is only detectable in the male germline. Although the SXL protein is known to control expression post-transcriptionally in other contexts, the observation that germ cells lacking SXL protein show defects in phf7 transcription argues that Sxl is likely to indirectly control phf7 promoter choice. Thus, how this sex-specific gene expression program is stably maintained remains to be determined.

Here, we report our discovery that female germ cell fate is maintained by an epigenetic regulatory pathway in which SETDB1 (aka EGGLESS, KMT1E, and ESET) is the required chromatin writer and phf7 is one of the critical SETDB1 target genes. SETDB1 trimethylates H3K9 (H3K9me3), a feature of heterochromatin. Using tissue-specific knockdown approaches we establish that germ cell specific loss of SETDB1, its protein partner WINDEI (WDE, aka ATF7IP, MCAF1 and hAM), and the H3K9me3 reader, HETEROCHROMATIN BINDING PROTEIN 1a [HP1a, encoded by the Su(var)205 locus], leads to ectopic expression of euchromatic protein-encoding genes, many of which are normally expressed only in testis. We further find that H3K9me3 repressive marks accumulate in a SETDB1 dependent manner at 21 of these ectopically expressed genes, including phf7. Remarkably, SETDB1 dependent H3K9me3 deposition is highly localized and does not spread into neighboring loci. Regional deposition is especially striking at the phf7 locus, where H3K9me3 accumulation is restricted to the region surrounding the silent testis-specific TSS. Lastly, we find that H3K9me3 accumulation at many of these genes, including phf7, is dependent on Sxl. Collectively our findings support a model in which female fate is preserved by deposition of H3K9me3 repressive marks on key spermatogenesis genes.

**Fig. 1** H3K9me3 is required for germ cell differentiation. a–c Reduced H3K9me3 staining in germ cells upon SETDB1 and WDE depletion (GLKD, see text for details). Representative confocal images of a wild-type (WT), *setdb1 GLKD*, and *wde GLKD* germarium stained for H3K9me3 (green, white in a′–c′). Germ cells were identified by α-VASA staining (magenta). Scale bar, 25 μm. Insets show a higher magnification of a single germ cell, in which the nucleus is outlined by a white dashed line in a′–c′. d–g Undifferentiated germ cells accumulate in *setdb1 GLKD*, *wde GLKD*, and *hp1a GLKD* mutant germaria. Representative confocal images of wild-type and mutant germaria stained for DNA (magenta) and α-Spectrin (cyan) to visualize spectrosomes (sp), fusomes (fu), and somatic cell membranes. Scale bar, 25 μm. h Quantification of mutant germaria with 0, 1–5, and >5 round spectrosome-containing germ cells. The number of scored germaria (n) is indicated.
Results

SETDB1, WDE and HP1α loss blocks germ cell differentiation.

Of the three Drosophila enzymes known to methylate H3K9, only SETDB1 is required for germline development⁹. Several studies reported that loss of SETDB1 caused a block in germ cell differentiation, characteristic of a germ cell tumor phenotype⁰¹. Because of the known connection between the germ cell tumor phenotype and ectopic testis gene transcription, we wondered whether SETDB1 played a role in silencing the expression of testis genes in female germ cells. Previous studies established that SETDB1 is important for Piwi-interacting small RNA (piRNA) biogenesis and transposable element (TE) silencing in germ cells¹²¹⁶¹⁷. However, mutations that specifically interfere with piRNA production, such as rhino, complete oogenesis¹⁸–²¹. Furthermore, our analysis of published RNA-seq data confirmed ectopic expression of the testis-specific PHF7 isoform. Genome browser view of the phf7 locus. Tracks show RNA-seq reads aligned to the Drosophila genome (UCSC dm6). All tracks are viewed at the same scale. The screen shot is reversed so that the 5’ end of the gene is on the left. The reads that are unique to the mutant ovaries are highlighted in gray. The two phf7 transcripts, phf7-RA and phf7-RC, are indicated.

Depletion of H3K9me3 pathway members leads to ectopic expression of the testis-specific phf7-RC isoform. RT-qPCR analysis of the testis phf7-RC transcript in wild-type testis, wild-type and mutant ovaries. Expression, normalized to the total level of phf7, is shown as fold change relative to testis. Primers are shown in panel F. Error bars indicate standard deviation (s.d.) of three biological replicates.

SETDB1, WDE and HP1α mutant ovaries express testis genes.

