STONEWALL PREVENTS EXPRESSION OF TESTIS-ENRICHED GENES AND BINDS TO INSULATOR ELEMENTS IN D. MELANOGASTER

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Daniel Edward Zinshteyn

August 2019

STONEWALL PREVENTS EXPRESSION OF TESTIS-ENRICHED GENES AND BINDS TO INSULATOR ELEMENTS IN D. MELANOGASTER

Daniel Edward Zinshteyn, Ph.D.

Cornell University 2019

Germline stem cells (GSCs) are the progenitor cells for the entire population of an organism's germline. In *Drosophila*, these cells reside in a well-defined 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 maintenance factor Stonewall (Stwl) has undergone adaptive evolution across the *Drosophila* phylogeny and has been implicated in heterochromatin maintenance. In the hopes of identifying potential drivers of selection at Stwl, we investigated the molecular functions of Stwl.

We performed RNA-Seq on *stwl* mutant ovaries and testes to assay the transcript abundance of transposable elements in the absence of functional Stwl. We found that *stwl* mutant ovaries (but not testes) show significant de-repression of many transposon families, but that heterochromatic genes are not preferentially misregulated relative to euchromatic genes. We also found that testis-enriched genes, including the differentiation factor *bgcn* and a large cluster on chromosome 2, are upregulated in *stwl* mutant ovaries. This abnormal masculinization of the ovary was accompanied by ectopic expression of a number of testis- and somatic-enriched genes. Surprisingly, we also found that RNAi knockdown of *stwl* in S2 cells results in ectopic expression of these gene classes.

In order to understand how Stwl regulates transcription, we developed antibodies against

a Stwl epitope and thoroughly validated it prior to performing parallel ChIP-Seq and RNA-Seq experiments in S2 cells. Analysis of Stwl binding sites shows that Stwl binds upstream of transcription start sites and localizes to heterochromatic sequences. We also show that Stwl is enriched at repetitive sequences associated with telomeres. Finally, we identify binding motifs that are shared with known insulator binding proteins. We propose that Stwl affects gene regulation by binding insulators and establishing chromatin boundaries.

BIOGRAPHICAL SKETCH

Daniel E. Zinshteyn was born and raised in Northeast Philadelphia, where his parents had immigrated to less than two months before he was born. He grew up speaking Russian at home and in the playground but managed to pick up English from various duck-based animated cartoons. Dan's dalliance with biological research began in middle school, where he murdered countless flies in the pursuit of science. This continued in high school, where he was given the opportunity to work in an actual research setting in the Fox Chase Cancer Center, performing confocal microscopy on human and mouse cell lines.

Dan's needlessly exhaustive research record as a high school student aided in securing him a spot at the University of Pennsylvania for his undergraduate studies. It was there that he began to truly hone his passion for biological research, exploring new and exciting fields with world-class researchers. At Penn, he worked for two years in Dr. Tatyana Svitkina's lab studying actin filament organization in cells. After taking a class in population genetics, Dan then began working in Dr. Sarah Tishkoff's lab, studying the genetics of bitter taste perception in sub-Saharan Africa. Upon graduating from Penn, Dan worked for a year as a research assistant in Dr. Alana O'Reilly's lab, studying follicle stem cell maintenance in *D. melanogaster*. In 2012 he enrolled in the field of Genetics, Genomics, and Development at Cornell, with the intention of studying molecular evolution and functional genetics. He subsequently joined Dr. Daniel Barbash's lab and performed the research contained herein, with much help and guidance from his mentor, colleagues, friends and family. He always trusted the process.

Для тех, кто не смог быть здесь сегодня: дедушка и Леночка.

ACKNOWLEDGMENTS

The number of people who are directly or indirectly responsible for my success and sanity is really immense, but I will make an attempt to put into words how grateful I am for having had them in my life. Ithaca is a truly wonderful place to live, work, and play, but gorges and farmers markets can only go so far. The true magic of Ithaca and Cornell is that they are inhabited by wonderful people, many of whom I now call friends. I'll eventually shout a few out by name, but I'm not sure the graduate school will allow me to submit a document as long as the one I would have to write to properly thank everyone who I've broken bread and/or beer with on countless occasions.

I knew that I wanted to work with Dr. Daniel Barbash shortly after meeting him. I suspected we shared more than just a first name, and indeed over the years I came to find that his dry sense of humor (often-times dark) matches mine. Dan is a brilliant geneticist and a committed mentor: he's instilled in me a great appreciation for doing things the right way and doing them carefully. He also encouraged me to just shut up and write already, advice that I should have taken to heart more often. Most importantly, he's an incredibly kind and caring boss who always puts his mentees interests above all else. Thanks for all of that, Dan.

Studying at Cornell has been a tremendous privilege. I'm especially grateful for having had the opportunity to get to know so many brilliant faculty members in MBG. I would like to especially thank my committee members, Drs. Charles Aquadro and John Lis. Chip was one of the reasons I applied to GGD at Cornell, and his enthusiasm and warmth were among the reasons I ended up joining the program. I'm honored to have learned and taught pop gen under his tutelage. John was always generous with his time, and although I never took "The Nucleus" I'm pretty sure I've absorbed a good bit of that class just from osmosis in being proximal to John for committee meetings.

Outside of my committee, I would also like to thank Drs. Mariana Wolfner, Mike Goldberg, Kelly Liu, Paul Soloway, Eric Alani, and Cedric Feschotte. Kelly, Paul, and Eric were very helpful and supportive DGS's and/or department heads during their respective reins.

Eric and Paul are better scientists than they are skit actors. Mariana was the fourth committee member for my A-exam and gave some really helpful feedback that helped guide my thinking early on in the lifespan of my project. I've always been a huge admirer of her research and have become an even bigger fan of her as a human being. Mike has always been a well of wisdom to me, an unexpected personal and professional resource that I am immensely grateful for. Cedric arrived at Cornell years after I did, but he tracked me down to tell me how excited he was about my research at a time when I was stuck. His enthusiasm was (and is) infectious, and it helped to reinvigorate my interests in this project and move this research forward.

The Barbash lab was defined by the people that came in and out of it: each time a person left or joined, it became an entirely different place, and I loved coming to work at all of those places. The only constant personality in the lab (excluding Boss Dan) was our lab manager Shuqing Ji. Shuqing has a deep well of knowledge when it comes to biochemistry, molecular biology, genetics, and fashion. She has her hands in nearly every project in the lab and is extremely generous with her time and abilities. I can't begin to count the number of techniques she taught me.

In the first iteration of the Barbash lab as I knew it, Dr. Satyaki Rajavasireddy served as my wise elder brother (no offense to my wise, biological elder brother). He was and remains a great friend and a shoulder to lean on when needed. Dr. Kevin Wei was the first person to formally welcome me to the lab, but also the person who helped me truly feel at home in the lab. The year I went around with him playing pickup ball was probably the best shape I've ever been in; ball is life, Kevin. Dr. Tawny Cuykendall was a great friend in the lab. She taught me how to perform ChIP, but more importantly she taught me the importance of being meticulous and organized in molecular biology.

Dr. Mike McGurk (newly-minted) joined the lab a year after I did, but it's really hard to picture the place without him. I wouldn't tell it to his face, but the man is a genius and a shoo-in for greatness in academia. Nonetheless, his academic brilliance is secondary to his integrity as a person. At the risk of saying too many nice things about him, I'd like to add, "Shut up, Mike."

Phase II of the Barbash lab, while I was there, consisted of the arrival of three post-docs, Drs. Sarah Lower, Anne-Marie Dion-Côté, and Dean Castillo. They each made the lab a better place to be, and all have moved on to their dream jobs in academia. I miss having scones and watching Downton Abbey with Sarah, having a beer with AMDC (and making fun of her Canadian-ness), and talking NBA/NFL/Marvel movies with Dean. Also, I think I still have a book or two I borrowed from AMDC. Iskander Said has only recently joined the lab, but I can confirm that he's good people and the lab is better for having him in it.

Much of my project has found me venturing outside of the Barbash lab in search of assistance. I would like to thank Drs. Roman Spektor, Fabiana Duarte, Jacob Tome, Judhajeet Ray, and Abdullah Ozer for the protocols they shared with me and the number of times they held my hands through a biochemical protocol until I got it. Vanessa Bauer DuMont taught me the work that goes behind the scenes of population genetics.

Before I move on from thanking Cornellians, I have to thank some of the people who made Cornell home for me. Drs. Nithya Kartha and Peter Sullivan, as well as Drs. Florencia Schlamp and Adam Dolan did this in a literal sense, and I'll never forget their generosity in bringing me into their homes when I was between places. The following people have a lifetime ban from the karaoke bar on triphammer mall: Roman, Andrea, Adele, Jackie, Aaron, Maria. Roman: remember that time we made a movie? And that time we made a second movie? Ian, Zach, and Mike are better friends than they are poker players. Joo Hyun and Karl, I miss our dinner dates. Jen and Tom, you were the best neighbors in the history of neighbors. Jae: Ithaca was never the same once you left. I'm looking forward to showing you where the good whitefish salad is in Philly. We can send some to Roman.

Drs. Michael Campbell and Alessia Ranciaro were my mentors while I worked on my undergraduate research thesis. They saw something in me that I didn't see at that time, and I'm grateful for their encouraging me to pursue a PhD, and for convincing me to go to Cornell. Dr. Farida Korobova was my research mentor during my first two years of undergraduate research; although I found that the lab's focus didn't quite fit my interests, Farida taught me some

invaluable molecular biology and microscopy skills that I used throughout my PhD training. I still brag about some of the cool experiments I did in that lab.

At the risk of sounding obsequious, Dr. Alana O'Reilly was probably my greatest champion before grad school. She hired me as a summer intern with only a single college biology class under my belt, despite the fact that her lab was brand new and I would have to work under her directly. As a result, I was thrown headfirst into a project which at the outset sounded completely impossible given my knowledge and ability at the time. Alana spent hundreds of hours working with me and teaching me what I didn't know at the time would be the basis of almost every experiment I would conduct as a graduate student. For some reason Alana brought me back to continue that research over the course of the next four years. In that time, I became more and more independent in my work and thinking. This experience gave me the confidence that I could succeed in grad school, and for that I am grateful.

I would also like to thank some teachers who were incredibly important to my development as a scientist and critical thinker: Mr. Dennis Erlick, Mr. Dominic Fedele, and Ms. Resa Levinson. These three public school teachers have turned thousands of students into scientists and science-minded adults. I can only hope to make a tenth of the impact on the world that any single one of them has made.

In the most literal sense, I would not be here without my family. I also would not be here without their constant love and support. My parents, Larisa and Yan, left their friends, families and professions in the USSR to come to a country where they didn't know the language and their engineering degrees weren't worth the paper they were printed on. They made the choice to reduce their own quality of life so that their children would have greater opportunity in theirs. I can never thank them enough for their sacrifices. Luckily, they were not alone in raising us: they also had the help of my grandmother Tatyana and my grandfather Mark. Babushka has an endless reservoir of strength that is beyond comprehension and a constant inspiration to me. Dedushka passed away before I completed my degree, but the memory of his smiling face and quiet intelligence motivates me every day to be more like him. My sister Olga is a tornado of

unconditional love. My Aunt Elena probably needs help with her printer right now, which I will fix gladly because it's the least I can do for someone who's always there for me.

I cannot claim the distinction of being the first Dr. Zinshteyn: that honor belongs to my eggheaded elder brother. While I claim to be a first-generation American, I never really had to grapple with being the first in my family to do anything. Any path I desired to tread upon, Dr. Boris Zinshteyn had already blasted wide open for me. The passion I have for biology is merely whatever rubbed off on me from being in near proximity to him. He's been there for me at every step, from my first science fair project to the last chapter of my dissertation. In short, this is all his fault.

To my wife, Vilma: this is all you. You crossed mountains and paused your career just to be with me. I can never thank you enough, and you've never asked me to. Ithaca only ever felt like a home when you were there with me, not before and not after. You bring joy, clarity, sanity, spontaneity, and bunnies into my life. I love you.

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION AND LITERATURE REVIEW

The Germline Stem Cell (GSC) regulatory network in the Drosophila ovary is a frequent target for positive selection, despite its critical role in maintaining fertility in the organism (Choi and Aquadro, 2015; DuMont et al., 2007; Flores et al., 2015). GSCs are one of several of adult stem cell populations that inhabit the fruit fly ovary. Adult stem cells exist in tissues where there is constant turnover of cells, such as animal gonads where gametes are produced and released at a rapid rate. The stem cells of the Drosophila ovary reside at the anterior-most structure, the germarium (Figure 1).

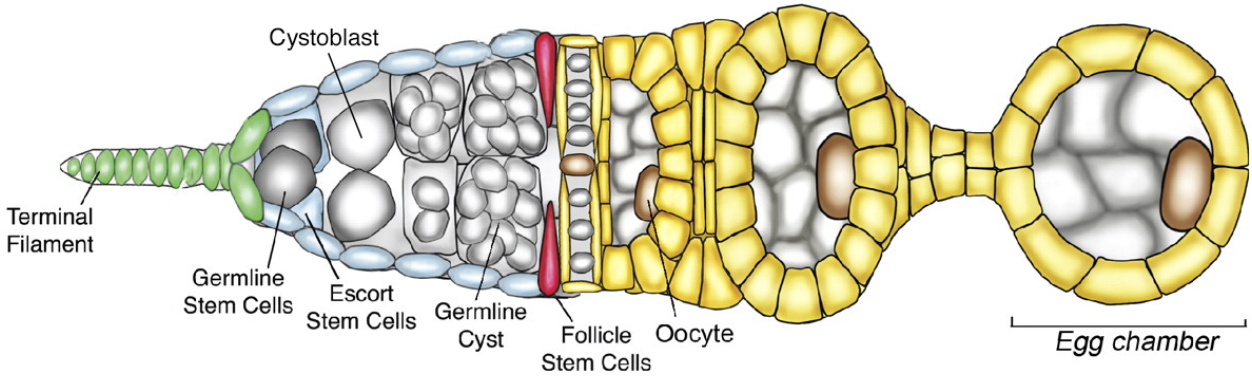


Figure 1 | Multiple adult stem cell populations reside in the Germarium. Adapted from Hartman et al.(Hartman et al., 2010)

Both the Germline and Follicle Stem Cells (FSCs) reside in a niche environment that is required to maintain the stem cell populations(Song and Xie, 2002; Xie and Spradling, 2000). Stem cells are classically defined by their ability to 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 GSCs, the differentiated daughter cell is the cystoblast, which then undergoes four progressive rounds of incomplete cell division to form a 16-cell cyst. Each germline cyst progresses through the germarium and becomes enveloped by follicle cells derived

from FSCs. Upon completing four rounds of mitosis, 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(Ting Xie, 2012). These factors can be broadly categorized as maintenance factors (required for self-renewal) or differentiation factors, which are often antagonistic to each other. The molecular mechanisms by which the GSC regulatory network acts are numerous and poorly understood. A complex interplay of intrinsic cell signaling, extrinsic (niche) signaling, spindle construction/orientation and chromatin remodeling is necessary for proper maintenance and differentiation of GSC populations. Further complicating matters (from the perspective of the researcher) is the fact that many, if not most, of the genes involved in GSC regulation are pleiotropic for oogenic and other functions. One such gene, *stonewall* (*stwl*), is the primary focus of this dissertation.

The *stwl* locus spans 7 KB on chromosome 3L, including a single 2.5 KB intron. The protein product is 1037 amino acids (~150 kD), including two evolutionary conserved functional domains: a N-terminal MADF (**M**yb-SANT like in **Adf**) domain and a C-terminal BESS (**BE**AF, **S**U-Var(3-7), **S**tonewall like) domain(Bhaskar and Courey, 2002; Clark and McKearin, 1996). This general pattern is conserved within the MADF-BESS family of proteins, which has expanded in the *Drosophila* lineage via gene duplication and subfunctionalization(Shukla et al., 2014). The MADF domain is capable of binding directly to DNA in a sequence-specific manner, while the BESS domain mediates protein-protein interactions(Bhaskar and Courey, 2002). Both

of these domains are characterized by a N-terminal α-helix followed by a helix-turn-helix motif, including amino acid residues that are highly conserved among this family of proteins. The secondary structure prediction algorithm JPRED identified ordered regions outside of these two functional domains, though data is lacking to identify the function of these putative domains(Drozdetskiy et al., 2015).

The *stwl* gene locus was discovered in a *P*-element mutagenesis screen for female sterility and subsequently identified as a germline-expressed gene in an enhancer trap screen(Berg and Spradling, 1991; Karpen and Spradling, 1992). The McKearin lab performed an extensive characterization of the *stwl* locus and its gene products(Clark, 1996). They found that homozygous and hemizygous mutant alleles of *stwl* result in female sterility, characterized by rudimentary ovaries and a failure to lay eggs. X-gal staining of the enhancer trap P[*lacZ*; ry^{+}] *stwl* showed reporter activity beginning in cystoblasts in region 1 of the germarium, with the signal decaying until it becomes readily apparent in nurse cells of stage 10 egg chambers. This pattern was recapitulated with RNA *in situ* against *stwl* transcript. The authors concluded based on *stwl*'s expression in germaria that it is likely involved in early germ cell differentiation and that its expression in stage 10+ nurse cells reflects maternal deposition in the developing embryo.

Further examination of the *stwl* null phenotype revealed that egg chamber growth arrested between stages 4 and 7, with germ cells of mutant ovaries displaying a "membrane blebbing" phenotype characteristic of programmed cell death. These cells were confirmed to be apoptotic via positive TUNEL staining. Despite this terminal phenotype of germ cells, they did not detect defects in cyst formation: BamC and Hts antibody staining demonstrated normal formation of 2-,4-, and 8-cell clusters, indicating that cystoblasts underwent 4 successive rounds of mitosis. Examining later stages of oogenesis, the investigators found that *stwl* mutant egg

chambers typically contained 16 polyploid nurse cells, indicating that in these ovarioles, the cystocyte-to-oocyte transformation does not occur. They confirmed this phenotype by showing that Orb protein and BicD protein fail to accumulate in *stwl* mutant cysts, thus impeding oocyte determination. Oocyte determination occurs at the conclusion of the transit-amplifying stage of cyst development: it is characterized by the formation of the synaptonemal complex, the appearance of double-strand breaks, and the transition of the specified cystocyte into a meiotic pro-oocyte. The authors concluded that the failure of Orb and BicD proteins to accumulate in the pro-oocyte was a result of a defect in asymmetric transport.

Clark and McKearin developed polyclonal antibodies against Stwl protein and found that Stwl localizes to germ cell nuclei, from GSCs through to stage 7 nurse cells and oocytes. They did not detect Stwl protein in subsequent stages of oogenesis or embryogenesis, re-enforcing that the RNA *in situ* signal in stage 10+ egg chambers is likely due to maternal deposition of RNAs. They detected faint signal for Stwl protein in somatic cell nuclei of the ovary and found that Stwl associated with polytene chromosomes in salivary glands. From these observations they concluded that Stwl is a chromosomal protein, and likely a transcription factor. Stwl also localized to first-instar larval gonads, as well as third instar male/female larval gonads, suggesting that it may be required for normal gonad formation during development. They also found that Stwl localized to germ cell nuclei in adult testis, but they failed to identify any phenotype for *stwl* null males. Finally, their antibody failed to detect Stwl protein in S2 cells and is a poor reagent for immunoprecipitation experiments, thereby limiting their ability to perform biochemical assays on the Stwl protein.

stwl was not actually identified as a GSC maintenance factor until it was independently isolated by Akiyama, who was carrying out a mutagenesis screen for female sterility(Akiyama,

2002). They noted that *stwl* null germaria are frequently lacking in Germline Stem Cells, and that this phenotype is exacerbated with age: specifically, 64% of examined 1-day old germaria lacked GSCs, while 99% of 7-day old germaria lacked GSCs. Thus, the agametic phenotype described in Clark and McKearin is caused not only by the oocyte determination defect, but also by perturbation of GSC maintenance. Furthermore, the study found that pupal ovarioles are often devoid of germ cells, corroborating the prediction (based on protein expression in larval gonads) that Stwl is required for normal gonad development. The author also observed that *stwl* mutant germaria contained an abundance of low-complexity fusomes, suggesting that cystoblasts failed to complete 4 rounds of incomplete mitosis; this contradicts the finding of Clark and McKearin, which stated that transit-amplification appeared normal in *stwl* null ovaries.

The stwl locus became a focal point of another study spawned from a forward genetic screen. This time, investigators were searching for genes whose induction led to an apoptosis phenotype in various somatic cells (recall that loss of Stwl results in apoptosis of germ cells, though it is unclear if this is a direct or indirect result)(Brun et al., 2006). They found that induction of UAS-stwl in eye, wing, and larval wing discs by gmr-, vg-, and en-gal4, respectively, resulted in phenotypes ranging from rough eye patches and wing notches at low temperatures, to lethality at room temperature. This lethality had been previously observed when full-length stwl was induced in imaginal discs(Toba et al., 1999). They further demonstrated that affected tissues were replete with cells exhibiting hyperproliferation, as evidenced by overexpansion and entry into mitotic S phase, followed by apoptosis (confirmed by visualization of proapoptotic gene transcripts). Induction of UAS-stwl via vg¬-gal4 during development led to pupal death; this lethality was rescued by co-expression of pro-apoptotic genes, including bax, rpr, hid, sickle, p53, and grim. Forced expression of stwl in salivary glands with fkh-gal4

inhibited endoreduplication of DNA in salivary glands, causing a reduction in DNA content and a shrinking of affected glands. They concluded that Stwl, like its distant cousin dMyb (more on this later), is involved in the control of cell cycle progression. This ties in nicely with the two requirements that Stwl has in the ovary: oocyte specification (i.e. transitioning from a mitotic cystocyte to a meiotic oocyte, rather than an endoreduplicating polyploid nurse cell) and stem cell maintenance.

Not to be outdone, the McKearin lab published another study on Stwl, this time confirming that it was required intrinsically for GSC maintenance(Maines et al., 2007). They showed that GSC clones mutant for *stwl* were rapidly depleted; no TUNEL staining was observed in these cells or their progenitor cystoblasts, indicating that stem cell loss was a consequence of terminal differentiation, not apoptosis. Furthermore, egg chambers derived from *stwl* null GSCs exhibited the same defects as previously reported, including apoptosis of germ cells and an absence of oocyte determination. Egg chambers derived from *stwl* null follicle cells did not exhibit these defects, indicating that Stwl was required in germ cells for normal germline development.

The authors also found that ectopic Stwl expression (*UAS-stwl* driven by *nanos*-gal4) in GSCs resulted in an expansion of GSCs in otherwise normal-looking germaria. This was direct evidence that Stwl promoted GSC proliferation at the expense of cystoblast differentiation. It also provided a useful assay to determine which genes antagonized *stwl* activity in GSCs; they found that single-copy mutants (heterozygotes) of *nanos* (*nos*) and *pumilio* (*pum*) suppressed the GSC expansion phenotype caused by ectopic expression of *stwl*, while mutations of *mad, punt,* and *tkv* did not. *nos* and *pum* are translational inhibitors that act cell intrinsically to promote GSC maintenance and prevent differentiation, whereas *mad, punt,* and *tkv* are members of the *dpp*

signaling pathway that suppresses differentiation factors from somatic cells (the GSC niche). This was further evidence that *stwl* function is required in GSCs to promote oogenesis.

Since *bam* is considered the master switch for differentiation of germ cells, the investigators checked to see how it interacts with *stwl*. They found that *bam* remains transcriptionally silent in *stwl* null GSC clones, indicating that *stwl* it is not directly responsible for *bam* silencing. They then generated *stwl bam* double mutants and found that branched fusomes form; this is in direct contrast to *bam* mutants, where fusomes do not branch and ovarioles fill up with spectrosome-dotted germ cells. The same result was seen with *bam*'s binding partner, *bgcn* (Park, 2007). They concluded from these experiments that *stwl* is antagonistic to *bam* function by a mechanism other than transcriptional regulation of *bam*.

Because *bam* mutants and *stwl bam* double mutants were both agametic, the authors performed microarray profiling on ovaries isolated from these animals to identify misregulated transcripts. They found 500 genes differentially expressed (DE) due to *stwl* loss, acknowledging that some of these DE genes may be indicative of early cyst development that is abolished in *bam* mutants and partially rescued in *stwl bam* double mutants. They found that *lola* transcripts increased in *stwl bam* double mutants and confirmed that an increase in Lola protein abundance in *stwl* mutant ovaries and germ cell clones. Lola protein has been well characterized as a transcriptional regulator necessary for GSC maintenance and mitosis-to-meiosis transition in the male germline, making it plausible that the *stwl* mutant phenotype is partially due to mis-regulation of *lola*. The authors found that *lola* transcripts, along with a significant percentage of other upregulated transcripts from *stwl bam* double mutants, had multiple Nanos Response Elements (NREs) in their 3'-UTRs. These NREs promote binding of Nanos and Pumilio, which act to repress translation of these transcripts. Thus, the authors concluded that the genetic

interaction between Stwl, Nanos, and Pumilio may be due to overlapping targets of repression.

Perhaps the most interesting finding from the Maines et al. paper was that Stwl acts as a dominant suppressor of Position Effect Variegation (PEV). PEV occurs when an actively transcribed gene is inserted near a heterochromatic locus and becomes subject to transcriptional silencing. A variegating phenotype occurs when the gene is silenced in a subset of cells and active in others, as is the case with the mottled white (w^{m4}) allele. This allele presents as predominantly white eyes with occasional red specks of pigment, as the heterochromatin silencing machinery renders the w+ allele inert in most ommatidia. Dominant suppressors of PEV, or Su(Var)s, are genes whose perturbation results in a demonstrable loss of the variegating phenotype. These genes are typically involved in heterochromatin maintenance, so that loss of their activity results in a destabilization of heterochromatin that prevents silencing of the variegating gene, which in the case of the mottled white locus means a near uniform distribution of red pigment. Following up on the observation that Stwl protein is expressed in eye discs, the investigators performed assays to test whether Stwl influenced PEV. stwl was proven to be a Su(Var), as single mutations (heterozygotes) for the gene resulted in a dramatic loss of PEV for three different variegating eye color alleles (*DX1,* w^{m4}*,* bw^{D}).

A link between Stwl and heterochromatin silencing was the first hint at Stwl's molecular function, with the exception of homology that Stwl has with other known proteins (see below). This link was expanded upon by yet another study derived from a forward genetic screen (clearly, the power of forward genetics remains unmatched). The aim of the screen was to identify genes involved in DNA damage and checkpoint response(Yi et al., 2009). Briefly, DNA synthesis was moderately inhibited via chemical treatment; in a healthy fly, the checkpoint response pathway should initiate to prevent entry into S phase. The chemical treatment also

induces DNA damage; therefore mutants that are hypersensitive to this treatment may have mutant alleles in genes involved in DNA damage or checkpoint initiation. *stwl* mutants were identified as being hypersensitive to treatment with hydroxyurea (HU). Larvae homozygous for mutant stwl were more susceptible to HU-induced lethality in a dose-dependent manner. They went on to show that this HU sensitivity did not manifest in larval lethality, but rather in morphological abnormalities in pupal development. These affected larvae had smaller and more disorganized imaginal discs than control larvae treated with HU.

Suspecting that Stwl may be an important checkpoint response protein, they examined cell cycle progression in salivary glands, larval brains, and S2 cells, with and without Stwl protein. They found that cells lacking Stwl (including RNAi-treated S2 cells) failed to enter into mitosis after HU treatment, similar to wild-type cells treated with HU. On the other hand, larval brains lacking Grp/Dchk1, a known checkpoint protein, contained numerous cells entering into mitosis, indicating that the DNA checkpoint response failed.

Following from the observation that Stwl is not involved in checkpoint response, they focused on the published finding that *stwl* is a Su(Var). First, they confirmed this result themselves with their own *stwl* allele. They also confirmed that Stwl was present in interphase nuclei of salivary glands (using the McKearin lab polyclonal antibody), and corroborated that the resolution is not high enough to determine specific patterns of Stwl localization on polytene chromosomes. Contrary to the Clark and McKearin study, they found that Stwl was highly expressed in S2 cells and was present in interphase nuclei of these cells. Immunofluorescence imaging of S2 cells labeled with α-Stwl revealed that Stwl is typically present in one or more foci at the edge of the nucleolus, where the nucleolus and the nuclear envelope are proximal. These interphase, nucleolar foci often colocalized with a-HP1 signal, providing another crucial

piece of evidence that Stwl associates with heterochromatin. Mitotic S2 cells typically had very diffuse Stwl signal, lacking distinct puncta. They subsequently examined Stwl localization under Electron Microscopy (EM). They found that Stwl accumulates in a structure closely associated with the nucleolus, and in dense, transcriptionally inert, heterochromatin-like structures dispersed throughout the nucleus.

Following this observation, they next asked whether HP1 localization is dependent on Stwl. They found that *stwl* mutant salivary glands and *stwl* dsRNA-treated S2 cells both exhibited normal HP1 localization, indicating that Stwl is not required for HP1 localization to the nuclear periphery. Since heterochromatin is typically associated with markers for repressed chromatin, they tested for changes in levels of H3K9me3 and H3K27me3. They found that both of these repressed chromatin marks were reduced in *stwl* mutant larvae, relative to controls with H2A and H3K4. Additionally, they found a dramatic increase in the DNA damage marker **γ**-H2Av in HU-treated *stwl* mutant larvae. Finally, they showed that mammalian O23 cells (which lack endogenous Stwl) co-transfected with *Gal4-Stwl* (a fusion peptide) and *UAS-luciferase* plasmids exhibited a repression of the luciferase reporter; increasing the concentration of Gal4-Stwl fusion peptide led to a decrease in luciferase activity, relative to a Gal4 only control. They concluded from these experiments that Stwl is required for maintenance of heterochromatic marks H3K9me3 and H3K27me3, that loss of Stwl function (and these methylated histones) leads to an increase in DNA damage, and that Stwl is capable of repressing genes at ectopic loci. This latter observation has not been confirmed in an *in vivo* system.

Stwl cropped up in yet another forward screen, this time aimed at characterizing GFP protein traps that exhibit non-uniform subnuclear localization (Rohrbaugh et al., 2013). The authors used a Stwl-GFP protein trap line to determine that Stwl-GFP is present in distinct loci

on polytene chromosomes and is distributed in nuclear puncta in larval imaginal discs. They noted that both Stwl and Stwl-GFP proteins label the periphery of GSC nuclei diffusely, with localization becoming more punctate along the nuclear periphery in developing cystoblasts and nurse cells. These puncta occur at the nuclear lamina, as confirmed by overlap with a-Lamin Dm0. Since insulator proteins typically bind polytene chromosomes at multiple loci but localize to diploid nuclei in only a few puncta, the authors speculated that Stwl may be an insulator protein. They found that a-CP190 and a-Stwl co-localize in terminal filament cells, ovarian follicle cells, and larval imaginal disc cells, though this co-localization is not complete. Interestingly, they found that a-Stwl and a-CP190 do not overlap on polytene chromosomes, although a-Stwl bands along polytene chromosomes are often adjacent to a-CP190 bands. They also pointed out that protein interactome data identifies Stwl as an interactor with Mod(mdg4), which is a known component of the Su(Hw) insulator complex, and known lamina components Otefin, Bocksbeutel, and Man1. They suggest that the coalescence of a-Stwl signal into distinct puncta in developing cystoblasts is indicative of Stwl acting to modify chromatin architecture. It should be noted that the Yi et al. study found that a-Stwl signal is diffuse in mitotic S2 cells but is punctate in non-dividing cells. Thus, it may be that Stwl localization in the nucleus may be involved in changes in nuclear architecture that correspond to cellular differentiation.

In another study, Stwl was found to be a target of phosphorylation by Wee kinase (Sopko et al., 2014). Transfected HA-tagged Wee and FLAG-tagged Stwl were found to co-IP in S2 cells. The authors also showed that *wee*-deficient embryos exhibit a reduction of phosphorylated Tyr305 of Stwl and that overexpression of *wee* in S2 cells leads to an increase in phosphorylated precipitate. They also conducted *in vitro* kinase assays between human Wee1 and various fragments of *D. melanogaster* Stwl. They found that Wee1 directly phosphorylates these Stwl

fragments *in vitro*, although the BESS-domain containing fragment interferes with Wee1 autophosphorylation activity. They did not find changes in levels of H3K9me3 or H3K27me3 in *stwl* mutant 0-2 hour embryos, despite the finding from Yi et al. that these marks increase in *stwl* mutant larvae.

A phylogenetic analysis by Shukla et al. found that *stwl* was part of a lineage-specific expansion of MADF-BESS domain containing proteins in the *Drosophila* genus; they also demonstrated that *stwl* and its genomic neighbor, *CG3919*, are likely a pair of gene duplicates(Shukla et al., 2014). They found that a number of members of the MADF-BESS family, including *stwl*, are required for wing development. Specifically, RNAi knockdown of *stwl* in the dorsal region of the larval wing disc (*MS1096-gal4)* results in a complete loss of the alula on the adult wing blade and a reduction of wing size. This phenotype is exacerbated when *hng1*, another MADF-BESS family gene, is also knocked down using the same Gal4 driver.

