PLOS GENETICS

RESEARCH ARTICLE

Stonewall prevents expression of ectopic genes in the ovary and accumulates at insulator elements in *D. melanogaster*

Daniel Zinshteyn[¤], **Daniel A. Barbash***

Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York, United States of America

¤ Current address: Molecular Therapeutics Program,Fox Chase Cancer Center, Philadelphia, Pennsylvania, United States of America
* barbash@cornell.edu

Citation: Zinshteyn D, Barbash DA (2022) Stonewall prevents expression of ectopic genes in the ovary and accumulates at insulator elements in *D. melanogaster*. PLoS Genet 18(3): e1010110. https://doi.org/10.1371/journal.pgen.1010110

Editor: Gregory P. Copenhaver, The University of North Carolina at Chapel Hill, UNITED STATES

Received: May 24, 2021

Accepted: February 18, 2022

Published: March 24, 2022

Data Availability Statement: All raw sequence data files generated for this manuscript have been uploaded to NCBI Sequence Read Archive (SRA) in .fastq format under the BioProject ID PRJNA788954.

Funding: Research was supported by NIH R01-GM074737 to D.A.B. and NIH R01-GM095793 to D.A.B. and Charles Aquadro. Confocal microscope images were obtained at the Microscopy and Imaging Facility of the Cornell Institute of Biotechnology, supported by NIH grant S10RR025502. The funders had no role in study

Abstract

Germline stem cells (GSCs) are the progenitor cells of the germline for the lifetime of an animal. In *Drosophila*, these cells reside in a cellular niche that is required for both their maintenance (self-renewal) and differentiation (asymmetric division resulting in a daughter cell that differs from the GSC). The stem cell—daughter cell transition is tightly regulated by a number of processes, including an array of proteins required for genome stability. The germline stem-cell maintenance factor Stonewall (Stwl) associates with heterochromatin, but its molecular function is poorly understood. We performed RNA-Seq on *stwl* mutant ovaries and found significant derepression of many transposon families but not heterochromatic genes. We also discovered inappropriate expression of multiple classes of genes. Most prominent are testis-enriched genes, including the male germline sex-determination switch *Phf7*, the differentiation factor *bgcn*, and a large testis-specific gene cluster on chromosome 2, all of which are upregulated or ectopically expressed in *stwl* mutant ovaries. Surprisingly, we also found that RNAi knockdown of *stwl* in somatic S2 cells results in ectopic expression of these testis genes. Using parallel ChIP-Seq and RNA-Seq experiments in S2 cells, we discovered that Stwl localizes upstream of transcription start sites and at heterochromatic sequences including repetitive sequences associated with telomeres. Stwl is also enriched at *bgcn*, suggesting that it directly regulates this essential differentiation factor. Finally, we identify Stwl binding motifs that are shared with known insulator binding proteins. We propose that Stwl affects gene regulation, including repression of male transcripts in the female germline, by binding insulators and establishing chromatin boundaries.

Author summary

Stem cells are defined by their ability to divide asymmetrically, resulting in a differentiated cell and a stem cell daughter. In fruit flies, sperm and egg production begins with germline stem cells (GSCs). The ability of a GSC to differentiate or self-renew is tightly regulated by a myriad of factors. Some of these are transcription factors, which are responsible for

design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

activating or suppressing other genes to promote one state in favor of another. Stonewall is an ovarian nuclear protein required for GSC self-renewal, whose molecular function is poorly understood. Here we show that Stonewall is responsible for preventing the activation of "male" molecular programming in the fruit fly ovary. When Stonewall is absent from the ovary, egg production is terminated and testis-specific genes become highly expressed, including the male transcript of *Phf7*, which induces male sexual identity in female germ cells. We also show that Stonewall is likely localizing to genomic insulators, which are regions of the genome that shield genes from nearby regulators. Our findings suggest that Stonewall helps to organize the genome in ovarian germ cells and prevent expression of male genes.

Introduction

Adult stem cells exist in tissues where there is constant turnover of cells, such as gonads where gametes are continually produced and released. Germline Stem Cells (GSCs) are one of several adult stem cell populations that inhabit ovaries and testes. In the female ovary of the fruit fly, they reside in a niche environment that is required to maintain them [1,2]. Stem cells undergo asymmetric cell division, resulting in one differentiated daughter cell and one daughter cell that is identical to the parent, thus undergoing self-renewal. For ovarian GSCs, the differentiated daughter cell is the cystoblast, which then undergoes four rounds of incomplete mitosis to form a 16-cell cyst. Upon completing these four rounds, one of the cystocytes in the 16-cell cyst enters meiosis while the other 15 undergo endoreduplication. The meiotic cell will differentiate into an oocyte while the other 15 will become nurse cells that provide maternal factors to the oocyte.

The entire germ cell population of the ovary is derived from the 2–3 GSCs in each germarium. *Drosophila* have an intricate regulatory network of factors that are required for normal GSC function [3], which can be broadly categorized as maintenance factors required for self-renewal or differentiation factors required for cystoblast production. Many of the genes involved in GSC regulation are pleiotropic for other functions inside and outside of the ovary. For example, Piwi is required for both GSC maintenance and differentiation [4,5], as well as for silencing of transposable elements via the piRNA pathway [6]. Well-known differentiation factors include the translational repressors Bam, Bgcn, and Sxl. These proteins form a complex that represses mRNAs associated with GSC renewal, including the maintenance gene *nanos* [7]. Ovaries lacking in any of these differentiation factors exhibit a tumorous ovary phenotype, manifested as an overabundance of GSC-like cells that fail to become cystoblasts. Sxl is essential for the cell-autonomous sex determination of germ cells [8].

The *stwl* gene was discovered in a *P*-element mutagenesis screen for female sterility and subsequently identified as a germline-expressed gene in an enhancer trap screen [9,10]. It is primarily expressed in germline cells of the ovary, with weak expression in GSCs and increased expression in GSC progenitor cells (cystoblasts) and beyond [11]. *stwl* mutant egg chambers contain 16 polyploid nurse cells, indicating that the cystocyte-to-oocyte transformation does not occur and that *stwl* is required for oocyte determination. Egg chamber growth in *stwl* null ovaries arrests between stages 4 and 7, with germ cells undergoing apoptosis. *stwl* is also required for GSC maintenance: mutant ovaries typically lack GSCs, especially in older flies [12]. *stwl* mutant GSC clones are rapidly depleted from the ovary via differentiation into cystoblasts, and egg chambers derived from these clones exhibit oocyte determination defects as seen in *stwl* mutant animals [13]. Double mutants with *stwl* and either *bam* or *bgcn* make

differentiated ovarian germ cell cysts, unlike in *bam* or *bgcn* single mutants, suggesting that *stwl* function is downstream of *bam* and *bgcn* [13,14].

Stwl is also involved in heterochromatin maintenance. Heterochromatic regions are largely transcriptionally silent and populated by genomic parasites, including transposons. *stwl* mutations are dominant suppressors of position-effect variegation, suggesting that Stwl is required to promote the spreading of heterochromatin [13]. Stwl colocalizes with the heterochromatin-binding protein HP1a and dense, heterochromatin-like structures at the nucleolus in S2 cells, acts as a transcriptional repressor in *in vitro* experiments, and promotes the spreading of heterochromatic histone markers H3K9me3 and H3K27me3 in larvae [15].

Despite the essential function of *stwl* in *D. melanogaster*, the gene is novel to the *Drosophila* genus, part of a lineage-specific expansion of MADF-BESS domain genes [16,17]. Furthermore, the *stwl* locus has undergone recurrent positive selection at least once in the genus and has higher overall rates of substitution relative to closely related genes [18–20]. The driver of this evolutionary signature has not been identified, though some have speculated that the role of Stwl in heterochromatin maintenance may involve interaction with transposons [18,19]. Interestingly, both the narrow phylogenetic distribution and positive selection seen for *stwl* are shared with other essential germ line stem cell regulatory genes including *bam* and *bgcn* [19,21].

Stwl colocalizes with the insulator binding protein CP190 in terminal filament cells of the ovary and presents as puncti at the nuclear lamina [22]. Insulators are genomic regions that, when appropriately bound by insulator proteins, can prevent interaction between enhancers and their target promoters, or modulate the spreading of chromatin modifications [23]. We performed here a range of genomic experiments in order to identify genes regulated by Stwl and its possible molecular function. We report data from transcriptional profiling in mutant ovaries and ChIP-Seq in S2 cells which suggest that loss of Stwl results in the activation of male-specific genetic programming and misregulation of transposable elements, and that Stwl localizes to insulators.

Results

stwl deficient ovaries exhibit TE derepression

A screen for genes whose RNAi-induced knockdown (KD) in ovaries leads to misexpression of TEs found that germline KD of *stwl* results in moderate derepression of three TE transcripts (*Het-A*, *blood*, *and burdock*), as determined by qRT-PCR [24]. We performed our own qRT-PCR experiments to test for misexpression of the LTR retrotransposon *Copia*, and the non-LTR retrotransposons *Het-A*, *412*, and *I* element. *Het-A*, *Copia* and *I* element are germline-restricted, while *412* is expressed in both germline and ovarian follicle cells [25,26]. We tested for misexpression of these TEs in ovaries dissected from 2-day old trans-heterozygous null ($stwl^{j6c3}$/*Df(3L)Exel6122*), hemizygous ($stwl^{+}$/*Df(3L)Exel6122*), and wild-type ($stwl^{+}$/$stwl^{+}$) flies. We found that each of these TEs is derepressed in trans-heterozygous null ovaries, relative to hemizygous and wild-type ovaries (Fig 1A).

The *stwl* mutant phenotype presents challenges for interpreting assays of transcript abundance. *stwl* mutant ovaries are largely agametic as a consequence of GSC loss and defects in oocyte determination (S1 and S2 Figs) [11–13]. Nurse cells in *D. melanogaster* ovaries are polyploid and produce large quantities of mRNA that are maternally inherited by the developing oocyte. Differential expression between agametic mutant and wild-type ovaries might therefore reflect extensive differences in the cellular makeup of the ovaries rather than changes in transcript abundance specifically due to *stwl*. In order to account for differences in tissue composition, we chose the approach utilized by previous studies, which is to assay transcripts

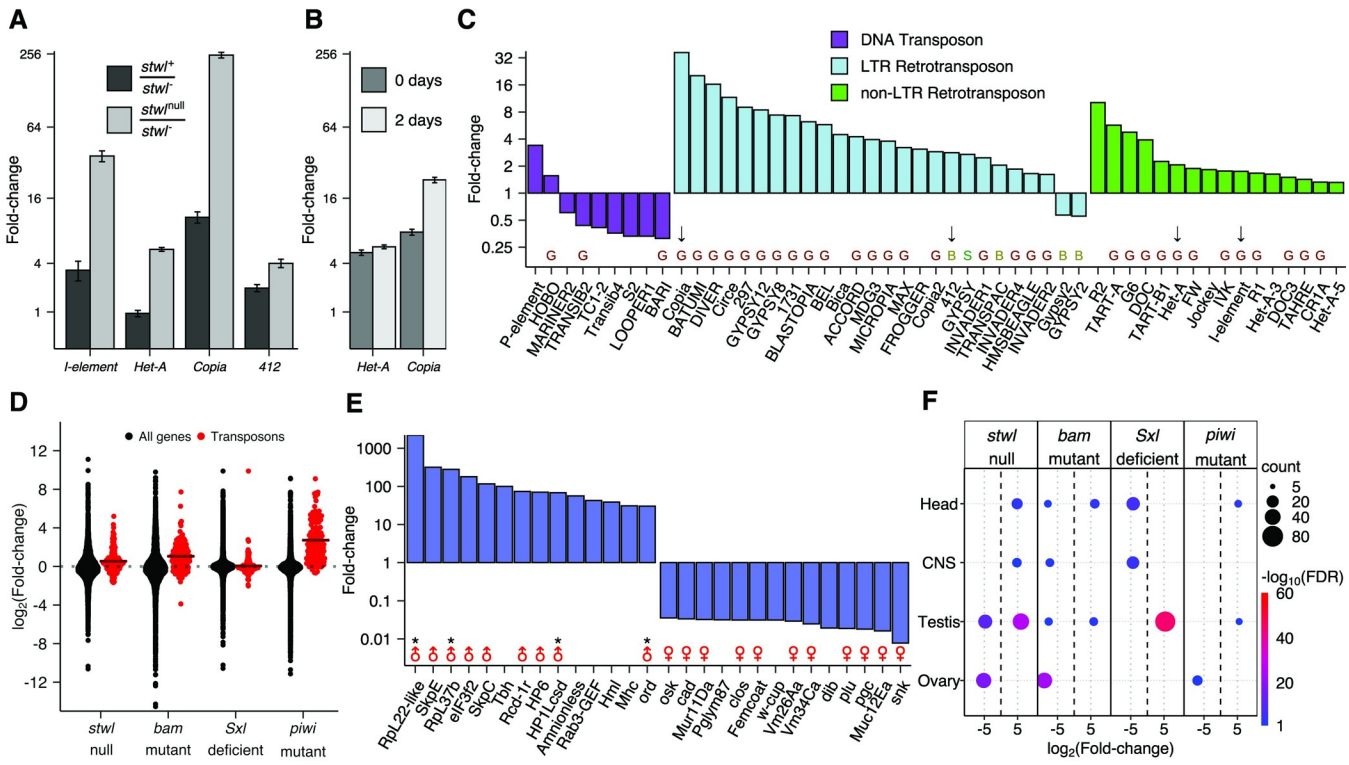


Fig 1. Loss of *stwl* results in upregulation of transposons and testis-enriched genes. (A) qRT-PCR of TEs from 2-day old ovaries, scaled to wild-type. $stwl^{null}$ is a mutant allele ($stwl^{j6c3}$), $stwl^{-}$ is a deficiency allele (*Df(3L)Exel6122*). (B) TEs are upregulated in both 0- and 2-day old ovaries. qRT-PCR of TEs from *stwl* null ($stwl^{j6c3}/stwl^{j6c3}$) ovaries, scaled to $stwl^{+}$/*Df(3L)Exel6122*. (A-B) Mean and SE plotted from 3 biological replicates, each with 3 technical replicates. (C) Fold-change of TEs in *stwl* null ($stwl^{j6c3}/stwl^{j6c3}$) relative to wild-type from RNA-Seq assay of 0- and 2-day old ovaries. Black arrows point to TEs validated with qRT-PCR data in Fig 1A and/or 1B. "G","S","B" indicates whether TE is typically expressed in germline, ovarian soma, or both, respectively [26]. (D) $\log_2$Fold-change (LFC) of TEs vs. all genes from *stwl*, *bam*, *Sxl*, and *piwi* mutant or deficient ovaries, relative to wild-type. Crossbars show the mean LFC for all TEs. (E) Fold-change of the top 14 and bottom 14 most affected annotated genes (based on FlyBase annotations) in *stwl* null ovaries relative to wild-type. Male and female symbols mark genes with testis- and ovary-enriched wild-type expression, respectively; "*" marks genes that are part of the 59C4-59D testis-specific cluster described in Fig 2. (F) Enriched tissue classes among the top and bottom 1% of misregulated genes in *stwl*, *bam*, *Sxl*, and *piwi* mutant or deficient ovaries. Average LFC is plotted for each set of tissue-enriched genes enriched among *stwl*, *bam*, *Sxl*, and *piwi* mutant or deficient ovaries relative to wild-type. Only gene sets with FDR <0.05 are plotted.

https://doi.org/10.1371/journal.pgen.1010110.g001

derived from extremely young ovaries [27,28]. These authors reasoned that dissection of ovaries from newly-eclosed individuals (dissected within 24 hours of eclosion) would limit the amount of late-stage egg chambers and eggs that are present. Our analyses confirm that ovaries from newly-eclosed *stwl* mutants more closely resemble wild-type in their morphology than older ones (S2–S4 Figs).