To investigate the possibility that the loss of H3K9me3 pathway members in female germ cells leads to ectopic testis gene expression, we first used RT-qPCR to assay for the presence of the testis-specific phf7-RC isoform in mutant ovaries. Using primer pairs capable of detecting phf7-RC, we found that phf7-RC is ectopically expressed in setdb1, wde, and hp1α mutant ovaries.
Depletion of H3K9me3 pathway members leads to ectopic expression of a similar gene set. a Comparisons of the deregulated gene expression profiles of setdb1, wde and hp1α GLKD ovaries reveals extensive similarities. Heat map comparing changes in gene expression in setdb1, wde, and hp1α GLKD ovaries compared to wild-type (WT) ovaries. Each row depicts a gene whose expression is deregulated at least 2-fold (FDR < 0.05) in all mutants when compared to wild-type. b Depletion of pathway members leads to ectopic expression of genes which are not normally expressed in ovaries. Scatter plots of significantly upregulated genes in setdb1, wde, and hp1α GLKD ovaries. The Log2 fold-change in gene expression is plotted against the Log2 of the FPKM values in wild-type ovaries. Colored points indicate ectopically expressed genes (Log2 FPKM < 0 in WT ovaries). c Venn diagram showing overlap of ectopically expressed genes in setdb1, wde, and hp1α GLKD ovaries. The amount of overlap is significantly higher than expected stochastically (P < 10^-6). The significance of each two-way overlap was assessed using Fisher’s exact test performed in R, in each case yielding P < 10^-6. The significance of the three-way overlap was assessed by a Monte Carlo simulation, yielding P < 10^-6. For each genotype, the number (n) of ectopically expressed genes is indicated.

To gain a genome-wide view of the expression changes associated with the loss of H3K9me3 pathway members in germ cells, we used RNA-seq to compare the transcriptomes of GLKD mutant ovaries with wild-type ovaries from newborn (0–24 h) females. In agreement with our RT-qPCR analysis, we find that the testis-specific phf7 transcript, phf7-RC, is ectopically expressed in setdb1, wde, and hp1α GLKD mutant ovaries (Fig. 2b–e). In addition to phf7, our differential analysis identified 1191 genes in setdb1 GLKD mutant ovaries, 904 in wde GLKD ovaries, and 1352 in hp1α GLKD ovaries that are upregulated at least 2-fold relative to wild-type (FDR < 0.05; Supplementary Data 1–3). Additionally, 657 genes in setdb1 GLKD mutant ovaries, 756 in wde GLKD ovaries, and 877 in hp1α GLKD ovaries are downregulated (Supplementary Data 4–6). Comparison of the differential gene expression profiles of setdb1 GLKD mutant ovaries with wde and hp1α GLKD mutant ovaries revealed extensive similarities, as expected for genes functioning in the same pathway (Fig. 3a).

Interestingly, all mutants express a set of upregulated genes that are normally not expressed in wild-type ovaries (FPKM < 1 in wild-type ovaries Fig. 3b, c, Supplementary Data 7–9). While we found that there was a significant overlap between the genes ectopically expressed in all three mutant backgrounds (Fig. 3c), we did not find that they were enriched for specific gene ontology terms. However, the pivotal role of phf7 in controlling germ cell sex identity suggested to us that many of the ectopically expressed genes might be normally expressed in testis. To test this idea, we compared our data with published RNA-seq analysis of wild-type testis and bsgc mutant testis23. In spermatogenesis, bsgc is required for the undifferentiated spermatogonia to stop mitosis and transition into the spermatocyte stage. In bsgc mutants this transition is blocked and the testis are enriched for dividing spermatogonial cells. The comparison of the wild-type and mutant expression profiles can therefore be used to identify genes preferentially expressed in early-stage spermatogonia (>2-fold increase in bsgc compared to wild-type, in blue) and in late-stage spermatocytes (>2-fold increase in bsgc compared to wild-type, in green).
Our studies raise the possibility that SETDB1 prevents lineage-peaks. Negative values indicate a reduction in H3K9me3 in mutant ovaries. GLKD

The ChIP-seq reads are in black, green and red. Genome browser view of CG34434 and its neighbors, rhoGAP5A and CG34435. RNA reads in wild-type (WT) and setdb1 GLKD are in blue and orange, respectively. The ChIP-seq reads are in black, green and red. CG34434 is shaded. Average enrichment profile indicates that the H3K9me3 peaks at the 21 SETDB1-regulated genes are localized over the gene body. The average H3K9me3 enrichment profile on the average gene body (transcription start site to transcription end site) scaled to 1500 base pairs, ±3 kb. In red, the average enrichment profile of euchromatic genes which display an H3K9me3 peak in wild-type ovaries. In blue, the average H3K9me3 profile of the 21 genes which are both ectopically expressed and display a loss of H3K9me3 enrichment in setdb1 GLKD ovaries.

This analysis also shows that 33–37% of the ectopically expressed genes are not normally expressed in gonads (gray, Fig. 3d), suggesting that the H3K9me3 pathway also represses somatic gene transcription. However, we did not identify a predominant tissue-specific signature amongst the remaining ectopically expressed genes (Supplementary Fig. 3). Furthermore, and despite ectopic expression of somatic genes, the mutant germ cells retained their germ cell identity, as evidenced by the presence of spectrosomes and fusomes, germ-cell specific organelles, as well as expression of the germ cell marker VASA (Fig. 1a–g). These results indicate that function of the H3K9me3 pathway in germ cells is not restricted to repressing the spermatogenesis gene expression program.

H3K9me3 islands correlate with sex-specific gene expression. Our studies raise the possibility that SETDB1 prevents lineage-inappropriate gene transcription by mediating the deposition of H3K9me3 on its target loci. To test this idea directly, we performed H3K9me3 chromatin immunoprecipitation followed by sequencing (ChIP-seq) on wild-type and setdb1 GLKD ovaries. By limiting the differential peak analysis to euchromatin genes, we identified 746 H3K9me3 enrichment peaks in wild-type that were significantly altered in setdb1 GLKD ovaries (Fig. 4a). Whereas a majority of the gene associated peaks show the expected decrease in H3K9me3 enrichment (84%, 630/746), we also observed regions with an increase in H3K9me3 enrichment (15%, 116/746). How loss of SETDB1 might lead to an increase in H3K9me3 is not known, but the effect is most likely indirect.