A recent follow-up to the MADF-BESS family study tested whether any of these family members were involved in oogenesis in *D. melanogaster*(Shukla et al., 2018). They tested whether RNAi knockdown of these genes in ovarian germline (*nos*-gal4) or ovarian somatic cells (*c587-gal4*) led to oogenesis defects similar to *stwl* knockdown. They found that one gene, which they dubbed *brickwall* (*brwl*), has mutant phenotypes similar to *stwl*. Like *stwl*, *brwl* mutant ovaries exhibit a reduction of GSCs and germline cysts, mislocalization of Orb protein, and germ cells undergoing apoptosis. They found that RNAi knockdown of *brwl* in germline (but not somatic) ovarian tissue phenocopies the *brwl* mutant defects, and that Brwl-GFP expresses in germline and somatic ovarian nuclei. Unlike *stwl*, they found that Brwl-deficient ovaries become defective over time, typically ~7 days post-eclosion. They also found that *brwl* mutant ovaries have a significant reduction in mitotically and meiotically active germ cells; for the latter, there

appears to be a near-complete loss of the synaptonemal complex in *brwl* mutant egg chambers. Germline expression of a *UAS-stwl-HA* transgene in *brwl* mutants partially rescues a number of these delayed-onset defects (ring canal deformities, oocyte specification defections, lack of mitotic activity; does not rescue apoptosis), indicating that they may have partially redundant function.

The authors also did phylogenetic and evolutionary analyses comparing *stwl* and *brwl* evolution in the *Drosophila* phylogeny. They found that both genes experienced purifying selection at the MADF and BESS domains across the 12 genomes, emphasizing that these domains are critical for normal protein function. They also found that the MADF domains of both proteins are positively charged, which according to previous research indicates that they likely bind DNA directly(Maheshwari et al., 2008). Despite these similarities, both functional domains of *stwl* appear to be more rapidly evolving (or deleterious alleles are less stringently selected against), and the protein-interacting BESS domains differ in charge (+2.2 in Stwl, -1.1 in Brwl). The authors proposed that the elevated rate of substitutions in Stwl across the *Drosophila* genus relative to Brwl may reflect positive selection acting on Stwl, and the difference in charge of the BESS domains could contribute to divergent functions of the two proteins. They also pointed out that the genes encoding MADF-BESS domain proteins are distributed across the genome, which suggests that the formation of this gene family was transposon-mediated. Notably, previous research has suggested that the MADF motif is likely transposon-derived(Casola et al., 2007).

The Shukla et al. group did not publish the first evolutionary analyses at the *stwl* locus. A previous study of GSC regulatory genes had found signatures of selection at the gene in *Drosophila ananassae*(Choi and Aquadro, 2014). Specifically, the authors found an elevated rate

of nonsynonymous divergence at *stwl* between *D. ananassae* and its sibling species, as well as evidence for selection on synonymous sites favoring the fixation of preferred mutations. A McDonald-Kreitman test on the protein-coding sequence also suggested that positive selection was causing an accumulation of nonsynonymous fixed differences. A maximum likelihood model also predicted positive selection at the gene body, with the strongest signal coming from the 3' end of the coding sequence. A follow-up study from the same group had similar findings in *D. melanogaster* and its sister species *D. simulans*(Flores et al., 2015). Flores et al. Sanger sequenced African strains of both species and analyzed the site-frequency spectrum at the *stwl* locus to identify signatures of selection. They found significant departures from neutrality at the locus compared to numerous demographic simulations. A McDonald-Kreitman test also rejected neutrality due to an excess of nonsynonymous fixed differences in the coding sequence. Thus, the data indicate that *stwl* has been a target of positive selection in multiple *Drosophila* lineages.

While the aforementioned studies are a thorough overview of most of the literature specifically focusing on the evolution and function of *stwl*, there is still valuable data we can analyze from other large datasets that may not explicitly focus on the gene. One study performed RNA *in situ* against hundreds of transcripts, including *stwl*; they found that *stwl* RNA is maternally loaded into early embryos, expressed throughout oogenesis, and localized to gonads in late embryogenesis(Tomancak et al., 2002, 2007). This observation supports the hypothesis in Clark and McKearin regarding maternal deposition of *stwl* transcript. Regarding Stwl function, one study found that induced expression of *stwl* in central motor neurons impacts dendrite morphology; another found that induced expression in the central nervous system leads to loss of adult-specific neurite projections(Ou et al., 2008; Zhao et al., 2008). Another study found that induction in the developing wing resulted in loss of bristles and changes in wing size and

patterning(Cruz et al., 2009). These findings add to the observations from Brun et al. that induction of *stwl* in various tissues leads to hyperproliferation and cell death; in other tissues the impact of *stwl* induction is less severe.

Other studies identified specific defects associated with germline knockdown of *stwl* transcript. Czech et al. found that *nos-Gal4->UAS-stwl-RNAi* knockdown of *stwl* results in moderate increases of *HetA*, *blood*, and *burdock* transposon transcripts(Czech et al., 2013). Yan et al. determined that *MTD-Gal4->UAS-stwl-RNAi* knockdown results in completely agametic ovaries with 100% phenotype penetrance and no egg production; *Mata-Gal4->UAS-stwl-RNAi* knockdown also results in full penetrance of complete sterility, but affected females lay eggs and ovaries appear to be phenotypically normal(Yan et al., 2014). These findings reflect the pleiotropic functions of *stwl* in oogenesis: the *MTD-Gal4-* driver forces expression beginning in GSCs, whereas *Mata-Gal4* expression begins in stage 1 egg chambers; both drivers induce expression in germ cells through to late stage egg chambers. The finding that *stwl* knockdown via both drivers causes sterility confirms a requirement for oocyte determination; the fact that the phenotype is more severe with the driver impacting GSC expression is due to the additional GSC maintenance function of *stwl*.

stwl also came up in a screen for genes involved in lymph gland development: expression of *stwl* in larval hemolymph (specifically lymph gland blood cells, e.g. hemocytes, and circulating blood cells) results in a dramatic expansion of the lymph gland and a reduction in the density of circulating blood cells(Mondal et al., 2014). Furthermore, RNAi knockdown in the same cells causes a reduction in lymph gland size and a loss of hemolymph progenitor cells. This result directly implicates *stwl* in the maintenance of a stem cell-like population outside of the germline, further indicating the pleiotropic function of the gene.

Stwl protein has been picked up in numerous proteomics assays. It was identified as a common contaminant in affinity capture using FLAG and StrepII purifications; it is therefore imperative to avoid utilizing these tags in biochemical assays involving Stwl(Rees et al., 2011). In a yeast one-hybrid screen, Stwl was found to interact with the enhancer for the *yellow* gene, which further supports Rohrbaugh et al.'s hypothesis that it is an insulator-binding protein(Kalay et al., 2016; Rohrbaugh et al., 2013). It was also identified in an assay for S2 cell proteins that interact with a stably transfected wasp-infecting polydnavirus protein(Salvia et al., 2017). This interaction with a protein from an invasive parasite may be indicative of Stwl having a role in defense against external or internal parasitic elements, as suggested by previous evolutionary studies.

The literature on *stwl* and its gene products is various and disconnected, making distinct conclusions on *stwl* function challenging. Our lab is interested in genes that are rapidly evolving within the *Drosophila* phylogeny, particularly those which interact with parasitic mobile elements or regulate repetitive DNA sequences. Therefore, our entry into investigating the *stwl* gene was rooted in our desire to understand the driving force behind the signatures of adaptive evolution in the protein sequence occurring along multiple lineages within the *Drosophila* genus. We predicted that, based on the published literature implicating *stwl* in maintenance of heterochromatin, the gene may be rapidly evolving in response to either satellite DNA expansion or transposon mobilization. Our findings have not supported this hypothesis.

In summary, here is what we know about *stwl*: its protein product is rapidly evolving in *D. melanogaster*, *D. simulans* and *D. ananassae*. The transcript is maternally deposited in embryos and expressed in germ cells throughout development. Protein localizes to germ cell and follicle cell nuclei, in a punctate manner; it appears to be associated with the nuclear lamina and

dense, heterochromatin-like structures. This localization is diffuse in GSCs and becomes more punctate in developed cells; this observation is similar to Stwl localization in S2 cells, where it presents as nuclear puncta that partly colocalize with HP1; these puncta disappear during mitosis. The protein also binds at specific chromosomal regions to polytene chromosomes in larval salivary glands, often adjacent to the insulator-binding protein CP190; interactome data shows that Stwl interacts physically with other known insulator binding-proteins and components of the nuclear lamina, and it physically binds to a known enhancer. These data point to Stwl's primary molecular function being that of an insulator-binding protein.

However, only a few studies have suggested that Stwl is an insulator binding protein. Most instead focus on the developmental or cellular consequences of Stwl gain and/or loss in an *in vivo* or *in vitro* context. We know that Stwl is required in germ cells for GSC maintenance and oocyte specification. Stwl may promote oocyte specification by inhibiting endoreduplication. The loss of Stwl results in agametic ovaries as a consequence of GSC loss and programmed cell death of affected germ cells; *stwl* mutants exhibit germline defects beginning at the pupal stage.

Overexpression of *stwl* typically results in over-proliferation of the affected cell; in otherwise wild type tissue, this leads to aberrant morphology resulting from programmed cell death (as a response to over-proliferation). It can also rescue mutant phenotypes that are caused by a failure of cells to proliferate. These observations could all be explained via a role in binding insulator sites. Genetic insulators are stretches of the genome that protein complexes bind to in a cell-specific manner. Binding of insulator sites typically acts to disrupt enhancer-promoter interactions, thus providing a critical component of developmentally regulated transcriptional control. The data presented in this dissertation will support the finding that Stwl is binding at genetic insulators, which in conjunction with its role as a rapidly evolving germline protein,

provide a potent model for studying evolutionary developmental biology.

References

1. Akiyama, T. (2002). Mutations of stonewall disrupt the maintenance of female germline stem cells in Drosophila melanogaster. Dev. Growth Differ. *44*, 97–102.

2. Berg, C.A., and Spradling, A.C. (1991). Studies on the Rate and Site-Specificity of P Element Transposition. Genetics *127*, 515–524.

3. Bhaskar, V., and Courey, A.J. (2002). The MADF–BESS domain factor Dip3 potentiates synergistic activation by Dorsal and Twist. Gene *299*, 173–184.

4. Brun, S., Rincheval-Arnold, A., Colin, J., Risler, Y., Mignotte, B., and Guénal, I. (2006). The myb-related gene stonewall induces both hyperplasia and cell death in Drosophila: rescue of fly lethality by coexpression of apoptosis inducers. Cell Death Differ. *13*, 1752–1762.

5. Casola, C., Lawing, A.M., Betrán, E., and Feschotte, C. (2007). PIF-like Transposons are Common in Drosophila and Have Been Repeatedly Domesticated to Generate New Host Genes. Mol. Biol. Evol. *24*, 1872–1888.

6. Choi, J.Y., and Aquadro, C.F. (2014). 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. *31*, 2457–2471.

7. Choi, J.Y., and Aquadro, C.F. (2015). Molecular Evolution of Drosophila Germline Stem Cell and Neural Stem Cell Regulating Genes. Genome Biol. Evol. *7*, 3097–3114.

8. Clark, K. (1996). The Drosophila Stonewall gene encodes a putative transcription factor that regulates oocyte differentiation and germ cell viability. PhD Thesis. University of Texas Southwestern Medical Center.

9. Clark, K.A., and McKearin, D.M. (1996). The Drosophila stonewall gene encodes a putative transcription factor essential for germ cell development. Development *122*, 937–950.

10. Cruz, C., Glavic, A., Casado, M., and de Celis, J.F. (2009). A Gain-of-Function Screen Identifying Genes Required for Growth and Pattern Formation of the Drosophila melanogaster Wing. Genetics *183*, 1005–1026.

11. Czech, B., Preall, J.B., McGinn, J., and Hannon, G.J. (2013). A Transcriptome-wide RNAi Screen in the Drosophila Ovary Reveals Factors of the Germline piRNA Pathway. Mol. Cell *50*,

749–761.

12. Drozdetskiy, A., Cole, C., Procter, J., and Barton, G.J. (2015). JPred4: a protein secondary structure prediction server. Nucleic Acids Res. *43*, W389–W394.

13. DuMont, B., L, V., Flores, H.A., Wright, M.H., and Aquadro, C.F. (2007). 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. *24*, 182–191.

14. Flores, H.A., DuMont, V.L.B., Fatoo, A., Hubbard, D., Hijji, M., Barbash, D.A., and Aquadro, C.F. (2015). Adaptive Evolution of Genes Involved in the Regulation of Germline Stem Cells in Drosophila melanogaster and D. simulans. G3 Genes, Genomes, Genetics *5*, 583–592.

15. Hartman, T.R., Zinshteyn, D., Schofield, H.K., Nicolas, E., Okada, A., and O'Reilly, A.M. (2010). Drosophila Boi limits Hedgehog levels to suppress follicle stem cell proliferation. J. Cell Biol. *191*, 943–952.

16. Kalay, G., Lusk, R., Dome, M., Hens, K., Deplancke, B., and Wittkopp, P.J. (2016). Potential Direct Regulators of the Drosophila yellow Gene Identified by Yeast One-Hybrid and RNAi Screens. G3 Genes, Genomes, Genetics *6*, 3419–3430.

17. Karpen, G.H., and Spradling, A.C. (1992). Analysis of Subtelomeric Heterochromatin in the Drosophila Minichromosome Dp1187 by Single P Element Insertional Mutagenesis. Genetics *132*, 737–753.

18. Maheshwari, S., Wang, J., and Barbash, D.A. (2008). Recurrent Positive Selection of the Drosophila Hybrid Incompatibility Gene Hmr. Mol. Biol. Evol. *25*, 2421–2430.

19. Maines, J.Z., Park, J.K., Williams, M., and McKearin, D.M. (2007). Stonewalling Drosophila stem cell differentiation by epigenetic controls. Development *134*, 1471–1479.

20. Mondal, B.C., Shim, J., Evans, C.J., and Banerjee, U. (2014). Pvr expression regulators in equilibrium signal control and maintenance of Drosophila blood progenitors. ELife *3*.

21. Ou, Y., Chwalla, B., Landgraf, M., and van Meyel, D.J. (2008). Identification of genes influencing dendrite morphogenesis in developing peripheral sensory and central motor neurons. Neural Develop. *3*, 16.

22. Park, J.K. (2007). Characterization of factors controlling germline stem cell maintenance in Drosophila melanogaster. PhD Thesis. University of Texas Southwestern Medical Center.

23. Rees, J.S., Lowe, N., Armean, I.M., Roote, J., Johnson, G., Drummond, E., Spriggs, H., Ryder, E., Russell, S., Johnston, D.S., et al. (2011). In Vivo Analysis of Proteomes and Interactomes Using Parallel Affinity Capture (iPAC) Coupled to Mass Spectrometry. Mol. Cell. Proteomics MCP *10*.

24. Rohrbaugh, M., Clore, A., Davis, J., Johnson, S., Jones, B., Jones, K., Kim, J., Kithuka, B., Lunsford, K., Mitchell, J., et al. (2013). Identification and Characterization of Proteins Involved in Nuclear Organization Using Drosophila GFP Protein Trap Lines. PLoS ONE *8*.

25. Salvia, R., Grossi, G., Amoresano, A., Scieuzo, C., Nardiello, M., Giangrande, C., Laurenzana, I., Ruggieri, V., Bufo, S.A., Vinson, S.B., et al. (2017). The multifunctional polydnavirus TnBVANK1 protein: impact on host apoptotic pathway. Sci. Rep. *7*.

26. Shukla, V., Habib, F., Kulkarni, A., and Ratnaparkhi, G.S. (2014). Gene Duplication, Lineage-Specific Expansion, and Subfunctionalization in the MADF-BESS Family Patterns the Drosophila Wing Hinge. Genetics *196*, 481–496.

27. Shukla, V., Dhiman, N., Nayak, P., Dahanukar, N., Deshpande, G., and Ratnaparkhi, G.S. (2018). Stonewall and Brickwall: Two Partially Redundant Determinants Required for the Maintenance of Female Germline in Drosophila. G3 Genes, Genomes,Genees Genetics *8*, 2027–2041.

28. Song, X., and Xie, T. (2002). DE-cadherin-mediated cell adhesion is essential for maintaining somatic stem cells in the Drosophila ovary. Proc. Natl. Acad. Sci. U. S. A. *99*, 14813–14818.

29. Sopko, R., Foos, M., Vinayagam, A., Zhai, B., Binari, R., Hu, Y., Randklev, S., Perkins, L.A., Gygi, S.P., and Perrimon, N. (2014). Combining Genetic Perturbations and Proteomics to Examine Kinase-Phosphatase Networks in Drosophila Embryos. Dev. Cell *31*, 114–127.

30. Ting Xie (2012). 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. *2*, 261–273.

31. Toba, G., Ohsako, T., Miyata, N., Ohtsuka, T., Seong, K.H., and Aigaki, T. (1999). The gene search system. A method for efficient detection and rapid molecular identification of genes in Drosophila melanogaster. Genetics *151*, 725–737.

32. Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S.E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S.E., et al. (2002). Systematic determination of patterns of gene expression during Drosophila embryogenesis. Genome Biol. *3*, RESEARCH0088.

33. Tomancak, P., Berman, B.P., Beaton, A., Weiszmann, R., Kwan, E., Hartenstein, V., Celniker, S.E., and Rubin, G.M. (2007). Global analysis of patterns of gene expression during Drosophila embryogenesis. Genome Biol. *8*, R145.

34. Xie, T., and Spradling, A.C. (2000). A niche maintaining germ line stem cells in the Drosophila ovary. Science *290*, 328–330.

35. Yan, D., Neumüller, R.A., Buckner, M., Ayers, K., Li, H., Hu, Y., Yang-Zhou, D., Pan, L., Wang, X., Kelley, C., et al. (2014). A regulatory network of Drosophila germline stem cell self-renewal. Dev. Cell *28*, 459–473.

36. Yi, X., de Vries, H.I., Siudeja, K., Rana, A., Lemstra, W., Brunsting, J.F., Kok, R.M., Smulders, Y.M., Schaefer, M., Dijk, F., et al. (2009). Stwl Modifies Chromatin Compaction and Is Required to Maintain DNA Integrity in the Presence of Perturbed DNA Replication. Mol. Biol. Cell *20*, 983–994.

37. Zhao, T., Gu, T., Rice, H.C., McAdams, K.L., Roark, K.M., Lawson, K., Gauthier, S.A., Reagan, K.L., and Hewes, R.S. (2008). A Drosophila Gain-of-Function Screen for Candidate Genes Involved in Steroid-Dependent Neuroendocrine Cell Remodeling. Genetics *178*, 883–901.

CHAPTER 2

INFERRING STWL FUNCTION FROM RNA-SEQ OF *STWL* DEFICIENT OVARIES

Introduction

Stwl is rapidly evolving in *D. melanogaster*, *D. simulans*, and *D. ananassae* (Choi and Aquadro, 2014; Clark and McKearin, 1996; Flores et al., 2015a; Shukla et al., 2018). Genes involved in gamete production and GSC maintenance have been identified as potential targets of parasitic biological agents, including wolbachia and transposons (Flores et al., 2015b; Ma et al., 2014). Germline components involved in transposon regulation and defense, such as *piwi*, perform this critical function both by directly targeting transposable element (TE) transcripts, and by maintaining these sequences in a repressive chromatin environment (Le Thomas et al., 2013). Previous studies have shown that Stwl is involved in heterochromatin maintenance; this has been supported by the following findings: 1) Stwl mutations are dominant suppressors of position effect variegation (Maines et al., 2007), 2) Stwl co-localizes with HP1a and dense, heterochromatin-like structures at the nucleolus in S2 cells (Yi et al., 2009), 3) Stwl acts as a repressor in *in vitro* experiments (Yi et al., 2009), and 4) Levels of H3K9me3 and H3K27me3 are diminished in *stwl* mutant larvae (Yi et al., 2009). We therefore chose to investigate whether *stwl* is required for TE defense.

Methods

Outcrossing *stwl* mutants into a homozygous background to minimize genetic background effects

stwl alleles $P^{\{w[+mC]=lacW\}}stwl^{j6C3}$ and $P^{y[+mDint2]}w^{[BR.E.BR]=SUPor\text{-}P}$ $stwl^{KG07971}$ were acquired from Bloomington Stock Center [#12087 and #14927, respectively]. Pilot experiments confirmed that each of these alleles is haploinsufficient in the presence of a *stwl* deficiency chromosome (*Df(3L)Exel6122*) [Bloomington Stock Center #7601]. Haploinsufficiency was determined by confirming female sterility and presence of ovaries characteristic of *stwl* mutant phenotypes (atrophy, loss of germline, lack of Orb accumulation in oocytes). Furthermore, qRT-PCR confirmed an increase of select transposon transcripts (Fig. 1) in ovaries extracted from $P^{\{w[+mC]=lacW\}}stwl^{j6C3}$/*Df(3L)Exel6122* and $P^{y[+mDint2]}w^{[BR.E.BR]=SUPor\text{-}P}$ $stwl^{KG07971}$/*Df(3L)Exel6122* individuals.

We also found that these alleles were homozygous lethal, suggesting an accumulation of lethal recessive mutations along the balanced chromosomes containing the *stwl* mutant alleles. In order to control for the impact of genetic background on gene expression, we outcrossed *stwl* 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 7-8 generations (for $stwl^{KG07971}$ and $stwl^{j6C3}$, respectively). Stocks were founded by balancing recombined 3rd chromosomes over *TM6b* from *w; Sp/cyO; TM2/TM6b* stock in single female matings. Presence of *P*-element insertions in *stwl* was confirmed by PCR. Resultant stocks produced viable homozygous males and females; females homozygous for both *stwl* mutant alleles were sterile.

Primers to confirm presence of P-element insertion in *stwl*; primers flank the insertion site, producing a 500 bp fragment:

Forward: CCACGGACATGCTAAGGGTT

Reverse: TCGCTTTTTCTTGCAACGCA

Preparation of tissue for RNA-seq

RNA-seq experiments were conducted only on the outcrossed (8 generations) *stwl*^{j6C3} allele. Homozygous males and females were collected from *y w* ; *stwl*^{j6C3} stock bottles; heterozygous females were collected from crosses between *y w* ; *stwl*^{j6C3} males and *y w* F10 females; *yw*F10 males and females were collected from *y w* F10 stock bottles. 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 (Wong and Schedl, 2006; Zamore and Ma, 2011). 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 were homogenized in 100-600 ul 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 and 2-day old *yw*; *stwl*^{j6C3}/+) was diluted 1:10 prior to RNA extraction, to prevent overloading of columns.

RNA was extracted according to previously published protocols (Rio et al., 2010). 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 (particularly important for removal of repetitive 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 ug 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 A-tailed and adaptor-ligated, followed by size selection, UDG digestion to eliminate the second strand, and PCR amplification. All libraries (21 in total) were sequenced on a single lane of Illumina NextSeq (single-end, 75 bp).

During preliminary analyses of sequencing reads we realized that 5 libraries (all 3 replicates from 2-day old *yw*; $stwl^{j6C3}$/+ females , 2 replicates from 0-day old *yw*; $stwl^{j6C3}/stwl^{j6C3}$ females) 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 for 0-day old *yw*; $stwl^{j6C3}/stwl^{j6C3}$ was increased from 60 ovaries to 90 ovaries per replicate). Stranded cDNA libraries were prepared as previously described and subsequently sequenced on a single lane of Illumina HiSeq 2500 High Output (single-end, 50 bp).

RNA-seq analysis of gonadal tissue

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) (Andrews, 2010; Bolger et al., 2014). We used FastQ Screen to identify non-*Drosophila* contaminants in our libraries (Wingett and Andrews, 2018).

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). Unmapped reads from this alignment were saved and aligned to the unmasked *Drosophila melanogaster* genome (r6.03) using bowtie2 default settings (Langmead and Salzberg, 2012). We used a custom perl script to count the number of reads aligning to repetitive sequences; we utilized HTSeq version 0.6.0 to count the number of reads aligning to exons in the genomic alignment (Anders et al., 2015). 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 effect, we normalized the read counts using EDAseq (Risso et al., 2011). First, we grouped samples according to tissue; group 1 contained WT and *stwl* null testes (6 samples), group 2 contained 2-day-old WT and 2-day-old *stwl* heterozygous ovaries (6 samples), group 3 contained 0- and 2-day-old WT and 0- and 2-day old *stwl* null ovaries (12 samples) and group 4 contained *LacZ*- and *stwl*-dsRNA-treated S2 cells (6 samples). All normalizations and differential expression analyses were performed within these groupings. Within each group, we removed counts for all genes with 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. We then used EDAseq to generate offset values for each element in the count matrix, so that raw counts could be analyzed for differential expression analysis.

We analyzed count data using DESeq2 (Love et al., 2014). We imported raw counts and offsets as described above from genes with mean read count >10 across all samples in a given tissue. For null testis and heterozygous ovary, we estimated differential expression of genes between mutant and wild type samples. For *stwl* null ovarian tissue, all samples were loaded together into a DESeq object factoring in age and genotype (age = 0, 2-days; genotype = +/+, $stwl^{j6C3}/stwl^{j6C3}$). We estimated differential expression of heterozygous and homozygous mutants against wild type. For *stwl* null ovary comparison, age of the samples was taken into account: genes were reported 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.

Subsequent to differential expression analysis, all $\log_2$(fold-change) estimates were transformed using apeGLM shrinkage estimator to reduce variability in LFC values among low-count genes (Zhu et al., 2018). 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 (Yu et al., 2012).

Results (Part 1 of 2)

Stwl deficient ovaries exhibit TE de-repression

A screen for genes whose knockdown (KD) in ovaries leads to mis-expression of TEs found that germline KD of *stwl* resulted in moderate de-repression of a few TE transcripts (*Het-A, blood, and burdock*), as determined by qRT-PCR (Czech et al., 2013). We performed our own qRT-PCR experiments to assay for mis-expression of *Het-A, Copia*, *412*, and *I* element. *Copia* is an LTR retrotransposon, while the other three are non-LTR retrotransposons. *Het-A, Copia* and *I* element are germline-restricted, while *412* is expressed in both germline and ovarian follicle cells (Li et al., 2009; Malone et al., 2009). We tested for misexpression of these TEs in ovaries dissected from 2-day old *stwl* null ($stwl^{-/-}$), heterozygous ($stwl^{+/-}$) and wild-type ($stwl^{+/+}$) flies. Four different genotypes of null flies were made with different alleles hemizygous to a deficiency, in order to determine which allele presents the strongest TE misexpression phenotype (Fig. 1). Among the four null alleles, only the 3 *P*-insertion alleles cause an increase in *Het-A* transcript abundance (one-tailed P>0.05). Among them, the $stwl^{j6c3}$ allele has the strongest *HetA* de-repression phenotype, with a 5-fold increase in abundance relative to wild-type. We found that *Het-A* expression is not altered in *stwl* heterozygous ovaries .

The other TEs are all consistently de-repressed (one-tailed P>0.05) across all genotypes (including the heterozygote) relative to the wild-type control. *412* de-repression in heterozygotes is 2-fold relative to wild-type, while abundance across all null genotypes is 3-4 fold higher relative to wild-type. *I* element and *copia* are much more strongly affected: null ovaries exhibit a ~37-67-fold increase in *I* element abundance relative to wild-type, while de-repression in heterozygous ovaries is more modest (3-fold over wild-type). *Copia* is most dramatically affected by *stwl* loss: transcript abundance increases to 10-fold over wild-type in the

heterozygous background and ranges from 85- to 250-fold increase across the homozygous null genotypes. The *stwl*^{j6c3} null causes the greatest increase in transcript abundance.

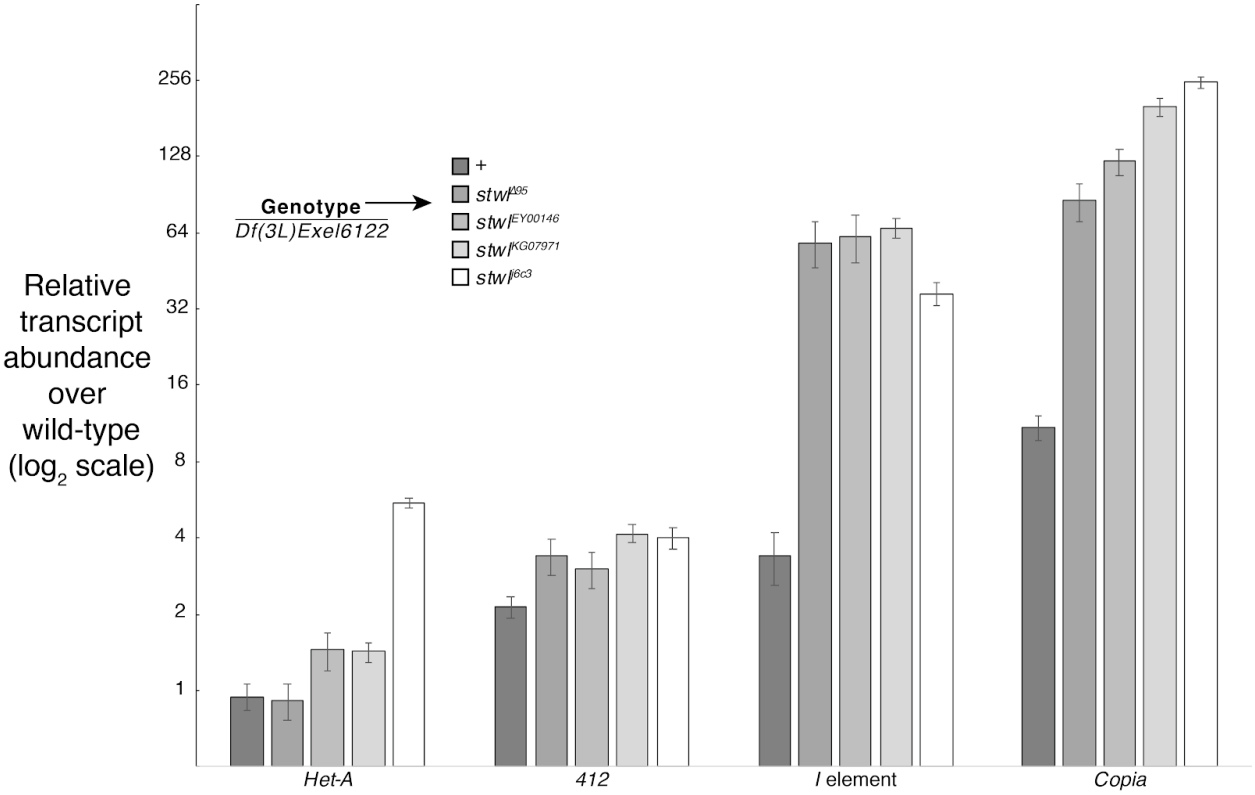


FIGURE 1. *stwl* mutant ovaries exhibit increased TE transcript abundance relative to WT
D. melanogaster ovaries that are hemizygous ('+' genotype) or homozygous null for *stwl* have an increase in TE expression relative to WT (*stwl*$^{+/+}$). Ovaries were extracted from females with the indicated *stwl* allele over the Df(3L)Exel6122a chromosome containing a deletion that spans the *stwl* gene locus. RNA was extracted once from 20-30 ovary pairs and quantified via qRT-PCR in technical triplicate. Transcript abundance of each technical replicate was normalized to average levels of *Rpl32* transcript in the source biological sample. For each TE transcript, we report normalized transcript abundance relative to the average normalized transcript abundance of the wild-type sample (*stwl*$^{+/+}$). Error bars represent +/- sum of means squared of (SEM across the 3 technical replicates for the indicated transcript in the relevant tissue) + (SEM across the 3 technical replicates for *Rpl32* in the relevant tissue), scaled to WT.

A common concern in genetic experiments is controlling for genetic background. In the qPCR assays above, we reduced background variation by assaying different *stwl*$^{-}$ alleles, each crossed to the same *stwl*$^{-}$ deficiency allele. We also found that in every case where the *stwl*$^{-/-}$

genotype exhibits significant TE de-repression relative to wild-type, it is also de-repressed relative to the $stwl^{+/-}$ control (one-tailed $P<0.05$).

Assaying transcript abundance in *stwl* mutant ovaries also presents a concern having to do with the phenotypic consequences of *stwl* loss. *stwl* mutant ovaries are largely agametic as a consequence of GSC loss and defects in oocyte determination (Akiyama, 2002; Clark and McKearin, 1996; Maines et al., 2007). We confirmed that these defects are apparent in all $stwl^{-/-}$ genotypes assayed 2+ days post-eclosion (Fig. 2a, 2b). Nurse cells in *D. melanogaster* ovaries are polyploid and produce large quantities of mRNA that are maternally inherited by the developing oocyte. Hence, agametic ovaries likely have drastically different transcript profiles relative to wild-type, independently of any *stwl*-dependent effects. Differential expression between agametic mutant and wild-type ovaries would hence reflect the cellular makeup of the ovaries rather than changes in transcript abundance specifically due to Stwl.