We assayed TE transcripts from newly-eclosed ovaries from both $stwl^{j6c3}$/*Df(3L)Exel6122* and $stwl^{+}$/*Df(3L)Exel6122*, as above. We found that the fold-increase of *Het-A* in trans-heterozygous nulls relative to hemizygotes is similar in newly-eclosed and 2-day-old ovaries (5-fold and 6-fold increase, respectively) (Fig 1B). *Copia* transcript is also derepressed in newly-eclosed *stwl* transheterozygous null ovaries (8-fold increase over hemizygotes), though this derepression phenotype is not as large as the one observed in 2-day-old ovaries (23-fold increase over hemizygotes). We conclude that the TE derepression phenotype we and others [24] have observed is likely due to loss of *stwl* activity, not to a general loss of germ cells.

In order to identify the genome-wide consequences of *stwl* loss, we performed RNA-Seq on ovaries dissected from newly-eclosed and 2-day old wild-type (*y w*) and *stwl* null (*y w;* $stwl^{j6c3}/stwl^{j6c3}$) individuals. The goal of this experiment was to identify and classify genes and TEs

which are consistently differentially expressed in *stwl* mutants. Therefore, we incorporated all four sample types (newly-eclosed wild-type, newly-eclosed mutant, two-day old wild-type, two-day old mutant) into a generalized linear model using DESeq2 [29]. A gene was only considered differentially expressed in *stwl* nulls if the transcript count for that gene significantly changed across both null samples relative to wild-type; that is, if the gene was differentially expressed between the two genotypes, regardless of age.

After accounting for potential batch effects and GC-content bias (see Methods), sample-to-sample distances for the resultant count matrices confirmed that the biological replicates for each sample type cluster together (S5 Fig). Principal Component Analysis (PCA) of the count data demonstrated that the samples are primarily stratified according to ovary maturity (S6 Fig). Principal Component 1 (PC1) accounts for 58% of the variance in the count matrix, which separates mature ovaries (2-day old $stwl^{+}/stwl^{+}$ wild-type) from immature ovaries (2-day old $stwl^{j6c3}/stwl^{j6c3}$ null, 0-day $stwl^{j6c3}/stwl^{j6c3}$ null, and 0-day old $stwl^{+}/stwl^{+}$wild-type). These trends in the PCA support the rationale behind our experimental design, in that comparing null and WT ovaries at two time-points more accurately identifies genes that are differentially expressed due to genotype. We found that analyzing 0-day and 2-day old ovaries separately yielded similar results to the combined approach in terms of upregulation of TEs and testis genes (S1 Table, S7 Fig). While *stwl* deficiency in the 2-day old ovaries resulted in a larger number of affected genes (ovarian transcripts were especially down-regulated, as expected), incorporating these data into a GLM with 0-day old ovaries increased the likelihood of identifying stwl-dependent genes.

We found that 4,839 genes (out of 10,165 genes with mean read count >10 across all ovary samples) are differentially expressed, with 2,147 genes upregulated in *stwl* null and 2,692 downregulated in *stwl* null (48%, 21%, and 26% of expressed genes) (S1 Table, S8 Fig). We also found that *P-element* transcript increases ~4-fold in *stwl* null ovaries; this can be explained by the *P-element* insertion into the *stwl* locus that created the $stwl^{j6c3}$ allele and serves as an internal validation for the presence of the $stwl^{j6c3}$ allele. The RNA-Seq data showed that repetitive elements are strongly impacted by loss of *stwl* (Fig 1C). These repeats include the *Copia*, *Het-A*, *412*, and *I element* elements we identified by qRT-PCR.

Loss of *stwl*, *bam*, and *piwi*, but not *Sxl*, results in TE derepression

Although *stwl* is expressed in somatic cells of the ovary, we found that RNAi-mediated knockdown of *stwl* using follicle cell drivers had no effect on fertility, while germline-specific knockdown strongly reduces or eliminates fertility (S9 Fig). These results are consistent with previous findings that the terminal phenotype of *stwl* mutant ovaries is sterility caused by loss of germline stem cells and apoptosis of differentiated germ cells (S1 and S2 Figs) [11–13]. DNA damage is also apparent in these sterile ovaries, possibly due to *stwl*'s requirement for maintenance of heterochromatin [15]. It is also possible that TE derepression is a consequence of these defects, rather than reflecting a direct role of Stwl in TE silencing. To help distinguish between these possibilities, we analyzed published RNA-seq data generated from ovaries for mutations in various genes affecting GSC maintenance or differentiation. These included the differentiation factor *bag-of-marbles* (*bam*), the sex-determination master regulator *Sex-Lethal* (*Sxl*), and the GSC maintenance factor and piRNA targeting protein *piwi* (*piwi*). *bam* and *Sxl* deficient ovaries exhibit a "bag of marbles" phenotype that is characteristic of disruption of differentiation factors that results in over-proliferation of GSC-like cells [27,30,31]. *piwi* mutants exhibit GSC maintenance defects similar to those in *stwl* mutants [4,6,32,33]. We found that loss of *bam* and *piwi*, but not *Sxl*, results in upregulation of TEs, particularly LTR retrotransposons and germline-expressed TEs (Figs 1D and S10, S2 Table). To our knowledge, TE

derepression has not previously been observed in *bam* mutants. These results suggest that the *stwl* TE derepression phenotype is not an indirect reflection of the loss of GSCs, since it also occurs in *bam* mutants, which have the opposite phenotype of GSC overproliferation. It is also notable though that the magnitude of effects in *stwl* mutants is substantially lower than in *piwi* mutants, suggesting that *stwl* may not be a direct repressor of TEs as *piwi* is. We suggest instead that loss of *stwl* may lead to widespread but moderate derepression of TEs through its role in insulator function described below.

A subset of testis-enriched genes are highly upregulated in *stwl* null ovaries

Other GSC regulatory genes, including *Sxl*, are required in ovaries for silencing of testis-specific transcripts [27]. We tested whether *stwl* mutant ovaries also exhibit abnormal derepression of testis-specific genes. We utilized RPKM values from the modENCODE anatomy RNA-Seq dataset to classify all genes according to tissue-biased expression [34]. We found that testis-enriched genes are among the most upregulated genes in *stwl* null ovaries, while ovary-enriched genes are among the most downregulated (Fig 1E and 1F). Genes consistently upregulated in *stwl*, *bam*, *Sxl*, and *piwi* deficient ovaries were not biased for expression in any tissue (S2 and S3 Tables, S11 Fig).

While testis-biased transcripts are among the most upregulated in *stwl* null ovaries, Gene Set Enrichment Analysis (GSEA) found that on the whole they are downregulated in *stwl* null ovaries (S12 Fig). This reflects a limitation in GSEA performance previously noted when attempting to perform analyses on large and potentially complex gene sets [35,36]. The essential problem is that these sets include genes that are misregulated in both directions, presumably because members of the same pathway may be either down- or upregulated in response to misexpression of an upstream activator or suppressor. We therefore performed an over-representation test for tissue-enriched genes among the top and bottom 1% of expressed genes in each RNA-Seq experiment, to identify strong biases at the extremes of the most differentially expressed genes (Fig 1F). Our top/bottom percentile over-representation tests confirmed that testis-enriched genes are over-represented within the top 1% of most highly upregulated genes in agametic ovaries, but they are much more prevalent in *stwl* and *Sxl* null or deficient ovaries, relative to *bam* or *piwi* mutant ovaries. We also found that transcripts enriched in adult head, as well as pharate and larval stage central nervous system (CNS) are upregulated in *stwl* null ovaries (Fig 1F).

Loss of *stwl* results in ectopic expression of a testis-specific gene cluster

In order to test whether specific regions of the genome are misregulated, we plotted LFC by genomic location (Fig 2A). We found a striking pattern of expression at 59C4-59D on chromosome 2R, where 11 genes clustered within 227.5 Kb are strongly upregulated in *stwl* null ovaries. Four of the genes in this cluster are among the most strongly upregulated genes in *stwl* null ovaries (Fig 1E). Coexpressed gene clusters are common in many species, adding a dimension of organization to the genome by allowing groups of adjacent genes to be regulated simultaneously. Testis-specific gene clusters are particularly common in *D. melanogaster*, with 59C4-59D being the largest, and their expression is tightly regulated to prevent somatic expression [37]. We confirmed that the 59C4-59D cluster described by Shevelyov et al. [37] is composed mostly (28/34 total genes) of testis-enriched genes, most of which are absent from the wild-type ovarian transcriptome (S4 Table).

Loss of the H3K9me3 pathway components results in ectopic expression of testis-enriched genes [27,31]. Similarly, we found that the genes of the 59C4-59D cluster are transcriptionally inert in ovaries and become ectopically expressed in *stwl* null ovaries (Fig 2B, S4 Table). We

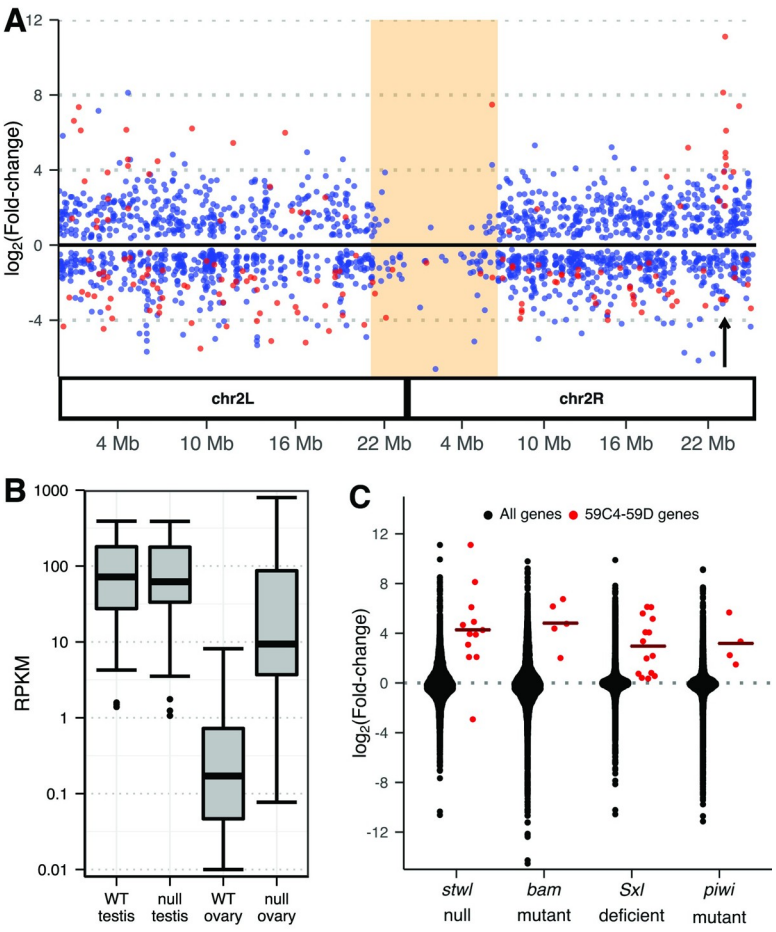


Fig 2. A cluster of testis-specific genes is derepressed in *stwl* null ovaries. (A) Differential expression (DE) of genes in ovaries (*stwl* null/WT) along chromosome 2. Only DE genes (FDR <0.01) are plotted. Testis-enriched genes are red (see Methods). Shaded orange area marks pericentromeric heterochromatin; arrow points to the testis-enriched cluster at 59C4-59D. (B) Reads Per Kilobase of transcript per Million mapped reads (RPKM) of genes in the 59C4-59D cluster in wild-type and *stwl* null gonads. Low-count outliers are not plotted. (C) $\log_2$Fold-change (LFC) of 59C4-59D cluster vs. all genes from *stwl*, *bam*, *Sxl*, and *piwi* mutant or deficient ovaries, relative to wild-type. Crossbars show the mean cluster LFC.

https://doi.org/10.1371/journal.pgen.1010110.g002

also found that this cluster is upregulated in *bam*, *Sxl*, and *piwi* mutant or deficient ovaries (Fig 2C), making it a potentially useful transcriptional reporter for loss of sex-specific gene silencing.

Loss of *stwl* results in derepression of testis-enriched genes in S2 cells and ovaries

Even when assayed mutant tissues appear morphologically similar to wild-type, the pleiotropic functions of Stwl nonetheless make it challenging to identify which genes are specifically mis-regulated as a consequence of Stwl loss. In order to further address this concern, we performed RNA-seq on a homogeneous tissue, using S2 cells treated with *stwl* dsRNA (see Methods). Immunoblotting against Stwl protein showed that *stwl* dsRNA treatment reduced Stwl protein levels by at least 80%, and RNA-Seq confirmed that *stwl* transcript was reduced by ~85% (Fig 3A and 3B).

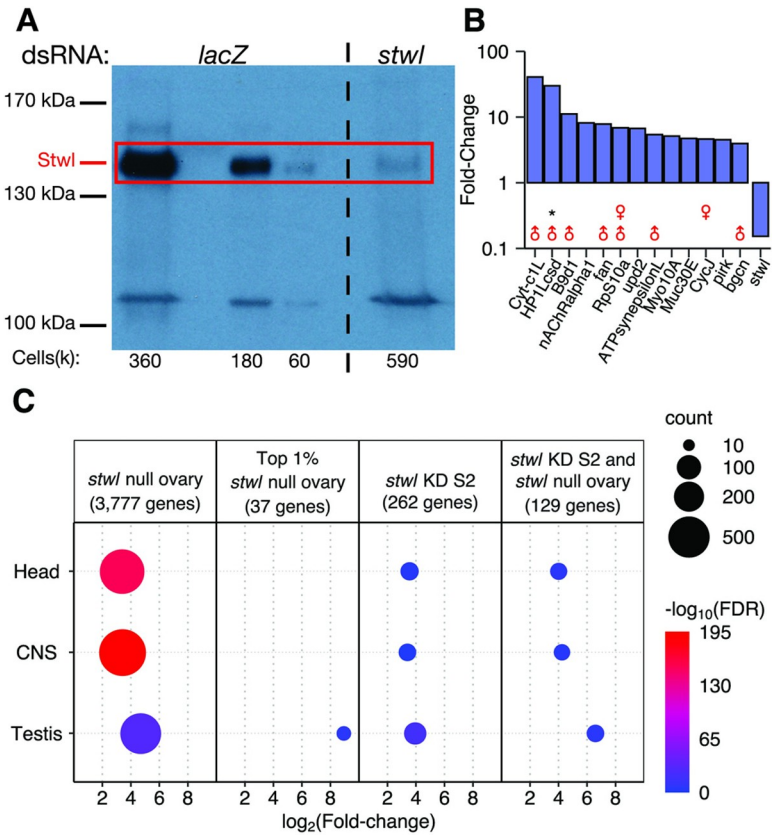


Fig 3. RNAi knockdown of *stwl* in S2 cells results in derepression of testis-enriched genes. (A) Western blot with anti-Stwl antibody of S2 cells treated with either control dsRNA (*lacZ*) or dsRNA targeting the *stwl* transcript. The estimated number of cells per lane (multiplied by 1000) is shown below the blot. (B) Fold-change of the 13 most affected annotated genes in *stwl* KD S2 cells relative to *lacZ* control. Male and female symbols indicate whether gene is testis- or ovary-enriched in wild-type. "*" marks genes that are part of the 59C4-59D testis-specific cluster. (C) FDR, count, and mean $\log_2$Fold-change (LFC) is plotted for each set of tissue-enriched genes that is overrepresented among ectopically expressed genes in *stwl* null ovaries and *stwl* dsRNA-treated S2 cells. Overrepresentation tests were also performed on the top 1% by LFC of ectopic genes in *stwl* null ovary, and of genes ectopic to both *stwl* null ovary and *stwl* dsRNA-treated S2 cells (for this intersect group, average LFC values in *stwl* null ovary are plotted). Only gene sets with FDR <0.05 are plotted.

https://doi.org/10.1371/journal.pgen.1010110.g003

Relative to loss of *stwl* in ovaries, *stwl* dsRNA treatment of S2 cells had a more subtle effect on transcript abundance and little effect on TEs (S13 Fig). S2 cells are male and hematopoietic-derived, and express fewer genes than ovaries [38,39]. Nonetheless, as in the *stwl* null ovaries we found that testis-enriched genes are among the most highly upregulated genes in *stwl* dsRNA-treated S2 cells, including a member of the 59C4-59D cluster (Fig 3B). Due to the very low average transcript abundance at this cluster in S2 cells, most of these genes were removed from the differential expression analysis performed by DESeq2 [29] (S1 Table). Our further analysis of ectopically expressed genes, which was not limited by low average counts, found that 5/12 of the 59C4-59D genes upregulated in *stwl* null ovary are ectopically expressed in *stwl* dsRNA-treated S2 cells (S4 Table). Testis-enriched genes are over-represented among the top 1% of ectopically expressed genes in *stwl* dsRNA-treated S2 cells (Fig 3C).