At the phf7 locus, H3K9me3 is concentrated over the silenced testis-specific TSS in wild-type ovaries (Fig. 4b). Reduction of this peak in setdb1 GLKD ovaries correlates with aberrant testis-specific transcription (Fig. 2f). These data suggest a functional link between the presence of repressive H3K9me3 chromatin and the silencing of phf7 testis-specific isoform transcription.

We identified an additional 24 normally silenced euchromatic genes at which reduction of the H3K9me3 peak in setdb1 GLKD ovaries correlates with ectopic expression. Of these genes, 4 contain transposable element (TE) sequences. Prior studies have shown that H3K9me3 is enriched around euchromatic TE insertion sites, suggesting the possibility that transcriptional repression might result from spreading of H3K9me3 from a silenced TE26–28. The absence of TE sequences at phf7 and the other 20 genes suggests H3K9me3 deposition is controlled by a different mechanism. Like phf7, the majority of these 20 SETDB1-regulated genes are normally expressed in spermatogenesis (Table 1). Furthermore, examination of their expression pattern in adult tissues, as reported in FlyAtlas27, indicates that 8 of these...
genes express at least one testis-specific isoform. Of the remaining genes, 7 express isoforms in the testis and other tissues, and 5 are not normally expressed in the adult testis. Further studies are needed to determine whether repression of these genes is as important for female germ cell development as *phf7*.

Previous studies have shown that ectopic PHF7 protein expression is sufficient to disrupt female fate and give rise to a germ cell tumor4. We therefore asked if ectopic PHF7 contributes to the *setdb1* GLKD mutant phenotype by generating double mutant females. We found that while loss of *phf7* did not restore oogenesis, there was a shift in the distribution of mutant phenotypes such that the majority of *phf7ΔN18, setdb1* GLKD double mutant ovarioles contained no germ cells (Supplementary Fig. 4). While these data indicate that *phf7* is a critical target of SETDB1 silencing, our finding that the phenotype was not fully rescued suggests that ectopic expression of one or more of the other SETDB1 target genes we identified in this study also contribute to the tumor phenotype.

Co-regulated genes are often clustered together in the genome. In *Drosophila*, about a third of the testis-specific genes are found in groups of three or more genes28,29. However, the 21 SETDB1-regulated genes we have identified do not fall within the previously identified testis gene clusters, nor are they clustered together in the genome (Fig. 5, Supplementary Table 1). Even on the X chromosome, where 11 of the 21 genes are located, the closest two genes are 100 kb away from each other. Thus, the SETDB1-regulated genes are not located within co-expression domains.

In fact, the H3K9me3 peaks at the 21 SETDB1-regulated genes are highly localized and do not spread into the neighboring genes (e.g. Fig. 4b, c, Supplementary Fig. 5). Averaging the H3K9me3 distribution over these 21 genes, scaled to 1.5 kb and aligned at their 5′ and 3′ ends, demonstrates a prominent enrichment over the gene body (Fig. 4d, in blue). In contrast, the average enrichment profile over all euchromatic genes with H3K9me3 peaks showed a broader pattern extending both upstream and downstream of the gene body (Fig. 4d, in red). Together these results clearly show that silencing involves formation of gene-specific blocks of H3K9me3 islands at a select set of testis genes.

### SXL loss in germ cells interferes with H3K9me3 accumulation

Previous studies have established that sex-specific *phf7* transcription is controlled by the female sex determination gene *Sxl*. Together with studies demonstrating that SXL protein is expressed in *setdb1* mutant germ cells, these data suggest that SXL acts upstream, or in parallel to SETDB1 to control *phf7* transcription. To assess the potential of a *Sxl*-mediated
mechanism, we asked whether the loss of SXL in germ cells affects H3K9me3 accumulation. As with our earlier studies, we take advantage of the viable sans-fille148 (snf148) allele to selectively eliminate SXL protein in germ cells without disrupting function in the surrounding somatic cells. SNF, a general splicing factor, is essential for Sxl autoregulatory splicing.7 The viable snf148 allele disrupts the Sxl autoregulatory splicing loop in female germ cells, leading to a failure in SXL protein accumulation, masculinization of the gene expression program (including ph77), and a germ cell tumor phenotype30–32. All aspects of the snf148 mutant phenotype described to date are restored by germ cell-specific expression of a Sxl cDNA. Therefore, the analysis of snf148 mutant ovaries directly informs us of Sxl function in germ cells. Interestingly, we found that the intense H3K9me3 staining foci observed in wild-type germ cells was reduced in snf148 mutants (Fig. 6a, b, Supplementary Fig. 2d, e). However, two lines of evidence indicate that SETDB1 protein expression, measured by an HA tag knocked into the endogenous locus33, was not disrupted in snf148 mutant germ cells. First, in whole mount immunostaining of both wild-type and snf148 mutant germ cells, SETDB1 showed diffuse cytoplasmic staining and punctate nuclear staining (Fig. 6c, d arrow head). Second, Western blot analysis of ovarian extracts showed that snf148 tumors and wild-type ovaries have a similar level of SETDB1 protein (Fig. 6e, Supplementary Fig. 6). Our finding that H3K9me3 staining is disrupted, even though SETDB1 protein accumulation appears normal in snf148 mutant ovaries, leads us to conclude that SXL and SETDB1 collaboratively promote H3K9me3-mediated silencing (Fig. 6f).