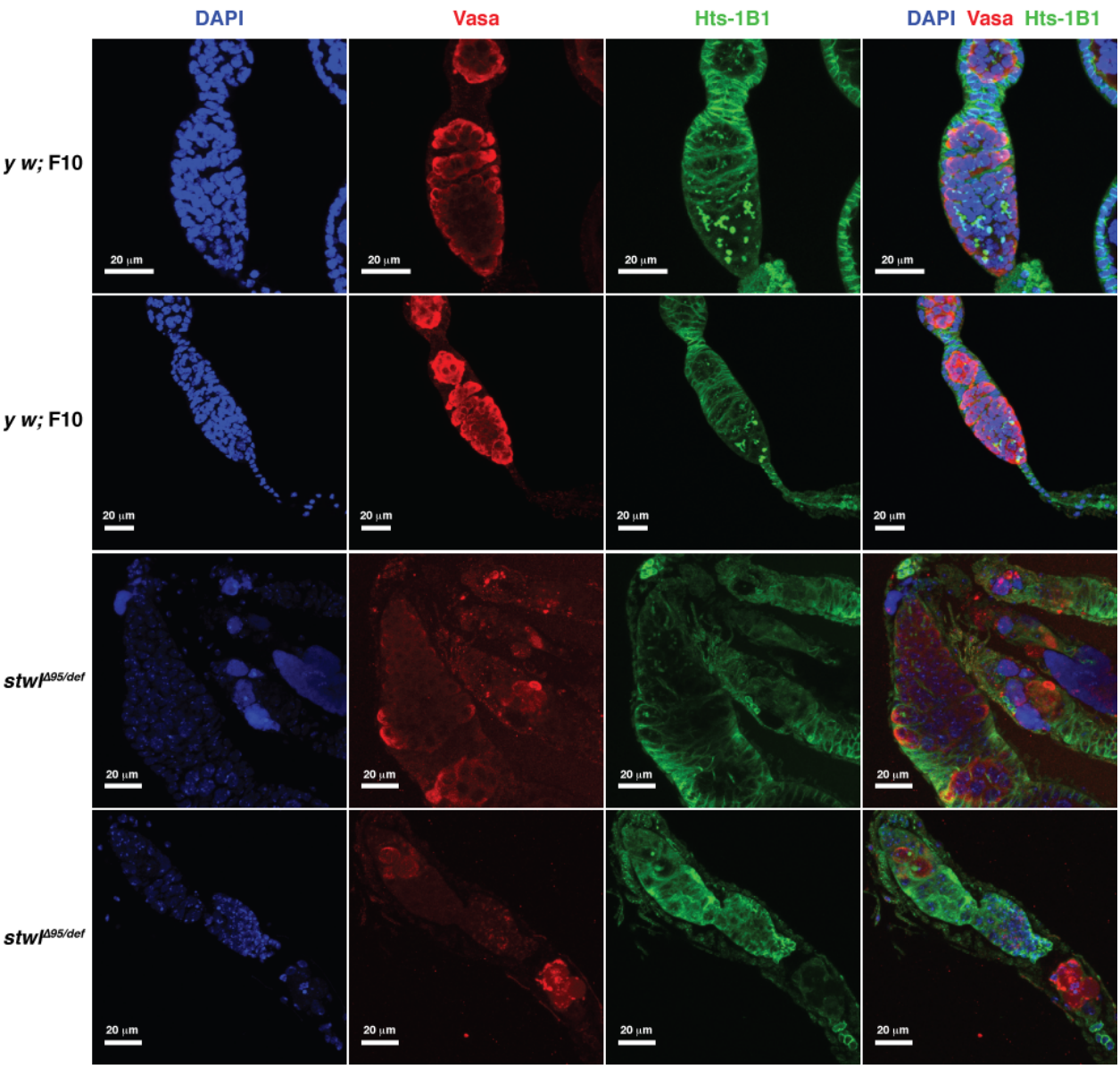


FIGURE 2a. *stwl* mutant ovaries have defective germline cysts

D. melanogaster ovaries were dissected from females of the indicated genotype 3-6 days post-eclosion. *stwl*$^{-/-}$ ovaries typically lack germ cells or contain severely disordered germline cysts. α-Vasa labels germ cells, α-Hts-1B1 labels branched fusomes or spectrosomes, as well as follicle cell membranes. Wild-type ovaries contain self-renewing GSCs 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. Two examples from each genotype are shown.

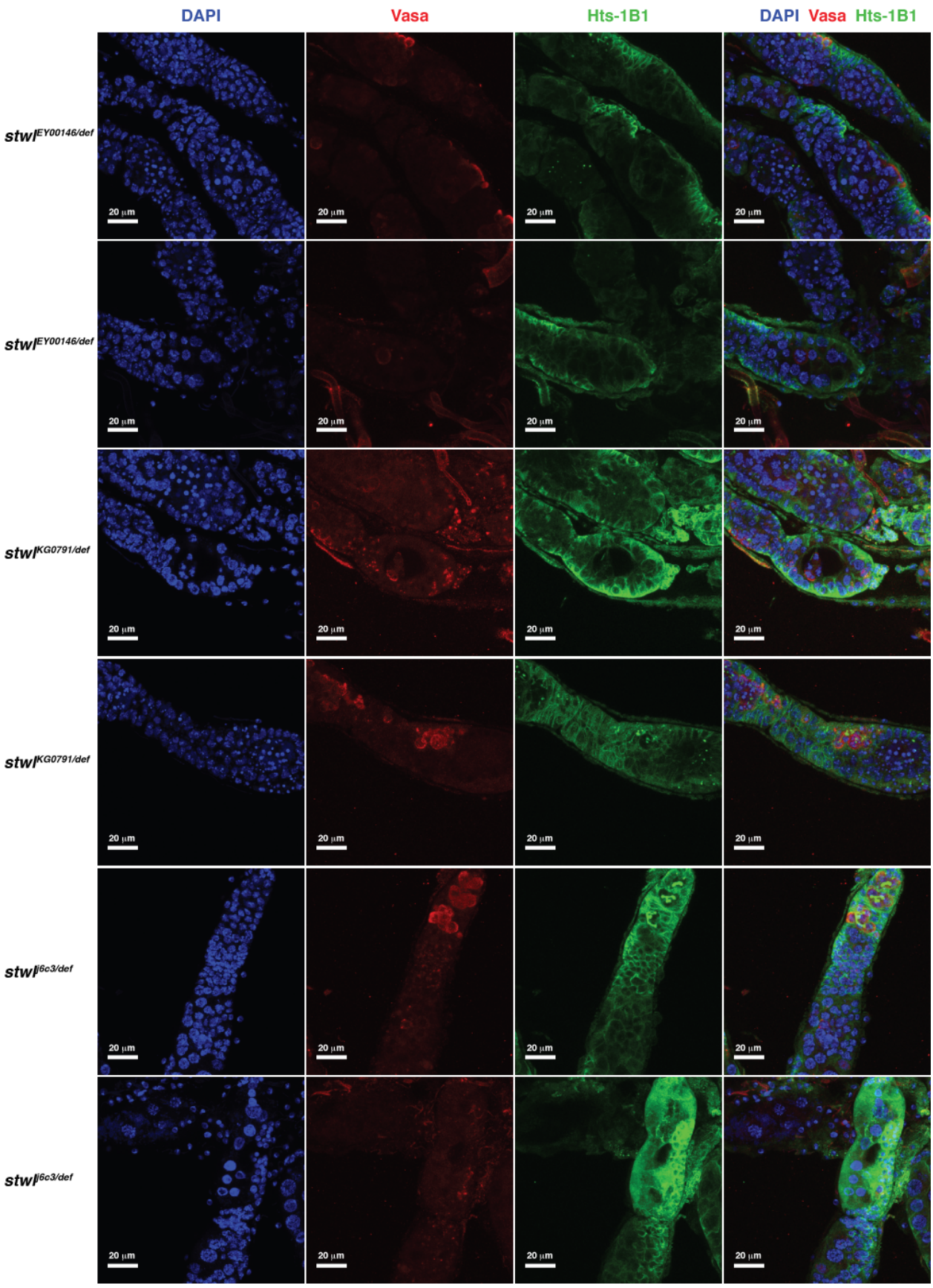

DAPI
Vasa
Hts-1B1
DAPI Vasa Hts-1B1
stwl EY00146/def
stwl EY00146/def
stwl KG0791/def
stwl KG0791/def
stwl j6c3/def
stwl j6c3/def
20 μm

FIGURE 2b. (page above) *stwl* mutant ovaries have defective germline cysts
D. melanogaster ovaries were dissected from females of the indicated genotype 3-6 days post-eclosion. *stwl*$^{-/-}$ ovaries typically lack germ cells or contain severely disordered germline cysts. α-Vasa labels germ cells, α-Hts-1B1 labels branched fusomes or spectrosomes, as well as follicle cell membranes. Wild-type ovaries contain self-renewing GSCs 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. Two examples from each genotype are shown.

Previous studies conducted on agametic ovaries have dealt with this concern in a number of ways. Maines et al. assayed transcript abundance in *bam*$^{-/-}$ *stwl*$^{-/-}$ double-mutant ovaries relative to *bam*$^{-/-}$ *stwl*$^{+/+}$ ovaries; they found that double mutant ovaries form germ cell cysts which are absent in *bam* mutants, therefore concluding that *stwl* is epistatic to *bam* (Maines et al., 2007). They reasoned that "both of these genetic backgrounds provided a nearly homogenous population of cell types, because *bam* mutant cells failed to differentiate into cystoblasts and *stwl bam* germ cells arrested as partially formed cysts". Rangan et al. utilized a similar approach in assaying small RNA abundance in *dSETDB1* mutant ovaries by comparing these to *bam* mutant ovaries which display a similar morphological phenotype (Rangan et al., 2011). Following these approaches, we could have either crossed a *stwl*$^{-/-}$ allele into an agametic mutant background and compared to the *stwl*$^{+/+}$ in the same background or compared *stwl*$^{-/-}$ to agametic mutants of another gene. We chose not to pursue these approaches because of the concern that *stwl* and whichever gene we chose to compare it to could have shared downstream targets that would not be identified in such an assay.

We instead chose the approach utilized by Sun and Cline, Shapiro-Kulnane et al. and others, which is simply to assay transcripts derived from extremely young ovaries (Shapiro-Kulnane et al., 2015; Sun and Cline, 2009). 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. We examined *stwl* mutant and wild-type ovaries from newly-eclosed (<12-hours old) individuals and compared them to *stwl* mutant and wild-type ovaries from older (2-days old to 10-days old) individuals to identify major morphological differences and confirm that the younger ovaries are appropriate for comparison (Fig. 2-6).

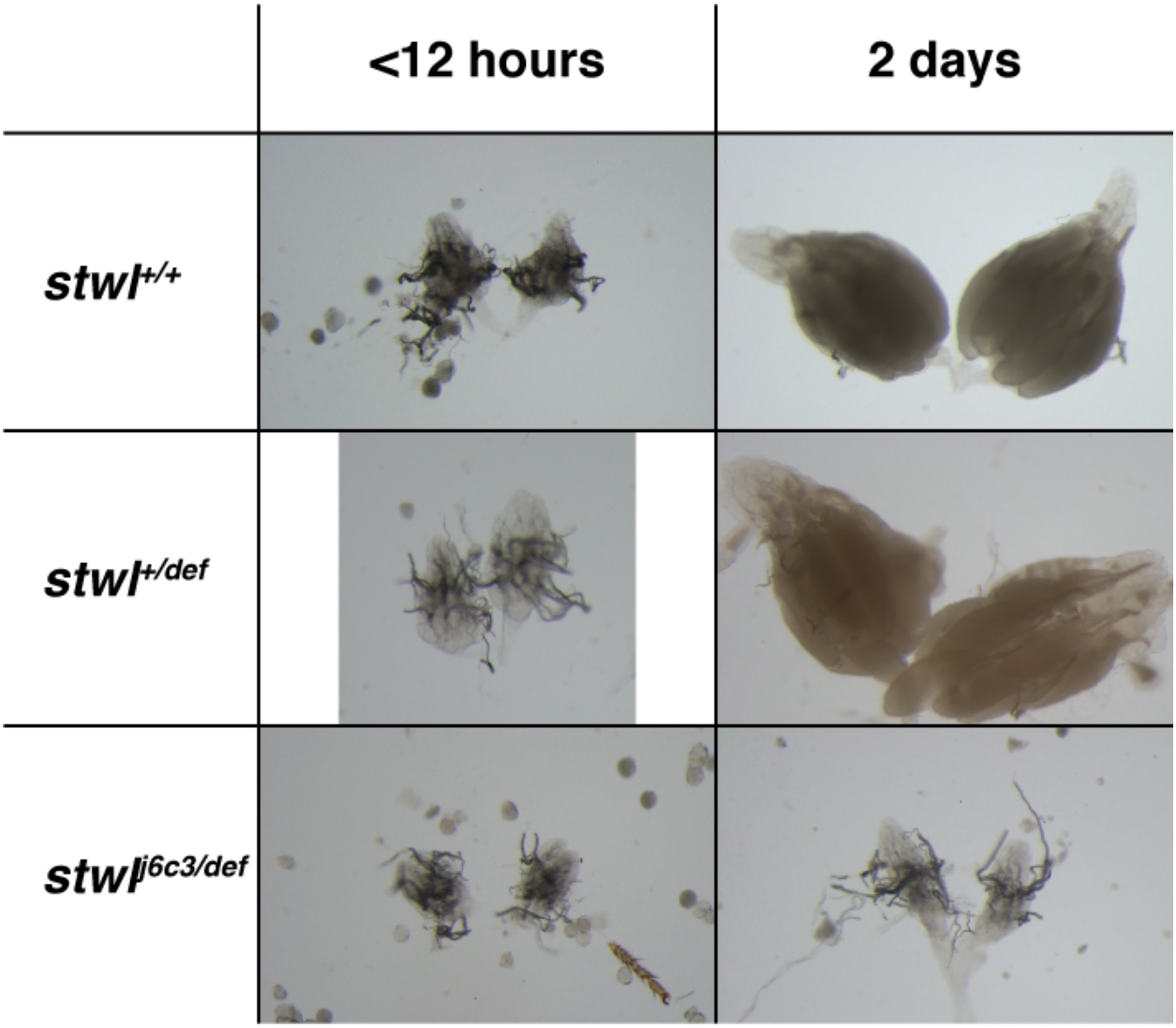


FIGURE 3. *stwl* null and WT ovaries from newly-eclosed individuals are more similar in size relative to ovaries from older flies

Ovaries were dissected from newly-eclosed and 2-day old females of the indicated genotypes. *stwl* hemizygotes and heterozygotes (not shown) do not exhibit obvious morphological changes compared to wild-type ovaries. *stwl* null ovaries are rudimentary, but more closely resemble wild-type ovaries from newly-eclosed individuals.

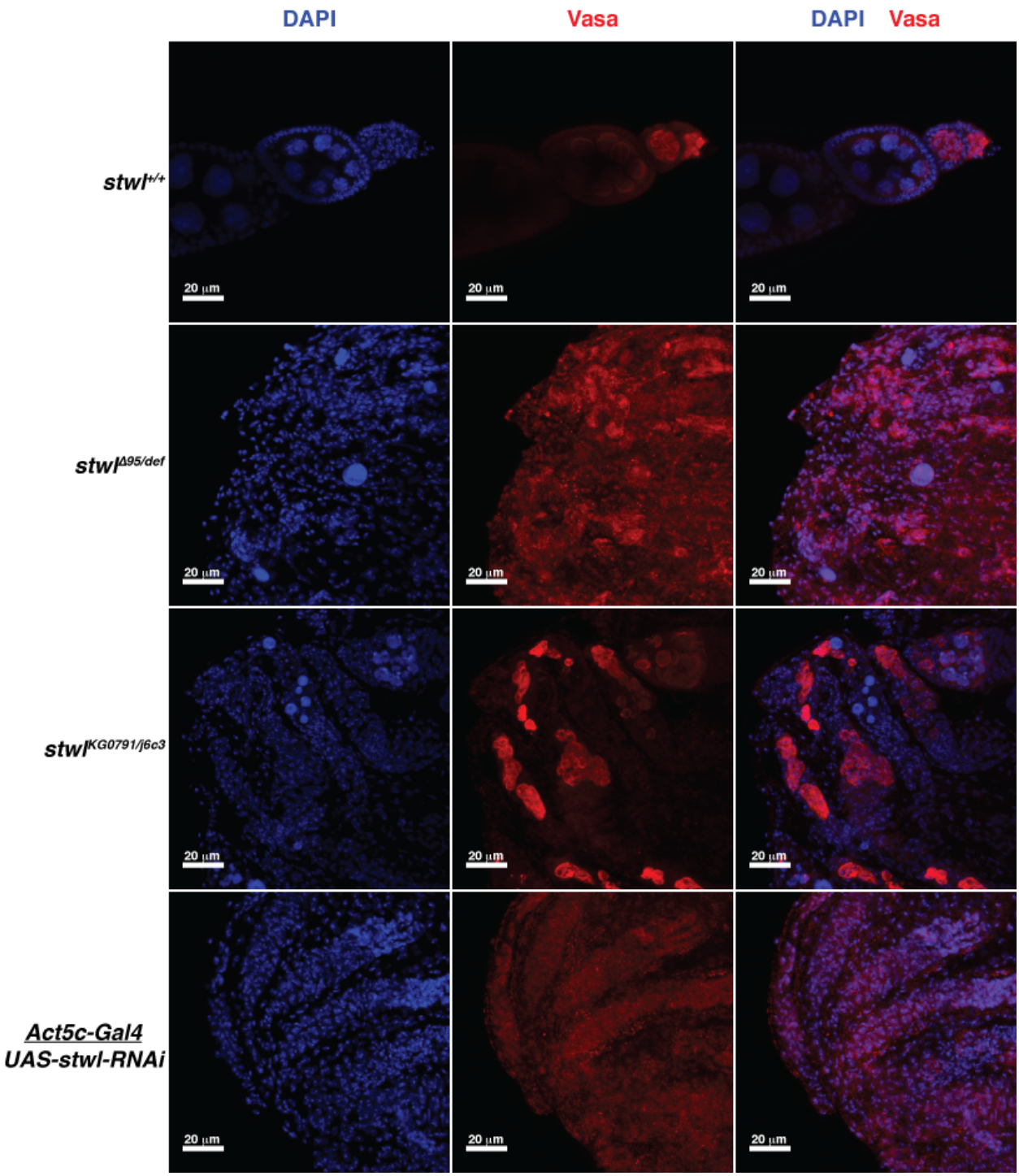
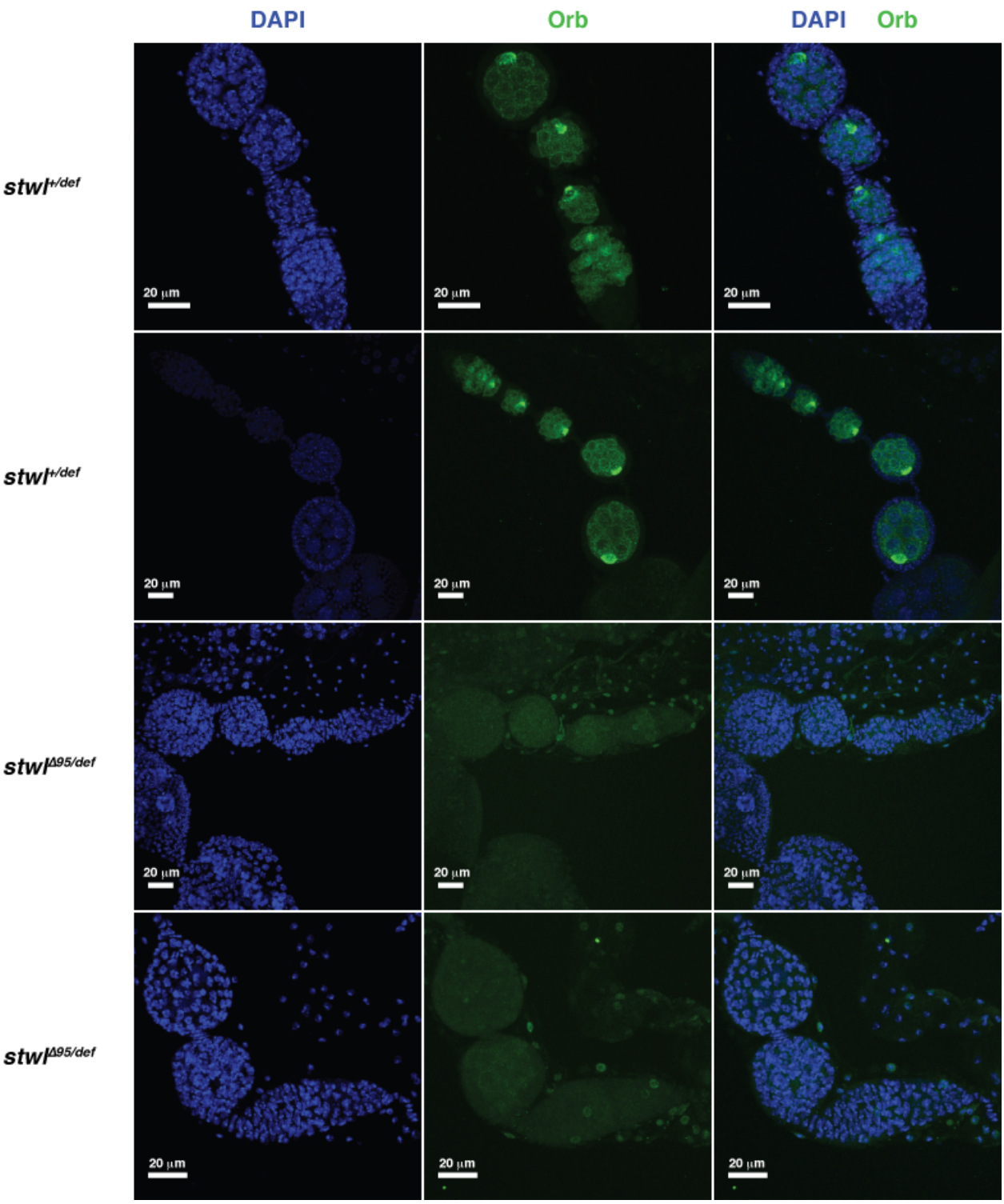


FIGURE 4. *stwl* deficient ovaries fail to retain germline cells as they age

D. melanogaster ovaries were dissected from females of the indicated genotypes 10 days post-eclosion. α-Vasa labels germ cells, which are typically not retained in older mutant ovaries. All images are maximum-intensity projections from a z-series representing a depth of 10 microns.

We confirmed that *stwl* mutant ovaries fail to retain germ cells as they age: 10-day old *stwl*$^{-/-}$ ovaries labelled with **α**-Vasa antibody are often completely devoid of germ cells, whereas 2-day old ovaries retain germ cells despite other demonstrable defects, including aberrant germline cysts and a lack of Orb localization to the developing oocyte (Fig. 2-5). These defects are dramatically lessened in ovaries dissected from newly-eclosed individuals (Fig. 3, 6, 7). In terms of gross morphology, the newly-eclosed ovaries have a mostly intact germ-cell population, and cyst formation is similar to wild-type (Fig. 6-7). What is especially notable is that newly-eclosed wild-type ovaries only produce egg chambers up to stage 7 or 8, while *stwl* mutant ovaries maintain egg chambers up to about stage 6 or 7. Thus, we concluded that careful dissection of newly-eclosed ovaries helps to ameliorate the downstream effects of *stwl* loss which would hinder our understanding of *stwl*'s molecular function.



<u>FIGURE 5A</u>. *stwl* mutant ovaries fail to undergo oocyte determination

D. melanogaster ovaries were dissected from females of the indicated genotypes 3-5 days post-eclosion. α-Orb labels germ cells; Orb localization increases dramatically before and during oocyte determination, so that the differentiation oocyte is clearly labelled with Orb in wild-type

cells. *stwl*$^{-/-}$ ovaries do not contain oocyte with obvious Orb accumulation. All images are maximum-intensity projections from a z-series representing a depth of 10 microns.

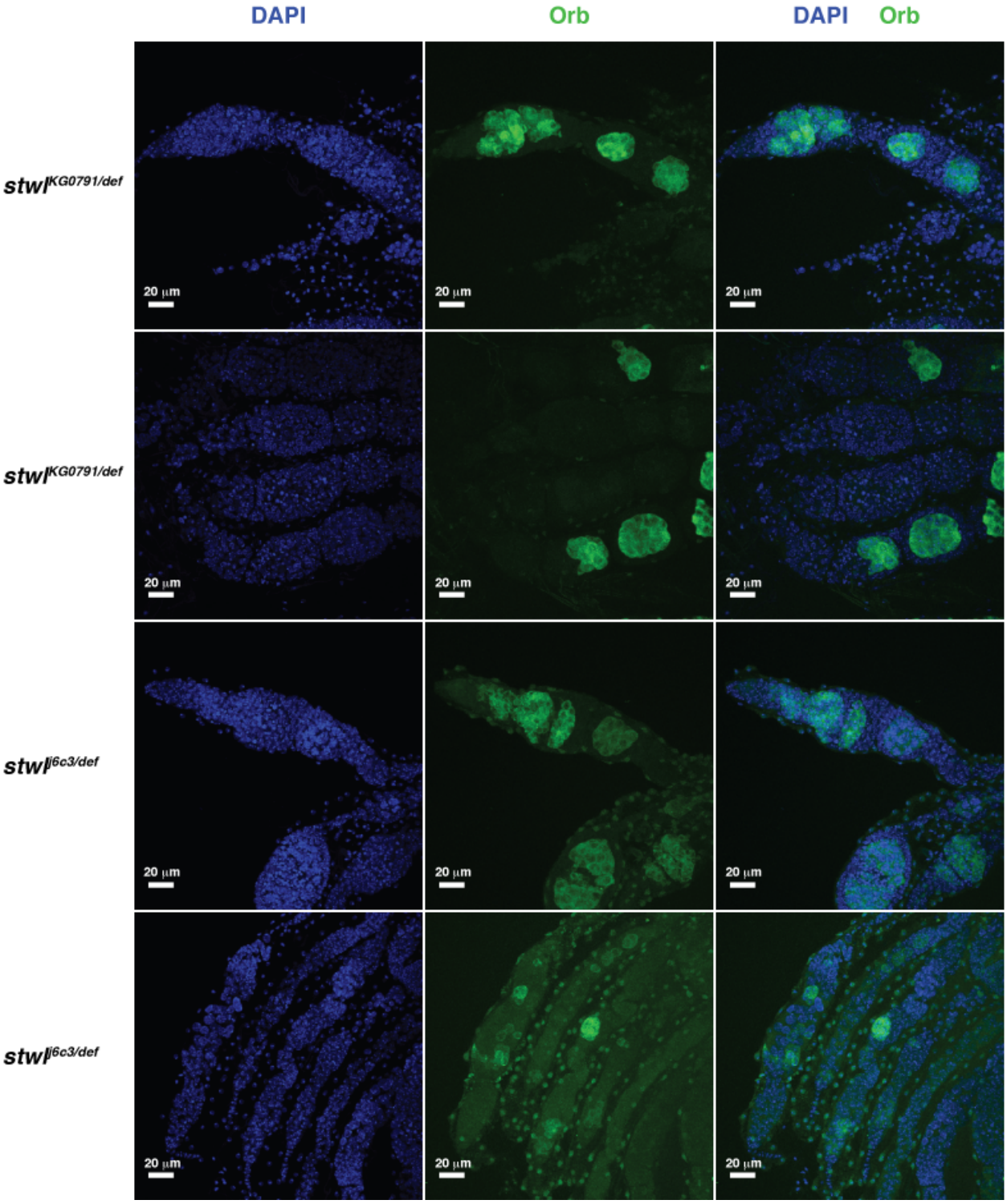


FIGURE 5B. *stwl* mutant ovaries fail to undergo oocyte determination

D. melanogaster ovaries were dissected from females of the indicated genotypes 3-5 days post-eclosion. α-Orb labels germ cells; Orb localization increases dramatically before and during oocyte determination, so that the differentiation oocyte is clearly labelled with Orb in wild-type cells. *stwl*$^{-/-}$ ovaries do not contain oocyte with obvious Orb accumulation. All images are maximum-intensity projections from a z-series representing a depth of 10 microns.

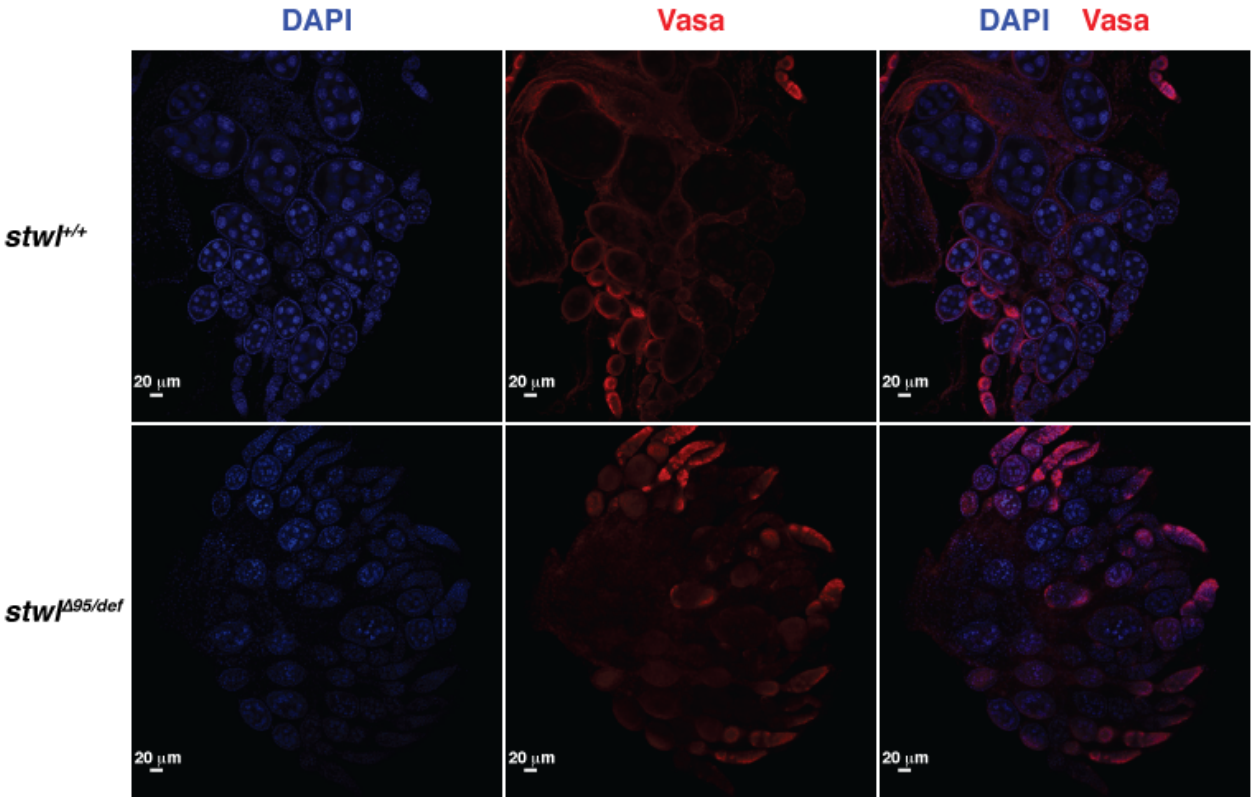


FIGURE 6. Newly-eclosed ovaries from *stwl* mutants resemble WT ovaries
D. melanogaster ovaries were dissected from females <12-hours post-eclosion. α-Vasa labels germ cells, which are typically not retained in older mutant ovaries. Wild-type ovaries only produce egg chambers up to stage 7 or 8, while *stwl* mutant ovaries maintain egg chambers up to about stage 6 or 7. Each image is a single confocal slice.