For comparison, we applied the same methodology to our ovary data and found that 3,777 genes are ectopically expressed in *stwl* null ovaries. Head- and CNS-enriched genes are highly overrepresented among these ectopically enriched genes, as are testis-enriched genes, though to a lesser degree (Fig 3C). However, testis-enriched genes are the only upregulated tissue

classes among the top 1% of ectopically expressed genes by LFC (Fig 3C). We found that 49% (129/262) of ectopically expressed genes in *stwl*-dsRNA treated S2 cells are also ectopically expressed in *stwl* null ovaries. This subset of genes is highly enriched for testis, head, and CNS transcripts (Fig 3C). We conclude that *stwl* functions to repress genes with male-enriched expression, including in somatic tissue culture cells and ovaries.

Stwl regulates key sex-determination and differentiation transcripts

Similar phenotypes of ectopic expression and upregulation of non-ovarian genes in female gonads have been reported for the female sex-determination gene *Sxl*, the H3K9me3 pathway members *egg*, *wde*, and *hp1a*, and the differentiation factor *bam* [27,31,40]. *Sxl* is required in ovaries for female sex determination; one of its critical functions is silencing (via the H3K9me3 pathway) of the male germline determining protein PHF7, a histone reader whose expression is necessary and sufficient for induction of spermatogenesis in the germline [41,42]. PHF7 induction in female germ cells is also necessary for induction of the tumorous germ cell phenotype of *Sxl* deficient ovaries [27]. In wild-type female germ cells, transcription of *Phf7* is initiated from a TSS in the second exon, which results in truncation of the 5' UTR of the female-specific transcript and absence of Phf7 protein in ovaries (Fig 4A). We found that the male-specific 5' UTR is consistently and ectopically expressed in *stwl* null ovaries, regardless of age, but not in *stwl*-dsRNA treated S2 cells (Fig 4A). Therefore, *stwl* is required for silencing of male-specific programming in ovaries.

Since the overlap of genes ectopically expressed in *stwl* deficient ovaries and S2 cells is so striking, we predicted that upregulation of these genes may be consistent among ovaries exhibiting germline defects. Indeed, we found that these genes are generally upregulated in *bam*, *piwi*, and *Sxl* mutant or deficient ovaries, as well as *egg*, *wde*, and *hp1a* germline knockdown (GLKD) ovaries (Fig 4C). Of note, we find that *bgcn* transcripts are highly upregulated in each of the mutant and GLKD ovary datasets we examined. Bgcn binds to Bam and suppresses mRNAs associated with germline stem cell renewal, and both proteins are required to promote differentiation of developing cystoblasts.

Pleiotropy and substantial changes in ovary composition create challenges for interpretation of data generated from ovaries deficient for GSC maintenance and differentiation genes. In *bam* and *Sxl* mutant or deficient ovaries, GSC-like cells overproliferate to form structures with tumor-like and stem-cell-like qualities and gene expression patterns [43,44]. They also express transcripts associated with early gametogenesis in both wild-type sexes, many of which are testis-enriched. One possible explanation for the perceived "masculinization" of the ovary as a result of *stwl*, *Sxl* or *bam* mutations is that it reflects an overabundance of GSC-like cells and thus of transcripts that are normally expressed during the early stages of wild-type gametogenesis [27]. Single cell data make it possible to address these concerns [45]. We found that some of the most dramatically upregulated transcripts, including *bgcn*, are expressed in GSCs (S20 Fig). However, the overlap of testis-enriched transcripts and GSC transcripts is quite small, and cannot account for the majority of ectopically expressed testis transcripts we observed in *stwl* null ovaries. Furthermore, we find that *stwl*-dsRNA treated S2 cells ectopically express many of the same testis-enriched genes that we identified in *stwl* null ovaries. S2 cells are male-derived, but our analysis nonetheless finds that the affected genes are almost completely silent in the control S2 cells.

In order to determine whether the effects of Stwl on gene expression that we discovered are direct or indirect, we developed two ChIP-grade antibodies against the protein and performed ChIP-Seq experiments on S2 cells. Both Stwl antibodies met our antibody validation criteria, including recognition of the target protein in immunoblotting (S14 Fig) and

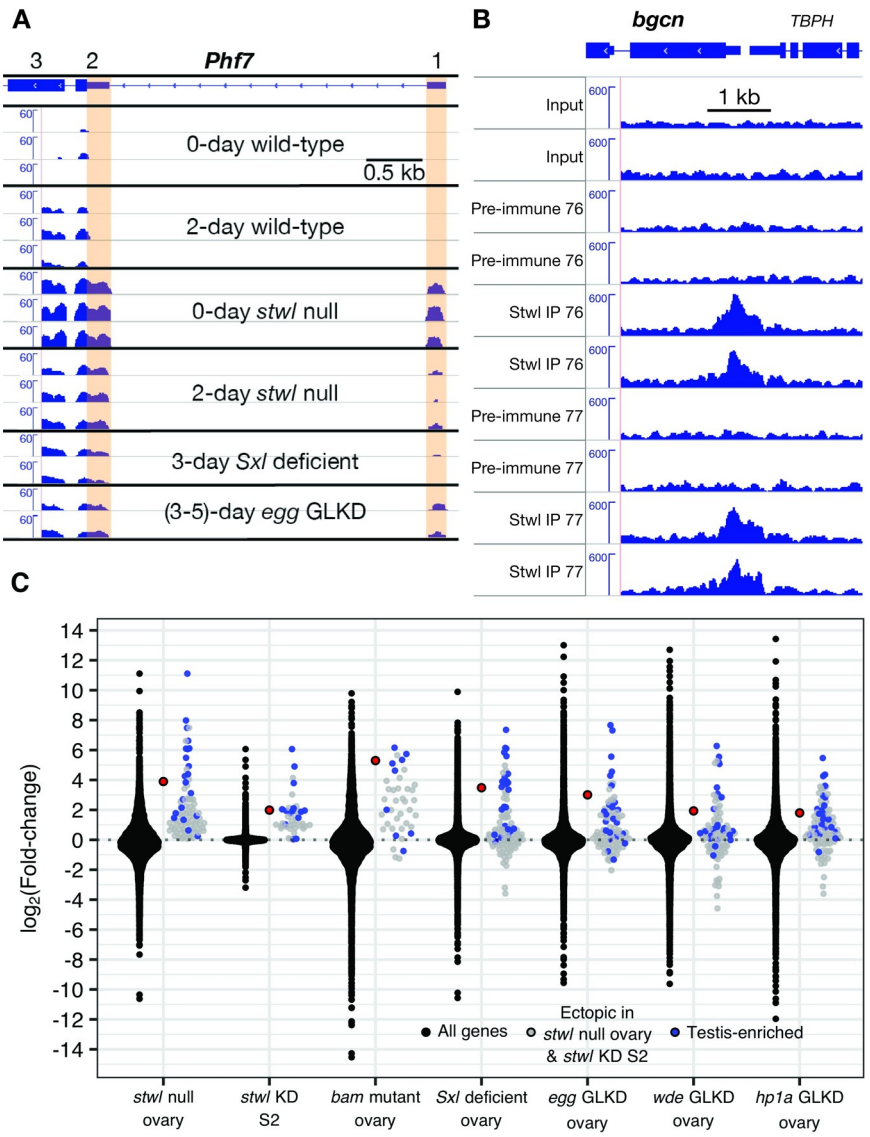


Fig 4. *Phf7* and *bgcn* are regulated by Stwl. (A) The male-specific 5' UTR of *Phf7* (indicated by orange shading) is expressed in *stwl* deficient ovaries and S2 cells, as well as ovaries lacking *Sxl* and its downstream targets *egg*, *wde* (not shown), and *hp1a* (not shown) [31]. Reads were normalized to 1x depth of coverage and visualized in IGV, with *Phf7* shown in 3' to 5' orientation. Exons 1, 2, and 3 are indicated; exon 1 and part of exon 2 are male-specific. (B) *bgcn* is bound by Stwl in ChIP-Seq with two different anti-Stwl antibodies (76 and 77). Two independent replicates of each condition are shown. Reads were corrected for GC-bias, scaled to RPKM, and visualized in IGV. (C) Genes ectopically expressed in both *stwl* null ovary and *stwl*-dsRNA treated S2 cells (including *bgcn*, in red) are also upregulated in ovaries mutant or deficient for other GSC genes.

https://doi.org/10.1371/journal.pgen.1010110.g004

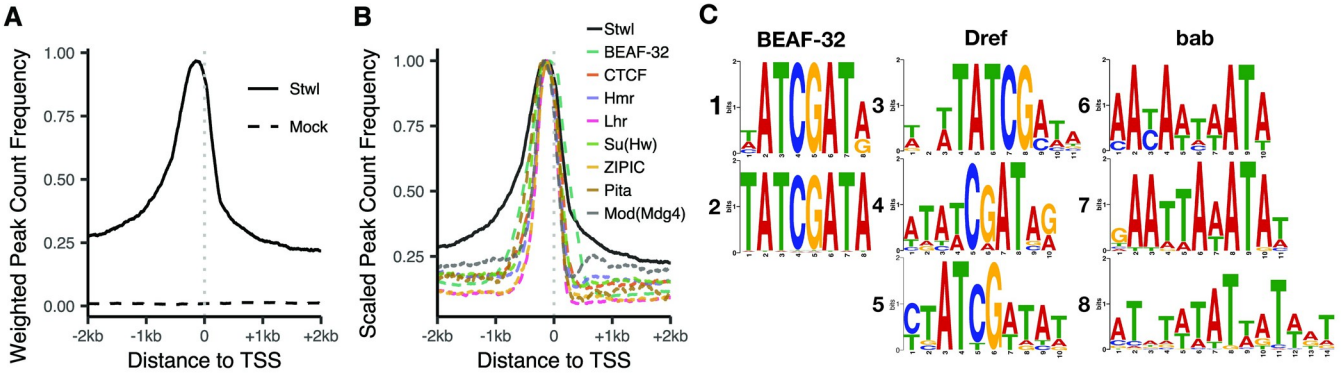
immunofluorescence (S15 and S16 Figs) experiments, successful immunoprecipitation of the target protein (S17 Fig), and low background in immunoblots of S2 cells (Figs S17 and 3A). Our ChIP-Seq experiments produced a robust set of peaks when compared to both input and mock samples (S18 Fig). We identified 3,265 genes whose transcription start sites (TSS) were within 1 Kb of a Stwl peak, and therefore could be considered putatively bound by Stwl (S2 Table).

While we found some evidence of Stwl enrichment at the *Phf7* locus (1.7-fold over input, IDR = 0.048), we found very strong fold-enrichment of Stwl at the *bgcn* promoter (4.1-fold

over input, IDR = 1.3e^{-5}) (Fig 4B). The peak at *bgcn* is among the top 1% of Stwl peaks when ranked according to fold-enrichment. *bgcn* transcript is expressed at very low levels in *Drosophila* ovaries; its expression is limited largely to GSCs, where it is critical for promoting asymmetrical division into cystoblast daughters [46]. Loss of *bgcn* results in a tumorous ovary phenotype, as GSCs proliferate without differentiating into cystoblast daughters. While overexpression of Bam, the binding partner of Bgcn, results in GSC maintenance defects, this defect is not observed when Bgcn overexpression is driven in early germ cells [46,47]. We conclude that Stwl directly regulates expression of *bgcn* in the ovary and posit that aberrant expression of *bgcn* caused by Stwl loss results in activation of male-specific programming in the female germline. This finding is inconsistent with a simple model that *stwl* acts downstream of *bgcn* (and *bam*), as suggested by previous analyses of double mutants [13,14]. We therefore also examined *bgcn stwl* double mutants and corroborated the previous result that they are able to make differentiated germline cysts (S19 Fig). We also found that the double mutants are fully sterile, which is unsurprising, given that both single mutants are each sterile.

Stwl binding peak profiles are similar to known insulator binding proteins

We annotated 2,143 Stwl binding sites across the genome using ChIPseeker [48]. Stwl is highly enriched at promoters, centered ~150 bp upstream of transcription start sites (Fig 5A). To understand this pattern more deeply we compared our Stwl ChIP-Seq profile to the ModERN consortium data of 475 ChIP-Seq experiments on Gfp-tagged DNA and chromatin binding proteins in *D. melanogaster* embryos and larvae [49] using the Genomic Association Tester (GAT) program [50]. Briefly, GAT simulates a null distribution of peaks based on the size of each peakset, then estimates the number of overlaps expected by chance and compares this to the number of observed overlaps. We examined the most similar binding profiles according to fold-enrichment and % overlap. Reassuringly, ChIP-Seq against Stwl-GFP from ModERN had the most similar



D

	1	2	1|2	3	4	5	3|4|5	6	7	8	6|7|8
TF Match	BEAF-32	BEAF-32	BEAF-32	Dref	Dref	Dref	Dref	bab	bab	bab	bab
# Matches	277	521	538	498	513	519	749	301	854	473	1022
% Matches	20	38	39	36	37	38	54	22	62	34	74

Fig 5. Stwl binding sites overlap with insulator-protein bindings sites. (A) Peak density of promoters bound by Stwl and mock antibodies. Frequency for each condition is weighted by the number of peaks present in the displayed 4 kb space. (B) Peak density of promoters bound by Stwl and known insulator binding proteins. Frequency for each protein is scaled to a maximum of 1. (C) Enriched motifs identified in narrow Stwl peaks using Meme Suite [51]. (D) For each motif from (C), we include the transcription factor that motif most closely matches with and the number of Stwl peaks that contain the given motif (# Matches). The % Matches identify the percent of the given motifs found in 1379 narrow Stwl peaks. Union columns (for example, 1|2) describe the number and % of narrow Stwl peaks that match to 1 or more of the indicated motifs.

https://doi.org/10.1371/journal.pgen.1010110.g005

binding profile to our Stwl peakset, according to fold-enrichment (S5 Table). We also found that Stwl ChIP-Seq profiles were highly similar to a number of established and putative insulator binding proteins, including BEAF-32, CTCF, Su(Hw), Hmr, and Lhr (Fig 5B).

We utilized the Meme Suite to identify enriched binding motifs in S2 cell Stwl ChIP-Seq [51]. We found that Stwl peaks are enriched for DNA sequence motifs that are common to BEAF-32, Dref, and bab (Fig 5C). Dref is an insulator binding protein that is additionally required for telomere maintenance [52]. bab, which we identified as ectopically expressed in *stwl* null and *stwl*-dsRNA-treated cells, plays an important role in female sex differentiation [53]. The occurrence of insulator motifs in Stwl ChIP-Seq combined with the above binding profiles provides strong evidence that Stwl binds to insulators.

Stwl localizes to repetitive DNA, including telomeric repeats, chromosome 4, and pericentromeric heterochromatin

Previous studies have shown that Stwl is required for heterochromatin maintenance and colocalizes with HP1 at heterochromatin-like structures at the nuclear periphery [13,15]. We found that Stwl is highly enriched across the dot chromosome (chromosome 4), which is highly repetitive and mostly heterochromatic [54] (Fig 6A). Stwl is also enriched at pericentromeric heterochromatin on chromosome 2, especially at the heterochromatin-euchromatin boundary. Finally, we saw a marked increase in coverage at cytological region 31 on chromosome 2L. Each of these regions is also enriched in Hmr ChIP-Seq in S2 cells [55]. Hmr localization at chromosome 2 has also been identified via immunofluorescence [56].