To directly test whether Sxl plays a role in controlling H3K9me3 deposition, we profiled the distribution of H3K9me3 by ChIP-seq in snf148 mutant ovaries and compared it to the distribution in wild-type ovaries. By limiting the differential peak analysis to within 1 kb of euchromatic genes, we identified 1,039 enrichment peaks in wild-type that were significantly altered in snf148 mutant ovaries, 91% of which show the expected decrease in H3K9me3 enrichment (Fig. 6g). When we compared the changes in snf148 and setdb1 GLKD mutant ovaries, we found a close correlation ($R^2 = 0.6$; Fig. 6h). The strong overlap between the regions displaying decreased H3K9me3 enrichment in snf148 and setdb1 GLKD mutant ovaries suggests that SXL and SETDB1 influence H3K9me3 accumulation on the same set of genes.

Fig. 6 H3K9me3 distribution is affected in snf148 mutant ovaries. a, b Reduced H3K9me3 staining in snf148 mutant germ cells. Confocal images of a wild-type (WT) and snf148 germline stained for H3K9me3 (green, white in inset). Germ cells were identified by α-VASA staining (magenta). Scale bar, 25 μm. Insets show a higher magnification of a single germ cell, outlined by a dashed line. c, d SETDB1 protein localization is not altered in snf148 mutant germ cells. Confocal images of a germline from wild-type and snf148 females carrying a copy an endogenously HA-tagged allele of setdb1 stained for HA (green, white in inset). Germ cells were identified by α-VASA staining (magenta). Scale bar, 12.5 μm. Insets show a higher magnification of a single germ cell, outlined by a dashed line. e Western blot of ovarian lysates from wild-type and snf148 females carrying a copy an endogenously HA-tagged allele of setdb1 probed with an antibody against the HA tag, α-tubulin the loading control. f Diagram of genetic pathway controlling H3K9me3 accumulation in female germ cells. In WT, SXL collaborates with SETDB1 to regulate H3K9me3 accumulation. snf148, by virtue of the fact that it interferes with Sxl splicing, leads to germ cells without SXL protein. This in turn leads to a defect in H3K9me3 accumulation without interfering with SETDB1 protein accumulation. g Differential analysis of paired H3K9me3 ChIP-seq data identifies peak changes in snf148 mutant ovaries. Scatter plot of significantly altered (FDR < 0.05) H3K9me3 peaks in snf148 ovaries relative to wild-type (WT) ovaries. The x-axis is the log2 H3K9me3 enrichment peaks subtracted from the log2 H3K9me3 enrichment in WT ovaries. Negative values indicate a decreased H3K9me3 peak in mutant ovaries. h setdb1 GLKD and snf148 influence H3K9me3 accumulation on a similar set of genes, including the genes ectopically expressed in setdb1 GLKD ovaries. Plot comparing the significantly altered H3K9me3 peaks observed in snf148 ovaries to setdb1 GLKD ovaries. Genes that are not normally expressed in ovaries but are ectopically expressed in setdb1 GLKD are labeled in red or in blue. Blue indicates genes which are normally expressed in testis.
**Discussion**

This study reveals a role for H3K9me3 chromatin, operationally defined as facultative heterochromatin, in maintaining female identity by silencing lineage-inappropriate transcription. We show that H3K9me3 pathway members, the H3K9 methyltransferase SETDB1, its binding partner WDE, and the H3K9 binding protein HP1a, are required for silencing testis gene transcription in female germ cells. Our studies further suggest a mechanism in which SETDB1, in conjunction with the female fate determinant SXL, controls transcription through deposition of highly localized H3K9me3 islands on a select subset of these genes. The male germ cell sexual identity gene \textit{phf7} is one of the key downstream SETDB1 target genes. H3K9me3 deposition on the region surrounding the testis-specific TSS guarantees that no PHF7 protein is produced in female germ cells. In this model, failure to establish silencing leads to ectopic PHF7 protein expression, which in turn drives aberrant expression of testis genes and a tumor phenotype (Fig. 7).

Prior studies have established a role for SETDB1 in germline Piwi-interacting small RNA (piRNA) biogenesis and TE silencing14,16,17. However, piRNAs are unlikely to contribute to sexual identity maintenance as mutations that specifically interfere with piRNA production, such as \textit{rhino}, do not cause defects in germ cell differentiation18–21 or lead to global changes in gene expression (Supplementary Fig. 1). These findings, together with our observation that \textit{rhino} does not control sex-specific \textit{phf7} transcription, suggests that the means by which SETDB1 methylates chromatin at testis genes is likely to be mechanistically different from what has been described for piRNA-guided H3K9me3 deposition on TEs.