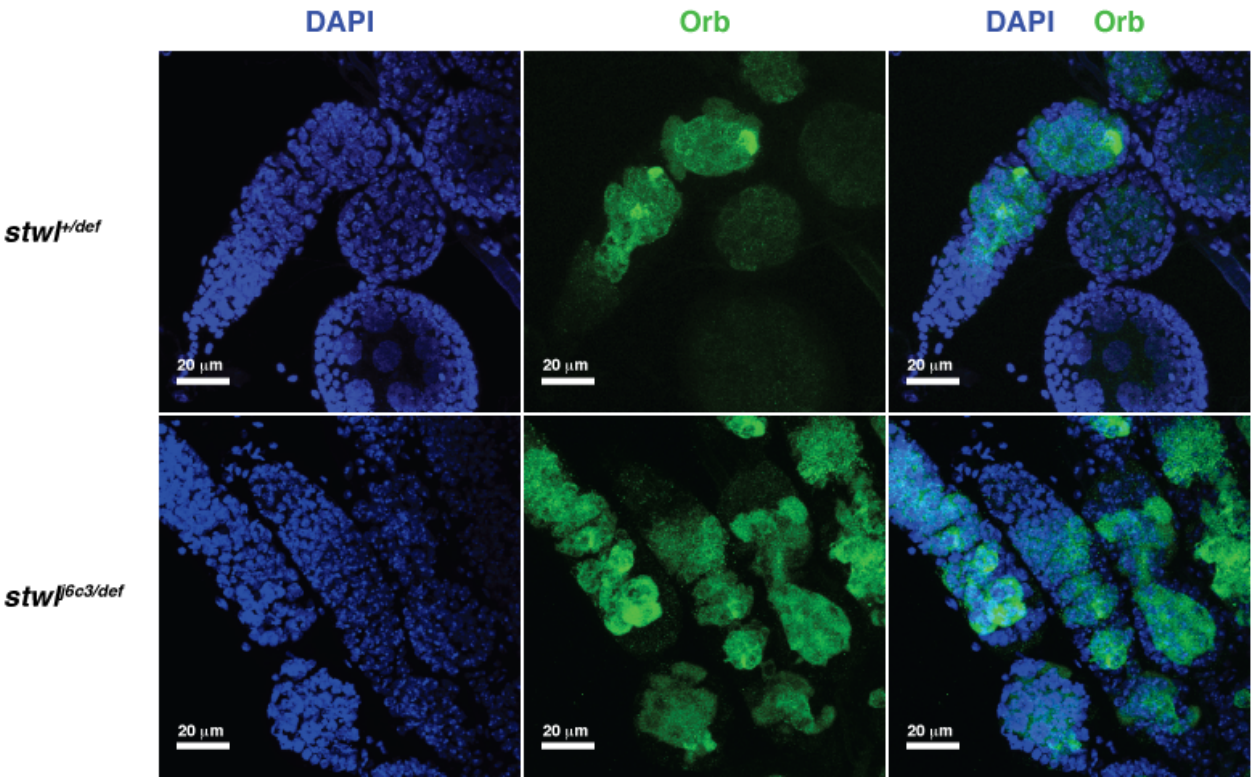


FIGURE 7. Newly-eclosed *stwl* mutant ovaries fail to undergo oocyte determination
D. melanogaster ovaries were dissected from females of the indicated genotype <12-hours post-eclosion. α-Orb labels germ cells; Orb localization increases dramatically before and during oocyte determination, so that the differentiating oocyte is clearly labelled with Orb in wild-type cells. $stwl^{-/-}$ ovaries do not contain oocytes with obvious Orb accumulation. Images of $stwl^{+/-}$ hemizygous ovaries (top row) and $stwl^{-/-}$ homozygous null ovaries (bottom row) are maximum-intensity projections from a z-series representing a depth of 30 microns and 20 microns, respectively.

Since the *stwl* phenotype is much less severe in ovaries from newly-eclosed individuals, we investigated whether there are still obvious defects present. We found an accumulation of Orb in developing wild type oocytes but not in newly-eclosed *stwl* mutants (Fig. 7). This is consistent with the effect observed in older *stwl* mutant ovaries and demonstrates that *stwl* is required for oocyte determination very early in adulthood (or before). We also tested for TE mis-expression. We assayed newly-eclosed ovaries from $stwl^{j6c3}/stwl^{Deficiency}$ and $stwl^{j6c3}/+$, as in the previously described assay. We found that the fold-increase of *Het-A* in homozygous nulls

relative to heterozygotes is similar in newly-eclosed and 2-day-old ovaries (5-fold and 6-fold increase, respectively) (Fig. 8). *Copia* transcript is also de-repressed in newly-eclosed *stwl* null ovaries (8-fold increase over heterozygote), though this de-repression phenotype is not as large as the one observed in 2-day-old ovaries (23-fold increase over heterozygote).

RNA-seq of *stwl* mutant ovaries requires careful controls

In order to assess genomic consequences of Stwl loss *in vivo*, we performed RNA-seq on ovaries dissected from newly-eclosed and 2-day old wild-type (*y w*; F10) and *stwl* null (*y w; stwl*$^{j6c3/j6c3}$) individuals, as shown in Fig. 3. We also dissected ovaries from 2-day old heterozygous individuals (*y w; stwl*$^{+/j6c3}$). The goal of this experiment was to identify and classify genes and TEs which are *consistently* differentially expressed in *stwl* mutants. Rather than identify genes which are differentially expressed only among newly-eclosed or two-day old individuals, we incorporated all 4 sample types (newly-eclosed wild-type, newly-eclosed mutant, two-day old wild-type, two-day old mutant) into a generalized model using DESeq2 (Love et al., 2014). The model was constructed as follows: design ~ age + genotype. Thus, 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.

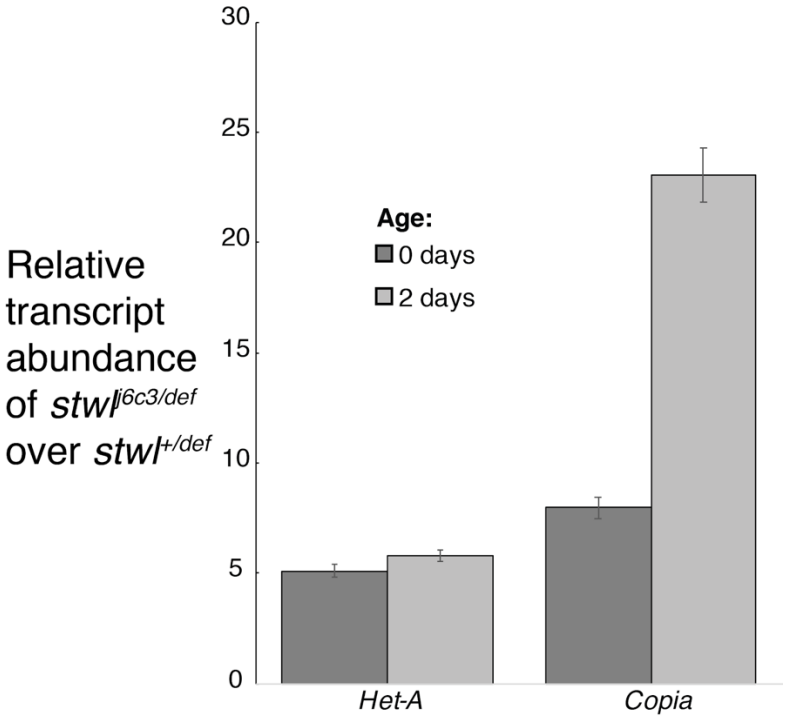


FIGURE 8. TE derepression phenotype is present in *stwl* mutant ovaries from newly-eclosed individuals

D. melanogaster ovaries that are homozygous null for *stwl* (*stwl*[*j6c3*]/Df(3L)*stwl*[-]) have an increase in TE expression relative to hemizygotes (*stwl*[+]/Df(3L)*stwl*[-]). Ovaries were extracted from newly-eclosed females (<12-hours) having either 0 ($stwl^{j6c3/def}$) or 1 ($stwl^{+/def}$) functional *stwl* alleles over a deficiency chromosome containing a deletion that spans the *stwl* gene locus. RNA was extracted from three sets of 20-30 ovary pairs and quantified via qRT-PCR in triplicate. Transcript abundance of each technical replicate was normalized to average levels of *Rpl32* transcript in the source biological sample. For each TE transcript, we report normalized transcript abundance across 3 biological replicates relative to the average normalized transcript abundance of the hemizygous sample ($stwl^{+/def}$). Error bars represent +/- sum of means squared of (SEM across the 3 biological replicates for the indicated transcript in the relevant tissue) + (SEM across the 3 biological replicates for *Rpl32* in the relevant tissue), scaled to the hemizygote. Data for the 2-day old samples are from the same experiment described in Figure 1, scaled to hemizygotes instead of WT.

All reads were trimmed before alignment to an index of consensus repeat sequences and to release 6 of the unmasked *D. melanogaster* genome (see Methods). *stwl* null ovary libraries contain a higher proportion of reads aligning to consensus repeat sequences than their wild-type counterparts (Table 1). In order to account for potential batch effects and GC-content bias during library preparation, we normalized gene-level read counts for GC-content within each sample (Fig. 9). We then normalized gene-level read counts between samples to account for differences in sequencing depth (Fig. 9). Sample-to-sample distances for the resultant count matrices confirm that the biological replicates for each sample type cluster together (Fig. 10). Principal Component Analysis of the count data demonstrated that the samples are primarily stratified according to ovary maturity (Fig. 11). Principal Component 1 (PC1) accounts for 67% of the variance in the count matrix, which separates mature ovaries (2-day old $stwl^{+/-}$ heterozygotes and 2-day old $stwl^{+/+}$ wild-type) from immature ovaries (2-day old $stwl^{-/-}$ null, 0-day $stwl^{-/-}$ null, and 0-day old $stwl^{+/+}$ 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.

Sample_ID	Sequencing Batch	Tissue	Genotype	Age (Days)	Total Reads	Post-Trimming	% Trimmed	Aligned to Repeat (1x)	Aligned to Repeat (>1x)	% of Trimmed Reads Aligned to Repeats	Aligned to Reference Genome (1x)	Aligned to Reference Genome (>1x)	% of Trimmed Reads Aligned to Reference Genome	Unmapped	% of Unmapped Trimmed Reads
Null_0day_A	1	Ovary	Null	0	13,409,128	12,186,382	9.1%	197,943	29,076	1.9%	10,658,032	289,567	89.8%	1,011,764	8.3%
Null_0day_B	2	Ovary	Null	0	59,895,069	57,326,635	4.3%	1,176,809	297,562	2.6%	50,314,949	1,268,553	90.0%	4,268,762	7.4%
Null_0day_C	2	Ovary	Null	0	55,382,560	53,066,432	4.2%	1,061,700	340,555	2.6%	45,498,476	1,170,543	87.9%	4,995,158	9.4%
WT_0day_A	1	Ovary	WT	0	17,897,414	16,157,302	9.7%	102,221	24,692	0.8%	14,148,775	314,473	89.5%	1,567,141	9.7%
WT_0day_B	1	Ovary	WT	0	15,071,584	13,616,399	9.7%	86,629	12,226	0.7%	11,780,224	288,393	88.6%	1,448,927	10.6%
WT_0day_C	1	Ovary	WT	0	21,607,696	19,208,795	11.1%	218,497	73,287	1.5%	16,380,526	310,202	86.9%	2,226,283	11.6%
Null_2day_A	1	Ovary	Null	2	25,542,667	22,246,847	12.9%	316,553	58,403	1.7%	19,620,971	512,632	90.5%	1,738,288	7.8%
Null_2day_B	1	Ovary	Null	2	17,199,769	15,515,711	9.8%	225,996	45,111	1.7%	13,720,852	343,030	90.6%	1,180,722	7.6%
Null_2day_C	1	Ovary	Null	2	17,820,450	16,022,868	10.1%	235,408	41,275	1.7%	14,069,580	356,911	90.0%	1,319,694	8.2%
WT_2day_A	1	Ovary	WT	2	24,454,495	22,040,597	9.9%	229,319	24,709	1.2%	19,638,436	419,113	91.0%	1,729,020	7.8%
WT_2day_B	1	Ovary	WT	2	25,968,008	23,328,551	10.2%	175,759	25,672	0.9%	20,688,326	459,931	90.7%	1,978,863	8.5%
WT_2day_C	1	Ovary	WT	2	21,386,122	19,185,771	10.3%	138,036	19,012	0.8%	17,105,029	387,300	91.2%	1,536,394	8.0%
Het_2day_A	2	Ovary	Het	2	43,511,252	41,580,830	4.4%	286,689	139,977	1.0%	37,500,015	805,650	92.1%	2,848,499	6.9%
Het_2day_B	2	Ovary	Het	2	58,084,038	55,561,079	4.3%	345,869	168,191	0.9%	50,113,268	1,073,489	92.1%	3,860,262	6.9%
Het_2day_C	2	Ovary	Het	2	69,761,405	66,726,358	4.4%	463,277	186,962	1.0%	60,392,154	1,242,388	92.4%	4,441,577	6.7%
WT_A	1	Testis	WT	2	14,856,393	13,292,674	10.5%	114,839	19,656	1.0%	11,626,546	502,167	91.2%	1,029,466	7.7%
WT_B	1	Testis	WT	2	16,401,422	14,552,061	11.3%	117,447	43,307	1.1%	12,773,382	534,632	91.5%	1,083,293	7.4%
WT_C	1	Testis	WT	2	33,320,610	30,081,604	9.7%	192,824	44,559	0.8%	26,426,034	1,172,265	91.7%	2,245,922	7.5%
Null_A	1	Testis	Null	2	17,394,114	15,573,807	10.5%	130,879	24,026	1.0%	13,670,526	513,745	91.1%	1,234,631	7.9%
Null_B	1	Testis	Null	2	15,560,458	13,846,911	11.0%	93,435	20,947	0.8%	12,196,414	494,262	91.6%	1,041,853	7.5%
Null_C	1	Testis	Null	2	18,942,597	17,035,877	10.1%	112,848	29,678	0.8%	15,022,663	608,801	91.8%	1,261,887	7.4%

Table 1. Summary of Bowtie2 Alignments

Raw reads were trimmed for quality and aligned to the consensus sequences of 313 repetitive DNA elements. Unmapped reads leftover from this initial alignment were then mapped to the unmasked *D. melanogaster* genome (release 6).

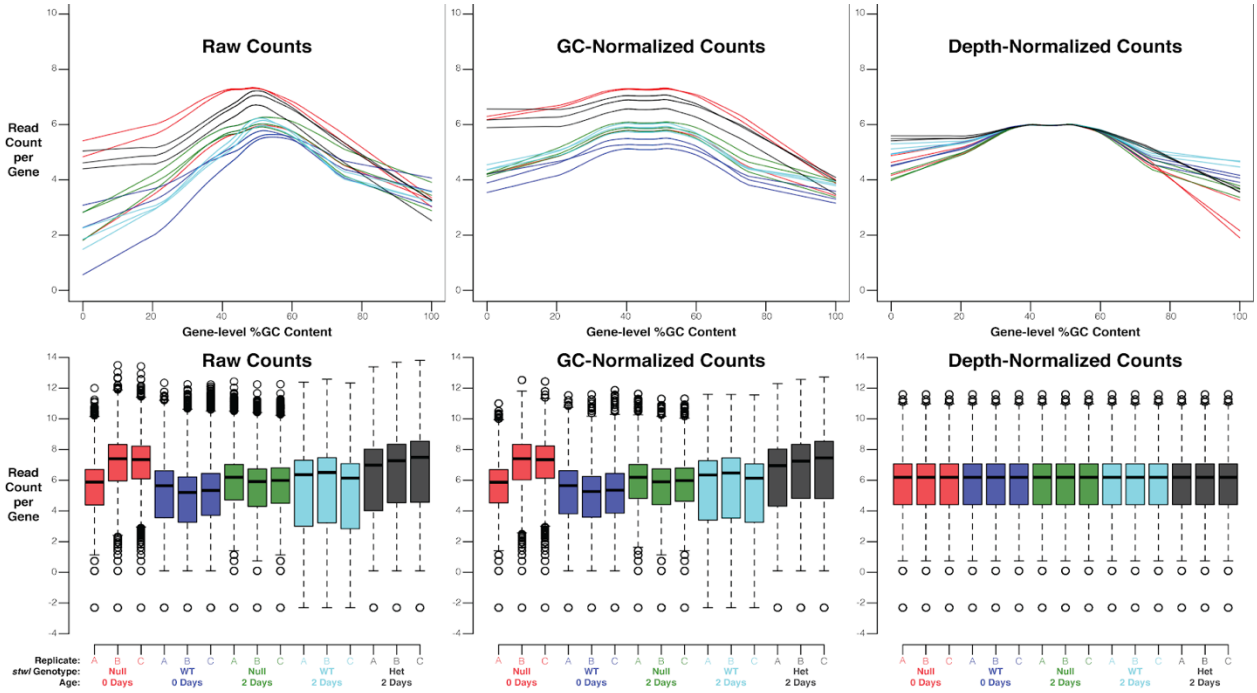
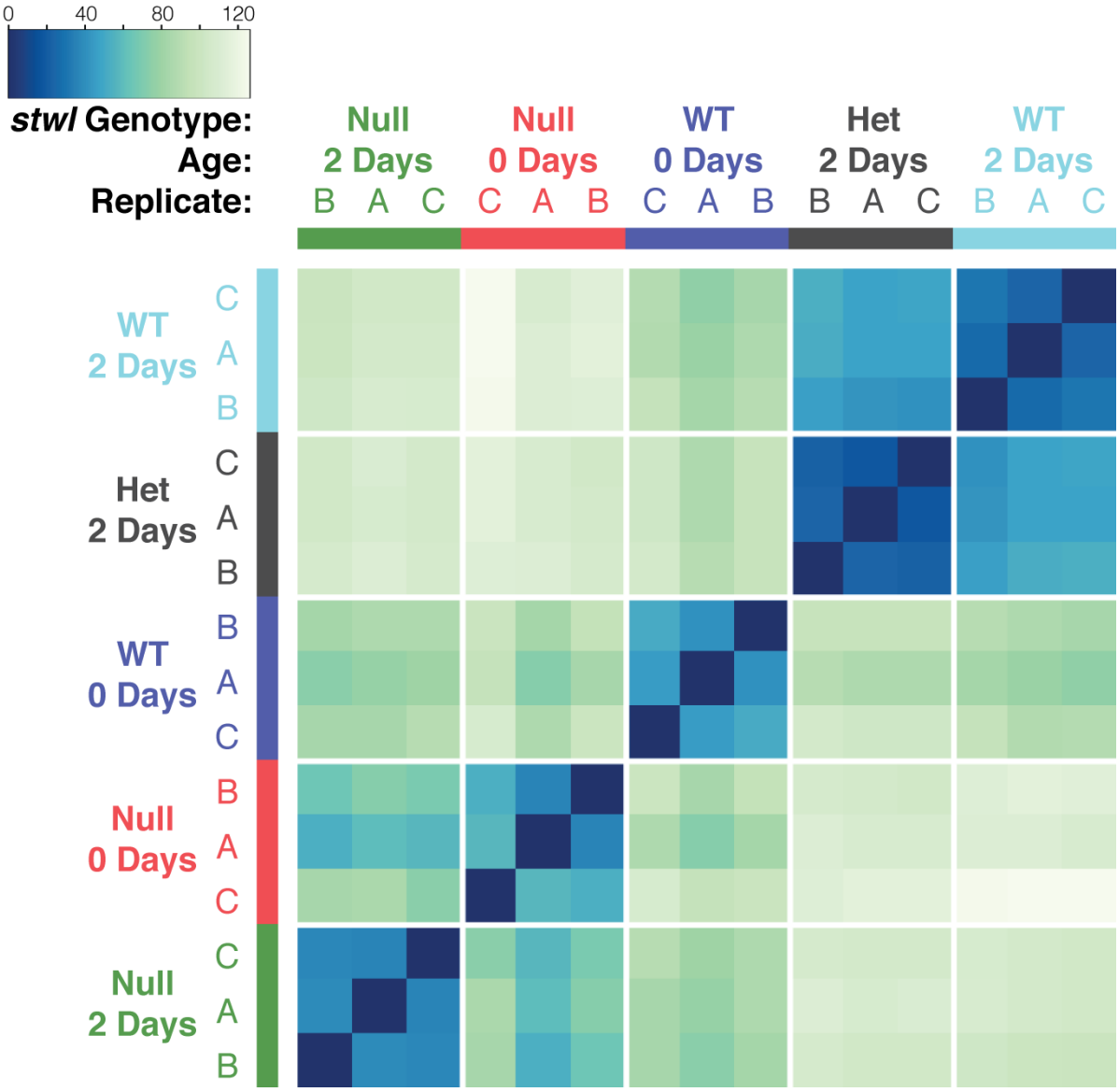


Figure 9. Within- and between-sample normalizations account for biases in library production

To account for potential biases in amplification of transcripts with variable GC content, full-quantile normalization was performed against %GC for all expressed genes. These GC-normalized counts were then normalized between samples to account for sequencing depth. Colors at the bottom serve as a legend that applies to both rows.

Top: Plot of per-gene read count against GC-content. The skew at the edges in the plots of the raw counts indicates a bias against counts appearing in low and high %GC transcripts. These plots are smoothened after GC-normalization and shifted along the y-axis after depth-normalization.

Bottom: Bar plots showing the distribution of gene counts within each sample library. Note that the distributions of read counts are identical between samples after normalization for sequencing depth. All y-axes are log-normalized.



<u>Figure 10</u>. Sample-to-sample distance matrix

Read counts were regularized log transformed using 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). Samples within the same group (identical age and genotype) occur together and form blue clusters.

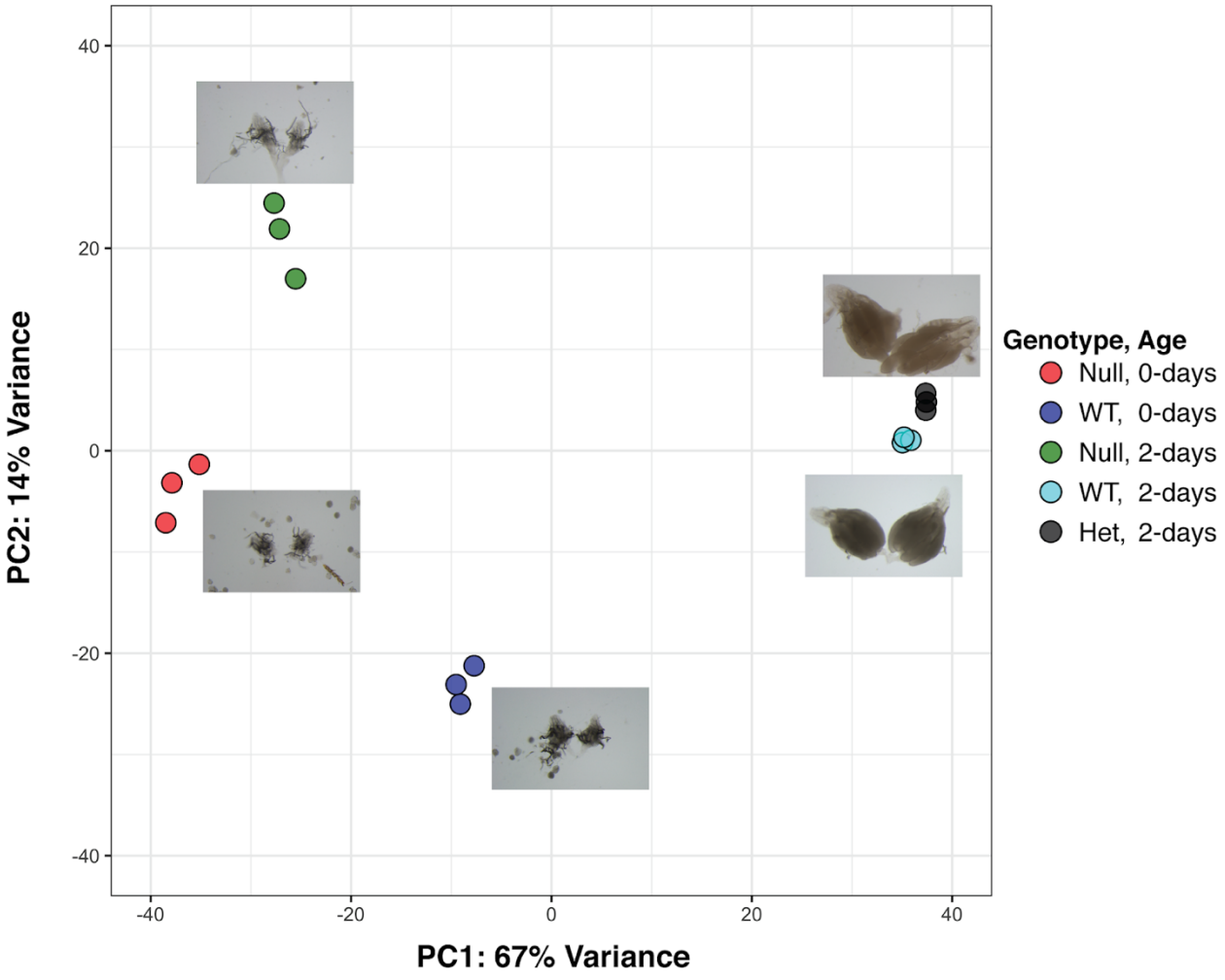


Figure 11. Principal component analysis of count matrices

Read counts were regularized log transformed using DESeq. PCA was performed on the 500 most variable genes in the count matrix. Samples within the same group (identical age and genotype) cluster together, indicating minimal contribution of batch effect to the dataset.

Differential Expression analysis was performed using the DESeq function of the DESeq2 package (Love et al., 2014). We identified genes as differentially expressed if they met an adjusted p-value (False Discovery Rate) cutoff of 0.01. We found that 4,839 genes (out of 10,165 genes with mean count >10 across all ovary samples) met this criteria, with 2,147 genes upregulated in *stwl*$^{-/-}$ null and 2,692 downregulated in *stwl*$^{-/-}$ null (48%, 21%, and 26% of expressed genes) (Fig. 12, Table 2). When analyzed separately, we found that 49% of genes were differentially expressed in 2-day old ovaries, 56% in 0-day old ovaries, and 75% in the union of both datasets (Table 2).

***stwl* null ovaries are enriched for LTR and non-LTR retrotransposon transcripts**

The RNA-Seq data showed that repetitive elements are strongly impacted by loss of functional *stwl* (Fig. 13-17). We found that *P-element* transcript increases ~4-fold in *stwl* null ovaries (Fig. 14); this can be explained by the *P-element* insertion into the *stwl* locus that created the *stwl*^{j6c3} allele used in this experiment. The significant increase in *P-element* transcript serves as an internal validation for the presence of the *stwl*^{j6c3} allele. We also confirmed our qRT-PCR experiments that showed an increase in abundance of *Copia, Het-A, 412*, and *I element* in *stwl*^{j6c3} ovaries, though the magnitude was more muted in the RNA-Seq experiment.

We ranked all genes for which we had expression data according to their shrunken $\log_2$(Fold-change) in *stwl* null ovaries relative to wild-type (Fig. 16-17). When these same values are separated according to various repeat-associated classifications (i.e. transposon, satellite, etc.), it is apparent that certain classes (e.g. active transposons) are upregulated relative to the full dataset ("All Genes"). This upregulation results in a mean LFC greater than 0 for affected classes. This simple analysis is naive to the biological reality that classes of genes are not

uniformly affected and that the most interesting genes are often found at the extremities of these ranges.

In order to identify which classes of repetitive transcripts are enriched in *stwl* null ovaries, we performed Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005). GSEA calculates enrichment of a particular class or "set" of genes at the extremes of a ranked list of all genes in the genome. Rather than focusing on the differences in the spread of data between an experiment and background group, GSEA identifies a core group of genes of a particular class that are enriched in a dataset. This is done by calculating a running Enrichment Score (ES) by walking from the top (most upregulated) gene on down. Each time a gene in the full list is present in the queried set, the ES increases; each time a gene is not in the queried set, the ES decreases. The final ES is the maximum deviation from 0, where a positive ES indicates that the gene set is enriched at the top of the list (if, for example, the top-10 genes by LFC were all transposons), while a negative ES means the gene set is associated with downregulation. The group of genes above or below the maximum ES are the ones which contribute the most to the enrichment; these will be referred to as the "core-enriched" genes.

GSEA was performed on the entire dataset of expressed genes (e.g. the "All Genes" category) to identify enrichment of repetitive genes. GSEA was implemented using the R package ClusterProfiler (Yu et al., 2012). GSEA against repeat-associated classes found that LTR retrotransposons and germline-restricted TEs are highly upregulated in *stwl* null ovaries (Fig. 18). Active transposons, non-LTR retrotransposons, and TEs belonging to the Gypsy, Bel-Pao and Ty1-Copia superfamilies are also upregulated in this tissue, albeit at a lower magnitude; the upregulation of these seven classes is driving the bulk of the upregulation signal among transposons and repeats as a whole (Fig. 19).

Tissue	Numerator	Reference	DESeq Design Formula	Age (days)	Total Genes (Mean Read Count >10)	D.E. (%)	Up (%)	Down (%)
Ovary	Null	WT	~Age + Genotype	0, 2	10165	4839 (48%)	2147 (21%)	2692 (26%)
Ovary	Null	WT	~Genotype	0	10165	5696 (56%)	2575 (25%)	3121 (31%)
Ovary	Null	WT	~Genotype	2	10165	4960 (49%)	2206 (22%)	2754 (27%)
Ovary	Het	WT	~Genotype	2	9039	4370 (48%)	1911 (21%)	2459 (27%)
Testis	Null	WT	~Genotype	2	11851	356 (3%)	197 (2%)	159 (1%)
S2 Cells	*stwl*	*lacZ*	~RNAi	NA	7807	451 (6%)	275 (4%)	176 (2%)

Table 2. Summary of differential gene expression experiments

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 <.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).

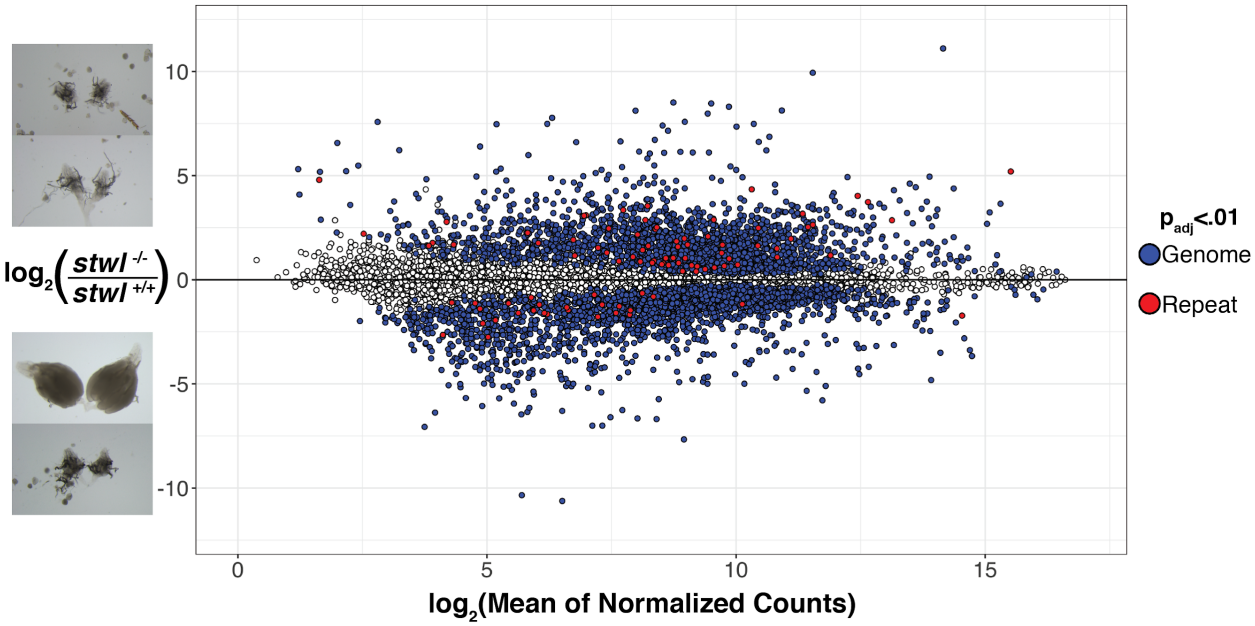


Figure 12. *stwl* null ovaries show large changes in transcript abundance

Fold-change for each gene is plotted against its average transcript abundance across assayed ovarian samples (wild-type and null). Transcript abundance is represented by counts normalized according to GC-content and library size. The $\log_2$(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 <.01) in this comparison. Red points represent entries from Repbase, blue points are from the genomic annotation.