We next asked whether repetitive DNA, including satellite and transposable element sequences, are enriched among Stwl peaks. Since peak calling methods are not robust to repetitive DNA, we re-analyzed our ChIP-Seq data and instead calculated differential enrichment of reads in IP samples relative to mock (see Methods). This differential enrichment analysis identified repetitive DNAs enriched in Stwl IP samples (Fig 6B). All of these repeats passed a FDR threshold of 0.05, but the fold-changes of significantly enriched repeats were all less than 2. We note, however, that peak-calling algorithms can robustly identify enriched regions of DNA where fold-change of IP/mock is very low. In our own peakset, Stwl-bound sites passed IDR thresholding and were replicated in both antibodies, despite fold-change values as low as 1.2; the median fold-change for enrichment among all Stwl-bound peaks was 2.0. We are therefore confident that our Stwl ChIP-Seq has identified binding to repetitive DNA.

We found that Stwl ChIP-Seq is enriched for LTR retrotransposons, specifically members of the *copia*, *gypsy*, and *bel* superfamilies (Fig 6B). *Copia* elements are among the most highly upregulated transcripts in *stwl* null ovaries, and Stwl was enriched on both the *Copia* LTR and internal sequences. However, for other LTR retrotransposons such as *Bel* and *gypsy* this enrichment only occurred along the LTR component and not the internal region. We note that we did not detect motifs indicative of *gypsy* insulator binding in our motif enrichment analysis. These results suggest that Stwl might be involved in regulating these retrotransposons via their LTR regions. Alternatively, Stwl may be binding to heterochromatic fragments rather than regulating full-length active elements.

We were surprised to find that telomere-associated sequences are consistently enriched in Stwl IP (Fig 6B). With the exception of the *Jockey* family element *Doc6*, all enriched satellites and non-LTR retrotransposons are telomeric. These include telomeric satellite sequences and each of the members of the telomeric HTT array, *Het-A*, *TAHRE*, and *TART*. Furthermore, we found that Stwl peaks are highly similar to peaks generated from ChIP-Seq against the transcription factor pzg, and that Stwl shares DNA-binding motifs with Dref (S5 Table, Fig 5C). Each of these factors localizes to and is necessary for telomere maintenance [57,58]. Lastly, we

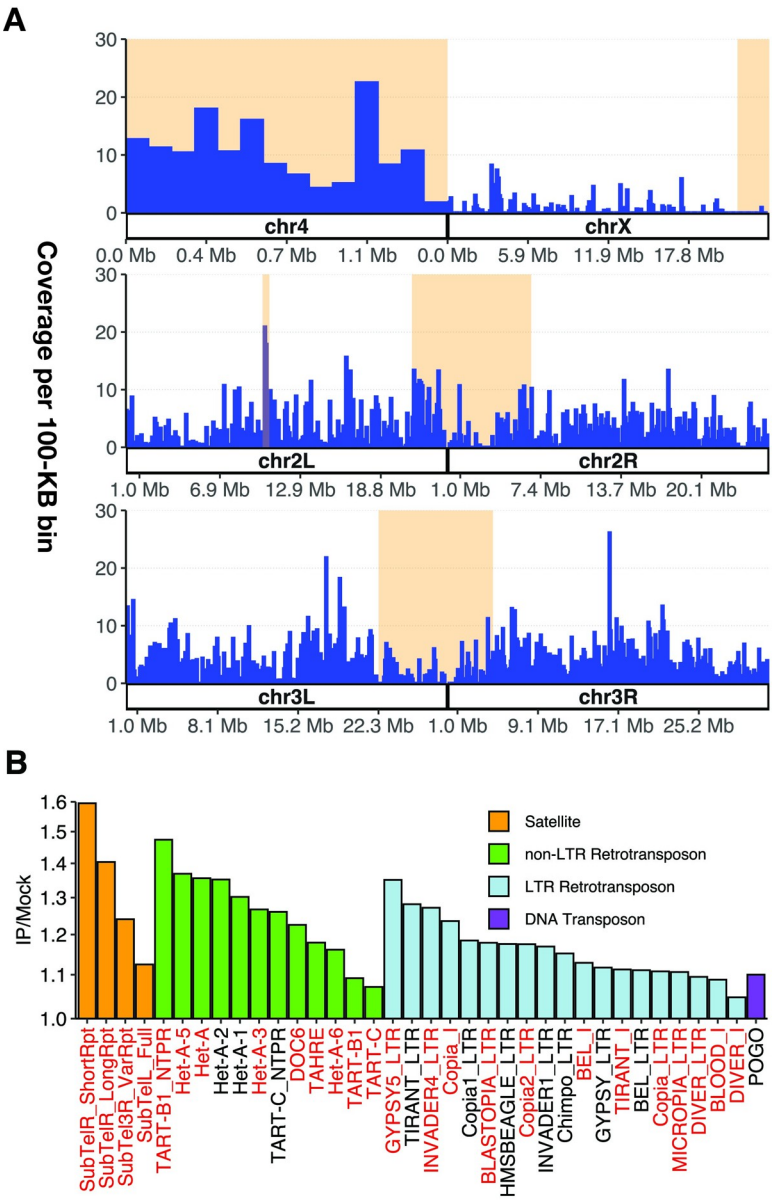


Fig 6. Stwl binds to repetitive DNA. (A) Percent (%) coverage of Stwl peaks (Y axis) per 100-kb bins of the genome (X axis). Shaded orange areas represent constitutive heterochromatin (pericentromeric regions and chromosome 4). Cytological region 31 on chromosome 2L is also highlighted. (B) Fold-change of read count abundance of repetitive elements for IP/mock comparison. Y-axis is in log2-scale. All significantly enriched elements (adj $p < .01$) are plotted. Red elements are upregulated in *stwl* null ovaries. All of the satellite and non-LTR retrotransposon sequences (except DOC6) are telomeric.

https://doi.org/10.1371/journal.pgen.1010110.g006

found that a majority of Stwl-bound telomeric repeat sequences are also upregulated in *stwl* null ovaries (gene names in red in Fig 6B). These findings suggest that Stwl localizes to telomeres and represses expression of telomeric repeats.

Discussion

Identifying the molecular functions of Stwl is especially challenging due to its pleiotropic activity, including in GSC maintenance, oocyte determination, DNA damage response, and TE repression.

Inferring Stwl function is further complicated by the consequences of *stwl* loss, e.g. apoptosis and eventual loss of the female germline. The resulting alteration of cellular content could lead to the identification of misregulated transcripts in *stwl* mutants that do not correspond to actual targets of wild-type *stwl*. We sought to tease apart direct versus indirect effects when analyzing steady-state RNA profiles of tissues affected by *stwl* loss. First, we assayed ovaries at two stages of development, thereby incorporating ovarian developmental status as a factor in the generalized linear model for differential expression. Second, we looked at differential expression in S2 cells after *stwl* knockdown in order to assay Stwl function in a homogeneous tissue. By combining these two assays, we were able to identify genes that are consistently upregulated due to *stwl* loss.

Stwl represses male-specific transcripts

We found that testis-enriched genes show a mixed pattern of both up- and downregulation in *stwl* null ovaries. Strikingly, testis-enriched genes are consistently among the most highly upregulated genes in ovaries. Most of these misregulated genes are not directly bound by Stwl, suggesting that derepression of their transcripts may be a downstream consequence of *stwl* loss. We found that *stwl* null ovaries express the male-specific transcript of the master sex determination factor *Phf7*, but the evidence for it being bound by Stwl in S2 cells is tentative. *Phf7* remains an attractive candidate as a direct target of Stwl regulation in ovaries.

We identified in *stwl* null ovaries a single cluster of highly upregulated, testis-enriched genes on chromosome 2R. Genes in this cluster are among the top 1% of upregulated genes in ovaries, including the GSC genes *HP1Lcsd*, *ord*, *RpL37b*, and *RpL22-like*. The 59C4-59D cluster is located within a lamina-associated domain (LAD). Such structures are thought to specifically repress expression of testis-specific genes by tightly binding these gene clusters to the nuclear lamina and preventing their expression. With the exception of this cluster, we did not find an association between *stwl* loss and misregulation of testis-enriched gene clusters, or LADs. We also do not find that Stwl is binding to this region, or overlapping with LADs.

We found that multiple genes within the 59C4-59D cluster are also derepressed in *stwl*-dsRNA treated S2 cells, as well as *Sxl* and *piwi* mutant or deficient ovaries, and *egg*, *wde*, and *hp1a* germline-knockdown ovaries. In each of these cases, many of the genes in the cluster are ectopically expressed relative to wild-type ovaries. We conclude that ectopic expression of the 59C4-59D cluster and other testis-enriched genes is a consistent reporter of the "masculinization" defect associated with *stwl*, *Sxl*, *bam*, *hp1a*, *wde* and *egg* mutants.

Stwl regulates Bgcn

While male transcripts are upregulated and/or ectopically expressed in *stwl* mutants, our ChIP data suggest that Stwl does not bind at these loci in S2 cells, suggesting that the masculinization defect is an indirect consequence of *stwl* loss. One possibility is that these phenotypes are associated with ectopic expression of *bgcn*, which is typically restricted to GSCs and cystoblasts in ovaries, but widely and highly expressed throughout spermatogenesis [46,59]. *bgcn* is a strong candidate for Stwl regulation: its promoter is bound by Stwl in S2 cells, and it is highly upregulated in *stwl* null ovaries and *stwl* dsRNA-treated S2 cells. Furthermore, the expression of *bgcn* transcripts in ovaries is anti-correlated to Stwl, and females expressing a *hs-bgcn* transgene are sterile [46]. The ectopic expression of *bgcn*, however, is unlikely to be the sole cause of the sterility of *stwl* mutants because *bgcn*; *stwl* mutants remain sterile (S19 Fig).

The molecular pathways in which Stwl functions to maintain oogenesis, either at the stage of germline stem cell retention or oocyte determination, may overlap significantly with pathways in which Bam and Bgcn are crucial actors. Stwl acts antagonistic to and downstream of the GSC differentiation functions of both Bam and Bgcn: *bam* and *bgcn* mutants present with

GSC-tumorous ovarioles, while *stwl*; *bam* and *stwl*; *bgcn* double mutants form rudimentary germline cysts [13,14]. Despite the fact that Bgcn has a defined and important role in the ovary, it is nonetheless largely silent throughout oogenesis [45]. It is possible that one of Stwl's functions in the female germline is to restrict expression of *bgcn* to GSCs. Considering all of these results, we propose that *stwl* is downstream of *bgcn* GSC differentiation function while also being required to repress its later expression. However, Stwl binding sites in ovaries are not known, and may differ from those we identified in S2 cells. ChIP-Seq against Stwl in ovaries may further elucidate the relationship between Stwl and male transcripts, including *bgcn*.

Stwl accumulates at genomic insulators and heterochromatin

Loss of *stwl* results in derepression of repetitive elements, a phenotype that is also observed in *bam* and *piwi* mutant ovaries. While Piwi is a known regulator of TEs via the piRNA pathway, it is unclear whether upregulation of TEs in *bam* and *stwl* mutant ovaries reflects a direct role in TE silencing. In order to answer whether Stwl directly targets repetitive DNA and how it may be involved in TE control, we developed antibodies to Stwl and assayed Stwl binding in S2 cells. Our analyses indicate that Stwl localizes to insulator elements. Most Stwl peaks are located just upstream of promoters; this binding profile is common among insulator-bound proteins. More directly, we identified strong sequence similarity between Stwl peaks and peaks from a number of insulator binding proteins, including BEAF-32, Dref, ZIPIC, Pita, Hmr and Su(Hw). In addition, we found that *stwl* peaks accumulate at heterochromatic loci, specifically pericentromeric heterochromatin boundaries, telomeres and the dot chromosome.

These data suggest that Stwl is an insulator binding protein. Future biochemical studies will be required to confirm this, and it will also be important to determine whether the localization pattern we discovered from ChIP-Seq data in S2 cells also occurs in the germline. Nevertheless, our proposal is consistent with previous work showing that Stwl associates with insulator complexes in immunofluorescence experiments [22]. Insulators have multiple functions including blocking enhancer-promoter interactions and establishing boundaries to prevent the spread of chromatin modifications and to separate differentially expressed promoter pairs. Insulator-binding proteins, such as CP190, can also mediate long-range interactions [60,61]. If Stwl is involved in the formation of long-range interactions, it may promote tethering of euchromatic insulator sites to heterochromatic regions. We speculate that Stwl-bound sites are located adjacent to regions of repressed chromatin, and that loss of Stwl results in spreading of these repressed chromatin marks to neighboring loci. It is likely that Stwl performs its function as an insulator by establishing boundaries, in conjunction with other insulator-binding or heterochromatin-associated proteins, that ensure proper expression of nearby genes.

The *D. melanogaster* ovary is a complex mixture of cell types in the adult fly. Differentiated somatic cells function as support cells to shepherd germ cells towards their ultimate fate of producing viable gametes. Furthermore, each of these cellular lineages is derived from a small population of self-renewing stem cells. We suggest that insulators allow genes to have pleiotropic functions during development of complex tissues such as ovaries. Insulators add a layer of genomic complexity to gene regulation by disrupting enhancer-promoter interactions. The detailed interplay of promoters, enhancers, and insulators during oogenesis is poorly understood but is likely key to explaining pleiotropic gene regulation in the developing ovary.

Methods

Drosophila stocks

P{w[+mC] = lacW}stwl^{j6C3} (*stwl*^{j6C3}) was acquired from the Bloomington Stock Center [BDSC #12087]. This allele is female sterile and shows *stwl* mutant phenotypes (ovarian atrophy, loss

of germline, lack of Orb accumulation in oocytes) when trans-heterozygous to a *stwl* deficiency chromosome (*Df(3L)Exel6122*) [BDSC #7601]. We observe no nuclear signal with anti-Stwl antibodies in $stwl^{j6C3}$ homozygotes, suggesting that it is a null allele (S2 Fig).

We found that the $stwl^{j6C3}$ chromosome is homozygous lethal, suggesting an accumulation of lethal recessive mutation(s). In order to remove these lethal mutations and homogenize the genetic background, we outcrossed $stwl^{j6C3}$ mutant females to males from an inbred *y w* strain (10 generations of inbreeding; strain will be subsequently referred to as *y w* F10) for 8 generations. Stocks were founded by balancing recombined 3rd chromosomes over *TM6b* from *w*; *Sp*/*CyO*; *TM2*/*TM6b* stock in single female matings. Presence of the *P*-element insertion in $stwl^{j6C3}$ was followed by its w^+ marker and confirmed by PCR. The resultant stock produced viable and fertile homozygous males and viable but sterile homozygous females. The *y w* F10 stock was used as a wild-type ($stwl^+$) control.

In Figs 1A, S4, and S19, the *stwl* null genotype corresponds to $stwl^{j6C3}$/*Df(3L)Exel6122* as explained in the figure legends. In S2 and S3 Figs, we utilized the $stwl^{\Delta 95}$ null mutant ($stwl^{\Delta 95}$, *ry*/*TM3*, *Sb e ry*), kindly provided by Prof. D. McKearin (HHMI, Washington, DC). This allele contains a 5' deletion at the *stwl* locus that results in a frameshift mutation and *stwl* null phenotype [11]. In all other experiments, the *stwl* null genotype corresponds to $stwl^{j6C3}$/$stwl^{j6C3}$.

For fertility assays in S9 Fig, assayed females were generated by crossing Gal4-containing females to *UAS-stwl-RNAi* ($y^1 sc^* v^1$; *P{y[+t7.7] v[+t1.8] = TRiP.GL00337}attP2*) [BDSC #35415] males. The following Gal4 lines were used: *Actin5c-Gal4*: y^1 *w*; P{Act5C-GAL4} 17bFO1/TM6B*, Tb^1 [BDSC #3954], *nos-Gal4*: w^{1118}; *P{GAL4::VP16- nos.UTR}CG6325*^{MVD1} [BDSC #4937], *C355-Gal4*: *P{w[+mW.hs] = GawB}c355*, w^{1118} [BDSC #3750], *T155-Gal4*: *P{w [+mW.hs] = GawB}T155* [BDSC #5076], *GR1-Gal4*: *w**; *P{w[+mW.hs] = GawB}GR1* [BDSC #36287], *C306-Gal4*: *P{w[+mW.hs] = GawB}c306*, w^{1118} [BDSC #3743], *Matalpha-Gal4*: *w**; *P {w[+mC] = matalpha4-Fal-VP16}V37* [BDSC #7063], y^1 *w* P{w[+mC] = bam-Gal4:VP16}1* [BDSC #80579]. *Tub-Gal4/TM3* was a gift from the Wolfner lab (Cornell University, Ithaca, NY).