We find that H3K9me3 accumulation at many of these genes, including \textit{phf7}, is dependent on the presence of SXL protein. Thus, our studies suggest that SXL is required for female-specific SETDB1 function. SXL encodes an RNA binding protein known to regulate its target genes at the posttranscriptional levels7. However, piRNAs are unlikely to contribute to sexual identity maintenance as mutations that specifically interfere with piRNA production, such as \textit{rhino}, do not cause defects in germ cell differentiation18–21 or lead to global changes in gene expression (Supplementary Fig. 1). These findings, together with our observation that \textit{rhino} does not control sex-specific \textit{phf7} transcription, suggests that the means by which SETDB1 methylates chromatin at testis genes is likely to be mechanistically different from what has been described for piRNA-guided H3K9me3 deposition on TEs.

We find that H3K9me3 accumulation at many of these genes, including \textit{phf7}, is dependent on the presence of SXL protein. Thus, our studies suggest that SXL is required for female-specific SETDB1 function. SXL encodes an RNA binding protein known to regulate its target genes at the posttranscriptional levels7. However, piRNAs are unlikely to contribute to sexual identity maintenance as mutations that specifically interfere with piRNA production, such as \textit{rhino}, do not cause defects in germ cell differentiation18–21 or lead to global changes in gene expression (Supplementary Fig. 1). These findings, together with our observation that \textit{rhino} does not control sex-specific \textit{phf7} transcription, suggests that the means by which SETDB1 methylates chromatin at testis genes is likely to be mechanistically different from what has been described for piRNA-guided H3K9me3 deposition on TEs.

We find that H3K9me3 accumulation at many of these genes, including \textit{phf7}, is dependent on the presence of SXL protein. Thus, our studies suggest that SXL is required for female-specific SETDB1 function. SXL encodes an RNA binding protein known to regulate its target genes at the posttranscriptional levels7. However, piRNAs are unlikely to contribute to sexual identity maintenance as mutations that specifically interfere with piRNA production, such as \textit{rhino}, do not cause defects in germ cell differentiation18–21 or lead to global changes in gene expression (Supplementary Fig. 1). These findings, together with our observation that \textit{rhino} does not control sex-specific \textit{phf7} transcription, suggests that the means by which SETDB1 methylates chromatin at testis genes is likely to be mechanistically different from what has been described for piRNA-guided H3K9me3 deposition on TEs.

Thus, our studies suggest that SXL is required for female-specific SETDB1 function. SXL encodes an RNA binding protein known to regulate its target genes at the posttranscriptional levels7. However, piRNAs are unlikely to contribute to sexual identity maintenance as mutations that specifically interfere with piRNA production, such as \textit{rhino}, do not cause defects in germ cell differentiation18–21 or lead to global changes in gene expression (Supplementary Fig. 1). These findings, together with our observation that \textit{rhino} does not control sex-specific \textit{phf7} transcription, suggests that the means by which SETDB1 methylates chromatin at testis genes is likely to be mechanistically different from what has been described for piRNA-guided H3K9me3 deposition on TEs.

We find that H3K9me3 accumulation at many of these genes, including \textit{phf7}, is dependent on the presence of SXL protein. Thus, our studies suggest that SXL is required for female-specific SETDB1 function. SXL encodes an RNA binding protein known to regulate its target genes at the posttranscriptional levels7. However, piRNAs are unlikely to contribute to sexual identity maintenance as mutations that specifically interfere with piRNA production, such as \textit{rhino}, do not cause defects in germ cell differentiation18–21 or lead to global changes in gene expression (Supplementary Fig. 1). These findings, together with our observation that \textit{rhino} does not control sex-specific \textit{phf7} transcription, suggests that the means by which SETDB1 methylates chromatin at testis genes is likely to be mechanistically different from what has been described for piRNA-guided H3K9me3 deposition on TEs.

We find that H3K9me3 accumulation at many of these genes, including \textit{phf7}, is dependent on the presence of SXL protein. Thus, our studies suggest that SXL is required for female-specific SETDB1 function. SXL encodes an RNA binding protein known to regulate its target genes at the posttranscriptional levels7. However, piRNAs are unlikely to contribute to sexual identity maintenance as mutations that specifically interfere with piRNA production, such as \textit{rhino}, do not cause defects in germ cell differentiation18–21 or lead to global changes in gene expression (Supplementary Fig. 1). These findings, together with our observation that \textit{rhino} does not control sex-specific \textit{phf7} transcription, suggests that the means by which SETDB1 methylates chromatin at testis genes is likely to be mechanistically different from what has been described for piRNA-guided H3K9me3 deposition on TEs.

We find that H3K9me3 accumulation at many of these genes, including \textit{phf7}, is dependent on the presence of SXL protein. Thus, our studies suggest that SXL is required for female-specific SETDB1 function. SXL encodes an RNA binding protein known to regulate its target genes at the posttranscriptional levels7. However, piRNAs are unlikely to contribute to sexual identity maintenance as mutations that specifically interfere with piRNA production, such as \textit{rhino}, do not cause defects in germ cell differentiation18–21 or lead to global changes in gene expression (Supplementary Fig. 1). These findings, together with our observation that \textit{rhino} does not control sex-specific \textit{phf7} transcription, suggests that the means by which SETDB1 methylates chromatin at testis genes is likely to be mechanistically different from what has been described for piRNA-guided H3K9me3 deposition on TEs. Further studies will be necessary to clarify how the sex determination pathway feeds into the heterochromatin pathway.