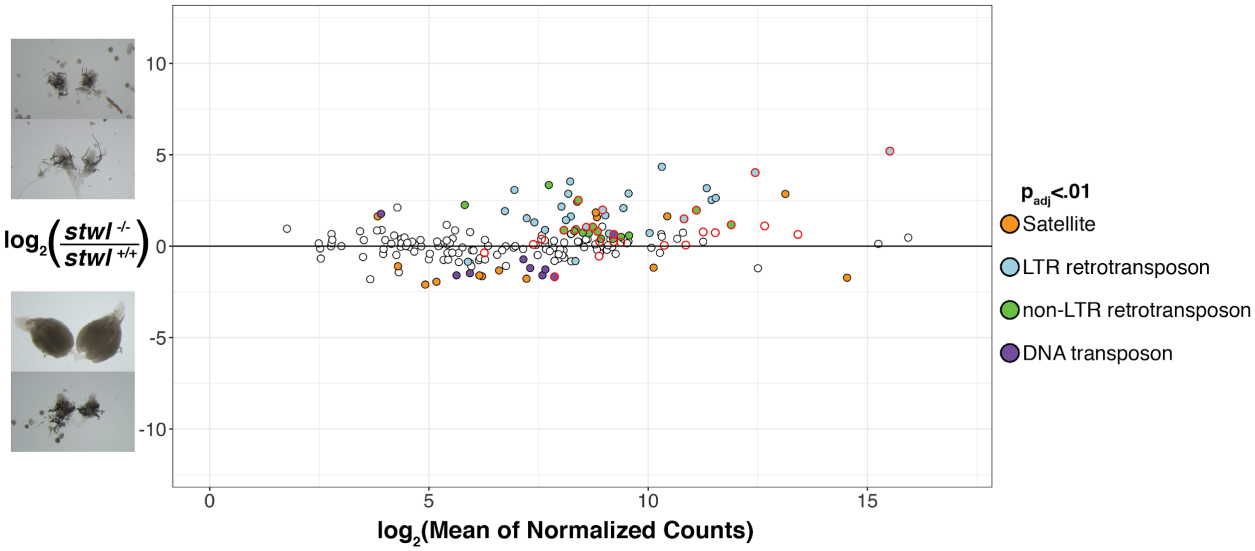
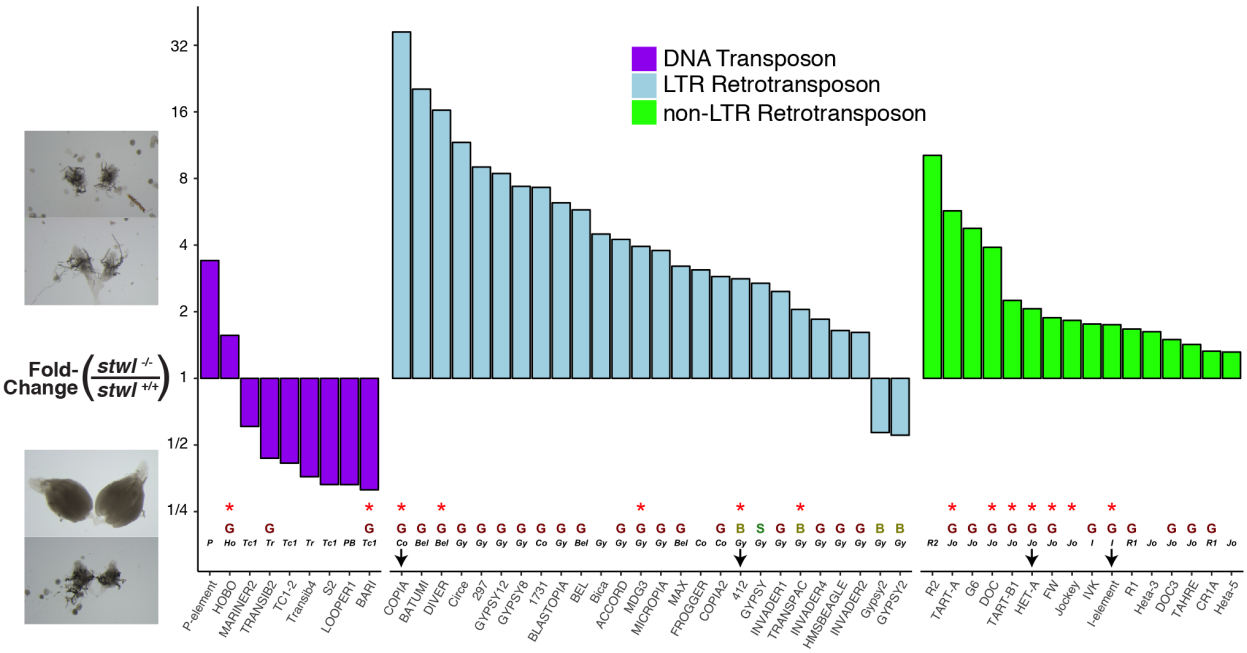
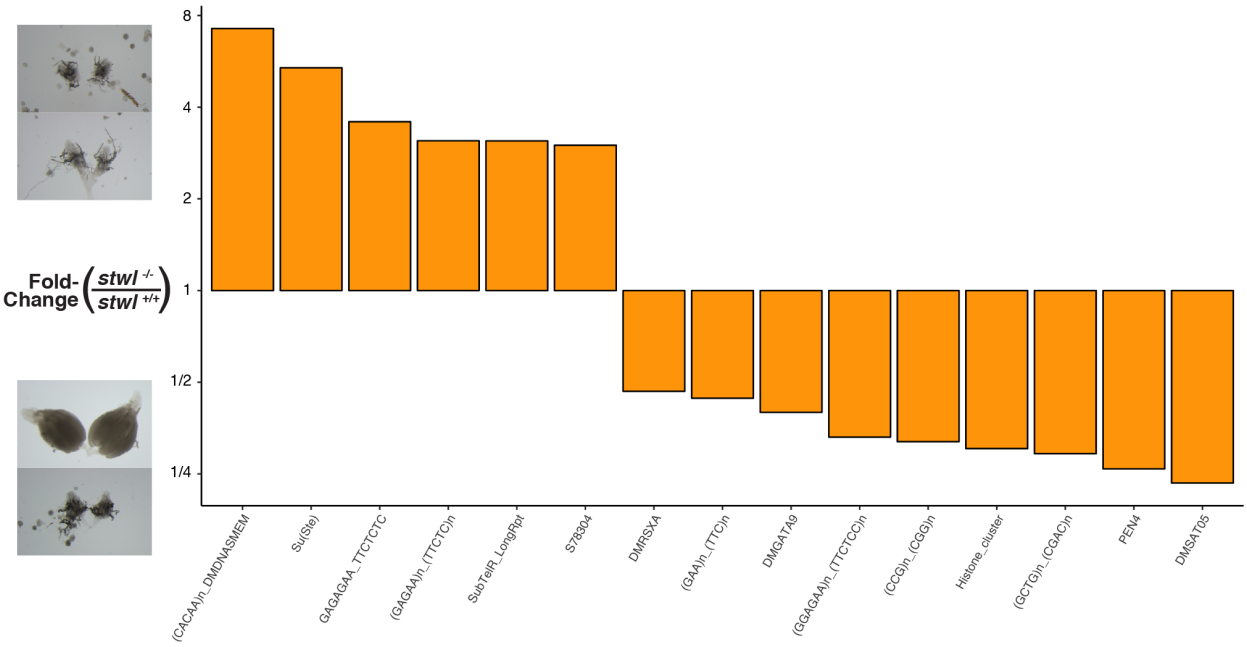


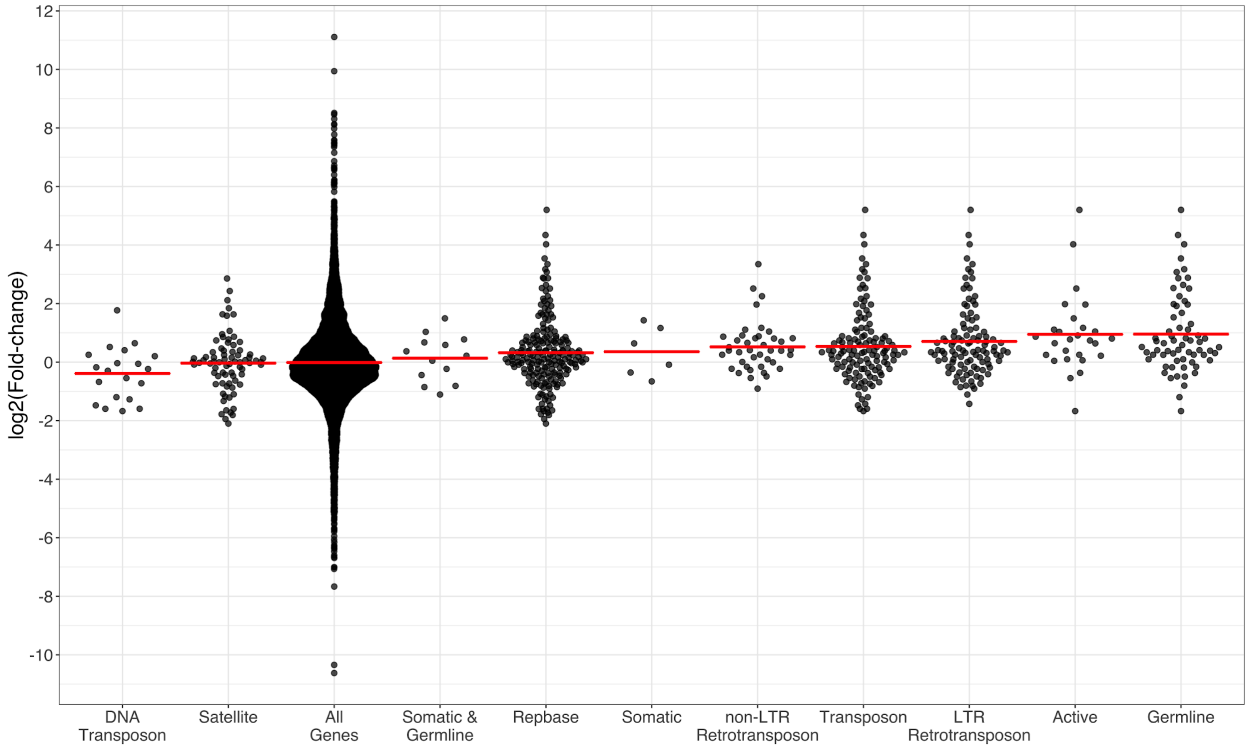
Figure 13. Transcripts from repetitive elements are upregulated in *stwl* null ovaries. Fold-change for each element in the repeat index is plotted against its average transcript abundance across assayed ovarian samples (wild-type and null). Transcript abundance is represented by counts normalized according to GC-content and library size. The $\log_2$(Fold-change) values (LFC) were "shrunk" to minimize the variance associated with low-count repeats. Filled points identify repetitive elements which are differentially expressed (adjusted p-value <.01) in this comparison. Points with red outlines represent transposons that are actively transposing in *D. melanogaster* (Kelleher and Barbash, 2013). All repeats from repbase-derived index are plotted.



<u>Figure 14</u>. Active transposable elements are upregulated in *stwl* null ovaries.
Fold-change of transcript abundance of transposable elements in *stwl* null/wild-type ovaries. Y-axis is in $\log_2$-scale. Red asterisks indicate elements that are actively transposing in *D. melanogaster* (Kelleher and Barbash, 2013). "G","B","S" indicate whether transposons are typically expressed in germline cells, somatic cells, or both. TE super-families are reported as follows: *Bel=Bel-Pao*, *Co=Ty1-Copia*, *Gy=Gypsy*, *Ho=Hobo*, *I=I-element*, *Jo=Jockey*, *P=P-element*, *PB=PiggyBac*, *Po=Pogo*, *Tr=Transib*. Black arrows point to TEs validated with qRT-PCR data (Figure 1).



<u>Figure 15</u>. Non-TE repetitive elements are misregulated in *stwl* null ovaries.
Fold-change of transcript abundance of satellites and other repetitive elements in *stwl* null/wild-type ovaries. Y-axis is in $\log_2$-scale.



<u>Figure 16</u>. Spread of LFC by repeat class in *stwl* null ovaries.
log2(Fold-change) of transcript abundance of repetitive elements in *stwl* null/wild-type ovaries. Each data point represents the LFC of a single gene. Genes are plotted according to their repeat classification, with genes appearing in multiple classes (e.g. all repetitive genes are plotted in the "Repbase" category). "All Genes" class plots the LFC of all 10,165 genes in the DESeq analysis. A subset of TEs are classified according to their expression in ovarian germline, ovarian somatic cells, or both. Red crossbars show the mean LFC for each class.

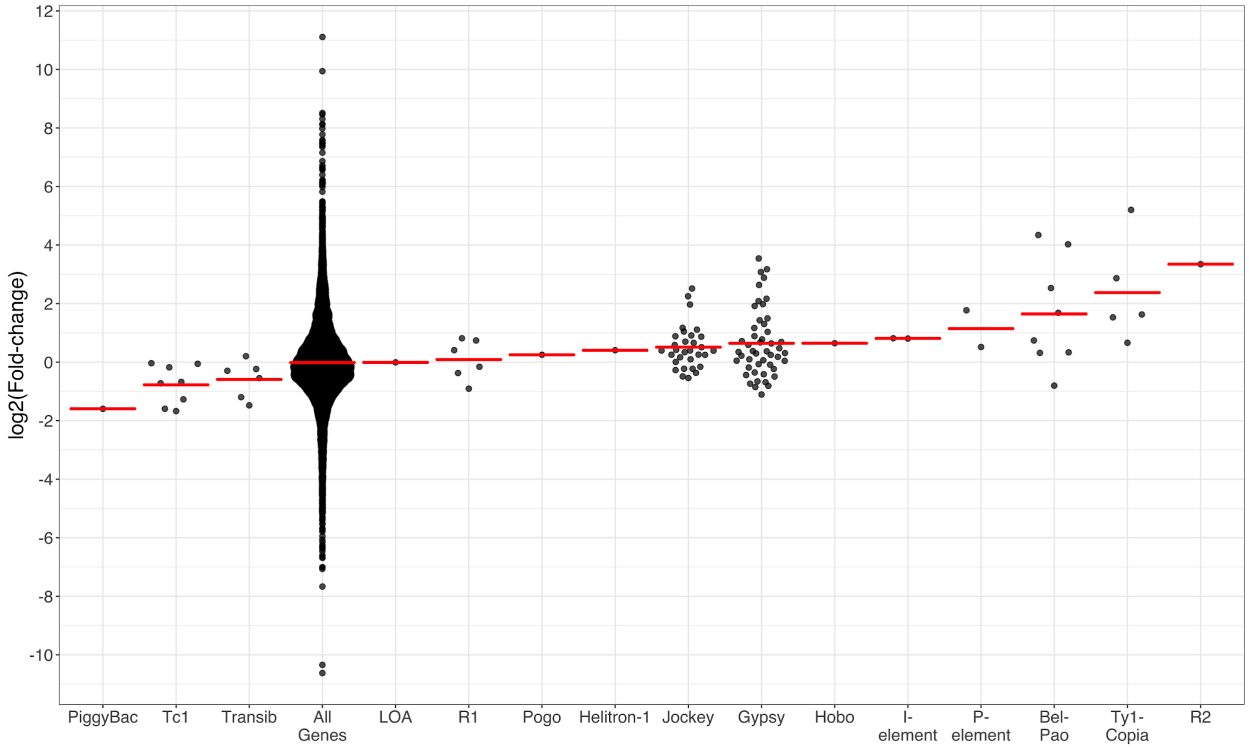


Figure 17. Spread of LFC by transposable element superfamily in null/WT ovaries. log2(Fold-change) of transcript abundance of repetitive elements in *stwl* null/wild-type ovaries. Each data point represents the LFC of a single gene or repeat. TEs are plotted according to their superfamily. "All Genes" class plots the LFC of all 10,165 genes in the DESeq analysis. Red crossbars show the mean LFC for each group.

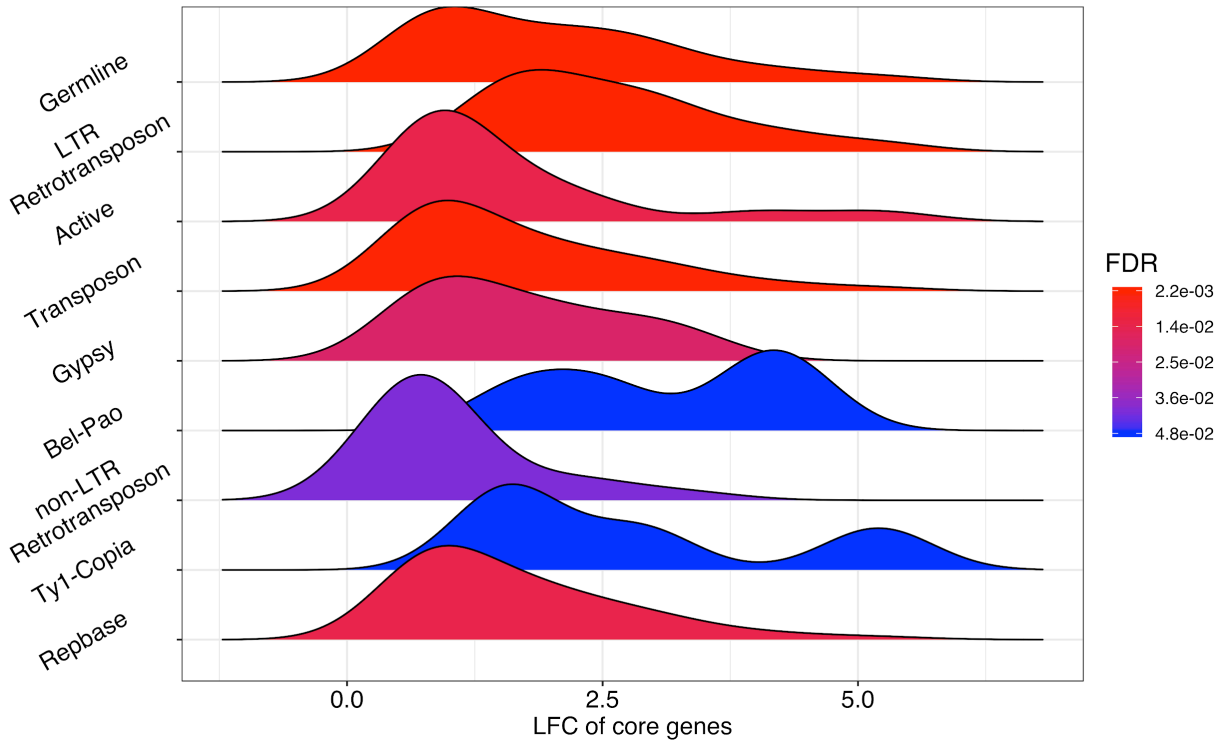


Figure 18. Retrotransposons are upregulated in *stwl* null ovaries

GSEA was performed on shrunken LFC values derived from comparing *stwl* null ovaries to wild-type ovaries. Repeat sets with FDR <.05 were called enriched with respect to LFC. For each significantly enriched gene set, a core set of genes was identified as contributing to the enrichment. The density of the shrunken LFC of core enriched genes for each set is displayed, ordered from top to bottom by Normalized Enrichment Score. The "Repbase" gene set represents all entries from the repeat index.

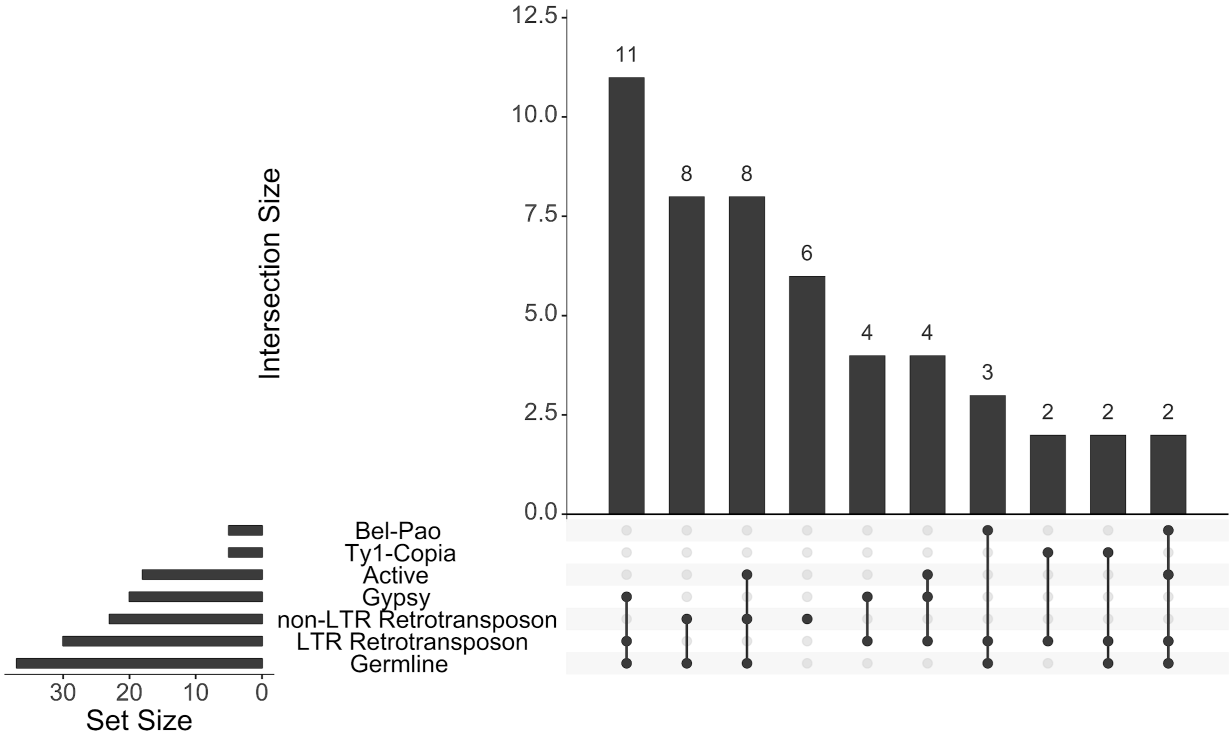


Figure 19. Germline retrotransposons are upregulated in *stwl* null ovaries

UpSet plot of core enriched repeats from GSEA on *stwl* null ovaries vs wild-type ovaries. Horizontal bars indicate "Set Size", which is the total number of core-enriched genes belonging to a particular set, with many genes belonging to multiple sets. Vertical bars indicate "Intersection Size", which is the total number of core-enriched genes belonging to a specific intersection of sets (e.g. 11 TEs are members of the Gypsy superfamily and predominantly expressed in ovarian germline cells; 4 Gypsy elements are NOT predominantly expressed in ovarian germline; all Gypsy elements are LTR Retrotransposons). Intersections of sets are non-overlapping.

Satellite-derived transcripts are upregulated in *stwl* heterozygous ovaries

stwl is a dominant suppressor of Position-Effect Variegation (Maines et al., 2007; Yi et al., 2009). We therefore predicted that heterozygous ovaries, despite manifesting as phenotypically normal in microscopy-based assays, might have altered transcription of some repeats that reflects a loss of heterochromatin maintenance. qRT-PCR assays described above supported this expectation for *412*, *Copia*, and *I element* (Fig. 1). Differential Expression Analysis was performed on 2-day old heterozygous ovaries relative to 2-day old wild-type ovaries. While the magnitude of differential expression is dramatically reduced relative to a similar analysis with *stwl* null ovaries (Fig. 12, Fig. 20), we nonetheless identified mis-regulation of 48% of expressed genes (Table 2). The contrast in the severity of the misexpression phenotypes in heterozygotes versus homozygotes may be indicative of an increase in "noise" caused by loss of germline tissue, induction of apoptosis, and absence of mature egg chambers in *stwl* nulls. Alternatively, the different patterns in each condition may reflect perturbations in distinct biological processes.

The effect size (i.e. range of log2 fold change or LFC values) in *stwl* heterozygous ovaries is much lower than in *stwl* null ovaries for most protein-coding genes and TEs (Fig. 12, Fig. 20). With the exception of a ~40-fold increase in expression of the *Agoriino* element, we did not identify large-scale changes in transposon abundance (Fig. 21, Fig. 22), although we confirmed our prior observation from rt-PCR that *Copia* and *I* element transcripts are derepressed in heterozygous ovaries (Fig. 1). We also found that the *Gypsy* LTR retrotransposons *Blood* and *Burdock* are de-repressed in heterozygotes; these TEs were previously found to be similarly affected in germline KD experiments in ovaries (Czech et al., 2013). *P-element* transcript increased ~4-fold in heterozygous ovaries, due to the P-element

insertion in the *stwl*^{j6c3} allele. This increase in *P-element* abundance is similar to the increase in *stwl* null ovaries, which have two copies of the insertion. This genetic perturbation does not result in loss of *stwl* transcript in heterozygous ovaries. We do find that *stwl* transcript is reduced ~2.5-fold in *stwl* null ovaries, which may be a result of the reduction of germline content in these ovaries (indeed, *stwl* transcript is reduced ~5-fold in 2-day old null ovaries vs 2-day old WT ovaries and only ~1.5-fold in 0-day old null ovaries vs 0-day old WT ovaries).

Satellite-derived transcripts were significantly derepressed in *stwl* heterozygous, but not in *stwl* null ovaries (Fig. 21, Fig. 23). This observation was confirmed via GSEA, which showed that *stwl* heterozygous ovaries are highly enriched for satellite transcripts (Fig. 24, Fig. 26). This phenotype may reflect misregulation of heterochromatic sequence caused by *stwl* haploinsufficiency, which is either distinct from the biological processes impacted in *stwl* null ovaries or masked by other genetic perturbations in the apoptotic, germline-deficient ovaries (see Discussion).

With the exception of non-LTR retrotransposons, we found that all other transposon classes that are enriched in *stwl* null ovaries are also enriched in *stwl* heterozygous ovaries (Fig. 28). These include LTR retrotransposons, actively transposing TEs, germline-restricted TEs, and Gypsy superfamily elements. The strength of the enrichment signal (represented by Normalized Enrichment Score) for LTR retrotransposons and germline-restricted TEs is higher in the null ovaries. The enrichment of germline-restricted TEs in both *stwl* mutant samples is a useful validation of our experimental design; the strong upregulation of germline-restricted TEs is unlikely to be caused by an imbalance of germline/somatic tissue in null ovaries, since *stwl* null ovaries contain fewer germline cells.

We identified 25 repeats whose transcripts are mis-regulated in both *stwl* null and heterozygous ovaries (Fig. 29). These include *I element* and *Copia*, each of which has been validated by qRT-PCR (Fig. 1). Most of these similarly mis-regulated transcripts have similar fold-changes in the null and heterozygous genotypes, but some exhibit an additive effect from *stwl* loss. These include the retrotransposons *Doc*, *Diver, Gypsy8*, *Copia*, and *Blastopia*. In contrast, the DNA transposon *Looper1* and the *GAGAGAA_TTCTCTC* satellite transcript are more strongly mis-regulated in *stwl* heterozygotes versus nulls.

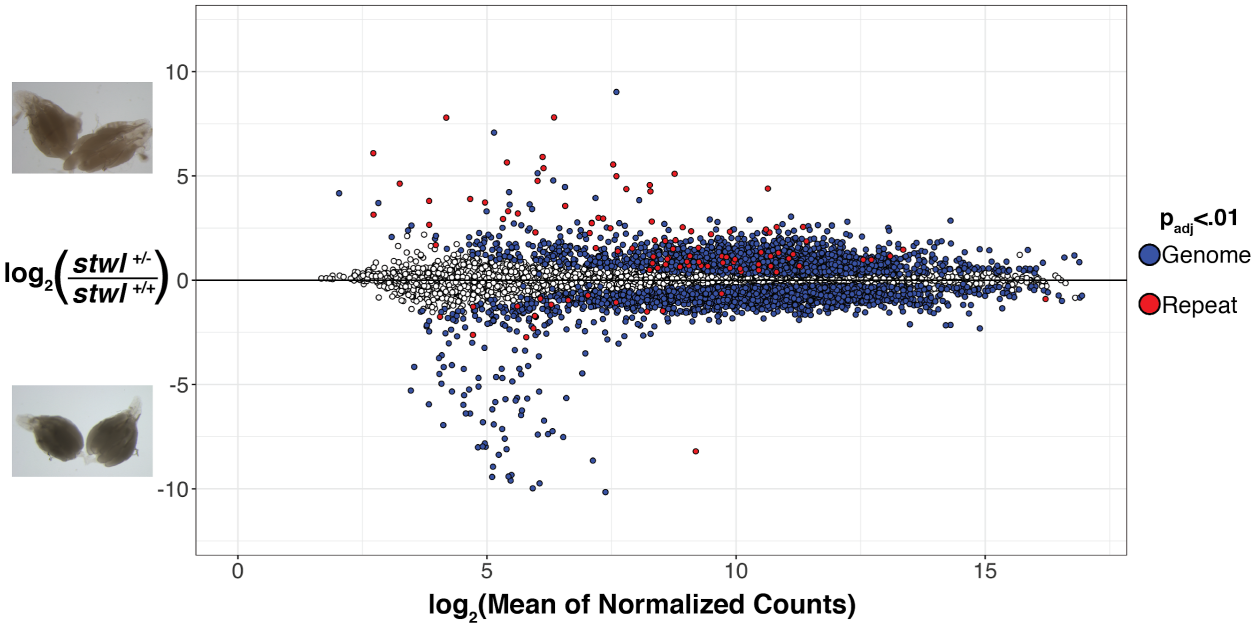


Figure 20. *stwl* heterozygous ovaries have thousands of mis-expressed genes
Fold-change for each gene is plotted against its average transcript abundance across assayed ovarian samples (wild-type and heterozygous). Transcript abundance is represented by counts normalized according to GC-content and library size. The $\log_2$(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 <.01) in this comparison. Red points represent entries from Repbase, blue points are from the genomic annotation.

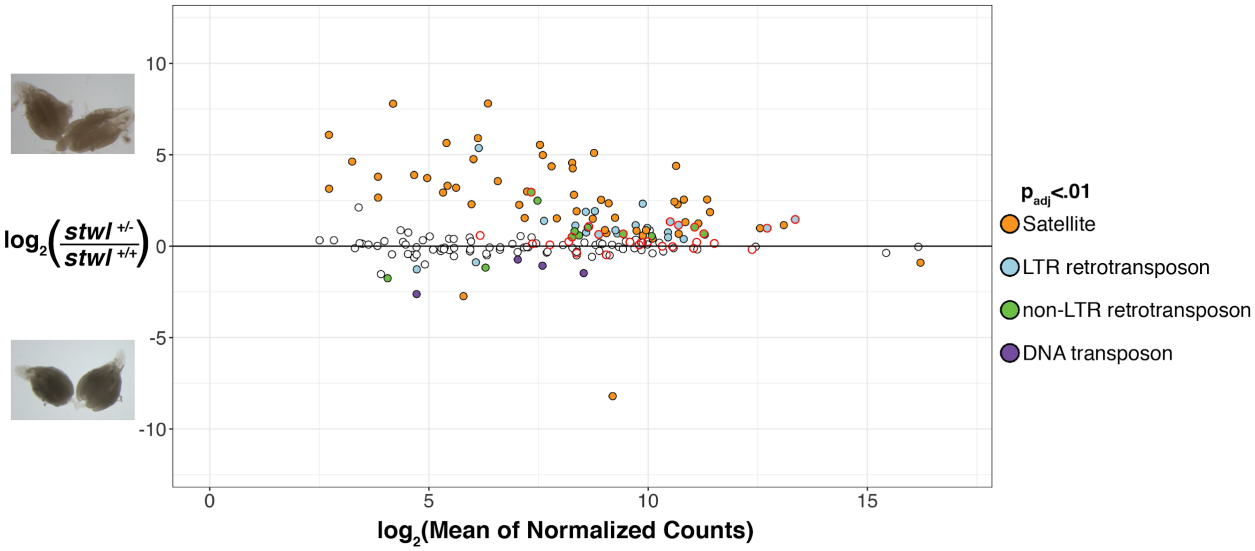


Figure 21. Transcripts from satellites are upregulated in *stwl* heterozygous ovaries. Fold-change for each element in the repeat index is plotted against its average transcript abundance across assayed ovarian samples (wild-type and heterozygous). Transcript abundance is represented by counts normalized according to GC-content and library size. The $\log_2$(Fold-change) values (LFC) were "shrunk" to minimize the variance associated with low-count genes. Filled points identify repetitive elements which are differentially expressed (adjusted p-value <.01) in this comparison. Points with red outlines represent transposons that are actively transposing in *D. melanogaster* (Kelleher and Barbash, 2013). All repeats from repbase-derived index are plotted.