To generate *bgcn* transheterozygous mutants, we crossed *y w*; $bgcn^1$/*CyO* to *cn* $bgcn^{z2\text{-}1748}$ *bw/CyO*. Both *bgcn* stocks were gifts from the Aquadro lab.

Preparation of gonadal tissue for qRT-PCR and RNA-seq

All flies were raised at 25˚ C. Virgin males and females of each genotype were collected and aged for two days for the "older" samples; for the newly-eclosed samples, virgin females were collected and dissected immediately (<4 hours post-eclosion). Testis and ovary dissections were performed according to previously published protocols [62,63]. Briefly, 15–30 flies at a time were sedated using CO_2 and stored on ice. Gonads were extracted in ice-cold 1x PBS using sharp forceps, separated from gut tissue (and accessory glands, in males) and stored in ice-cold 1x PBS for ~30 minutes. PBS was aspirated and tissues homogenized in 100–600 μl of Trizol (depending on total volume of dissected tissue) prior to snap-freezing in liquid NO_2 and storage at -80˚ C. All sample replicates for RNA-seq consisted of ~30 ovary/testis pairs, most of which were collected in single dissections at approximately the same time of day over a span of 23 days. Trizol homogenate from phenotypically "large" ovaries (2-day old *y w* F10) was diluted 1:10 prior to RNA extraction, to prevent overloading of columns.

RNA was extracted according to previously published protocols [64]. Briefly, Trizol-homogenized tissue samples were thawed at room temperature and treated with 0.2 volumes chloroform to promote phase separation. RNA was extracted from the aqueous phase using Qiagen RNeasy Plus Mini Kit. This included application of aqueous phase to Qiagen gDNA Eliminator spin columns to limit carryover of genomic DNA. DNA contamination was also

addressed by on-column DNAse digestion (Promega RQ1 DNAse). RNA quality and concentration was validated via Agilent Bioanalyzer; RNA quality for all samples was confirmed to have an RQN > = 7.0 and at least 1.0 μg of starting material.

Stranded cDNA library preparation was performed by Polar Genomics (Ithaca, NY). mRNA was isolated and fragmented from total RNA pools, followed by 1st and 2nd (dUTP incorporated) strand synthesis. dsDNA was subsequently dA-tailed and adaptor-ligated, followed by size selection, UDG digestion to eliminate the second strand, and PCR amplification. All libraries (18 in total) were sequenced on a single lane of Illumina NextSeq (single-end, 75 bp). Raw sequence reads have been uploaded to NCBI Sequence Read Archive (SRA) in.fastq format under the BioProject ID PRJNA788954.

During preliminary analyses of sequencing reads we found that 2 libraries (2 replicates from 0-day old *y w*; $stwl^{ji6C3}/stwl^{ji6C3}$ ovaries) were of insufficient quality, likely due to contamination during sample recovery or library preparation. We therefore discarded these reads and prepared new samples. Ovaries were collected, dissected and homogenized in Trizol, as described above (with the exception that the ovary pool was increased from 30 to 45 ovary pairs per replicate). Stranded cDNA libraries were prepared as described above and subsequently sequenced on a single lane of Illumina HiSeq 2500 High Output (single-end, 50 bp).

For qRT-PCR, ovaries were dissected and processed as above, with three biological replicates of ~30 ovary pairs each per sample. cDNA was synthesized with Invitrogen oligo-dT primers and reverse transcriptase using standard protocols. All qRT-PCR assays were performed with three technical replicates. Transcript abundance of each technical replicate was normalized to average levels of Rpl32 transcript in the source biological sample.

Production and validation of polyclonal antibody against Stwl

A DNA fragment coding for amino acids 911–1037 from the C-terminus of the Stwl protein was amplified from *D. melanogaster* ovarian cDNA extracted from ~20 *y w* F10 individuals. This region is lacking in predicted interaction domains, making it more likely to be accessible for immunoreactivity. The fragment was cloned into a N-terminal tagging MBP fusion vector (Genbank: AF097412.1) using NEB Gibson Assembly Kit [65] and transformed into chemically competent *E. coli* (One Shot TOP10). Successful assembly and transformation were confirmed via PCR and Sanger sequencing.

MBP-Stwl antigen was purified from induced bacterial culture using Amylose Resin (NEB: E8021L). Briefly, antigen expression was induced in 1 L of bacterial culture containing MBP-Stwl plasmid at log phase (OD 600 = 0.6) with 0.2 mM IPTG, then shaken for ~18 hours at 18˚ C. Bacteria were pelleted and resuspended in lysis buffer (50 mM Tris pH 8.8, 200 mM NaCl, 2 mM DTT, 1 mM PMSF, 1 mg/ml lysozyme, 1x Roche cOmplete EDTA-free Protease Inhibitor Cocktail) at 4˚ C. Lysate was sonicated on ice to ensure thorough lysis, then spun at 20,000x g for 45 minutes to pellet debris. Supernatant was then applied to Amylose Resin on column. Stwl-MBP bound resin was washed 4 times with 1 column volume of low salt buffer (50 mM Tris pH 8.8, 200 mM NaCl, 2 mM DTT), followed by 4 washes with 1 column volume of high salt buffer (50 mM Tris pH 8.8, 1.5 M NaCl, 2 mM DTT) and another 4 washes with 1 column volume of low-salt buffer. Stwl-MBP was eluted with 10 mM maltose in low-salt buffer. Presence of 57.5 kDa MBP-stwl protein was confirmed using Coomassie stain on 10% SDS PAGE; concentration was estimated using Bradford assay. Protein-containing fractions were pooled using Amicon Ultra-4 10K Centrifugal Filter Devices to a final concentration of 1.0 mg/ml.

Purified protein was submitted to Pocono Rabbit Farm & Laboratory Inc. for injection. Two guinea pigs (henceforth referred to as GP 76 and GP 77) were selected for antigen

injection based on absence of background signal in pre-immune sera (determined by probing wild-type *D. melanogaster* ovaries with sera in immunofluorescence assays).

We found that both antibodies recognize a ~135 kDa protein in wild-type ovaries (S14 Fig) and S2 cells (Fig 3A). This signal is absent in null and RNA knockdown samples, as well as wild-type lysates probed with pre-immune sera. The primary *stwl* transcript is predicted to produce a 112.9 kDa protein; previous work has shown that antibodies against Stwl recognize a similarly sized protein [11].

We performed immunofluorescence (IF) experiments to confirm that the Stwl antibodies target a nuclear protein and to compare to IF experiments done with other Stwl antibodies. *D. melanogaster* ovaries were dissected from wild-type (*y w* F10) individuals in cold 1x PBS and fixed in 4% paraformaldehyde with 0.1% Triton X-100 in PBS. Tissue was then washed 3x in PBT (1x PBS with 0.1% Tween 20), followed by 4x washes in PBTA (PBT with 1.5% BSA). Samples were incubated overnight at 4˚C with primary antibody at a concentration of 1:200 for Stwl antisera and 1:200 for Rabbit Vasa from Santa Cruz Biotechnology, Inc. Following 3x washes in PBT and 4x washes in PBTA, tissue was incubated for 2 hours with secondary antibodies (1:500 Goat anti-Guinea Pig Rhodamine Red-X, 1:500 Goat anti-Rabbit Alexa 488). Following 3x washes in PBT, tissue was mounted in vectashield with DAPI and imaged on a Zeiss Confocal. We found that both antibodies specifically labeled germ cell nuclei in testis and ovaries (S15 Fig). Furthermore, ectopic expression of HA-tagged Stwl [FlyORF stock #F001844] colocalized with signals from both Stwl antisera [66] (S16 Fig).

Cell culture and RNAi

S2 cells were cultured in M3+BPYE medium, made as directed from Shields and Sang Powdered Medium (Sigma S-8398), supplemented with 0.5 g $KHCO_3$, 1 g yeast extract and 2.5 g bactopeptone per liter, pH adjusted to 6.6 and sterile-filtered. 100x Antibiotic-Antimycotic (Thermo-Fisher 15240062) and Fetal Bovine Serum (Sigma F2442 Lot # 078K8405) were added to concentrations of 1x and 10%, respectively. Cells were maintained at 25˚C and passaged every 3–4 days for 7 passages prior to use for RNAi experiments.

For dsRNA-induced knockdown, cells were plated in serum-free medium at a concentration of 2.5 million cells/ml, then treated with 30 μg/ml of *lacZ*- or *stwl*-dsRNA for 60 minutes before addition of M3/BPYE medium containing 13% FBS (final concentration, 10% FBS, 7.5 μg/μl dsRNA). Cells were chemically cross-linked and frozen after 3 days.

RNA was synthesized using NEB HiScribe T7 High Yield RNA Synthesis Kit (E2040S) from PCR products generated from YEp365 plasmid (*lacZ* control) or genomic DNA extracted from S2 cells. For efficient *stwl* KD we generated three distinct dsRNAs from reference, each targeting the second exon of *stwl*, which is present in all *stwl* transcripts [67]. S2 cells were treated with 10 μg/ml of each dsRNA.

Chromatin immunoprecipitation in S2 cells

Subsequent to dsRNA treatment for 3 days, cells were centrifuged for 5 minutes at 1000xg followed by removal of media, washed once in 1x PBS, and resuspended in 1x PBS and cross-linked via addition of 16% paraformaldehyde to 1% final concentration for 2 minutes at room temperature. Cross-linking was quenched by addition of 2.5 M glycine in 1x PBS (final concentration 0.15 M) for 5 minutes at room temperature. Cells were nutated for 15 minutes at 4˚C, spun and washed in 1x PBS brought to 4˚C, and pelleted and flash-frozen in liquid nitrogen.

Cells were thawed on ice and lysed in RIPA buffer containing 0.1% SDS, 1% Nonidet P-40, and 1 tablet/10 ml Pierce Protease Inhibitor Mini Tablets, EDTA-free (A32955), for 20

minutes. Lysates were then sonicated at high intensity in a Bioruptor (Diagenode) water bath to shear DNA to desired size range (300–500 bp) for 45 minutes total with cycles of 20 seconds on, 1-minute rest, with quick spins of lysates every 15 minutes to settle samples and re-fill the Bioruptor with ice-cold water.

6 μl of freshly thawed Stwl antisera and pre-immune sera were added to 300 μl of cell lysate (1:50 dilution) and incubated overnight at 4˚C. Cell lysates were prepared with approximately 34,000 cells per μl, so that each IP experiment was performed on roughly 10 x 10^6 cells. IP complexes were immunoprecipitated with Invitrogen Dynabeads Protein A for Immunoprecipitation (10001D). Prior to use, beads were washed 2x 10 minutes in blocking buffer containing 1 mg/ml BSA, 1 mg/ml propyl vinylpyrrolidone blocking agent, and 1 tablet/10 ml Pierce Protease Inhibitor Mini Tablets, EDTA-free (A32955), and 1x 10 minutes in chilled RIPA buffer (also with protease inhibitor). IP samples were added to blocked beads and incubated at 4˚C for 2 hours; 50 μl of beads were used for each IP.

Beads were washed 1x in low-salt buffer, 2x in high-salt buffer, 1x in LiCl buffer, and 2x in TE buffer. IP complexes were eluted from beads in 10% SDS, 1M $NaHCO_3$ elution buffer for 30 minutes at 65˚C. Cross-linking was reversed by addition of 5 M NaCl to 0.2 M NaCl final concentration and overnight incubation at 65˚C. DNA was treated with RNAse A for 2 hours at 37˚C and Proteinase K for 2 hours at 55˚C, then cleaned using Qiagen QIAquick Gel Extraction Kit. Input samples were frozen following sonication, then thawed and reverse-crosslinked as above. DNA libraries were prepared using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB# E7645), using Ampure XP beads for cleanup, without size selection.

RNA was extracted in triplicate from S2 cells originating from the same populations used for ChIP-Seq experiments, using Qiagen RNEasy plus extraction kit, which includes additional elimination of gDNA from samples. All RNA samples had RQN>7.0, as determined by Bioanalyzer instrument (Agilent). cDNA libraries were prepared using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (E7760G), using Ampure XP beads for cleanup. All cDNA and ChIP libraries (22 in total) were pooled together and sequenced on a single lane of Illumina NextSeq (single-end, 75 bp). Raw sequence reads have been uploaded to NCBI Sequence Read Archive (SRA) in.fastq format under the BioProject ID PRJNA788954. RNA--Seq data processing, QC, and analysis of S2 cell samples was performed as described in the "Read processing, alignment, and normalization" section.

Read processing, alignment, and normalization

We assayed quality of raw reads in fastq format using FastQC (version 0.11.6) and trimmed reads for adapter sequences and quality using Trimmomatic (version 0.32); (java -jar trimmomatic-0.32.jar SE [raw_reads.fq] [trimmed_reads.fq] ILLUMINACLIP:TruSeq3-SE.fa:2:30:10 SLIDINGWINDOW:4:20 MINLEN:50 AVGQUAL:20) [68,69]. We used FastQ Screen to identify non-*Drosophila* contaminants in our libraries [70].

For RNA-Seq, we aligned reads to a curated list of consensus sequences for repetitive elements using relaxed bowtie2 settings (bowtie2 -x [repetitive_consensus_sequences.fasta] -U [trimmed_reads.fq] -S [repetitive_alignment.sam]—un-gz [unmapped_reads.fq.gz]—score-min L,0,-1.5 -L 11 -N 1 -i S,1,.5 -D 100 -R 5) [71]. Unmapped reads from this alignment were saved and aligned to the unmasked *Drosophila melanogaster* genome (r6.03) using bowtie2 default settings [72]. We counted the number of reads mapping to each repetitive element; we utilized HTSeq version 0.6.0 to count the number of reads aligning to exons in the genomic alignment [73]. We concatenated the read counts into a single file for each sample.

In order to normalize for sequencing bias resulting from GC-content bias or batch effects, we normalized the read counts using EDAseq [74]. We removed counts for all genes with a

mean read count less than or equal to 10 across all samples. We then performed within-lane normalization for GC-content: read counts within individual samples were transformed via full-quantile normalization between feature strata to normalize for GC-content of assayed genes. Between-lane normalization was then performed (again using full-quantile normalization between feature strata) to account for differences in sequencing depth, and offset values were generated for each transcript in the count matrix so that raw counts could be analyzed for differential expression analysis.

In order to provide context for our biological observations, we compared *stwl* null ovary data to available data from genotypes that reduce or eliminate *bam*, *egg*, *wde*, *hp1a*, *Sxl*, and *piwi* in ovaries. The *bam* data comes from $bam^{1}/bam^{114\text{-}97}$ females (SRA PRJNA117723); *piwi* data from $piwi^{1}/piwi^{2}$ females (SRA PRJNA289709); *egg*, *wde* and *hp1a* data from germline-specific RNAi knockdown females (SRA PRJNA432192); and *Sxl* data from snf^{148} homozygous females, a mutant that phenocopies germline-specific loss of *Sxl* (SRA PRJNA275434) [27,30,31,33]. Alignment and normalization for all datasets were as described above.

For ChIP-Seq, trimmed reads were aligned to the unmasked *Drosophila melanogaster* genome (r6.03) using bowtie2 default settings [72]. All reads with mapping quality <20 were removed using SAMtools [75]. All alignment files were corrected for GC bias using the deepTools commands computeGCbias and correctGCbias [76]. Briefly, the distribution of GC content per read is assessed over the contents of each alignment file, typically revealing overamplification of high-GC content sequences. The correctGCbias command generates an alignment file identical to the original, except with reads artificially removed or duplicated at biased regions to eliminate GC bias.