\textit{phf7} stands out among the cohort of genes regulated by facultative heterochromatin because of its pivotal role in controlling germ cell sexual identity4,5. Because ectopic protein expression is sufficient to disrupt female fate, tight control of \textit{phf7} expression is essential. \textit{phf7} regulation is complex, employing a mechanism that includes alternative promoter usage and TSS selection. We report here that SETDB1/H3K9me3 plays a critical role in controlling \textit{phf7} transcription. In female germ cells, H3K9me3 accumulation is restricted to the region surrounding the silent testis-specific transcription start site. Dissolution of the H3K9me3 marks via loss of SXL or SETDB1 protein is correlated with transcription from the upstream testis-specific site and ectopic protein expression, demonstrating the functional importance of this histone modification. Together, these studies suggest that maintaining the testis \textit{phf7} promoter region in an inaccessible state is integral to securing female germ cell fate.

Although the loss of H3K9me3 pathway members in female germ cells leads to the ectopic, lineage-inappropriate transcription of hundreds of genes, our integrative analysis identified only 21 SETDB1/H3K9me3 regulated genes. Given that one of these genes is \textit{phf7} and that ectopic \textit{phf7} is sufficient to destabilize female fate4,5, it is likely that inappropriate activation of a substantial number of testis genes is a direct consequence of ectopic \textit{phf7} protein expression. How \textit{phf7} is able to promote testis gene transcription is not yet clear. \textit{phf7} is a PHD-finger protein that preferentially binds to H3K4me2, a mark associated with poised, but inactive genes and linked to epigenetic memory36–38. Thus, one simple model is that ectopic \textit{phf7} binds to H3K4me2 marked testis genes to tag them for transcriptional activation.

It will be interesting to explore whether any of the other 20 SETDB1/H3K9me3 regulated genes also have reprogramming activity. In fact, ectopic fate-changing activity has already been described for the homeobox transcription factor Lim1 in the eye-antenna imaginal disc39. However, whether Lim1 has a similar function in germ cells is not yet known. Intriguingly, protein prediction programs identify three of the uncharacterized testis-specific genes as E3 ligases ([http://www.gene2f.org)]40. SkpE is a member of the SKP1 gene family, which are components of the Skp1-Cullin-F-box type ubiquitin ligase. CG12477 is a RING finger domain protein, most of which are believed to have ubiquitin E3 ligases activity. CG42299 is closely related to the human small ubiquitin-like modifier (SUMO) E3 ligase NSMCE2. Given studies that have linked E3 ligases to the regulation of chromatin remodeling41,42, it is tempting to speculate that ectopic expression of one of more of these E3 ligases will be sufficient to alter cell fate. Future studies focused on this diverse group of SETDB1/H3K9me3 regulated genes and their role in reprogramming may reveal the multiple layers of regulation required to secure cell fate.

The SETDB1-mediated mechanism for maintaining sexual identity we have uncovered may not be restricted to germ cells. Recent studies have established that the preservation of sexual identity is essential in the adult somatic gut and gonadal cells for tissue homeostasis43–47. Furthermore, the finding that loss of HP1a in adult neurons leads to masculinization of the neural circuitry and male specific behaviors48 suggests a connection between female identity maintenance and H3K9me3 chromatin. Thus, we speculate that SETDB1 is more broadly involved in maintaining female identity.

Our studies highlight an emerging role for H3K9me3 chromatin in cell fate maintenance49. In the fission yeast \textit{S. pombe}, discrete facultative heterochromatin islands assemble at meiotic genes that are maintained in a silent state during vegetative growth50,51. Although less well understood, examples in
mammalian cells indicate a role for SETDB1 in lineage-specific gene silencing. Thus, silencing by SETDB1 controlled H3K9 methylation may be a widespread strategy for securing cell fate. Interestingly, H3K9me3 chromatim impedes the reprogramming of somatic cells into pluripotent stem cells (iPSCs). Conversion efficiency is improved by depletion of SETDB1. If erasure of H3K9me3 via depletion of SETDB1 alters the sexually dimorphic gene expression profile in reprogrammed cells, as it does in Drosophila germ cells, the resulting gene expression differences may cause stem cell dysfunction, limiting their therapeutic utility.