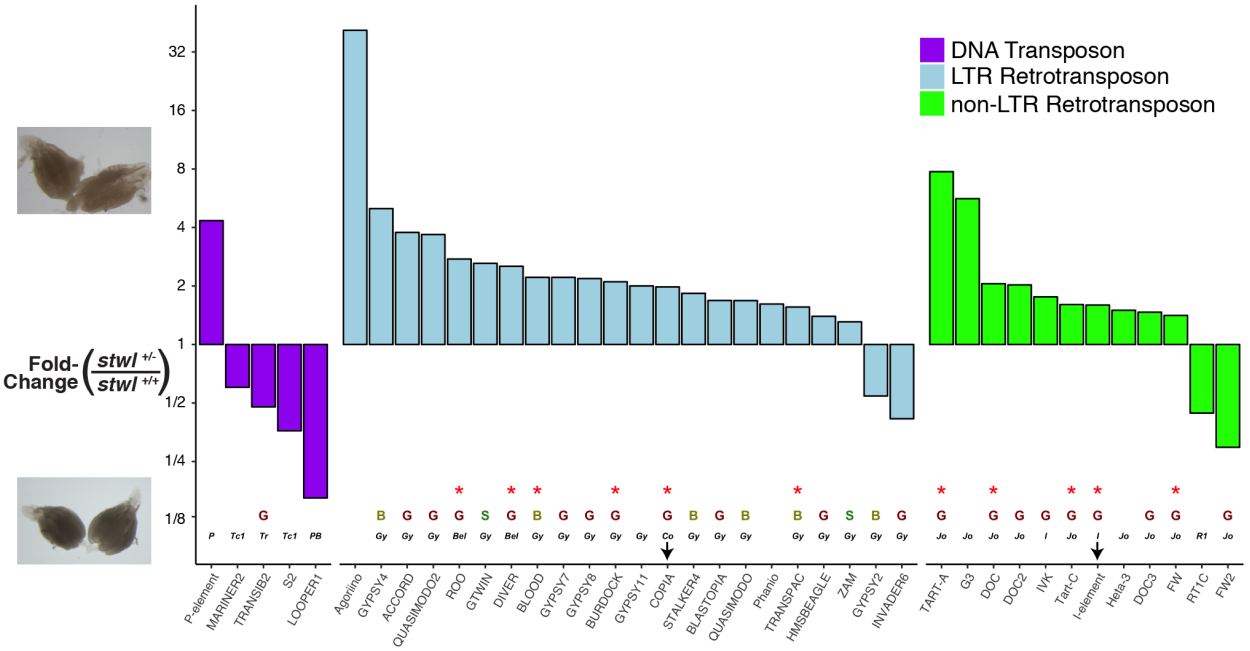
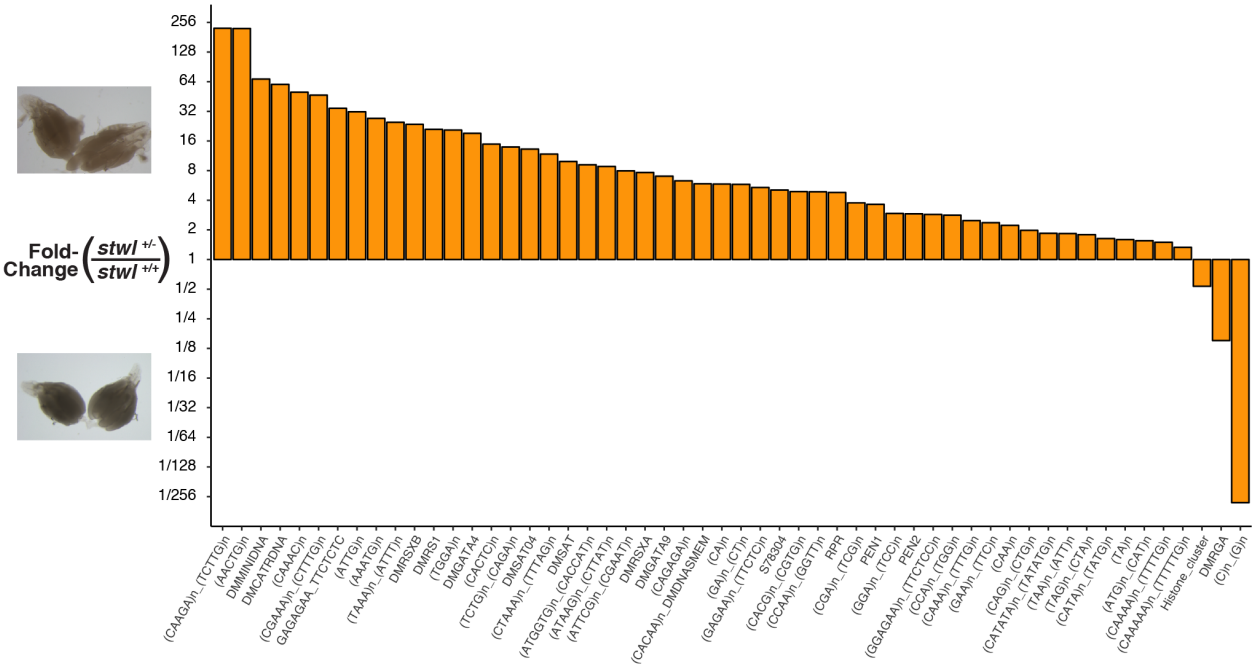


Figure 22. Active transposable elements are upregulated in *stwl* heterozygous ovaries. Fold-change of transcript abundance of transposable elements in *stwl* heterozygous/wild-type ovaries. Y-axis is in $\log_2$-scale. Red asterisks indicate elements that are actively transposing in *D. melanogaster* (Kelleher and Barbash, 2013). "G","B","S" indicate whether transposons are typically expressed in germline cells, somatic cells, or both. TE super-families are reported as follows: *Bel=Bel-Pao*, *Co=Ty1-Copia*, *Gy=Gypsy*, *Jo=Jockey*, *P=P-element*, *PB=PiggyBac*, *Tr=Transib*. Black arrows point to TEs validated with qRT-PCR data (Fig. 1).



<u>Figure 23</u>. Non-TE repetitive elements are up-regulated in *stwl* heterozygous ovaries. Fold-change of transcript abundance of satellites and other repetitive elements in *stwl* heterozygous/wild-type ovaries. Y-axis is in $\log_2$-scale.

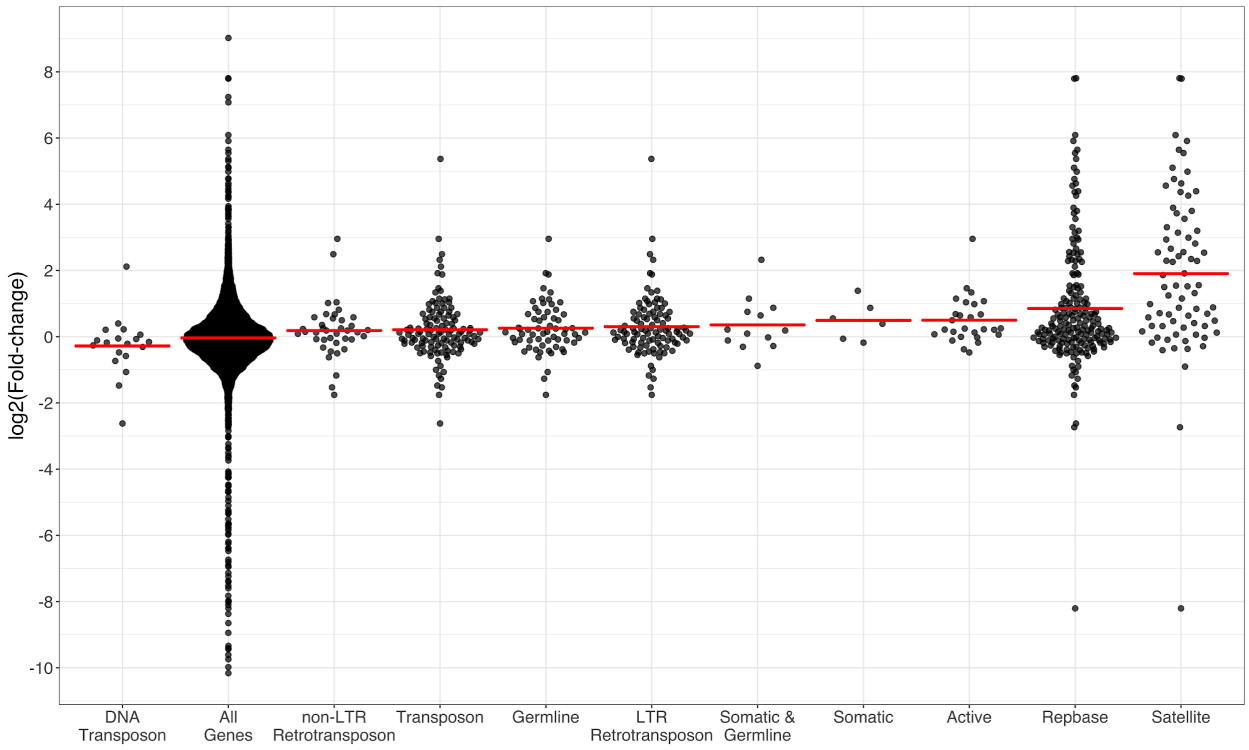
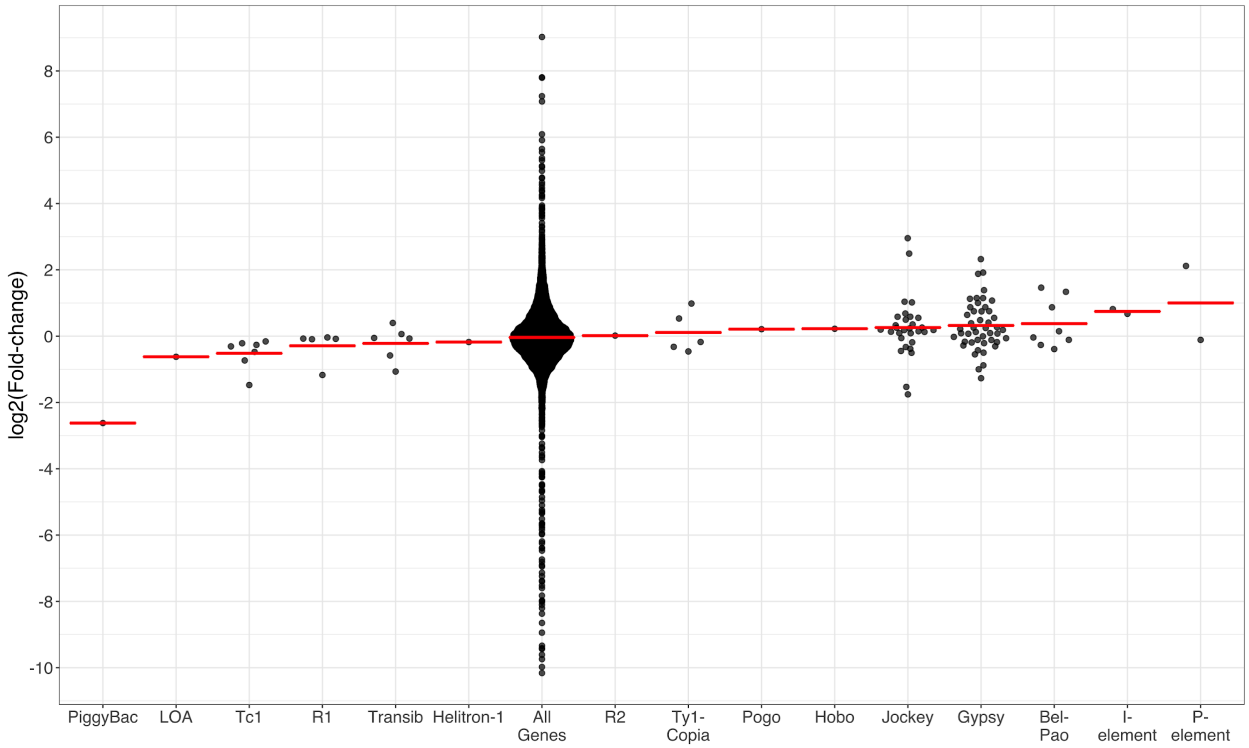


Figure 24. Spread of LFC by repeat class demonstrates upregulation in *stwl* heterozygous ovaries.

log2(Fold-change) of transcript abundance of repetitive elements in *stwl* heterozygous/wild-type ovaries. Each data point represents the LFC of a single gene. Genes are plotted according to their repeat classification, with genes appearing in multiple classes (e.g. all repetitive genes are plotted in the "Repbase" category). "All Genes" class plots the LFC of all 9039 genes in the DESeq experiment. A subset of TEs are classified according to their expression in ovarian germline, ovarian somatic cells, or both. Red crossbars show the mean LFC for each class.



<u>Figure 25</u>. Spread of LFC by transposable element superfmaily, heterozygous/WT ovary. log2(Fold-change) of transcript abundance of repetitive elements in *stwl* heterozygous/wild-type ovaries. Each data point represents the LFC of a single gene. Genes are plotted according to their superfamily. "All Genes" class plots the LFC of all 9039 genes in the DESeq experiment. Red crossbars show the mean LFC for each group.

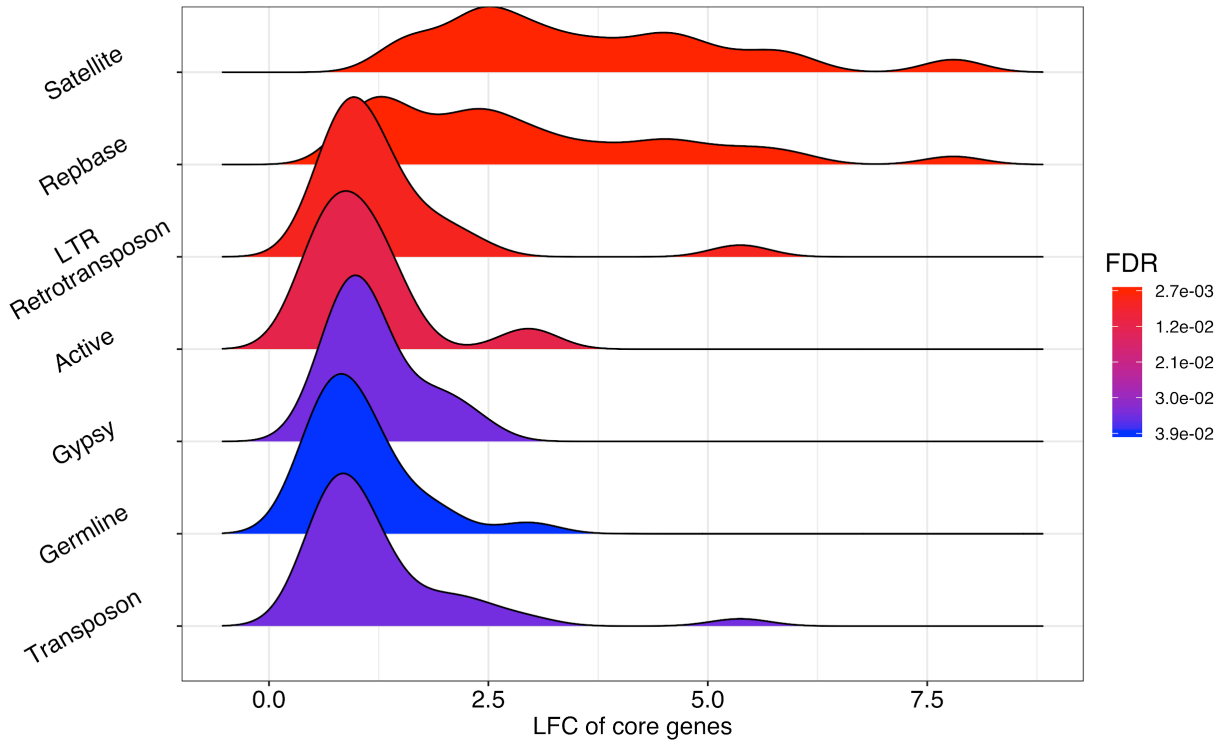
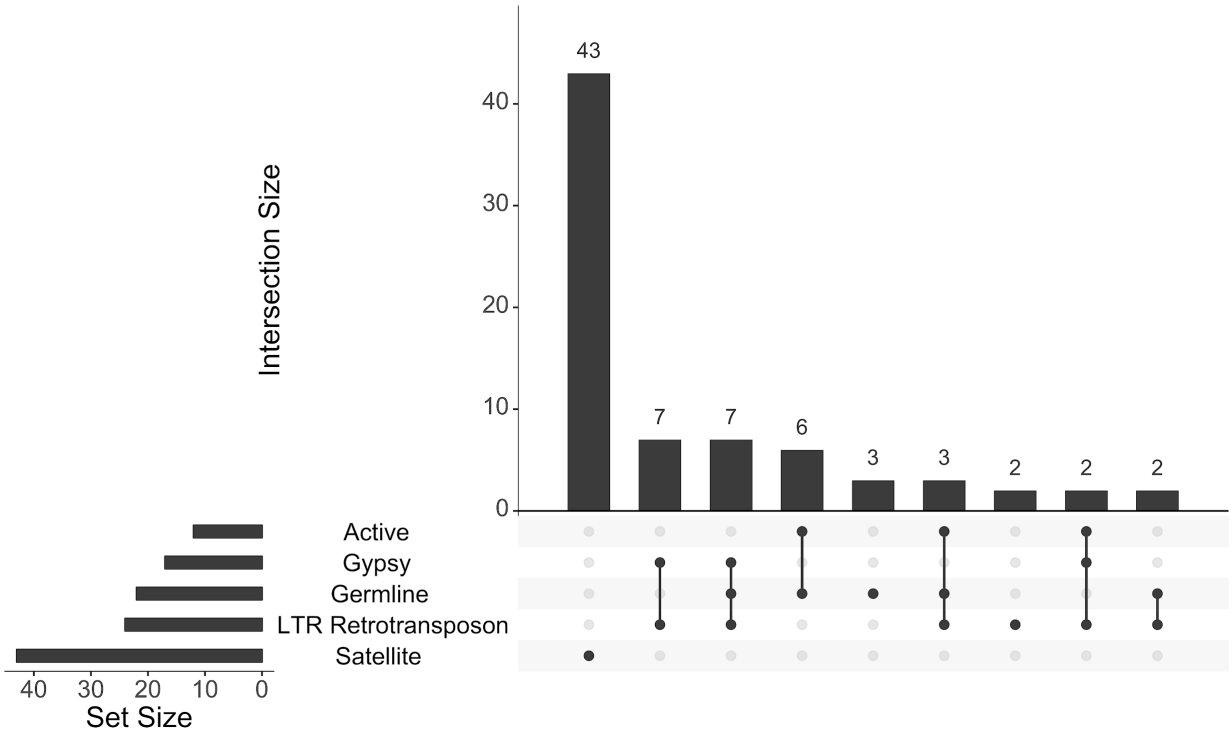


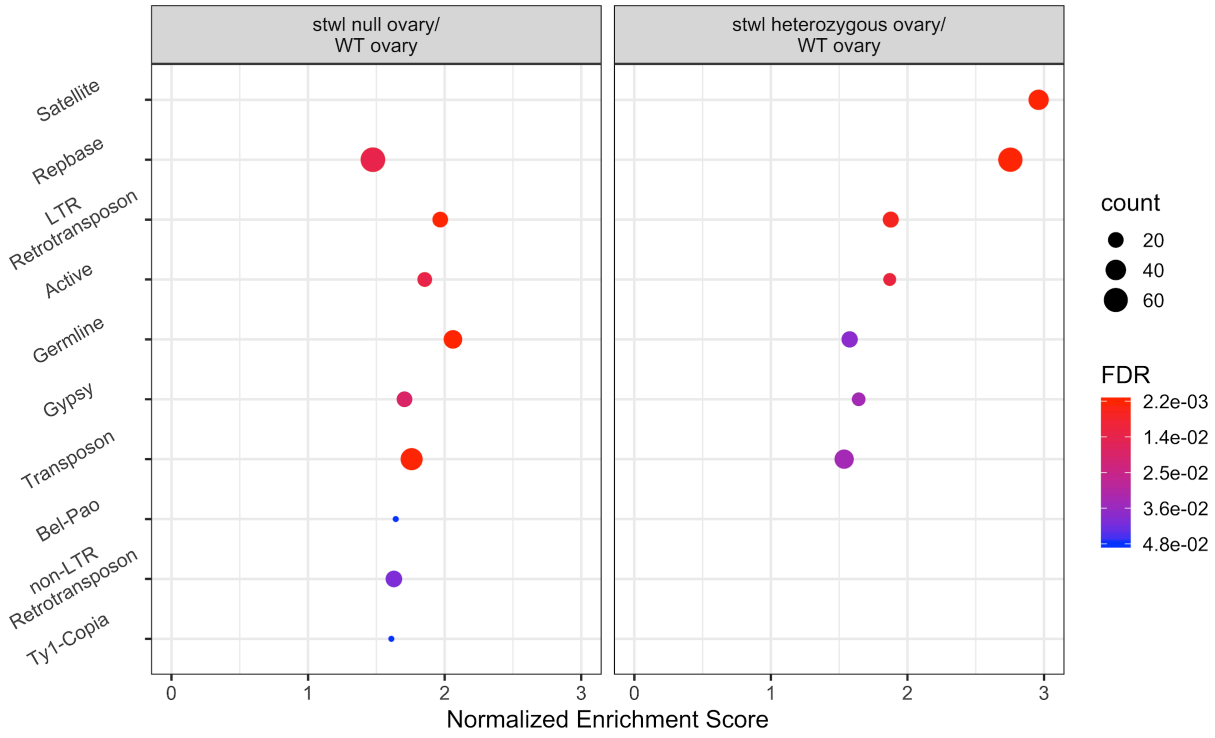
Figure 26. Satellites and LTR retrotransposons are upregulated in *stwl* heterozygous ovaries

GSEA was performed on shrunken LFC values derived from comparing *stwl* heterozygous ovaries to wild-type ovaries. Gene sets with FDR <.05 were called enriched with respect to LFC. The density of the shrunken LFC of core enriched genes for each set is displayed, ordered from top to bottom by Normalized Enrichment Score.



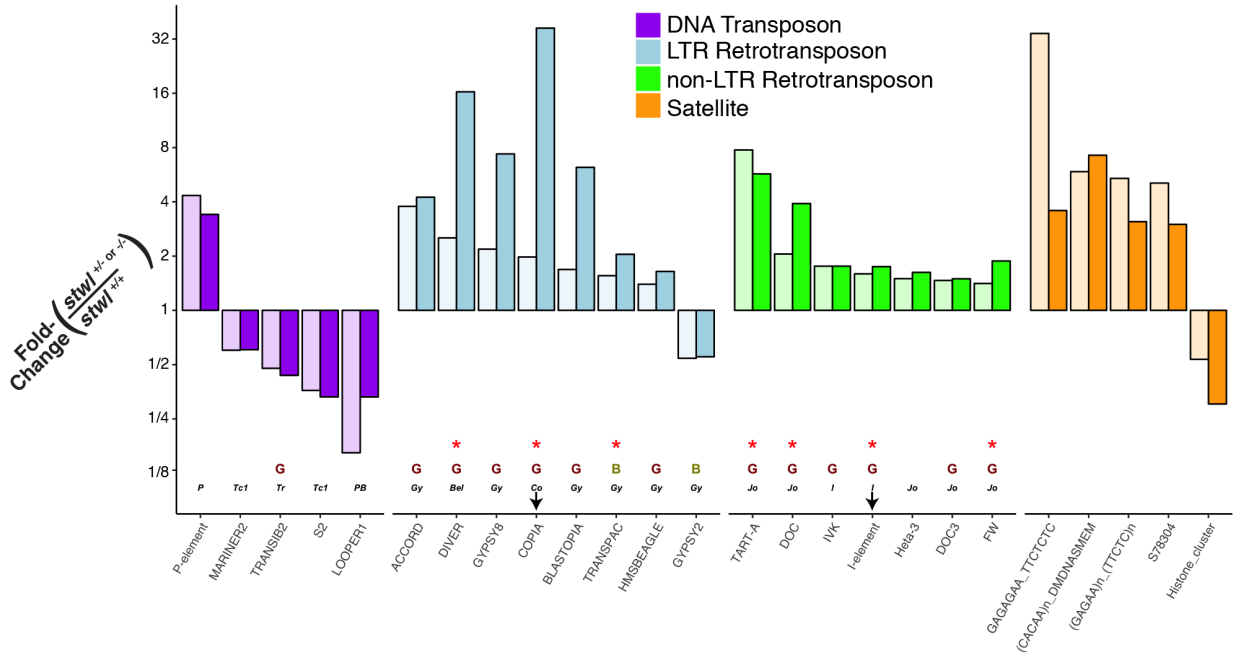
<u>Figure 27</u>. Satellites and LTR retrotransposons are upregulated in *stwl* heterozygous ovaries

UpSet plot of core enriched repeats from GSEA on *stwl* heterozygous ovaries vs wild-type ovaries. Horizontal bars indicate "Set Size", which is the total number of core-enriched genes belonging to a particular set, with many genes belonging to multiple sets. Vertical bars indicate "Intersection Size", which is the total number of core-enriched genes belonging to a specific intersection of sets (e.g. 7 TEs are members of the Gypsy superfamily and predominantly expressed in ovarian germline cells; 7 Gypsy elements are NOT predominantly expressed in ovarian germline; all Gypsy elements are LTR Retrotransposons). Intersections of sets are non-overlapping.



<u>Figure 28</u>. Loss of *stwl* results in upregulation of similar repetitive element classes in null and heterozygous ovaries.

Comparison of GSEA results from Fig. 18 and Fig. 26. Normalized Enrichment Score is plotted for each set of repetitive elements enriched among null/WT or heterozygous/WT ovaries. Only gene sets with FDR<.05 are plotted. Most genes are present in multiple sets (e.g. all Gypsy elements are in the LTR Retrotransposon, Transposon, and Repbase gene sets).



<u>Figure 29</u>. Some repetitive elements are consistently mis-regulated in both *stwl* null and heterozygous ovaries

Fold-change of transcript abundance of transposable elements in *stwl* heterozygous/wild-type ovaries and *stwl* null/wild-type ovaries. Fold-change in heterozygous ovaries is represented by lightly shaded bars. Red asterisks indicate elements that are actively transposing in *D. melanogaster* (Kelleher and Barbash, 2013). "G","B","S" indicate whether transposons are typically expressed in germline cells, somatic cells, or both. TE super-families are reported as follows: *Bel=Bel-Pao*, *Co=Ty1-Copia*, *Gy=Gypsy*, *Jo=Jockey*, *P=P-element*, *PB=PiggyBac*, *Tr=Transib*. Black arrows point to TEs validated with qRT-PCR data (Fig. 1).

Repetitive transcripts are rarely mis-regulated in *stwl* null testes

We also performed RNA-Seq on 2-day old *stwl* null and wild-type testes, using the same genotypes as for the *stwl* null ovaries. Despite exhibiting a sperm-exhaustion defect, *stwl* null testes are transcriptionally similar to their wild-type counterparts (Fig. 30). Only 3% of genes in *stwl* testes are mis-expressed, as opposed to 48% in *stwl* null and heterozygous ovaries (Table 2). Only ten repetitive genes are mis-expressed in null testes, the majority of them germline-restricted TEs (Fig. 31-34). GSEA did not identify significant enrichment for any repetitive gene

classes or families at the top or bottom of the distribution of the null testes data. Among the handful of repetitive genes that are mis-expressed in *stwl* testes, we found that four are consistently de-repressed among all three *stwl* mutant genotypes sampled: *Copia*, *Accord, Blastopia*, and *Tart-A* (Fig. 35). *Tart-A* is consistently the 1st or 2nd most highly up-regulated non-LTR retrotransposon in all three samples, while the remaining LTR retrotransposons are among the most highly up-regulated TEs in *stwl* null ovaries (Fig. 14, Fig. 22).

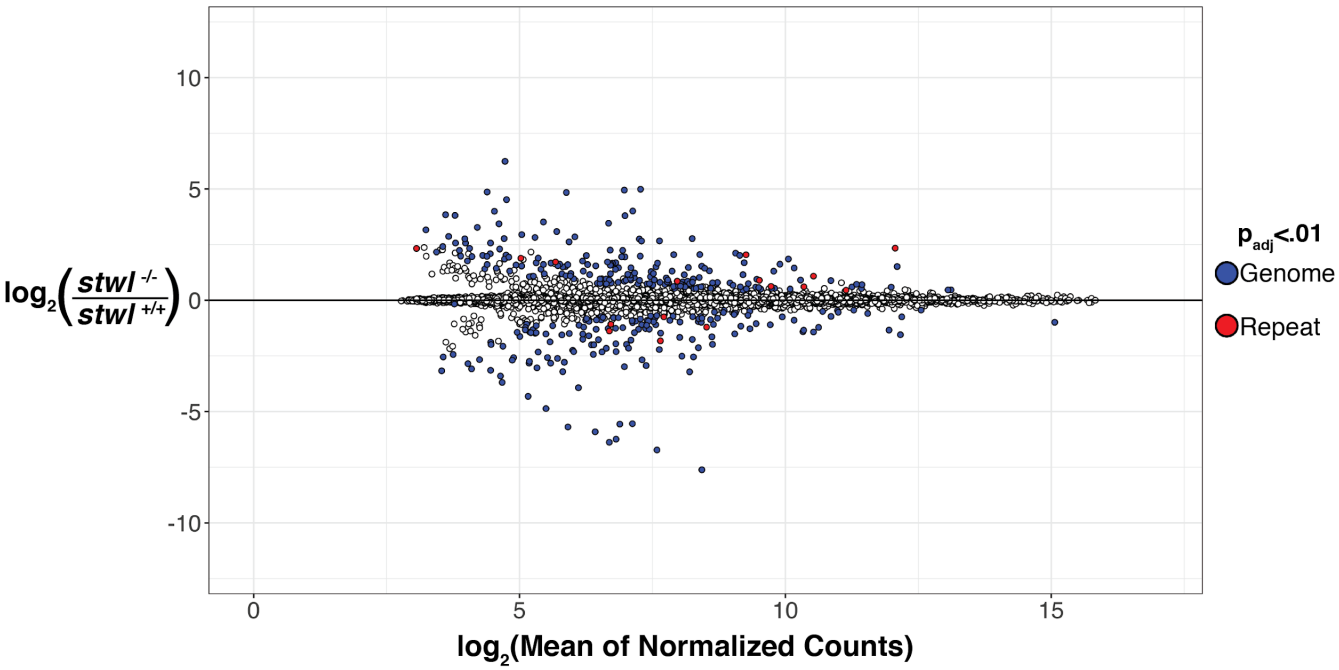


Figure 30. *stwl* null testes are transcriptionally similar to wild-type.
Fold-change for each gene is plotted against its average transcript abundance across assayed testis samples (wild-type and null). Transcript abundance is represented by counts normalized according to GC-content and library size. The $\log_2$(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 <.01) in this comparison. Red points represent entries from Repbase, blue points are from the genomic annotation.

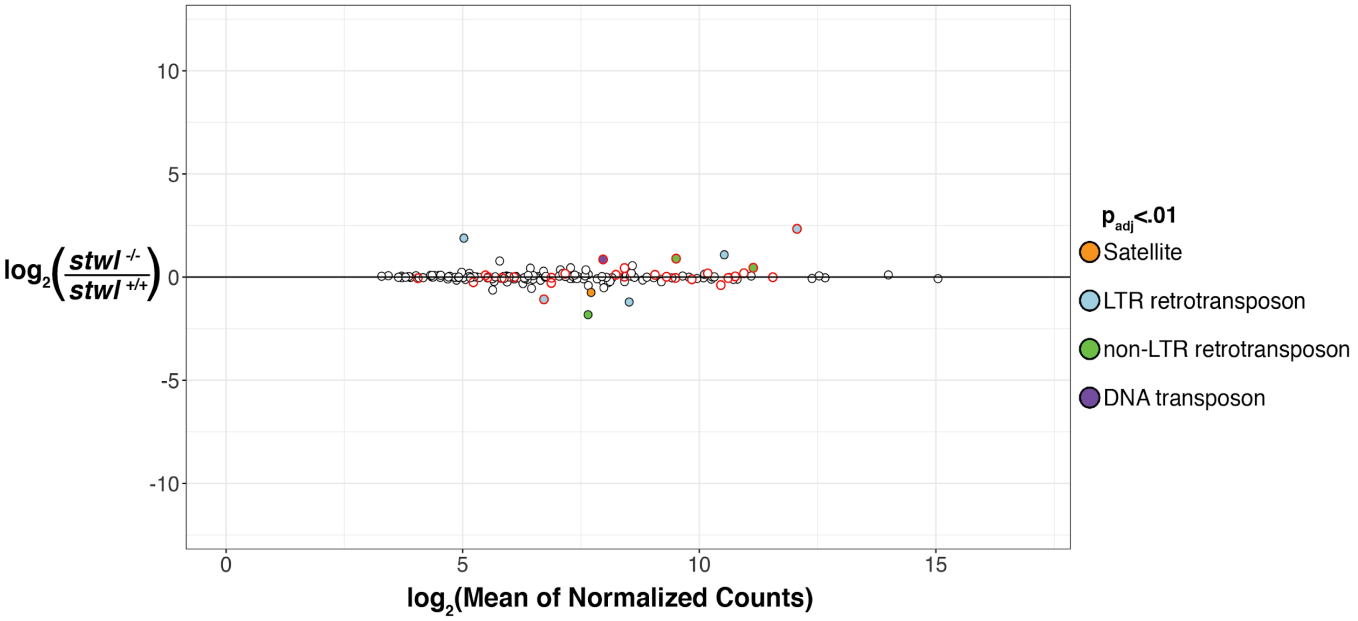
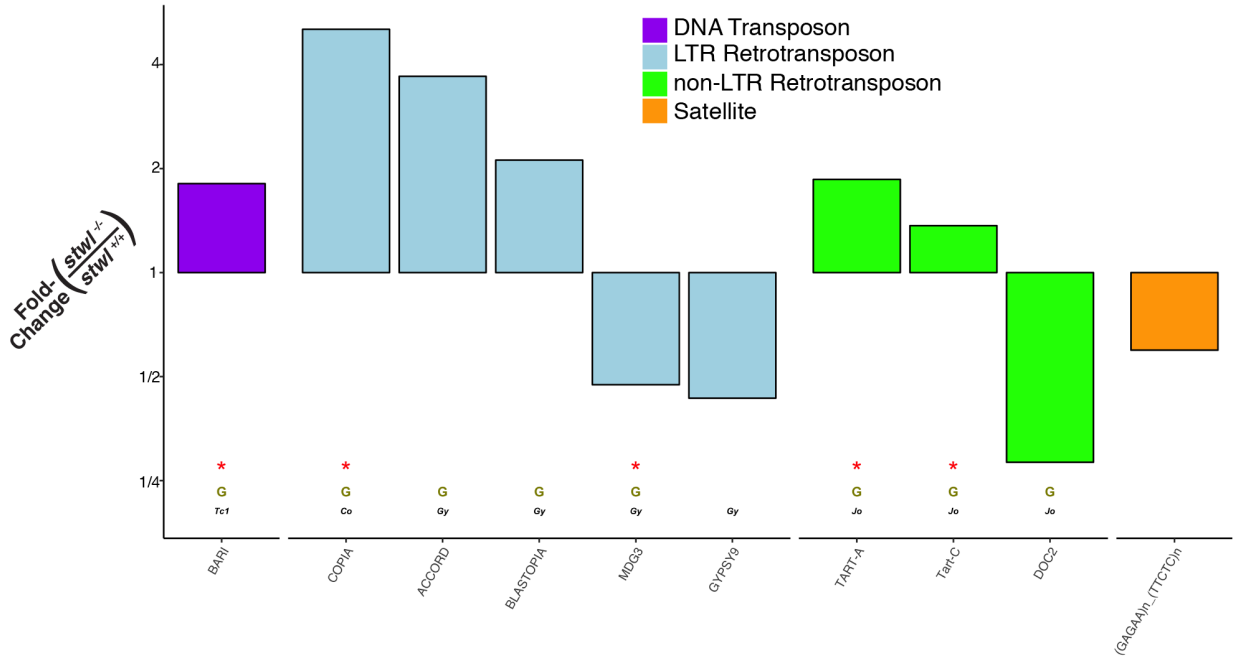
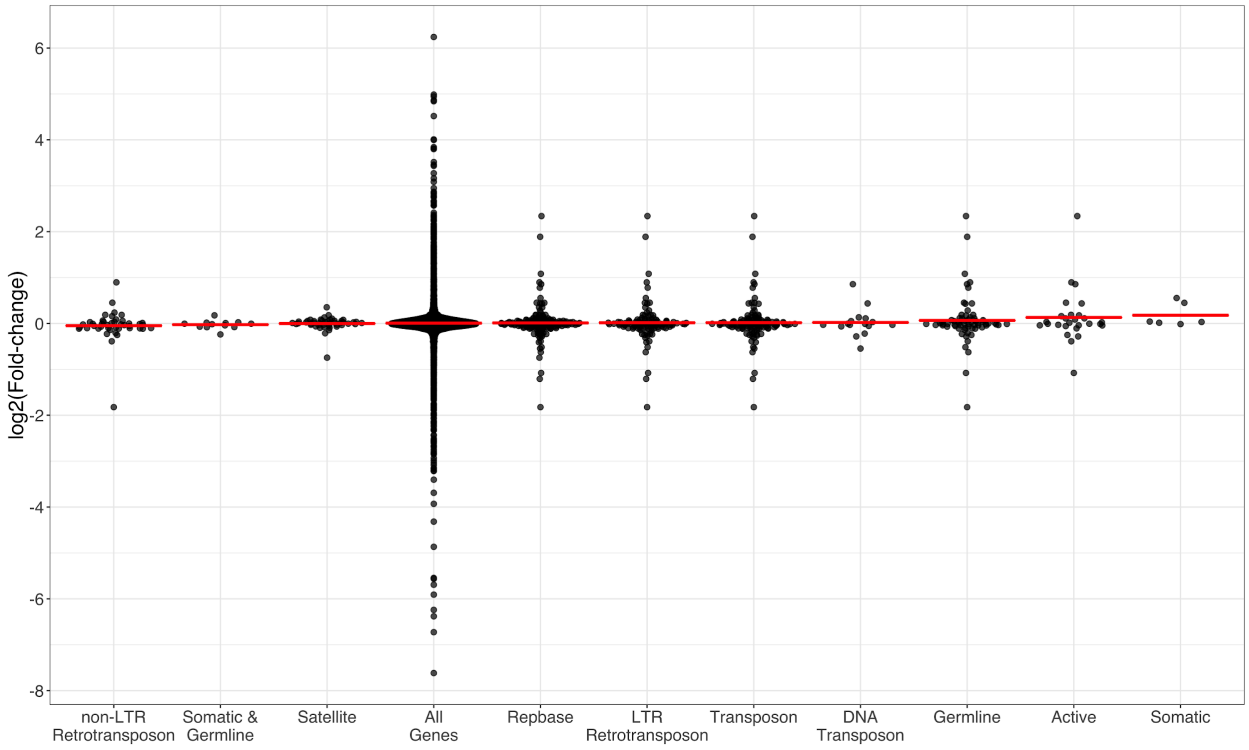


Figure 31. Only a handful of repetitive transcripts are mis-expressed in *stwl* null testes. Fold-change for each element in the repeat index is plotted against its average transcript abundance across assayed testes samples (wild-type and null). Transcript abundance is represented by counts normalized according to GC-content and library size. The $\log_2$(Fold-change) values (LFC) were "shrunk" to minimize the variance associated with low-count genes. Filled points identify repetitive elements which are differentially expressed (adjusted p-value <.01) in this comparison. Points with red outlines represent transposons that are actively transposing in *D. melanogaster* (Kelleher and Barbash, 2013).