Repetitive DNA alignment and analysis for ChIP-Seq

Limitations and challenges of identifying enriched repetitive elements from ChIP-Seq data have been well documented [77–79]. With relatively short (75 bp) single-end reads, it is nearly impossible to identify the genomic origin of most reads coming from repetitive DNA, and therefore enrichment cannot be called against a true background signal. We therefore instead calculated differential enrichment of repetitive DNAs in IP samples relative to mock samples, normalized against genomic reads. The process is explained below in greater detail.

For analysis of repetitive DNA, reads were trimmed and aligned as described for RNA-Seq reads. For genomic reads, rather than counting reads aligned to gene bodies as we did for RNA-Seq analysis, we calculated the number of reads aligned to each 1 kb bin of the genome. We concatenated the repetitive and genomic read counts into a single file for each sample.

Differential expression/enrichment analysis

We analyzed count data using DESeq2 [29]. We imported raw counts and offsets as described above from genes with mean read count >10 across all samples. For testis, we estimated differential expression of genes between mutant and wild-type samples. For ovary comparisons, age of the samples was taken into account: genes were called as differentially expressed if the normalized read counts from the experimental genotype were consistently different from those in the control genotype, excluding those cases where genes were found to be differentially expressed between 0- and 2-day old samples.

For ChIP-Seq, we estimated differential enrichment of genomic regions between IP and mock samples, taking into account the antibody used (GP 76 or GP 77) and the source S2 population (replicate 1 or 2). A genomic region was reported as differentially expressed only when the normalized read counts for that region were consistently greater in IP samples than in mock samples, and not due to differences in IP conditions (source animal antibody or source

cell population). PCA on the count matrix confirmed that the majority of the variance in the count data was explained by the variance between mock and IP samples (S21 Fig).

Subsequent to differential expression/enrichment analysis, all $\log_2$(fold-change) estimates were transformed using apeGLM shrinkage estimator to reduce variability in LFC values among low-count genes [80]. Shrunken LFC values were used for all subsequent analyses, including overrepresentation tests, gene set enrichment analysis, and Gene Ontology analyses, implemented using the R package ClusterProfiler [81].

ChIP-Seq peak calling and analysis

Each IP experiment was performed with 2 biological replicates; each biological replicate originated from a single 150 cm^2 flask of *lacZ*-dsRNA-treated S2 cells. From each flask, we immunoprecipitated chromatin using Stwl antisera 76 and 77 (IP), pre-immune sera 76 and 77 (mock), and also sequenced input DNA. Therefore each ChIP-Seq experiment could be called against enrichment from its own input DNA, and mock datasets against both antibodies could be used to exclude spurious peaks.

We followed peak calling standards established by the ModERN consortium [49]. We performed peak calling on all mock and IP samples using the peak calling algorithm MACS2 [82]. In order to generate an intentionally noisy set of peaks for downstream IDR (irreproducible discovery rate) analysis, peaks were called with low stringency (FDR<0.75) as follows: macs2 callpeak -t {IP/Mock.bam} -c {Input.Bam} -g 142573024—tsize 75 -n output.file -m 2 50 -q 0.75—keep-dup all. This command generates a large set of statistically insignificant peaks which can be fed into the IDR algorithm [83]. Confident peak sets were identified by performing IDR analysis on peaksets between biological replicates, using an IDR cutoff of 0.05; significant peaks passing this IDR threshold co-occur in the same genomic location at similar intensities. IDR was also done on mock samples with a much looser restriction (IDR<0.25), in order to create a more expansive list of peaks that could potentially be generated from biological noise (S18 Fig).

After IDR, any peaks in the IDR 76 and IDR 77 IP peaksets that overlapped with spurious peaks from either mock were removed using BEDTools subtract [84]. These filtered peaksets were then merged together using BEDTools merge so that the final Stwl peakset contained the union of peaks confidently called in IPs from either antibody. Finally, peak calling was repeated using MACS2 broadpeak setting (mfold 2–50, q<0.50), and the same steps were followed as before. The final broadpeak and narrowpeak calls were merged together to form a single set of peaks in broadpeak format. Motif identification, which requires narrow, sharply defined peaks, was done on only the narrowpeak calls; all other analyses were performed on the broadpeak format calls.

For motif analysis, we extracted summits from 1,379 narrowpeak calls as described above. Symmetrical peaks were then extracted as 500 bp sequences centered at each summit. These sequences were loaded onto the MEME-ChIP web browser (version 5.0.5) and motifs were identified using MEME, DREME, and CentriMo programs on default settings [51].

Identifying tissue-enriched and ectopically expressed genes

We utilized RPKM values from the modENCODE anatomy RNA-Seq dataset to classify all genes according to tissue-biased expression [33]. The Tau metric is among the most simple and reliable tools for determining tissue-specific expression of a given gene [85,86]. Tau was calculated from $\log_2$(RPKM) values in a subset of available tissues (S6 Table). Tau values range from 0 to 1, corresponding to the range from ubiquitous expression to highly tissue-specific expression. Genes with Tau> = 0.7 were considered tissue-specific; to identify which tissue(s)

each of these genes are enriched in, we assigned any tissue where $\log_2$(RPKM) is greater than 1.5 standard deviations above the mean $\log_2$(RPKM) across all tissues to that gene (S7 Table). According to this classification, we found that 45.5% of annotated genes exhibit tissue-specific expression, meaning that transcripts for those genes are preferentially enriched in one or more of the represented tissues (S8 Table).

Ectopic gene expression refers to the expression of a gene in a tissue where it is silent under normal conditions. Due to the nature of ectopic expression (i.e. an increase in transcript abundance from a baseline of very low counts), it is challenging to capture accurate $\log_2$(Fold-change) values, especially since GC-normalization typically requires removal of low-count genes. Ectopic gene expression in ovaries, testes, and S2 cells was assayed from library size-corrected RPKM values calculated in DESeq2, without the removal of low-count genes. We defined ectopic gene expression by 1) identifying the gene as phenotypically silent in wild-type tissue and 2) finding that gene expression increases significantly by at least 2-fold in the mutant tissue. Genes with mean RPKM < 2.0 in a given WT tissue according to a Mann-Whitney ($p<0.1$) were considered transcriptionally inert. Inert genes where mean_RPKM_null/mean_RPKM_WT>2.0 were subjected to a BH-corrected Mann-Whitney test ($p<0.25$) to identify ectopic expression.

Supporting information

S1 Fig. *stw*^{lj6c3}/*stwl*^{lj6c3} ovaries display the known *stwl* null phenotype. Ovaries were dissected from females of the indicated genotype 3–6 days post-eclosion. *stwl* null ovaries typically lack germ cells or contain severely disordered germline cysts [11–13]. α-Vasa labels germ cells, α-Hts-1B1 labels branched fusomes or spectrosomes as well as follicle cell membranes. Germaria are positioned with anterior to posterior going (left to right). *stwl*$^{+}$ ovaries (*y w* F10) contain self-renewing GSCs (anterior-most germ cells) which differentiate into cystoblasts and become ordered, organized germline cysts. All images are maximum-intensity projections from a z-series representing a depth of 10 microns. Scale bars are 20 microns.
(TIF)

S2 Fig. *stwl* deficient ovaries fail to retain germline cells as they age. *D. melanogaster* ovaries were dissected from females 10–15 days post-eclosion, immunostained with α-Stwl sera from GP 76 (Methods). α-Vasa labels germ cells, which are typically not retained in older mutant ovaries. Germaria are positioned with anterior to posterior going left to right. All images are maximum-intensity projections from a z-series representing a depth of 10 microns. In *stwl* mutant images, the green channel is overexposed to demonstrate the absence of Stwl signal. Scale bars are 20 microns. The *stwl*$^{\Delta 95}$/*Df(3L)Exel6122* ovary (middle row panels) is displayed at 0.4x magnification relative to other images, to demonstrate loss of germline across the ovary.
(TIF)

S3 Fig. Newly-eclosed ovaries from *stwl* mutants resemble WT ovaries. *D. melanogaster* ovaries were dissected from females <12-hours post-eclosion and immunostained with α-Stwl sera from GP 76 (Methods). α-Vasa labels germ cells, which are typically not retained in older mutant ovaries. Germaria are positioned with anterior to posterior going left to right. Wild-type ovaries produce egg chambers up to stage 7 or 8, while *stwl* mutant ovaries maintain egg chambers up to about stage 6 or 7. Low (0.24x) magnification images are of a single confocal slice (first and third rows), higher magnification images are maximum-intensity projections from a z-series representing a depth of 10 microns (second and fourth rows). In *stwl* mutant

images, green channel is overexposed to demonstrate absence of Stwl signal.
(TIF)

S4 Fig. *stwl* null and WT ovaries are more similar in size from newly-eclosed individuals relative to older flies. Ovaries were dissected from newly-eclosed and two-day-old females of the indicated genotypes. *stwl* deficient ovaries are rudimentary, but more closely resemble wild-type ovaries when from newly-eclosed individuals. Scale bars are 1 mm.
(TIF)

S5 Fig. Sample-to-sample distance matrix of RNA-Seq samples. Read counts were regularized log transformed in DESeq, and the distance between samples was calculated based on these transformed count values. The heatmap is sorted by similarity after hierarchical clustering and color-coded according to distance, where dark blue cells indicate a distance of 0 (completely self-similar) and white cells a maximal distance (completely dissimilar). Samples within the same group (identical age and genotype) occur together and form blue clusters.
(TIF)

S6 Fig. Principal Component Analysis (PCA) of RNA-Seq count matrices. PCA was performed on regularized log transformed read counts of the 500 most variable genes in the count matrix. Samples within the same group (identical age and genotype) cluster together, indicating minimal batch effects.
(TIF)

S7 Fig. Transposons and testis-enriched genes are consistently upregulated in *stwl* null ovaries. (A-B) Fold-change of TEs in *stwl* null ($stwl^{j6c3}/stwl^{j6c3}$) relative to wild-type from RNA-Seq assay of 0- and 2-day old ovaries. Black arrows point to TEs validated with qRT-PCR data in Fig 1A and/or 1B. "G","S","B" indicates whether TE is typically expressed in germline, ovarian soma, or both, respectively [26]. (C) $\log_2$Fold-change (LFC) of TEs vs. all genes from *stwl* null ovaries in the combined GLM, 2-day, and 0-day datasets, relative to wild-type. Crossbars show the mean LFC for all TEs. (D-E) Fold-change of the top 14 and bottom 14 most affected annotated genes (based on FlyBase annotations) in *stwl* null ovaries relative to wild-type. Male and female symbols mark genes with testis- and ovary-enriched wild-type expression, respectively; "*" marks genes that are part of the 59C4-59D testis-specific cluster described in Fig 2. (F) Enriched tissue classes among the top and bottom 1% of misregulated genes. Average LFC is plotted for each set of tissue-enriched genes enriched among *stwl* null ovaries relative to wild-type. Only gene sets with FDR <0.05 are plotted.
(TIF)

S8 Fig. MA plot of RNA-Seq data from ovaries. Fold-change for each gene is plotted against its average transcript abundance across all assayed ovarian samples (wild-type and null). Transcript abundance is represented by counts normalized according to GC-content and library size. The log2(Fold-change) values (LFC) were "shrunk" to minimize the variance associated with low-count genes. Filled points (blue and red) identify genes which are differentially expressed (adjusted p-value <0.01) in this comparison. Red points represent entries from Repbase, blue points are from the genomic annotation.
(TIF)

S9 Fig. *stwl* expression in ovarian follicle cells is dispensable for female fertility. Age-matched *Gal4*/*UAS-stwl-RNAi* transheterozygous virgin females were continuously mated to 1–5 day old *y w* males to determine the effect of *stwl* KD in follicle cells and germ cells. Progeny per day is the total number of progeny that emerged as in the indicated timeframe, divided by the number of days in that span. Each vial contained 10 females and 10 males. *C355-Gal4*,

T155-Gal4, and *C306-Gal4* drive expression in border cells and follicle cells from stage 9 onward (*C306-Gal4* additionally drives expression in stalk cells); *GR1-Gal4* drives expression in follicle stem cells of the germarium into later stages of oogenesis; *bam-Gal4* drives expression in germ cells starting at the cystoblast; *Mat-alpha-Gal4* drives expression in the post-GSC germline; *bam-Gal4* drives expression in early and late germ cells; *nos-Gal4* drives expression in GSCs; *Act5c-gal4* and *Tub-Gal4* drive ubiquitous expression in germline and somatic cells.
(TIF)

S10 Fig. TE de-repression in *stwl*, *bam*, and *piwi* mutant ovaries. Comparison of Gene Set Enrichment Analysis (GSEA) results. Normalized Enrichment Score (NES) is plotted for each set of repetitive elements enriched among mutant/WT ovaries. Higher NES indicates that the gene set is more upregulated in mutant ovaries. Count represents the number of genes in that set. Only gene sets with FDR<0.05 are plotted. *Sxl* deficient ovaries were also analyzed but are not shown because they were not enriched for any repeat classes.
(TIF)

S11 Fig. Overlap of genes upregulated in *stwl*, *bam*, *Sxl*, and *piwi* mutant or deficient ovaries. Venn diagram showing number of genes (from genomic annotation) that were upregulated (LFC>0, FDR<0.01) in each of the DESeq2 results outputs from S2 Table.
(TIF)

S12 Fig. Comparison of Gene Set Enrichment Analysis (GSEA) results for tissue-enrichment. Normalized Enrichment Score (NES) is plotted for each set of tissue-enriched genes enriched among mutant or deficient/WT ovaries. Higher/lower NES indicates that the gene set is highly upregulated/downregulated in mutant or deficient ovaries. Count represents the number of genes in that set. Only gene sets with FDR<0.05 are plotted.
(TIF)

S13 Fig. MA plot of RNA-Seq data from S2 cells. Fold-change for each gene is plotted against its mean transcript abundance across all assayed S2 cell samples (cells treated with *stwl* dsRNA and *lacZ* dsRNA as a control). Transcript abundance is represented by counts normalized according to GC-content and library size. The log2(Fold-change) values (LFC) were "shrunk" to minimize the variance associated with low-count genes. Filled points (blue and red) identify genes and repeats which are differentially expressed (adjusted p-value <0.01) in this comparison. Red points represent entries from Repbase, blue points are from the genomic annotation. Y-axis scale is identical to S8 Fig, for comparison.
(TIF)

S14 Fig. Stwl antibodies recognize a ~130 kDA fragment. Western blots on whole-fly lysates from ~10 $stwl^{+}$ (*y w* F10) and ~10 *stwl* null ($stwl^{j6c3}$/*Df(3L)Exel6122*) individuals aged 1–4 days. 6% SDS PAGE gel was loaded with lysates as indicated (row labelled "*stwl* genotype"), then transferred and probed with pre-immune or antibody sera of each animal, as indicated. The bottom panel shows the same membrane stripped and re-probed with a loading control (guinea pig α-Chromator). Final-bleed serum of each Stwl antibody recognizes a ~130 kDa fragment specific to $stwl^{+}$ lysates.
(TIF)

S15 Fig. α-Stwl sera label germ cell nuclei in $stwl^{+}$ testes (A, B) and ovaries (C, D). Tissues were dissected from *y w* F10 flies 10–15 days post-eclosion and immunostained with α-Stwl sera from GP 76 (A, C) and GP 77 (B, D). Vasa labels germ cells, DAPI labels cell nuclei. All images are maximum-intensity projections from a z-series representing a depth of 10 μm.