Methods

Drosophila stocks and culture conditions. Fly strains were kept on standard medium at 25 °C unless otherwise noted. Knockdown studies were carried out with the following lines generated by the Drosophila Transgenic RNAi Project (openst denti1-PTIRIP.HSM00112) [BDSC #34803, RRID: BDSC_34803, Su(var)205/HP1α-A [TRIP.GLO0531] (BDSC #36792, RRID: BDSC_36792), and wde-PTIRIP.HSM00205] [BDSC #33339, RRID: BDSC_33339]. Different conditions were used to maximize the penetration of the germ cell tumor phenotype. For knockdown of set1β and Hpla the nan-Gal4b-Gal80 driver was used [41], crosses were set up at 29 °C and adults were aged 3–5 days prior to gonad dissection. For wde knockdown the nan-Gal4 driver was used (BDSC #4937, RRID: BDSC_4937) [32], crosses were set up at 18 °C and adults were transferred to 29 °C for 2 days prior to gonad dissection. The following Drosophila stocks were also used in this study: su(TF2) [BDSC #7398, RRID: BDSC_7398], HA-setd1β [phf7ΔN18] and Phac3XHA-Phα7 [32]. Wild-type ovaries are either sibling control y w [BDSC #1495, RRID: BDSC_1495].

For each experiment described below, sample sizes were not predetermined by statistical calculations, but were based on the standard of the field. In a pool of control or experimental animals, specimens of the correct age and genotype were selected randomly and independently from different vials/bottles. Data acquisition and analysis were not performed blindly.

Immunofluorescence and image analysis. Drosophila gonads were fixed and stained according to standard procedures with the following primary antibodies: mouse α-Spectrin (1:100, Developmental Studies Hybridoma Bank [DSHB] A9, RRID: AB_528473), rabbit α-H3K9me3 (1:1,000, Active Motif Cat# 39162, RRID: AB_253212), rat α-HA (1:500, Roche Cat# 11867423001, RRID: AB_399019), mouse α-Sil (1:100, DSHB M18 RRID: AB_528464), rabbit α-Vasa (1:2,000, a gift from the Rangan lab), and rat α-Vasa (1:100, DSHB RRID: AB_760531). Staining was detected with FITC (1:20, Jackson Immunoresearch Labs) or Alexa Fluor 555 (1:200, Life Technologies) conjugated species appropriate secondary antibodies. TO-PRO-3 Iodide (Fisher, Cat# T3605) was used to stain DNA. Images were taken on a Leica SP8 confocal with 1024 × 1024 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. Immunoprecipitation (ChIP) was carried out with commercial kits. Image Manipulation Program (GIMP) and Microsoft PowerPoint. The GIMP Hue-saturation tool was used to assign appropriate colors to merged panels. All staining experiments were replicated at least two times. The “n” in the figure legends represents the number of germlar scored from a single staining experiment.

RNA-seq analysis and data analysis. For each experiment described below, sample sizes were not predetermined by statistical calculations, but were based on the standard of the field. In a pool of control or experimental animals, specimens of the correct age and genotype were selected randomly and independently from different vials/bottles. Data acquisition and analysis were not performed blindly.

RNA-seq and data analysis. For wild-type, set1β GLKD, wde GLKD, and HPla1 ovaries: Total RNA was extracted from dissected ovaries using standard TRIzol (Invitrogen, Cat# 15956026) and DNase (Promega, Cat# M601), Quantity and quality were measured using a NanoDrop spectrophotometer. cDNA was generated by reverse transcription using the SuperScript First-Strand Synthesis System Kit (Invitrogen, Cat# 11904018) using random hexamers. Quantitative real-time PCR was performed using Power SYBR Green PCR Master Mix (ThermoFisher, Cat# 437659) with the Applied Biosystems 7300 Real Time PCR system. PCR steps were as follows: 95 °C for 10 min followed by 40 cycles of 95 °C for 15 s, 60 °C for 60 s, 25 °C for 1 min. A melt curve was generated for each PCR run by an initial denaturation step of 95 °C for 15 s, 60 °C for 1 min, 95 °C for 15 s, and 60 °C for 15 s. Measurements were taken in biological triplicate with two technical replicates. of the phf7-ΔCt levels were normalized to the total amount phf7. Relative transcript amount was calculated using the 2-ΔΔCt method. Primer sequences to measure the expression of phf7, phf7ΔN18, and HP1α were designed (forward GACCTGATCGTGCTCTTGAG; for phf7-ΔCt forward AGTTGGGATTAACGCTT and reverse GATAGACCGCCCCAGCAAGA).

ChIP sequences were generated using the Rubicon Genomics Prep Kit (Cat# R440406) with 16 amplification cycles. DNA was cleaned and

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-06697-x | www.nature.com/naturecommunications
assessed for quality with Qubit and Agilent Bioanalyzer. Sequencing was completed on 2 biological replicates of each genotype with the Illumina HiSeq 2500 v2 with 50 bp single-end reads.

H3K9me3 reads were aligned to the Drosophila genome (UCSC dm6) using bowtie2 (2.2.6)65, and duplicate reads were removed with samtools (1.3)66. Peaks were called with MACS (2.1.20150731) using the broadpeaks option with all other parameters set to default66. Differential peak analysis on all replicates was completed with the DIFFBIND program (2.4.8), using summits = 500 and the DESeq2 package [http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf]. The mean peak concentration (log2) was assessed for quality with Qubit and Agilent Bioanalyzer. Sequencing was completed with the DIFFBIND program (2.4.8), using summits = 500 and the DESeq2 package [http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf]. The mean peak concentration (log2) was calculated by normalizing reads to the total library size and subtracting the corresponding input reads. Differential peak fold changes were calculated by subtracting wild-type mean concentrations from mutant mean concentrations. Mutant peaks were considered significantly altered relative to wild-type if they had a False Discovery Rate (FDR) < 5%. The average H3K9me3 deposition on genes in wild-type ovaries was generated with deepools (2.5.3), using normalized ChIP reads from wild-type reads46. Screen shots are from Integrated Genome Viewer (IGV).