<u>Figure 32</u>. Few transposable elements are mis-regulated in *stwl* null testes.
Fold-change of transcript abundance of transposable elements in *stwl* null/wild-type testes. Y-axis is in $\log_2$-scale. Red asterisks indicate elements that are actively transposing in *D. melanogaster* (Kelleher and Barbash, 2013). "G","B","S" indicate whether transposons are typically expressed in germline cells, somatic cells, or both. TE super-families are reported as follows: *Co=Ty1-Copia*, *Gy=Gypsy*, *Jo=Jockey*.



<u>Figure 33</u>. Spread of LFC by repeat class in *stwl* null testes.
log2(Fold-change) of transcript abundance of repetitive elements in *stwl* null/wild-type testes. Each data point represents the LFC of a single gene. Genes are plotted according to their repeat classification, with genes appearing in multiple classes (e.g. all repetitive genes are plotted in the "Repbase" category). "All Genes" class plots the LFC of all 9039 genes in the DESeq experiment. A subset of TEs are classified according to their expression in ovarian germline, ovarian somatic cells, or both. Red crossbars show the mean LFC for each class.

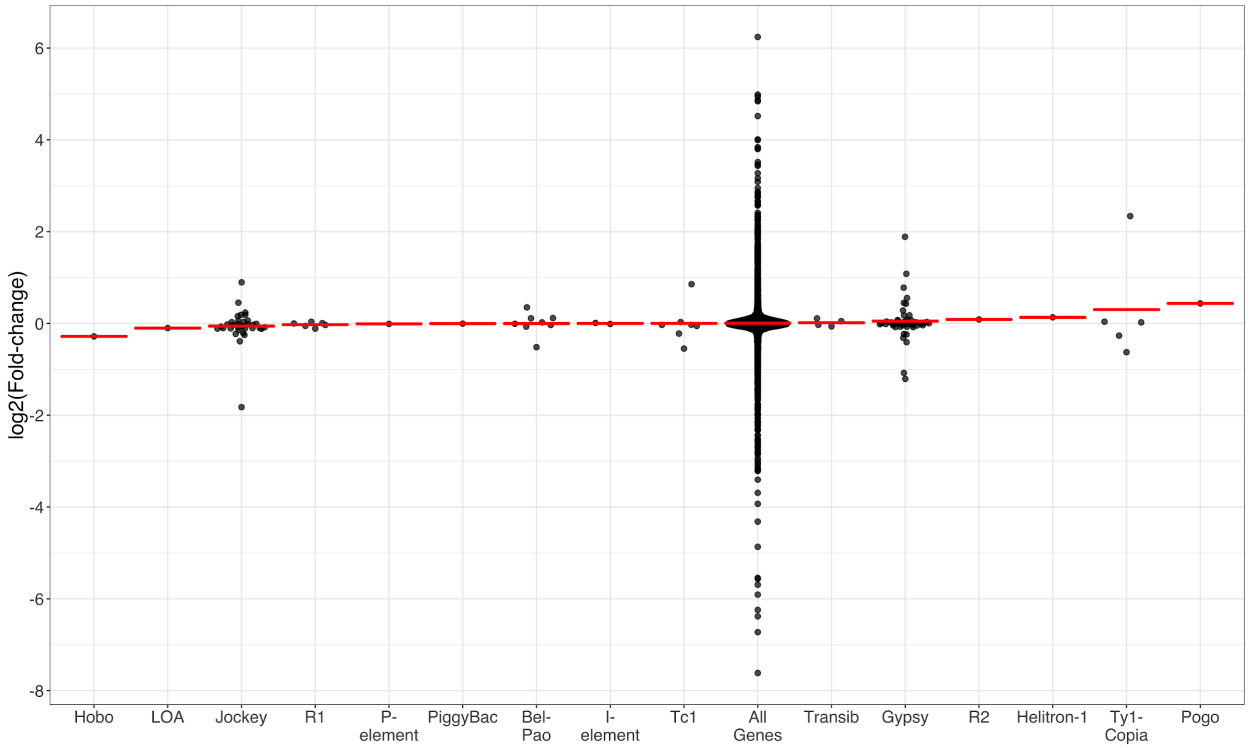


Figure 34. Spread of LFC by transposable element superfamily, null/WT testes.
log2(Fold-change) of transcript abundance of repetitive elements in *stwl* null/wild-type testes. Each data point represents the LFC of a single gene. Genes are plotted according to their superfamily. "All Genes" class plots the LFC of all 9039 genes in the DESeq experiment. Red crossbars show the mean LFC for each group.

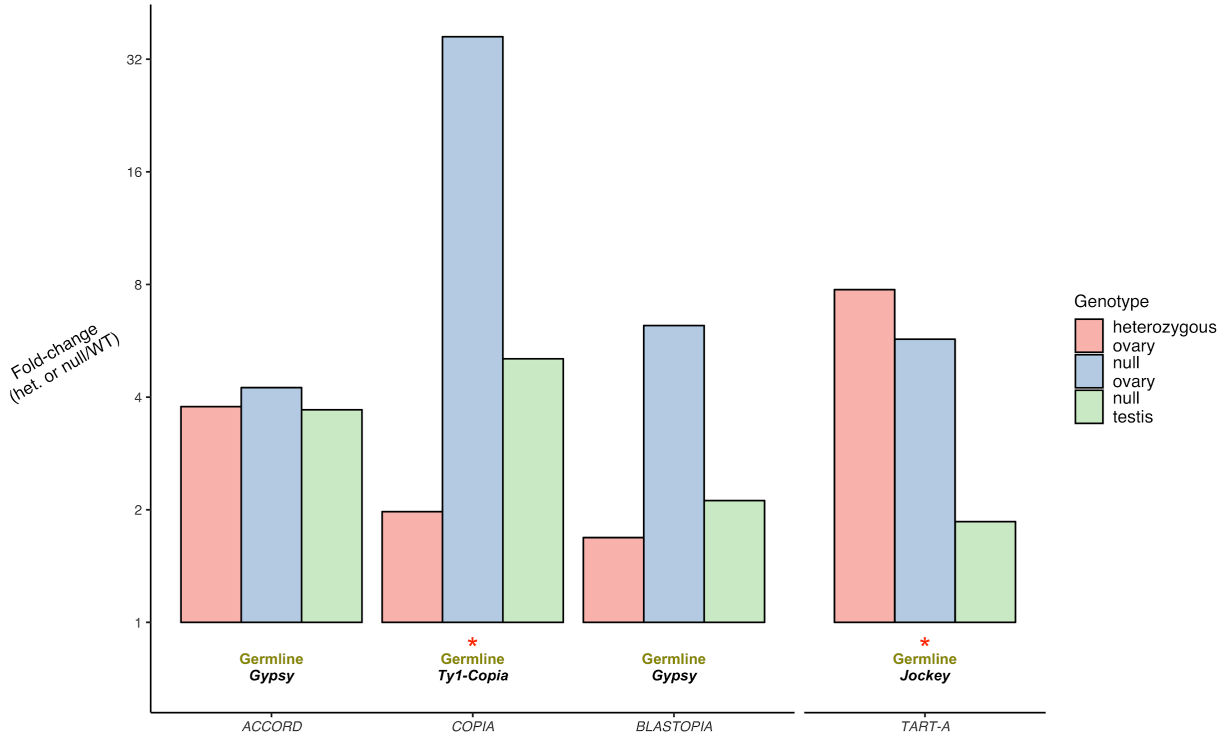


Figure 35. Few repetitive elements are consistently mis-regulated in *stwl* null ovaries, *stwl* heterozygous ovaries, and *stwl* null testis.
Fold-change of transcript abundance of transposable elements. Red asterisks indicate elements that are actively transposing in *D. melanogaster* (Kelleher and Barbash, 2013).

Stwl loss results in derepression of *Copia* transcript but not of *Copia*LTR-*lacZ* construct

We found that the *Copia* element is among the most strongly mis-expressed genes in *stwl* null ovaries. *Copia* is transcriptionally active in *D. melanogaster*, but is silent in *D. simulans* (Díaz-González et al., 2010). We considered the possibility that Stwl is a transcriptional repressor of *Copia* elements, and perhaps that positive selection at the *stwl* locus is driven by differences in *Copia* activity or variation in the *melanogaster* and *simulans* clades. To study this, we obtained genetic transformants carrying the *Copia*-LTR-*lacZ* construct on the X chromosome (Kalmykova et al., 2005; Morozova et al., 2004). This construct expresses *lacZ* under the control

of the *Copia*LTR promoter, but this expression is repressed in wild-type ovaries by the piRNA complex. Disruption of piRNA associated genes, such as *piwi* and *spn-E* results in expression of the *lacZ* reporter, which is revealed by staining the tissue with X-gal (Kalmykova et al., 2005). We found that loss of *stwl* did not result in loss of *CopiaLTR* repression, as it did with a *spn-E* mutant control (Fig. 36). This result suggests that Stwl is not necessary for repression of all *Copia* elements. It may be that Stwl functions in a context dependent manner at particular genomic regions, rather than as a global trans-acting repressor. We also cannot rule out the possibility that Stwl targets internal *Copia* sequence that is absent from this construct.

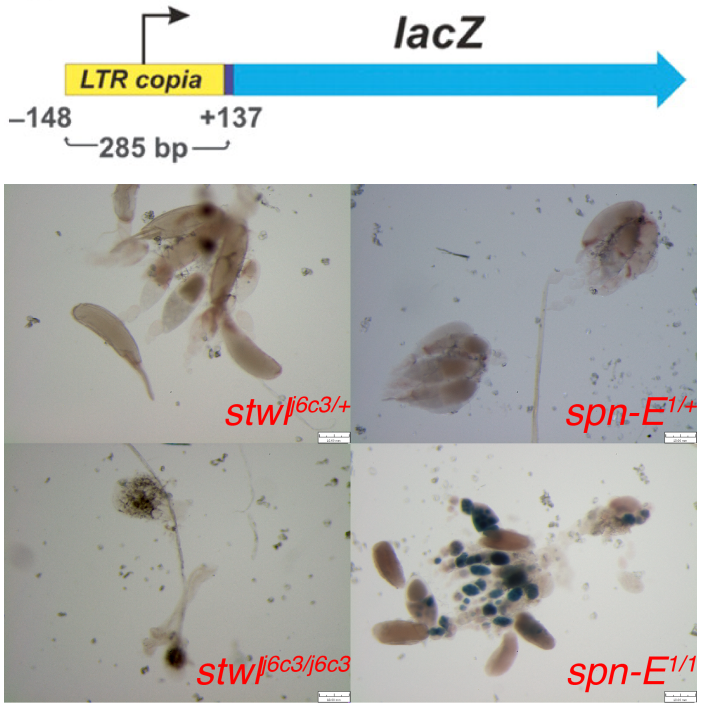


Figure 36. *copia*LTR-*lacz* construct is not mis-regulated in *stwl* null ovaries
Adapted from (Kalmykova et al., 2005). *copia*LTR-*lacZ* construct is a fusion of the *copia*LTR (and the first 137 bp of internal *copia* fragment) with *lacZ*. X-gal staining was conducted on ovaries dissected from the indicated genotypes (*stwl* heterozygotes, *stwl* nulls, *spn-E* heterozygotes and *spn-E* nulls). β-galactosidase activity was found only in the germline of *spn-E* null ovaries.

Comparison of *stwl* to genes required for germline maintenance, differentiation, and TE silencing

The terminal phenotype of *stwl* mutant ovaries is sterility caused by loss of germline stem cells and apoptosis of differentiated germ cells. DNA damage is also apparent in these sterile ovaries, possibly due to Stwl's requirement for maintenance of heterochromatin (Yi et al., 2009)**.** It is also possible that TE derepression is a consequence of these defects, rather than reflecting a direct role of Stwl in TE silencing. In order to compare *stwl* to other GSC maintenance and differentiation genes, we analyzed published RNA-seq data generated from ovaries (and in a single case, testes) for mutations in various proteins affecting GSC maintenance or differentiation. These included the differentiation factor Bam (*bam*), the sex-determination master regulator Sex-Lethal (*Sxl*), the GSC maintenance factor and piRNA targeting protein Piwi (*piwi*), the piRNA pathway components Rhino, Cutoff, and Deadlock (*rhi*, *cuff*, *del*), and the H3K9me3 pathway members HP1a, Setdb1, and Windei (*hp1a*, *egg*, *wde*). *bam*, *Sxl*, *egg*, and *hp1a* deficient ovaries all display the "bag of marbles" phenotype that is characteristic of disruption of differentiation factors that results in overproliferation of GSC-like cells(Gan et al., 2010; Mohn et al., 2014; Peng et al., 2016; Shapiro-Kulnane et al., 2015; Smolko et al., 2018). Mutations in *piwi, rhi*, *cuff* and *del* exhibit GSC maintenance defects similar to those in *stwl* mutants.

After performing normalization and differential expression analysis as described above, we calculated shrunken LFC values for each experimental condition, relative to its matched wild-type control. We found that many of the assayed mutations and gene knockdowns resulted in upregulation of transposable elements (Figs. 37-38). *bam* and *piwi* null ovaries in particular were enriched for most of the same classes of TEs as *stwl* null and heterozygous ovaries. An important

caveat is that *rhi*, *cuff, del*, *hp1a*, *egg*, and *wde* deficient ovaries were produced via germline-specific RNAi (germline-knockdown, or GLKD). Furthermore, these experiments had only one (*rhi/cuff/del*) or two (*hp1a/egg/wde*) replicates, which limits the confidence of the resultant LFC values.

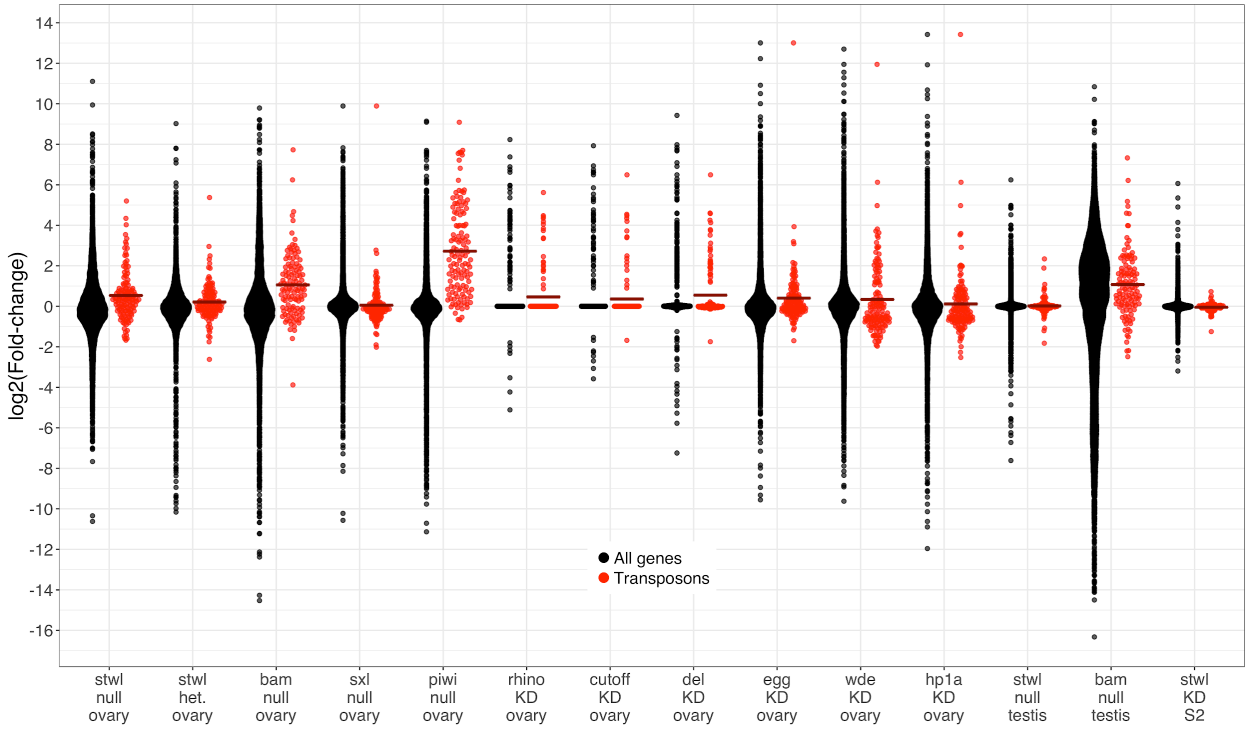
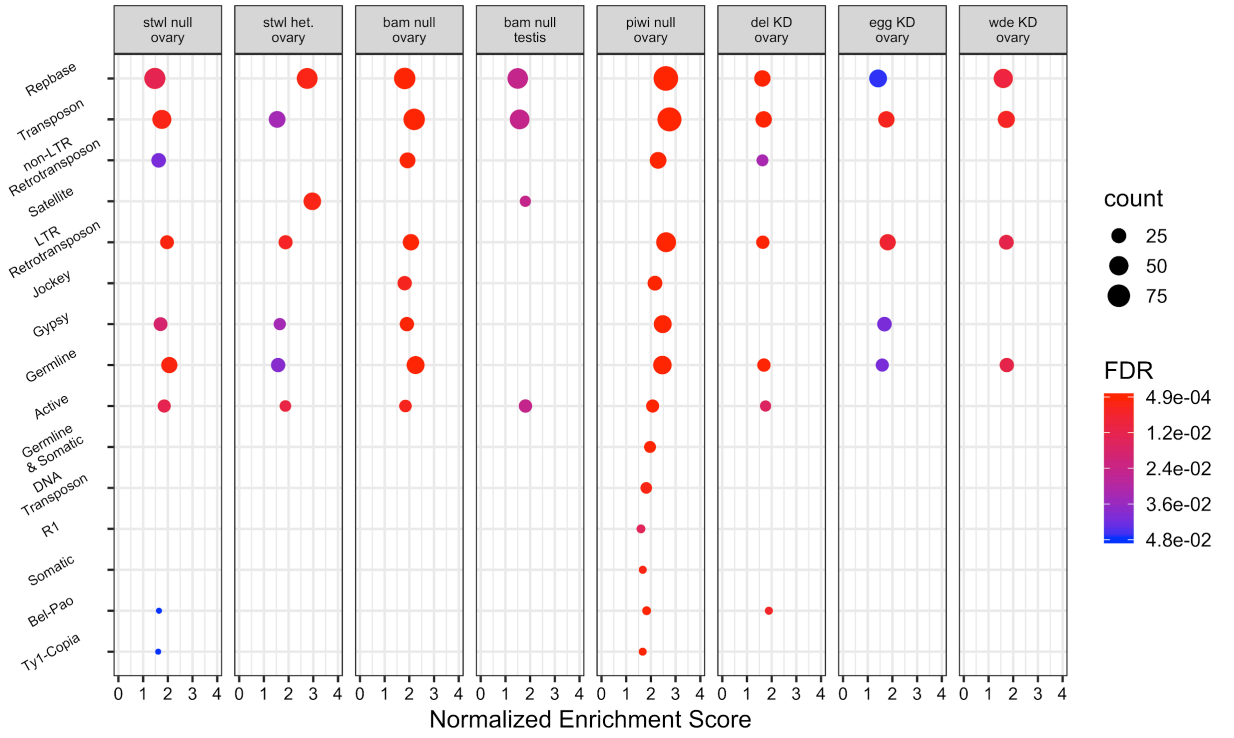


Figure 37. TE de-repression in *stwl* mutants compared to GSC differentiation and piRNA pathway mutants

log2(Fold-change) of transcript abundance of transposons (red) vs all transcripts, including transposons (black) for each differential expression analysis. Each data point represents the LFC of a single gene. Dark red crossbars show the mean LFC for all transposons in the given differential expression experiment.



<u>Figure 38</u>. TE de-repression in *stwl* null and heterozygous ovaries compared to GSC differentiation and piRNA pathway mutants

Comparison of GSEA results from our *stwl* datasets with RNA-seq data from published datasets. Normalized Enrichment Score is plotted for each set of repetitive elements enriched among null/WT, heterozygous/WT, or germline knockdown (KD)/WT ovaries and testes. Only gene sets with FDR<.05 are plotted. *stwl* null testes, *Sxl* null ovaries, *rhi* KD ovaries, *hp1a* KD ovaries, and *cuff* KD ovaries are not shown because they were not enriched for any repeat classes.

We also performed pairwise comparisons between shrunken LFC values of all mutant/knockdown experiments to determine how similar the changes in transcriptional profiles are between experimental conditions (Fig. 39). As expected, these comparisons revealed that data generated from the same bioproject are the most highly correlated: the H3K9me3 pathway components *egg*, *wde*, and *hp1a* vary in LFC similarity from 0.54 to 0.74, while the piRNA pathway components *rhi*, *cuff*, and *del* vary in LFC similarity from 0.57 to 0.73. For each of these two groups, all samples were generated and processed in the same laboratories, and LFC values were calculated against the same wild-type samples, thus reducing the amount of internal

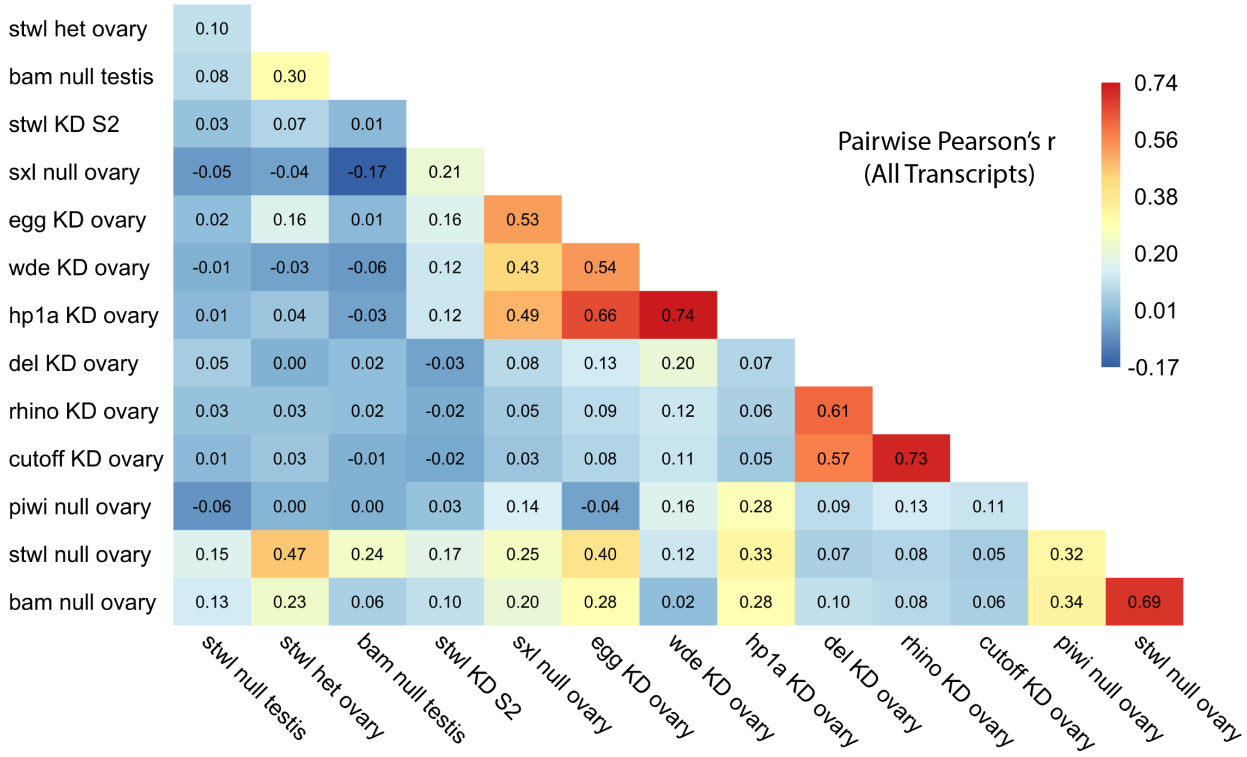
technical variability relative to other pairwise comparisons. *Sxl* functions upstream of the H3K9me3 pathway in ovarian germline; this is supported by the relative similarity of LFC values in *Sxl* mutant ovaries to *egg*, *wde*, and *hp1a* deficient ovaries (0.53, 0.43, 0.49, respectively). Based on these comparisons, we establish that correlation coefficients >0.5 suggest a relatively high degree of correlation in transcriptional changes, while a value between 0.3 and 0.5 suggests moderate correlation.

Surprisingly, *piwi* mutant ovaries shared very little similarity with *rhi*, *cuff*, and *del* GLKD ovaries, despite all genes being critical components of the piRNA pathway. While this can be partly attributed to technical variability (the samples were generated and processed in different labs, LFC was calculated against different wild-type samples, the number of replicates and manner of deficiency differs), it is notable that *piwi* has multiple functions in oogenesis, including GSC maintenance and differentiation, that are separate from its role in piRNA-mediated silencing (Klenov et al., 2011; Ma et al., 2014b; Peng et al., 2016). *piwi* and *bam* mutant ovaries are moderately correlated (Pearson's r = 0.34), which may reflect similarities in their requirement for stem cell differentiation.

Surprisingly, we found that transcriptional changes between *bam* and *stwl* null ovaries are highly correlated (Pearson's r = 0.69). *bam* mutant ovaries are replete with undifferentiated germline cells, while *stwl* mutant ovaries are generally devoid of them. The similarities in transcriptional changes between these two mutants may therefore reflect an as-yet unknown similarity in the function of the two gene products. We also found moderate similarities between *stwl* null ovaries and *piwi*, *egg*, and *hp1a* deficient ovaries (Pearson's r = 0.32, 0.40, 0.33, respectively). The observation that *piwi* and *egg* null ovaries have no correlation in LFC makes

this observation difficult to interpret, but subsetting the data by gene class may help to explain these correlations.

Correlations for *stwl* null testes were not significant with *bam* null testes or any of the ovarian samples tested. Nor were correlations with RNAi-induced knockdown of *stwl* in S2 tissue-culture cells (described in a further section). As expected, *stwl* heterozygous ovaries were moderately similar to *stwl* null ovaries, despite an absence of oogenesis defects in these tissues.



<u>Figure 39</u>. Pearson's r reveals strong correlation in LFC between *stwl* null and *bam* null ovary datasets across all genes and repeats.
Differential expression analysis was performed as described above for all datasets; all $\log_2$(Fold-Change) values were calculated as LFC(mutant/wild-type), using wild-type tissues sequenced with the given experiment. Sample-to-sample Pearson's r was calculated for all pairwise comparisons of shrunken $\log_2$(Fold-Change) values for the displayed datasets. Samples were clustered according to Euclidean distance.

We further performed comparisons restricted to TE transcripts (Fig. 40). We found that *rhi/cuff/del* and *egg/wde/hp1a* generally have higher within-group similarity for repeats than for all transcripts. This relative increase in correlation is indicative of the critical role that these piRNA pathway and heterochromatin silencing components play in TE repression (Fig. 41). TE transcript changes between *piwi* mutant ovaries and each of these six GLKD ovaries are also more strongly correlated than among all transcripts. As expected, genotypes that exhibit a strong TE derepression phenotype are more similar to each other in terms of TE transcript abundance than they are when all transcripts are taken into account, even when the transcript changes are relatively dissimilar as a whole (Fig. 41). Similarly, changes in the TE-derived transcriptome of *stwl* null ovaries are moderately similar to most of the ovary genotypes exhibiting a TE derepression phenotype, including *rhi/cuff/del*, *wde,* and *piwi* (Fig. 40). The strongest similarities are with *piwi* null and *wde* knockdown ovaries (Pearson's r = 0.35, 0.39, respectively), supporting the GSEA finding that mutations in these genes cause misexpression of similar classes of repeats (Fig. 38). Interestingly, correlations between *stwl* null ovaries and *rhi/cuff/del/wde* GLKD ovaries are much higher for repeat-derived transcripts relative to all transcripts (Fig 41).