Scale bars are 20 μm.
(TIF)

S16 Fig. α-Stwl sera detect ectopically expressed Stwl-HA protein. Ovaries were dissected from *Act5c-Gal4/UAS-stwl-HA* females 0–1 days post-eclosion. Ovaries were probed with α-Vasa (germ cells), α-HA, and either GP 76 or GP 77 α-Stwl serum. HA signal recognizes cells in which Stwl-HA is being expressed; in these examples, expression is mostly limited to somatic cells (follicle cells and stalk cells). α-Stwl signal for both antibodies clearly overlaps with HA signal, resulting in bright yellow foci in the composite image. All images are maximum-intensity projections from a z-series representing a depth of 10 μm. Scale bars are 20 μm.
(TIF)

S17 Fig. α-Stwl sera immunoprecipitate Stwl from S2 cell lysates. S2 cell nuclei were lysed in RIPA buffer (Input), then incubated with one of two α-Stwl serum or a control antibody (α-Chromator) at 1:50 and 1:100 dilutions. Antibody-Protein complexes were isolated with Protein-A Agarose beads. Western blot of input, flow-through (FT) and IP complexes (IP) probed with α-Stwl GP 76 serum (top panel), then stripped and probed with α-Chromator antibody (bottom panel). Stwl runs at ~130 kDa (as shown in Figs S14 and Fig 3A), as does Chromator. Both α-Stwl sera immunoprecipitate Stwl effectively at a concentration of 1:100 (Stwl protein is eliminated from flowthrough). α-Chromator antibody fails to immunoprecipitate Stwl (Stwl protein remains in flow-through), but successfully immunoprecipitates Chromator.
(TIF)

S18 Fig. Stwl IP replicates create reproducible peaksets. IDR plots show the distribution of peak scores in replicate 1 (x-axis) vs replicate 2 (y-axis). Grey dots are reproducible peaks that pass the given IDR threshold, red dots are irreproducible peaks. Each dot represents a ChIP--Seq peak called in both replicates of a single antibody (C, D) or mock (A, B) experiment. Peak scores reflect the fold-enrichment of reads in the IP or mock sample relative to input. IDR identifies peaks whose signal intensities (i.e. scores) are similar in both replicates. Peaks with low signal intensity in both replicates do not pass the IDR threshold, but are useful for generating a background dataset for IDR analysis. Very few peaks were identified in mock ChIP-Seq experiments, even with a relaxed IDR threshold of 0.25.
(TIF)

S19 Fig. *bgcn; stwl* double mutant ovaries resemble *stwl* mutant ovaries. Ovaries were dissected from females of the indicated genotype 1–3 days post-eclosion. α-Vasa labels germ cells, α-Hts-1B1 labels branched fusomes or spectrosomes as well as follicle cell membranes. Germaria are positioned with anterior to posterior going left to right. *stwl* mutants form rudimentary cysts, indicated by branched fusomes, while *bgcn* mutant ovaries are populated with GSC-like cells, as indicated by spectrosomes and lack of branched fusomes [11–13,46–47]. Arrows point to branched fusomes in *stwl* mutants and *bgcn; stwl* double mutants, and spectrosomes in *bgcn* mutants. The *bgcn; stwl* result is consistent with previous findings [14]. All images are maximum-intensity projections from a z-series representing a depth of 10 microns. Scale bars are 20 microns.
(TIF)

S20 Fig. GSC genes are enriched in *stwl* null ovaries. Single-cell data from Rust et al. 2020 [45] was used to identify GSC, undifferentiated germ cell and older germ cell transcripts in assayed datasets. These were the only three types of germline cells identified in the Rust et al. study. (A) Germline transcripts among the top and bottom 1% of misregulated genes in the

indicated mutant or deficient ovary. (B) Germline transcripts overrepresented among ectopically expressed genes in *stwl* null ovaries and *stwl* dsRNA-treated S2 cells. Overrepresentation tests were also performed on the top 1% by LFC of ectopic genes in *stwl* null ovary, and of genes ectopic to both *stwl* null ovary and *stwl* dsRNA-treated S2 cells (for this intersect group, average LFC values in *stwl* null ovary are plotted). (A-B) Average LFC in the indicated mutant or deficient ovary is plotted for each single-cell germline cluster. All gene sets with FDR >0.1 are shaded grey. (C) Overlap of genes belonging to the following categories: enriched in testis (S7 Table), expressed in GSCs, or ectopically expressed in *stwl* null ovaries and *stwl* dsRNA-treated S2 cells.
(TIF)

S21 Fig. Principal components analysis (PCA) of α-Stwl ChIP-Seq read counts separates mock from IP. PCA for read counts generated from alignment to genomic bins and repeat index. Experiments labeled as mock were performed with pre-immune sera, IPs were performed with Stwl antibodies. Antibodies were generated from two different animals (referred to as 76 and 77) using the same epitope. DNA was isolated from two pools of S2 cells (biological replicates). The majority of the variance in the data is contained in PC1 and is explained by differences between mock and IP conditions, not by differences in the source animal or replicate pools.
(TIF)

S1 Table. Summary of differentially expressed genes. For each experiment, differential expression analysis was performed by contrasting counts in the "numerator" samples to the "reference" samples. Genes were identified as Differentially Expressed (D.E.) if the adjusted p-value was <0.01 (Wald Test p-value with a Benjamini-Hochberg FDR correction). Upregulated and downregulated genes have Log Fold Change (LFC) values greater than 0 (Up) or less than 0 (Down). For the ovary dataset, newly-eclosed and two-day old ovary data were combined in a GLM as described in Methods.
(XLSX)

S2 Table. Results of DE analyses. Results of DESeq2 output for all RNA-Seq datasets used in this study. LFC = shrunken log2FoldChange values, FDR = adjusted p-value (Wald Test p-value with a Benjamini-Hochberg FDR correction). "Repeat info": for repeats, details of specified repeat are listed; genomic annotations are labelled "Genome". "Ectopic status in *stwl* mutant" specifies whether indicated gene is ectopically expressed in *stwl* null ovary, *stwl* KD S2 cells, or both. "Stwl_Bound" indicates whether there is a Stwl peak within 1 Kb of the TSS of the indicated gene, in either direction.
(XLSX)

S3 Table. Genes upregulated in *stwl*, *bam*, *Sxl*, and *piwi* mutant or deficient ovaries. Intersection of all genes (from genomic annotation) that were upregulated (LFC>0, FDR<0.01) in each of the DESeq2 results outputs from S2 Table.
(XLSX)

S4 Table. Summary of genes and expression profiles in the 59C4-59D cluster. Methods for estimation of tissue enrichment, wild-type expression, up-regulation and ectopic expression are described in the text. Distance between genes is calculated as the difference between the right-most coordinate of the indicated gene body and the left-most coordinate of the adjacent gene body, regardless of gene orientation.
(XLSX)

S5 Table. Transcription factors with ChIP-Seq overlap to Stwl. Comparison of Stwl ChIP--Seq peakset identified in this study to all peaksets compiled by the ModERN group (including their independent analysis of Stwl). This table represents a subset of the top 40 transcription factors whose binding profiles are most similar to Stwl (from a total of 475 TFs). Fold-enrichment refers to the number of observed peaks overlapping with Stwl relative to the number of expected peaks, based on the size of the peaksets and the mappable genome. Also shown are the number of peaks in the listed peakset, the number of these peaks that overlap with the Stwl peakset, the % of Stwl peaks that are in this overlap (from a total of 2153 peaks), the % of Stwl sites (bp) that are covered in the listed peakset, the % of overlap peaks in the listed peakset, and the % sites in the listed peakset covered by Stwl.
(XLSX)

S6 Table. Tissue groupings key for tissue-enrichment data, based on modENCODE anatomy RNA-Seq [34]. We selected a subset of the tissues assayed for the modENCODE tissues profile. Column 1 gives the name of each dataset as listed in the modENCODE database. Where relevant, tissues were grouped into broader categories (e.g. A_MateM_4d_head and A_MateF_4d_head are grouped into a single "Head" category). A—Adults; L3_Wand—Wandering 3rd instar larvae; WPP—White pre-pupae; P8—Pharate adult stage P8; MateM/MateF—for adults, mated males or females; 4d - for adults, 4 days post-eclosion.
(XLSX)

S7 Table. Tissue-specificity index (τ) calculations per gene, based on modENCODE anatomy RNA-Seq [34]. For each transcript, RPKM was downloaded from Flybase precomputed file, "gene_rpkm_report_fb_2018_05.tsv.gz". modENCODE tissues were grouped according to S6 Table, so that enrichment in any member of a given group was called as enrichment in that tissue group. All RPKM values were converted to log2(RPKM) in R, excluding 0 counts. log2(RPKM) is shown in each tissue column. log2(RPKM)< = 1 was considered not expressed. All transcripts that are not expressed in at least one tissue were coded as "Not Expressed". For each transcript, τ was calculated from log2-normalized RPKM values, as described in Yanai et al., 2005:

$$\tau = \frac{\sum_{i=1}^{n}(1-\hat{x}_i)}{n-1};\ \hat{x}_i = \frac{x_i}{max(x_i)}$$

where x^i is the expression of a transcript in tissue i. Transcripts where $\tau<0.70$ were called "Not enriched". Transcripts were considered enriched in a given tissue where $\tau> = 0.70$ and expression in that tissue was greater than the cutoff (mean expression + 1.5 standard deviations).
(XLSX)

S8 Table. Summary of tissue-enrichment based on modENCODE anatomy RNA-Seq [34]. Total number of genes with tissue-specific expression, by tissue category (S6 Table). Genes marked as "Not Expressed" are not expressed in any of the 18 selected modENCODE datasets (RPKM <2 in all samples). "Not Enriched" genes are expressed genes that do not pass the tau (τ) threshold specified in S7 Table. % of Genome is based on a total of 18,194 annotated genes.
(XLSX)

S9 Table. Primers used in this study.
(XLSX)

Acknowledgments

We thank Chip Aquadro, Shuqing Ji, Heather Flores, Satyaki Prasad, Kevin Wei, Michael McGurk, Jackie Bubnell, and Dean Castillo for their input at various stages of this project. We

also acknowledge Abdullah Ozer, Judhajeet Ray, Roman Spektor and Tawny Cuykendall for their invaluable guidance on ChIP-Seq library prep, Aaron Chen for his technical contributions, Helen Salz for helpful conversations, and Iskander Said for assistance with data management.

Author Contributions

Conceptualization: Daniel Zinshteyn, Daniel A. Barbash.

Data curation: Daniel Zinshteyn.

Formal analysis: Daniel Zinshteyn.

Funding acquisition: Daniel A. Barbash.

Investigation: Daniel Zinshteyn.

Methodology: Daniel Zinshteyn.

Project administration: Daniel Zinshteyn.

Supervision: Daniel A. Barbash.

Validation: Daniel Zinshteyn.

Visualization: Daniel Zinshteyn.

Writing – original draft: Daniel Zinshteyn.

Writing – review & editing: Daniel Zinshteyn, Daniel A. Barbash.

References

1. Song X, Xie T. DE-cadherin-mediated cell adhesion is essential for maintaining somatic stem cells in the *Drosophila* ovary. Proc Natl Acad Sci U S A. 2002; 99: 14813–14818. https://doi.org/10.1073/pnas.232389399 PMID: 12393817
2. Xie T, Spradling AC. A Niche Maintaining Germ Line Stem Cells in the *Drosophila* Ovary. Science. 2000; 290: 328–330. https://doi.org/10.1126/science.290.5490.328 PMID: 11030649
3. Xie Ting. Control of germline stem cell self-renewal and differentiation in the *Drosophila* ovary: concerted actions of niche signals and intrinsic factors. Wiley Interdiscip Rev Dev Biol. 2012; 2: 261–273. https://doi.org/10.1002/wdev.60 PMID: 24009036
4. Lin H, Spradling AC. A novel group of *pumilio* mutations affects the asymmetric division of germline stem cells in the *Drosophila* ovary. Development. 1997; 124: 2463–2476. https://doi.org/10.1242/dev.124.12.2463 PMID: 9199372
5. Ma X, Wang S, Do T, Song X, Inaba M, Nishimoto Y, et al. Piwi Is Required in Multiple Cell Types to Control Germline Stem Cell Lineage Development in the *Drosophila* Ovary. PLoS ONE. 2014; 9. https://doi.org/10.1371/journal.pone.0090267 PMID: 24658126
6. Klenov MS, Sokolova OA, Yakushev EY, Stolyarenko AD, Mikhaleva EA, Lavrov SA, et al. Separation of stem cell maintenance and transposon silencing functions of Piwi protein. Proc Natl Acad Sci. 2011; 108: 18760–18765. https://doi.org/10.1073/pnas.1106676108 PMID: 22065765
7. Li Y, Zhang Q, Carreira-Rosario A, Maines JZ, McKearin DM, Buszczak M. Mei-p26 cooperates with Bam, Bgcn and Sxl to promote early germline development in the *Drosophila* ovary. PloS One. 2013; 8: e58301. https://doi.org/10.1371/journal.pone.0058301 PMID: 23526974
8. Hashiyama K, Hayashi Y, Kobayashi S. *Drosophila Sex lethal gene* initiates female development in germline progenitors. Science. 2011; 333: 885–888. https://doi.org/10.1126/science.1208146 PMID: 21737698
9. Berg CA, Spradling AC. Studies on the Rate and Site-Specificity of *P* Element Transposition. Genetics. 1991; 127: 515–524. https://doi.org/10.1093/genetics/127.3.515 PMID: 1849859
10. Karpen GH, Spradling AC. Analysis of Subtelomeric Heterochromatin in the *Drosophila* Minichromosome Dp1187 by Single *P* Element Insertional Mutagenesis. Genetics. 1992; 132: 737–753. https://doi.org/10.1093/genetics/132.3.737 PMID: 1334894

11. Clark KA, McKearin DM. The *Drosophila stonewall* gene encodes a putative transcription factor essential for germ cell development. Development. 1996; 122: 937–950. https://doi.org/10.1242/dev.122.3.937 PMID: 8631271

12. Akiyama T. Mutations of *stonewall* disrupt the maintenance of female germline stem cells in *Drosophila melanogaster*. Dev Growth Differ. 2002; 44: 97–102. https://doi.org/10.1046/j.1440-169x.2002.00625.x PMID: 11940096

13. Maines JZ, Park JK, Williams M, McKearin DM. Stonewalling *Drosophila* stem cell differentiation by epigenetic controls. Development. 2007; 134: 1471–1479. https://doi.org/10.1242/dev.02810 PMID: 17344229

14. Park JK. Characterization of factors controlling germline stem cell maintenance in Drosophila melanogaster. PhD Thesis, University of Texas Southwestern Medical Center. 2007.