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. RNA-seq and ChIP-seq data sets generated during the course of this study are available from the National Center for Biotechnology Information’s GEO database under accession number GSE109852. RNAseq datasets from testis33 and rhino mutant ovaries31 are available from the National Center for Biotechnology Information’s GEO database under accession numbers GSE86974 and GSE55824. Tissue enrichment datasets are available from the FlyAtlas and FlyAtlas2.

References

Received: 9 May 2018 Accepted: 19 September 2018

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-06697-x | www.nature.com/naturecommunications
44. Regan, J. C. et al. Sex difference in pathology of the ageing gut mediates the greater response of female lifespan to dietary restriction. *eLifE* 5, e10956 (2016).

45. Ma, Q., de Cuevas, M. & Matunis, E. L. Chinmo prevents male to female transformation of adult *Drosophila* ovary. *Development* 143, 754–763 (2016).

46. Ma, Q., Wawersik, M. & Matunis, E. L. The Jak-STAT Target Chinmo Promotes Single-Neuron Sexual Dimorphism. *Science* 341, 474–486 (2014).

47. Grmai, L., Hudry, B., Miguel-Aliaga, I. & Bach, E. A. Chinmo prevents transformer alternative splicing to maintain male sex identity. *PLoS Genet.* 14, e1007203 (2018).

48. Ito, H. et al. Fruitless recruits two antagonistic chromatin factors to establish single-neuron sexual dimorphism. *Cell* 149, 1327–1338 (2012).

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

50. Zofall, M. et al. RNA elimination machinery targeting meiotic mRNAs promotes facultative heterochromatin formation. *Science* 335, 96–100 (2012).

51. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. & Grewal, S. I. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* 292, 110–113 (2001).

52. Jiang, Y. et al. The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain. *Nat. Genet.* 49, 1239–1250 (2017).

53. Du, D. et al. Smad3-mediated recruitment of the methyltransferase SETDB1/ ESET controls Snail1 expression and epithelial-mesenchymal transition. *EMBO Rep.* 19, 135–155 (2018).

54. Koide, S. et al. Sdtd1 maintains hematopoietic stem and progenitor cells by restricting the ectopic activation of nonhematopoietic genes. *Blood* 128, 638–649 (2016).

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

56. Tan, S.-L. et al. Essential roles of the histone methyltransferase ESET in the epigenetic control of neural progenitor cells during development. *Development* 139, 3808–3816 (2012).

57. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors’ initial engagement with the genome. *Cell* 151, 994–1004 (2012).

58. Sridharan, R. et al. Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1y in reprogramming to pluripotency. *Nat. Cell Biol.* 15, 872–882 (2013).

59. Chen, J. et al. H3K9 methylation is a barrier during somatic cell reprogramming into iPSC. *Nat. Genet.* 45, 34–42 (2013).

60. Hu, Y. et al. FlyRNAi.org–the database of the *Drosophila* RNAi screening center and transgenic RNAi project: 2017 update. *Nucleic Acids Res.* 45, D672–D678 (2017).

61. Matias, N. R., Mathieu, J. & Huynh, J.-R. Abscission Is Regulated by the ESCRT-III Protein Shrub in Drosophila Germline Stem Cells. *PLoS Genet.* 11, e1004653–e1004653 (2015).

62. van Doren, M., Williamson, A. L. & Lehmann, R. Regulation of zygotic gene expression in *Drosophila* primordial germ cells. *Carr. Biol.* 8, 243–246 (1998).

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

64. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *J. Gerontol.* 25, 1105–1111 (2009).

65. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-Seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578 (2012).

66. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359 (2012).

67. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 20, 2078–2079 (2009).

68. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137 (2008).

69. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44, W160–W165 (2016).

**Acknowledgements**

We thank Jean-Rene Huynh, Mark Van Doren, Prash Rangan, the Bloomington *Drosophila* Stock Center and the Iowa Developmental Studies Hybridoma Bank for fly stocks and antibodies; Alex Miron, Neil Molyneaux, Ricky Chan, Ulrich Ness, Olivia Corradin, Dan Factor, Alina Saakhaba for help with the bioinformatic analysis; Heather Broihier, Michelle Longworth, Ron Conlon, Peter Harte, and Peter Scacheri for helpful suggestions, discussions, and comments on the manuscript; and Jane Heatwole for fly food. This work was supported by the National Institutes of Health, R01GM102141 and R01GM129478 to H.K.S. and T32GM008056 to A.E.S. Imaging was performed using equipment purchased through NIH S100D016164.

**Author contributions**

Conceptualization: A.E.S and H.K.S.; Investigation, methodology, and analysis: A.E.S and L.S.K.; Manuscript writing, reviewing, and editing: A.E.S and H.K.S.; Supervision: H.K.S.

**Additional information**

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-06697-x.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

**Publisher’s note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2018