15. Yi X, de Vries HI, Siudeja K, Rana A, Lemstra W, Brunsting JF, et al. Stwl Modifies Chromatin Compaction and Is Required to Maintain DNA Integrity in the Presence of Perturbed DNA Replication. Mol Biol Cell. 2009; 20: 983–994. https://doi.org/10.1091/mbc.e08-06-0639 PMID: 19056684

16. Lespinet O, Wolf YI, Koonin EV, Aravind L. The Role of Lineage-Specific Gene Family Expansion in the Evolution of Eukaryotes. Genome Res. 2002; 12: 1048–1059. https://doi.org/10.1101/gr.174302 PMID: 12097341

17. Shukla V, Habib F, Kulkarni A, Ratnaparkhi GS. Gene Duplication, Lineage-Specific Expansion, and Subfunctionalization in the MADF-BESS Family Patterns the *Drosophila* Wing Hinge. Genetics. 2014; 196: 481–496. https://doi.org/10.1534/genetics.113.160531 PMID: 24336749

18. Choi JY, Aquadro CF. The Coevolutionary Period of *Wolbachia pipientis* Infecting *Drosophila ananassae* and Its Impact on the Evolution of the Host Germline Stem Cell Regulating Genes. Mol Biol Evol. 2014; 31: 2457–2471. https://doi.org/10.1093/molbev/msu204 PMID: 24974378

19. Flores HA, DuMont VLB, Fatoo A, Hubbard D, Hijji M, Barbash DA, et al. Adaptive Evolution of Genes Involved in the Regulation of Germline Stem Cells in *Drosophila melanogaster* and *D. simulans*. G3 GenesGenomesGenetics. 2015; 5: 583–592. https://doi.org/10.1534/g3.114.015875 PMID: 25670770

20. Shukla V, Dhiman N, Nayak P, Dahanukar N, Deshpande G, Ratnaparkhi GS. Stonewall and Brickwall: Two Partially Redundant Determinants Required for the Maintenance of Female Germline in *Drosophila*. G3 GenesGenomesGenetics. 2018; 8: 2027–2041. https://doi.org/10.1534/g3.118.200192 PMID: 29669801

21. DuMont B, L V, Flores HA, Wright MH, Aquadro CF. Recurrent Positive Selection at Bgcn, a Key Determinant of Germ Line Differentiation, Does Not Appear to be Driven by Simple Coevolution with Its Partner Protein Bam. Mol Biol Evol. 2007; 24: 182–191. https://doi.org/10.1093/molbev/msl141 PMID: 17056645

22. Rohrbaugh M, Clore A, Davis J, Johnson S, Jones B, Jones K, et al. Identification and Characterization of Proteins Involved in Nuclear Organization Using *Drosophila* GFP Protein Trap Lines. PLoS ONE. 2013; 8. https://doi.org/10.1371/journal.pone.0053091 PMID: 23341925

23. Dorman ER, Bushey AM, Corces VG. The role of insulator elements in large-scale chromatin structure in interphase. Semin Cell Dev Biol. 2007; 18: 682–690. https://doi.org/10.1016/j.semcdb.2007.08.009 PMID: 17919949

24. Czech B, Preall JB, McGinn J, Hannon GJ. A Transcriptome-wide RNAi Screen in the *Drosophila* Ovary Reveals Factors of the Germline piRNA Pathway. Mol Cell. 2013; 50: 749–761. https://doi.org/10.1016/j.molcel.2013.04.007 PMID: 23665227

25. Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, et al. Without Argonaute3, Aubergine-bound piRNAs collapse but Piwi-bound piRNAs persist. Cell. 2009; 137: 509–521. https://doi.org/10.1016/j.cell.2009.04.027 PMID: 19395009

26. Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, et al. Specialized piRNA Pathways Act in Germline and Somatic Tissues of the *Drosophila* Ovary. Cell. 2009; 137: 522–535. https://doi.org/10.1016/j.cell.2009.03.040 PMID: 19395010

27. Shapiro-Kulnane L, Smolko AE, Salz HK. Maintenance of *Drosophila* germline stem cell sexual identity in oogenesis and tumorigenesis. Development. 2015; 142: 1073–1082. https://doi.org/10.1242/dev.116590 PMID: 25758221

28. Sun S, Cline TW. Effects of *Wolbachia* Infection and *ovarian tumor* Mutations on Sex-lethal Germline Functioning in *Drosophila*. Genetics. 2009; 181: 1291–1301. https://doi.org/10.1534/genetics.108.099374 PMID: 19171941

29. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15. https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281

30. Gan Q, Chepelev I, Wei G, Tarayrah L, Cui K, Zhao K, et al. Dynamic regulation of alternative splicing and chromatin structure in *Drosophila* gonads revealed by RNA-seq. Cell Res. 2010; 20: 763–783. https://doi.org/10.1038/cr.2010.64 PMID: 20440302

31. Smolko AE, Shapiro-Kulnane L, Salz HK. The H3K9 methyltransferase SETDB1 maintains female identity in *Drosophila* germ cells. Nat Commun. 2018; 9: 4155. https://doi.org/10.1038/s41467-018-06697-x PMID: 30297796

32. Cox DN, Chao A, Baker J, Chang L, Qiao D, Lin H. A novel class of evolutionarily conserved genes defined by *piwi* are essential for stem cell self-renewal. Genes Dev. 1998; 12: 3715–3727. https://doi.org/10.1101/gad.12.23.3715 PMID: 9851978

33. Peng JC, Valouev A, Liu N, Lin H. Piwi maintains germline stem cells and oogenesis in *Drosophila* through negative regulation of Polycomb Group proteins. Nat Genet. 2016; 48: 283–291. https://doi.org/10.1038/ng.3486 PMID: 26780607

34. Brown JB, Boley N, Eisman R, May GE, Stoiber MH, Duff MO, et al. Diversity and dynamics of the *Drosophila* transcriptome. Nature. 2014; 512: 393–399. https://doi.org/10.1038/nature12962 PMID: 24670639

35. Hong G, Zhang W, Li H, Shen X, Guo Z. Separate enrichment analysis of pathways for up- and downregulated genes. J R Soc Interface. 2014; 11. https://doi.org/10.1098/rsif.2013.0950 PMID: 24352673

36. Warden CD, Kanaya N, Chen S, Yuan Y-C. BD-Func: a streamlined algorithm for predicting activation and inhibition of pathways. PeerJ. 2013; 1: e159. https://doi.org/10.7717/peerj.159 PMID: 24058887

37. Shevelyov YY, Lavrov SA, Mikhaylova LM, Nurminsky ID, Kulathinal RJ, Egorova KS, et al. The B-type lamin is required for somatic repression of testis-specific gene clusters. Proc Natl Acad Sci U S A. 2009; 106: 3282–3287. https://doi.org/10.1073/pnas.0811933106 PMID: 19218438

38. Schneider I. Cell lines derived from late embryonic stages of *Drosophila melanogaster*. J Embryol Exp Morphol. 1972; 27: 353–365. PMID: 4625067

39. Zhang Y, Malone JH, Powell SK, Periwal V, Spana E, MacAlpine DM, et al. Expression in Aneuploid *Drosophila* S2 Cells. PLoS Biol. 2010; 8. https://doi.org/10.1371/journal.pbio.1000320 PMID: 20186269

40. Salz HK, Dawson EP, Heaney JD. Germ cell tumors: Insights from the *Drosophila* ovary and the mouse testis. Mol Reprod Dev. 2017; 84: 200–211. https://doi.org/10.1002/mrd.22779 PMID: 28079292

41. Yang SY, Baxter EM, Van Doren M. Phf7 controls male sex determination in the *Drosophila* germline. Dev Cell. 2012; 22: 1041–1051. https://doi.org/10.1016/j.devcel.2012.04.013 PMID: 22595675

42. Yang SY, Chang Y-C, Wan YH, Whitworth C, Baxter EM, Primus S, et al. Control of a Novel Spermatocyte-Promoting Factor by the Male Germline Sex Determination Factor PHF7 of *Drosophila melanogaster*. Genetics. 2017; 206: 1939–1949. https://doi.org/10.1534/genetics.117.199935 PMID: 28588035

43. Chau J, Kulnane LS, Salz HK. Sex-lethal Facilitates the Transition From Germline Stem Cell to Committed Daughter Cell in the Drosophila Ovary. Genetics. 2009; 182: 121–132. https://doi.org/10.1534/genetics.109.100693 PMID: 19237687

44. McKearin D, Ohlstein B. A role for the Drosophila bag-of-marbles protein in the differentiation of cystoblasts from germline stem cells. Development. 1995; 121: 2937–2947. https://doi.org/10.1242/dev.121.9.2937 PMID: 7555720

45. Rust K, Byrnes LE, Yu KS, Park JS, Sneddon JB, Tward AD, et al. A single-cell atlas and lineage analysis of the adult Drosophila ovary. Nat Commun. 2020; 11: 5628. https://doi.org/10.1038/s41467-020-19361-0 PMID: 33159074

46. Ohlstein B, Lavoie CA, Vef O, Gateff E, McKearin DM. The *Drosophila* cystoblast differentiation factor, benign gonial cell neoplasm, is related to DExH-box proteins and interacts genetically with *bag-of-marbles*. Genetics. 2000; 155: 1809–1819. https://doi.org/10.1093/genetics/155.4.1809 PMID: 10924476

47. McCarthy A, Deiulio A, Martin ET, Upadhyay M, Rangan P. Tip60 complex promotes expression of a differentiation factor to regulate germline differentiation in female *Drosophila*. Montell D, editor. Mol Biol Cell. 2018; 29: 2933–2945. https://doi.org/10.1091/mbc.E18-06-0385 PMID: 30230973

48. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015; 31: 2382–2383. https://doi.org/10.1093/bioinformatics/btv145 PMID: 25765347

49. Kudron MM, Victorsen A, Gevirtzman L, Hillier LW, Fisher WW, Vafeados D, et al. The ModERN Resource: Genome-Wide Binding Profiles for Hundreds of *Drosophila* and *Caenorhabditis elegans* Transcription Factors. Genetics. 2018; 208: 937–949. https://doi.org/10.1534/genetics.117.300657 PMID: 29284660

50. Heger A, Webber C, Goodson M, Ponting CP, Lunter G. GAT: a simulation framework for testing the association of genomic intervals. Bioinformatics. 2013; 29: 2046–2048. https://doi.org/10.1093/bioinformatics/btt343 PMID: 23782611

51. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME Suite: tools for motif discovery and searching. Nucleic Acids Res. 2009; 37: W202–W208. https://doi.org/10.1093/nar/gkp335 PMID: 19458158

52. Tue NT, Yoshioka Y, Mizoguchi M, Yoshida H, Zurita M, Yamaguchi M. DREF plays multiple roles during *Drosophila* development. Biochim Biophys Acta Gene Regul Mech. 2017; 1860: 705–712. https://doi.org/10.1016/j.bbagrm.2017.03.004 PMID: 28363744

53. Williams TM, Selegue JE, Werner T, Gompel N, Kopp A, Carroll SB. The Regulation and Evolution of a Genetic Switch Controlling Sexually Dimorphic Traits in *Drosophila*. Cell. 2008; 134: 610–623. https://doi.org/10.1016/j.cell.2008.06.052 PMID: 18724934

54. Riddle NC, Elgin SCR. The *Drosophila* Dot Chromosome: Where Genes Flourish Amidst Repeats. Genetics. 2018; 210: 757–772. https://doi.org/10.1534/genetics.118.301146 PMID: 30401762

55. Gerland TA, Sun B, Smialowski P, Lukacs A, Thomae AW, Imhof A. The *Drosophila* speciation factor HMR localizes to genomic insulator sites. PLOS ONE. 2017; 12: e0171798. https://doi.org/10.1371/journal.pone.0171798 PMID: 28207793

56. Thomae AW, Schade GOM, Padeken J, Borath M, Vetter I, Kremmer E, et al. A Pair of Centromeric Proteins Mediates Reproductive Isolation in *Drosophila* Species. Dev Cell. 2013; 27: 412–424. https://doi.org/10.1016/j.devcel.2013.10.001 PMID: 24239514

57. Andreyeva EN, Belyaeva ES, Semeshin VF, Pokholkova GV, Zhimulev IF. Three distinct chromatin domains in telomere ends of polytene chromosomes in *Drosophila melanogaster Tel* mutants. J Cell Sci. 2005; 118: 5465–5477. https://doi.org/10.1242/jcs.02654 PMID: 16278293

58. Silva-Sousa R, Varela MD, Casacuberta E. The Putzig partners DREF, TRF2 and KEN are involved in the regulation of the *Drosophila* telomere retrotransposons, *HeT-A* and *TART*. Mob DNA. 2013; 4: 18. https://doi.org/10.1186/1759-8753-4-18 PMID: 23822164

59. Insco ML, Bailey AS, Kim J, Olivares GH, Wapinski OL, Tam CH, et al. A self-limiting switch based on translational control regulates the transition from proliferation to differentiation in an adult stem cell lineage. Cell Stem Cell. 2012; 11. https://doi.org/10.1016/j.stem.2012.08.012 PMID: 23122292

60. Vogelmann J, Gall AL, Dejardin S, Allemand F, Gamot A, Labesse G, et al. Chromatin Insulator Factors Involved in Long-Range DNA Interactions and Their Role in the Folding of the *Drosophila* Genome. PLOS Genet. 2014; 10: e1004544. https://doi.org/10.1371/journal.pgen.1004544 PMID: 25165871

61. Saha P, Sowpati DT, Mishra RK. Epigenomic and genomic landscape of *Drosophila melanogaster* heterochromatic genes. Genomics. 2019; 111: 177–185. https://doi.org/10.1016/j.ygeno.2018.02.001 PMID: 29432976

62. Wong LC, Schedl P. Dissection of *Drosophila* Ovaries. JoVE J Vis Exp. 2006; e52. https://doi.org/10.3791/52 PMID: 18704176

63. Zamore PD, Ma S. Isolation of *Drosophila melanogaster* Testes. JoVE J Vis Exp. 2011; e2641. https://doi.org/10.3791/2641 PMID: 21610676

64. Rio DC, Ares M, Hannon GJ, Nilsen TW. Purification of RNA using TRIzol (TRI reagent). Cold Spring Harb Protoc. 2010; 2010: pdb.prot5439. https://doi.org/10.1101/pdb.prot5439 PMID: 20516177

65. Sheffield P, Garrard S, Derewenda Z. Overcoming expression and purification problems of RhoGDI using a family of "parallel" expression vectors. Protein Expr Purif. 1999; 15: 34–39. https://doi.org/10.1006/prep.1998.1003 PMID: 10024467

66. Bischof J, Björklund M, Furger E, Schertel C, Taipale J, Basler K. A versatile platform for creating a comprehensive UAS-ORFeome library in *Drosophila*. Dev Camb Engl. 2013; 140: 2434–2442. https://doi.org/10.1242/dev.088757 PMID: 23637332

67. Hu Y, Comjean A, Roesel C, Vinayagam A, Flockhart I, Zirin J, et al. FlyRNAi.org—the database of the *Drosophila* RNAi screening center and transgenic RNAi project: 2017 update. Nucleic Acids Res. 2016. https://doi.org/10.1093/nar/gkw977 PMID: 27924039

68. Andrews S. FastQC A Quality Control tool for High Throughput Sequence Data. 2010. Available: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

69. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

70. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Research. 2018; 7: 1338. https://doi.org/10.12688/f1000research.15931.2 PMID: 30254741

71. McGurk MP, Barbash DA. Double insertion of transposable elements provides a substrate for the evolution of satellite DNA. Genome Res. 2018; 28: 714–725. https://doi.org/10.1101/gr.231472.117 PMID: 29588362

72. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9: 357–359. https://doi.org/10.1038/nmeth.1923 PMID: 22388286

73. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015; 31: 166–169. https://doi.org/10.1093/bioinformatics/btu638 PMID: 25260700

74. Risso D, Schwartz K, Sherlock G, Dudoit S. GC-Content Normalization for RNA-Seq Data. BMC Bioinformatics. 2011; 12: 480. https://doi.org/10.1186/1471-2105-12-480 PMID: 22177264

75. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinforma Oxf Engl. 2009; 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352 PMID: 19505943

76. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 2014; 42: W187–W191. https://doi.org/10.1093/nar/gku365 PMID: 24799436

77. Huang XA, Yin H, Sweeney S, Raha D, Snyder M, Lin H. A Major Epigenetic Programming Mechanism Guided by piRNAs. Dev Cell. 2013; 24: 502–516. https://doi.org/10.1016/j.devcel.2013.01.023 PMID: 23434410

78. Marinov GK, Wang J, Handler D, Wold BJ, Weng Z, Hannon GJ, et al. Pitfalls of mapping high-throughput sequencing data to repetitive sequences: Piwi's genomic targets still not identified. Dev Cell. 2015; 32: 765–771. https://doi.org/10.1016/j.devcel.2015.01.013 PMID: 25805138

79. Lin H, Chen M, Kundaje A, Valouev A, Yin H, Liu N, et al. Reassessment of Piwi Binding to the Genome and Piwi Impact on RNA Polymerase II Distribution. Dev Cell. 2015; 32: 772–774. https://doi.org/10.1016/j.devcel.2015.03.004 PMID: 25805139

80. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinforma Oxf Engl. 2018. https://doi.org/10.1093/bioinformatics/bty895 PMID: 30395178

81. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS J Integr Biol. 2012; 16: 284–287. https://doi.org/10.1089/omi.2011.0118 PMID: 22455463

82. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. Nat Protoc. 2012; 7. https://doi.org/10.1038/nprot.2012.101 PMID: 22936215

83. Li Q, Brown JB, Huang H, Bickel PJ. Measuring reproducibility of high-throughput experiments. Ann Appl Stat. 2011; 5: 1752–1779. https://doi.org/10.1214/11-AOAS466

84. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26: 841–842. https://doi.org/10.1093/bioinformatics/btq033 PMID: 20110278

85. Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinforma Oxf Engl. 2005; 21: 650–659. https://doi.org/10.1093/bioinformatics/bti042 PMID: 15388519

86. Kryuchkova-Mostacci N, Robinson-Rechavi M. A benchmark of gene expression tissue-specificity metrics. Brief Bioinform. 2017; 18: 205–214. https://doi.org/10.1093/bib/bbw008 PMID: 26891983