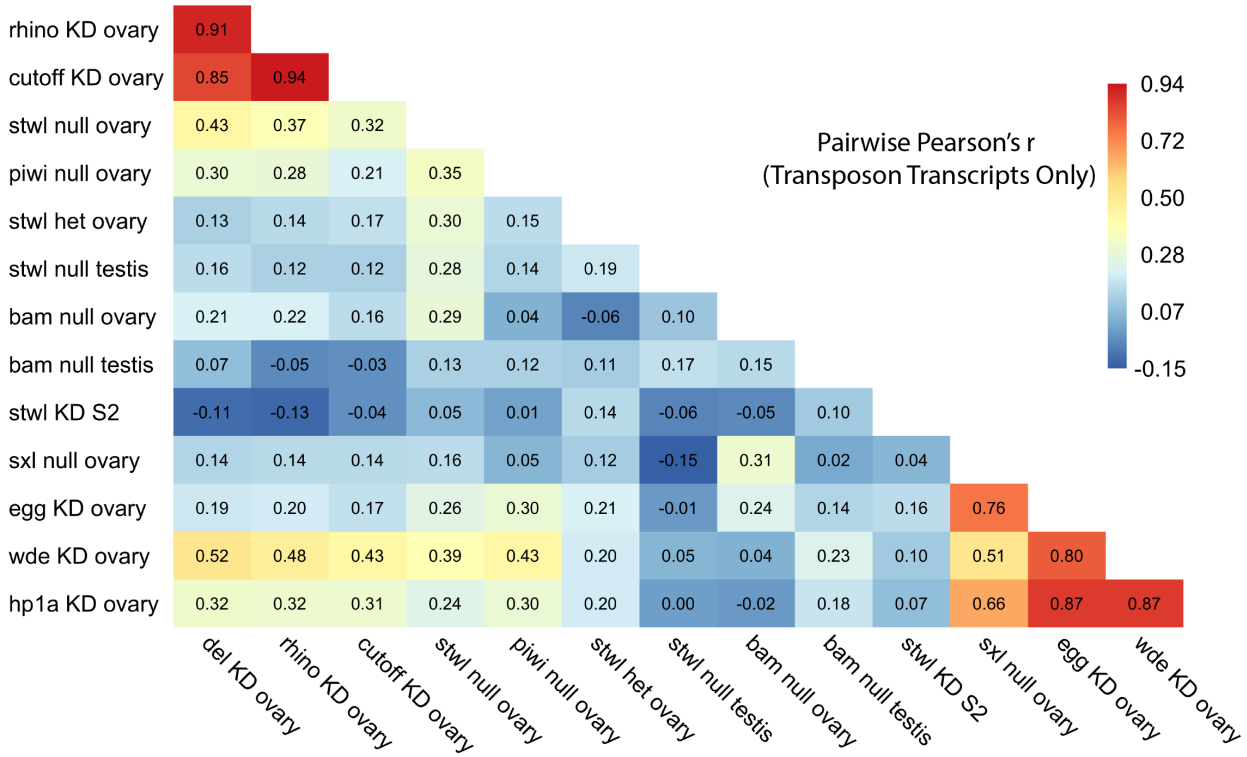


Figure 40. LFC of transposon transcripts in *stwl* mutant ovaries relative to GSC and piRNA pathway mutants

Differential expression analysis was performed as described above for all datasets; all $\log_2$(Fold-Change) values were calculated as LFC(mutant/wild-type), using wild-type tissues sequenced with the given experiment. Sample-to-sample Pearson's r was calculated for all pairwise comparisons of shrunken $\log_2$(Fold-Change) values for the displayed datasets. Here, Pearson's R was calculated only for transposable element transcripts. Samples are clustered according to Euclidean distance.

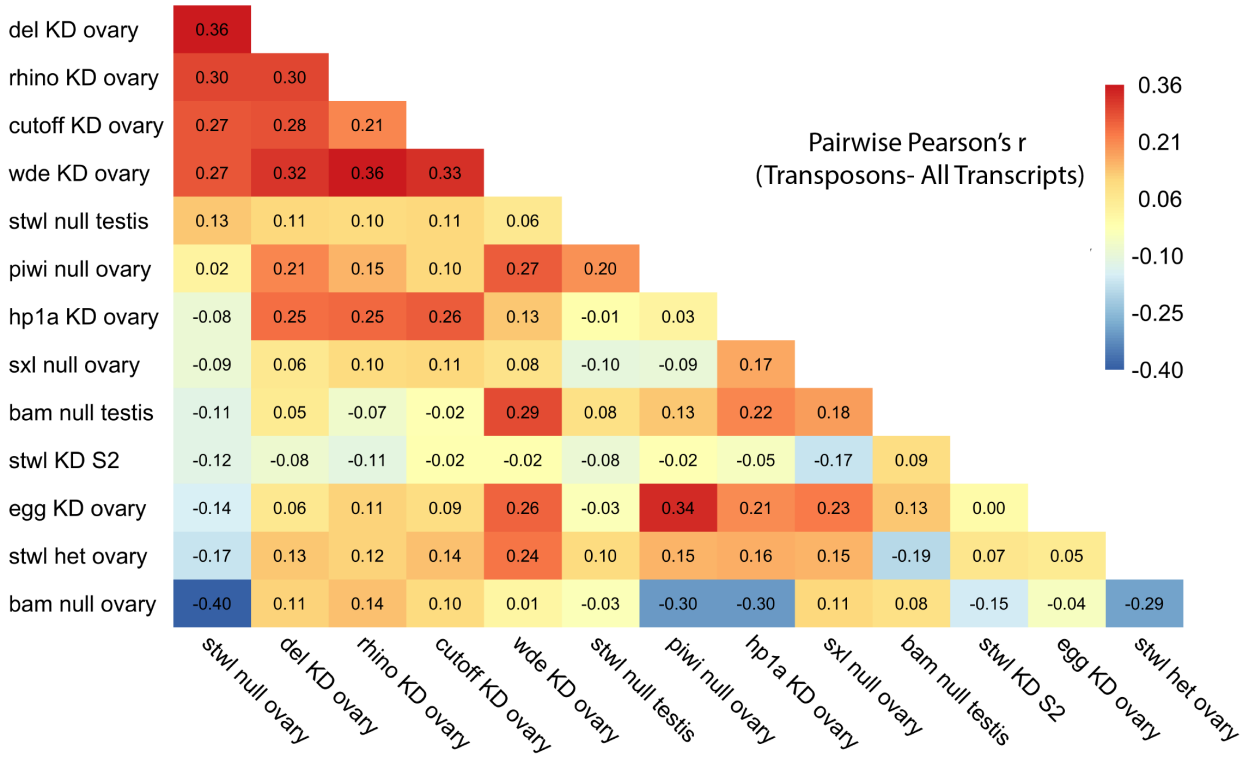


Figure 41. LFC of transposon transcripts relative to all transcripts
Pearson's r for transposable element transcripts minus Pearson's r for all transcripts is shown. Positive values identify pairwise comparisons where LFC of TE transcripts are more similar than LFC of all transcripts. The datasets are ordered by increase in similarity to *stwl* null ovary.

We also performed comparisons restricted to satellite-derived transcripts (Fig. 42). Similar to transposon-restricted comparisons, we found that *rhi/cuff/del* generally have higher within-group similarity for satellites than for all transcripts, which may be indicative of the critical role that piRNA pathway components play in satellite transcript repression (Fig. 43). Satellite transcript changes between *piwi* mutant ovaries and each of these three GLKD ovaries are also more strongly correlated than among all transcripts. Interestingly, *stwl* heterozygous ovaries are moderately correlated with *wde* GLKD ovaries among satellite-derived transcripts, despite the two having no correlation when all transcripts are considered. Correlations between *stwl* heterozygous ovaries and *sxl* null ovaries, as well as *wde/hp1a/egg* GLKD ovaries are much higher for satellite-derived transcripts relative to all transcripts (Fig 43).

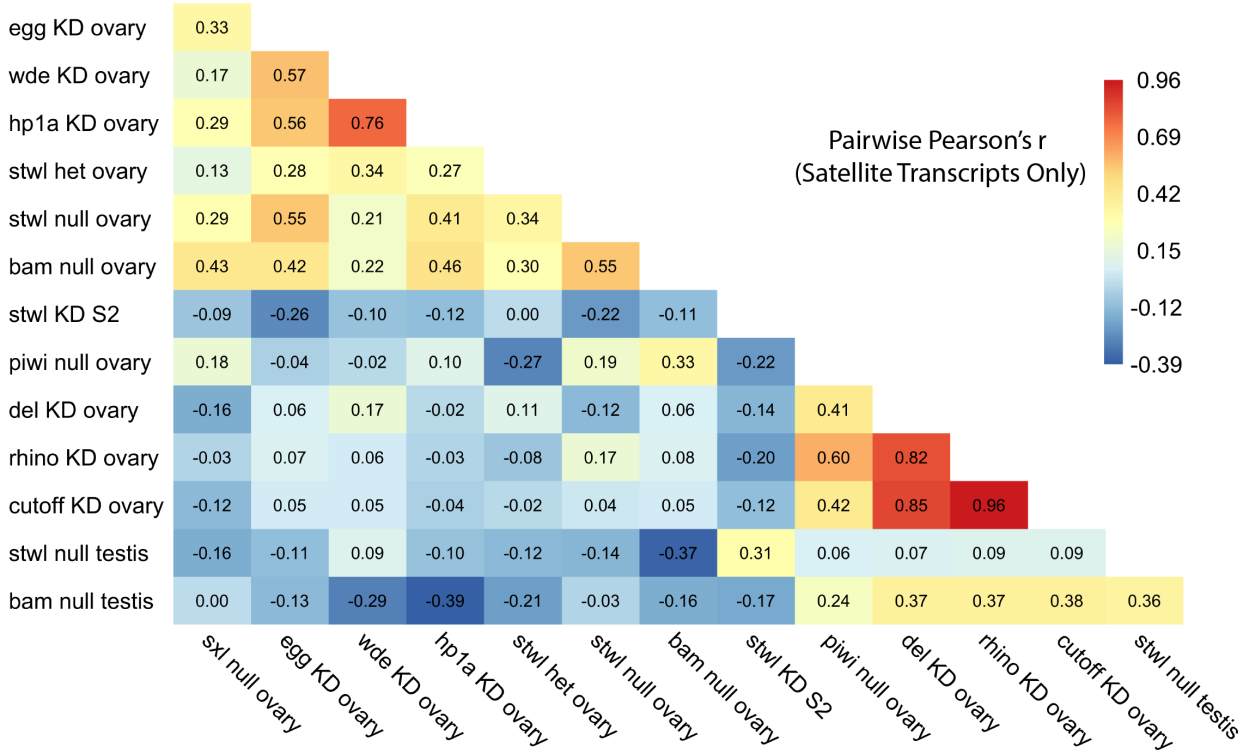


Figure 42. LFC of satellite transcripts in *stwl* mutant ovaries relative to GSC and piRNA pathway mutants

Differential expression analysis was performed as described above for all datasets; all $\log_2$(Fold-Change) values were calculated as LFC(mutant/wild-type), using wild-type tissues sequenced with the given experiment. Sample-to-sample Pearson's r was calculated for all pairwise comparisons of shrunken $\log_2$(Fold-Change) values for the displayed datasets. Here, Pearson's R was calculated only for Satellite-derived transcripts. Samples are clustered according to Euclidean distance.

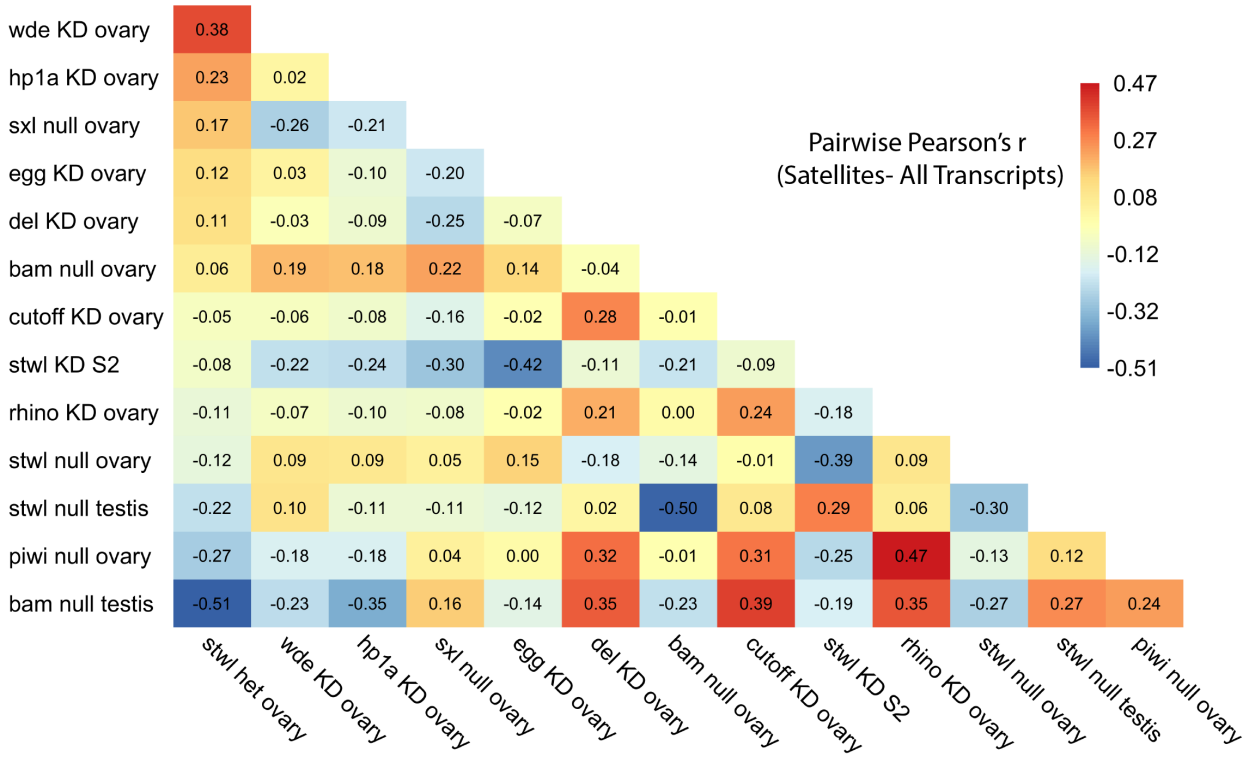


Figure 43. LFC of satellite transcripts relative to all transcripts

Pearson's r for satellite-derived transcripts minus Pearson's r for all transcripts is shown. Positive values identify pairwise comparisons where LFC of satellite transcripts are more similar than LFC of all transcripts. The datasets are ordered by increase in similarity to *stwl* heterozygous ovary.

Discussion (Part 1 of 2)

Stwl localizes to and is required for maintenance of heterochromatin (Maines et al., 2007; Yi et al., 2009). We performed qRT-PCR and RNA-Seq experiments on *stwl* heterozygous ovaries, *stwl* null ovaries and *stwl* null testes in order to test whether expression of repeats is affected by *stwl* loss. qRT-PCR identified several TEs that are derepressed in ovaries lacking two functional copies of *stwl* (Fig. 1). In order to account for differences in wild-type and mutant tissue related to aberrant germline cyst formation and germline stem cell loss, we incorporated newly-eclosed ovaries into our experimental design (Figs. 2-11). RNA-Seq analysis confirmed that TEs, including those assayed using qRT-PCR, are upregulated in *stwl* null and heterozygous ovaries (Figs. 12-15, 20-23).

In order to identify subsets of repetitive elements that are preferentially misregulated in *stwl* null and heterozygous ovaries, we categorized each element according to activity, class, germline/somatic expression bias, and TE superfamily (Figs. 16-19, 24-27). Gene Set Enrichment Analysis (GSEA) showed that active transposons, LTR retrotransposons, germline-expressed transposons and *Gypsy* superfamily transposons are all upregulated in *stwl* null and heterozygous ovaries. non-LTR retrotransposons, as well as *Copia* and *Bel-Pao* superfamily elements, are only misregulated in *stwl* null ovaries. As expected, the TE derepression phenotype is stronger in *stwl* null ovaries than in *stwl* heterozygous ovaries (Figs. 28-29).

Surprisingly, we found that transcripts derived from satellite DNA are strongly upregulated in *stwl* heterozygous, but not *stwl* null ovaries (Figs. 13, 15, 16, 21, 23, 24, 26, 28). An important caveat in interpreting this finding is that satellite transcripts are not typically polyadenylated and thus such transcripts should be preferentially removed during the oligoDT selection step of library preparation. In practice, while enrichment of polyadenylated sequences

via oligoDT selection dramatically increases the relative proportion of mRNA in the total RNA population, ribosomal RNAs and other non-polyadenylated transcripts are commonly found at high quantities in sequenced cDNA libraries. Since all of our samples were prepared in the same way, it is unlikely that the *stwl* heterozygous RNA samples were processed in such a way that fewer satellite transcripts were removed during oligoDT selection. We therefore surmise that the significant upregulation of satellite transcripts in *stwl* heterozygous ovaries reflects a dominant genetic phenotype of the $stwl^{j6c3}$ allele, as opposed to a technical artifact of library preparation.

Although *stwl* null testes do not display obvious morphological defects in germline production, we have shown that *stwl* deficient males are unable to retain fertility after successive matings. We profiled the transcriptomes of *stwl* null testes relative to wild-type in order to identify whether these changes in fertility are due to a loss of transposon repression. We found that *stwl* null testes are transcriptionally similar to wild-type, and that repetitive transcripts are largely unchanged (Figs. 30-34). We conclude that the reduced fertility of *stwl* null males (assayed under sperm exhaustion) is not caused by genome-wide loss of transposon and/or heterochromatin silencing.

Intriguingly, we found that a group of 20 transposable elements are consistently mis-regulated in *stwl* null and *stwl* heterozygous ovaries (Fig. 29). A majority (13/20) of these TEs are derepressed, germline-expressed retroelements; the magnitude of derepression among these TEs is either consistent between the two genotypes or shows an additive effect of *stwl* loss. This core group of upregulated transposons may reflect TEs that interact more directly with Stwl, relative to other TEs whose misexpression may be activated downstream of Stwl activity. We found that 4 of these TEs are consistently upregulated in each of the *stwl* null testes, *stwl* null

ovaries, and *stwl* heterozygous ovaries (Fig. 35). The *Copia* element is the single most upregulated repetitive transcript in each of the *stwl* null gonadal tissues.

In order to further investigate Stwl silencing of *Copia*, we examined a *Copia*LTR-lacZ construct in a *stwl* null background. We found that *stwl* null ovaries did not overexpress the *Copia*LTR-lacZ construct, as opposed to *spn-E* null ovaries which show a marked increase in *Copia*LTR-lacZ expression. *spn-E* is a critical component of the piRNA pathway, which is responsible for the targeting and suppression of repetitive elements in the genome. *Copia*LTR-lacZ silencing in the ovarian germline depends on piRNA pathway elements *spn-E*, *piwi*, and *aub* (Kalmykova et al., 2005; Klenov et al., 2007). While we cannot rule out the possibility that Stwl is responsible for targeting silencing machinery to internal *Copia* sequence absent from the *Copia*LTR-lacZ construct, we deem it unlikely that Stwl is involved in piRNA-mediated silencing.

To better understand Stwl function, we compared *stwl* mutant gonads to gonads deficient (e.g. mutant or GLKD) for proteins with better understood molecular functions. We found that many of these mutant gonads exhibit a TE derepression phenotype, the most dramatic being *piwi* null ovaries (Figs. 37-38). *bam* null ovaries have a very strong TE derepression phenotype, despite not being directly implicated in TE silencing or heterochromatin maintenance. Direct comparisons of $\log_2$(Fold-Change) (LFC) values between RNA-seq experiments found that the *stwl* differential expression profile is most similar to *bam* null ovaries (Fig. 39).

Surprisingly, we found that differential expression of TEs in *stwl* null ovaries, *piwi* null ovaries and *rhi/cuff/del* GLKD ovaries was moderately similar, despite the overall differential expression profile having no correlation (Figs. 40-41). That the degree of upregulation of TEs in *stwl* null ovaries is similar to those of known piRNA pathway components might suggest that

they are involved in a similar regulatory mechanism. Since other evidence suggests that Stwl is not directly involved in piRNA-mediated silencing, the similar differential expression profiles could reflect that both Stwl and piRNA genes are required for heterochromatin maintenance.

Our analysis showed that *stwl* null testes and *bam* null testes did not have similar differential expression profiles. We found that *bam* null testes show weak upregulation of TEs and satellite transcripts. *bam* mutant males are completely sterile as a consequence of the testes filling up with undifferentiated gonialblasts.

One of the more surprising outcomes of our analyses is that *stwl* heterozygous ovaries are enriched for satellite transcripts. While none of our analyses of other ovarian datasets identified specific enrichment of satellite transcripts, we found moderate similarity in differential expression of satellite transcripts between *stwl* heterozygous ovaries and *egg/wde/hp1a* GLKD ovaries. This suggests heterochromatin maintenance may be perturbed in *stwl* heterozygous ovaries via the H3K9me3 pathway. While *stwl* heterozygous females are fertile and their ovaries appear morphologically similar to wild-type, Stwl is a dominant suppressor of position effect variegation (Maines et al., 2007; Yi et al., 2009), which is typically interpreted as suggesting loss of a single functional copy of *stwl* results in heterochromatin spreading. It is therefore understandable that *stwl* heterozygotes exhibit an upregulation of normally heterochromatic sequences, including TEs and satellites. What remains unclear is why upregulation of satellite transcripts is observed in *stwl* heterozygotes but not *stwl* nul ovaries.

Identifying the molecular functions of Stwl is made especially challenging by its pleiotropic functions at various stages of development, including GSC maintenance and oocyte determination. This is further complicated by the downstream consequences of *stwl* loss, i.e. apoptosis of and eventual loss of the female germline. We must take into account that Stwl is

likely performing multiple molecular functions that could muddy the waters in terms of what to expect in steady-state RNA profiles. We sought to tease apart these differences by assaying ovaries at varying degrees of development and identifying those genes which are consistently upregulated between *stwl* null and wild-type ovaries. Nonetheless, it is difficult to infer which transcripts are directly regulated by Stwl and which are misregulated by indirect, downstream consequences of *stwl* loss.

The similarity in differential expression between *stwl* and *bam* mutant ovaries suggests that they may be involved in similar gentic pathways (Fig. 39). The correlation in differential expression profiles is not driven by similarities in TE or satellite transcript misexpression (Figs. 40-43). This supports the hypothesis that Stwl is involved in distinct molecular processes, as suggested by the different transcriptional profiles of *stwl* heterozygous and null ovaries. Such pleiotropic functions are common among GSC maintenance genes; for example, Piwi is able to promote GSC maintenance even when its TE silencing function is inhibited (Klenov et al., 2011). We propose that the upregulation of satellite transcripts and TEs in *stwl* heterozygotes reflects the heterochromatin maintenance function of Stwl, whereas disruption of oogenesis and fertility is caused by a distinct molecular process that functions normally with a single wild-type *stwl* allele. It stands to reason that the molecular pathways in which Stwl functions to maintain oogenesis, either at the stage of germline stem cell retention or oocyte determination, overlap significantly with pathways in which Bam is a crucial actor. Stwl acts independently of and antagonistic to Bam: *bam* mutants present with GSC-tumorous ovarioles, while *stwl bam* double mutants form rudimentary germline cysts (Maines et al., 2007). We will further investigate the relationship between Bam, its binding partner Bgcn, and Stwl in subsequent sections of this manuscript.

Results (Part 2 of 2)

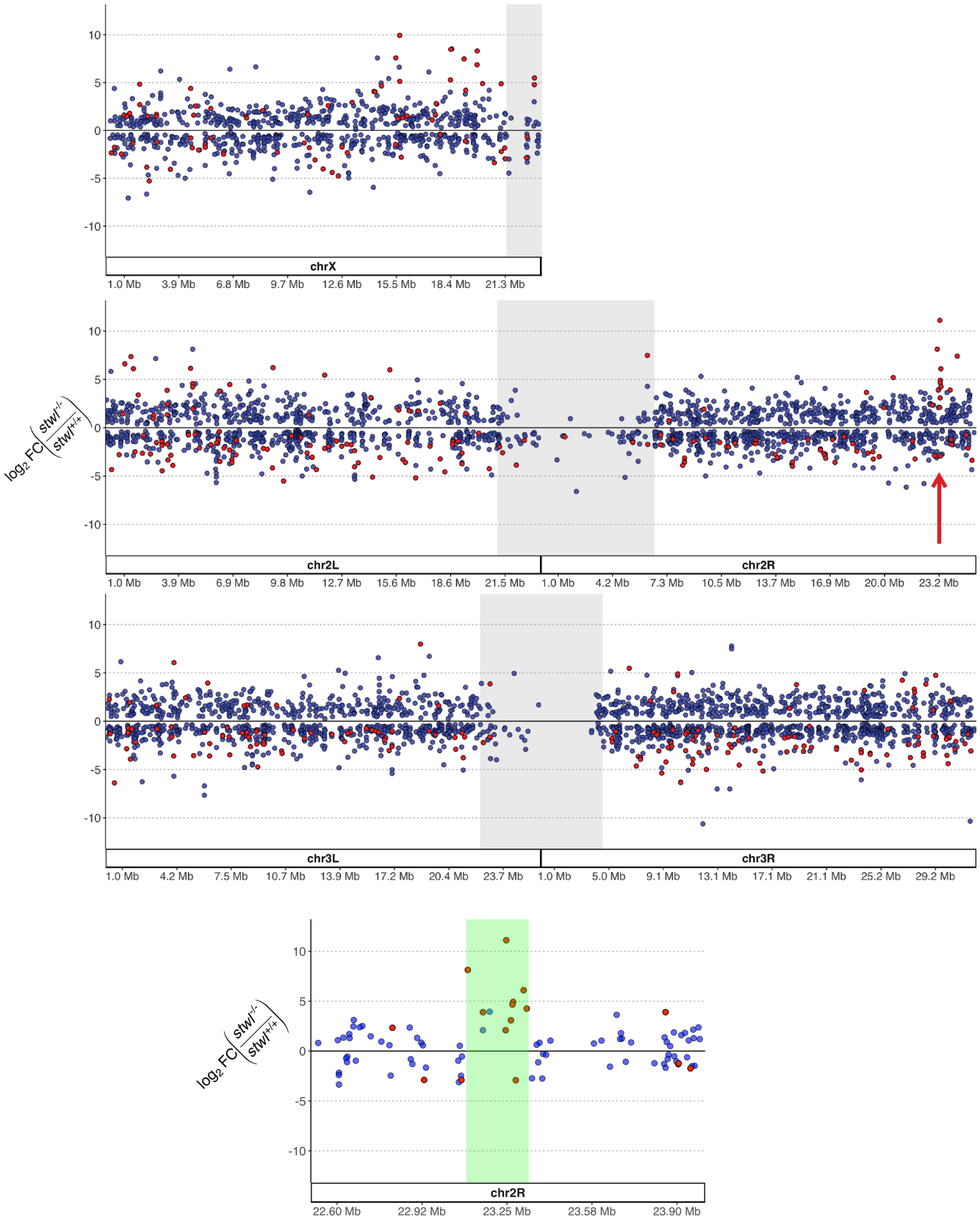
A single cluster of testis-specific genes is strongly upregulated in *stwl* null ovaries

Differential expression analysis in both *stwl* null and heterozygous ovaries identified many misexpressed genes. In order to test whether specific regions of the genome are misregulated, we plotted shrunken LFC by genomic location (Fig. 44). 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. 45).

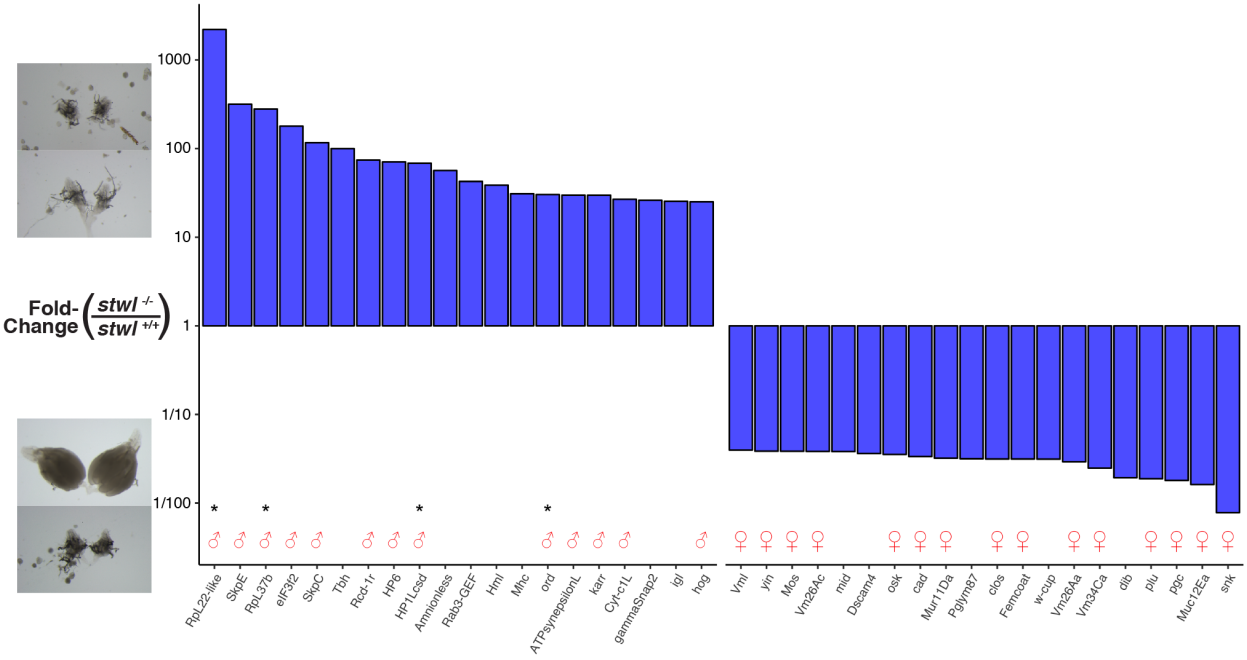
Coexpressed gene clusters are common in many species, adding a dimension of organization to the genome by allowing groups of adjacent genes to be regulated simultaneously. Testis-specific gene clusters are particularly common in *Drosophila*, and their expression is tightly regulated to prevent somatic expression (Shevelyov et al., 2009). Shevelyov et al. showed that the 59C4-59D cluster on chromosome 2R is the largest testis-specific gene cluster in *D. melanogaster*. The upregulation of this testis-specific gene cluster in *stwl* null ovaries led us to interrogate whether testis or other tissue-specific genes are preferentially misregulated in *stwl* null ovaries.

We utilized RPKM values from the modENCODE anatomy RNA-Seq dataset to classify all genes according to tissue-biased expression (Brown et al., 2014). Briefly, a gene was identified as tissue-biased based on the tissue-specificity metric tau, and the tissue(s) in which that gene is enriched were defined as each tissue where the RPKM for that gene is ≥1.5 SD above mean RPKM across all tissues (see Methods for further details). 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 (Tables 3-4).


log2 FC (stwl-/- / stwl+/+)
chrX
chr2L
chr2R
chr3L
chr3R
chr2R

<u>Figure 44</u> (page above) A cluster of testis-specific genes is derepressed in *stwl* null ovaries
Plot of log2(fold-change) of transcript abundance between *stwl* null ovary and wild type ovaries by genomic location (gene midpoint) at chromosome 2R. Only differentially expressed genes (FDR <.01) are plotted. Red dots represent genes identified as preferentially enriched in testis (see methods). Shaded grey areas represent pericentromeric heterochromatin. Top panel displays the X chromosome and autosomes, red arrow points to testis-enriched cluster at 59C4-59D; bottom panel magnifies the area around this cluster, which is highlighted in green..



<u>Figure 45</u>. Testis-specific genes are strongly upregulated in *stwl* null ovaries
Plot of fold-change (on a $\log_{10}$ scale) of transcript abundance between *stwl* null and wild type ovaries. The list excludes genes for non-coding RNA and those encoding uncharacterized proteins; the top 20 most upregulated and downregulated genes are shown. Red male/female symbols identify genes with testis- and ovary-specific expression, respectively; "*" marks genes that are part of the 59C4-59D testis-specific cluster.

	Head	Male Head	Female Head	Carcass	Dig. System	Fat	Salivary Gland	CNS	Larval CNS	Pharate CNS	Ovary	Imaginal Disc	Testis	Acc. Gland	S2R+ Cells
A_MateM_4d_head	x	x													
A_MateF_4d_head	x		x												
L3_Wand_carcass				x											
A_4d_carcass				x											
L3_Wand_dig_sys					x										
A_4d_dig_sys					x										
L3_Wand_fat						x									
WPP_fat						x									
P8_fat						x									
L3_Wand_saliv							x								
WPP_saliv							x								
L3_CNS								x	x						
P8_CNS								x		x					
A_MateF_4d_ovary											x				
L3_Wand_imag_disc												x			
A_MateM_4d_testis													x		
A_MateM_4d_acc_gland														x	
S2R+_cells															x

Table 3. modENCODE mRNA-seq tissue profiles used to identify tissue-enriched genes

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.

Enriched Tissue	Gene Number	% of Genome
Testis	3146	17.3%
Imaginal Disc	2260	12.4%
CNS	1211	6.7%
Pharate CNS	1004	5.5%
Larval CNS	464	2.6%
Fat	1084	6.0%
Head	1049	5.8%
Male Head	1008	5.5%
Female Head	444	2.4%
Digestive System	939	5.2%
Accessory Gland	934	5.1%
Carcass	852	4.7%
S2R+ Cells	744	4.1%
Ovary	649	3.6%
Salivary Gland	372	2.0%
Enriched	8280	45.5%
Not Enriched	6687	36.8%
Not Expressed	3227	17.7%

Table 4. Testis-enriched genes are abundant in *D. melanogaster*

Total number of genes with tissue-specific expression, by tissue category (see Table 3). 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 the methods. % of Genome is based on a total of 18,194 annotated genes.

Subsequent to classifying the genome according to tissue-enrichment, we confirmed that the 59C4-59D cluster described by Shevelyov et al. is composed mostly (28/34 total genes) of testis-enriched genes (Fig. 46). We then compared Reads per Kilobase per Million mapped reads (RPKM) across genes at this locus to more precisely determine how the *stwl* mutation affects gene expression at this locus in testis and ovaries (Fig. 47). Our own expression data confirmed that most (29/32) of these genes are expressed (though not necessarily enriched), in wild-type testis (mean RPKM >2), and that most (29/32) are not expressed in wild-type ovary (Table 5, Fig. 47).

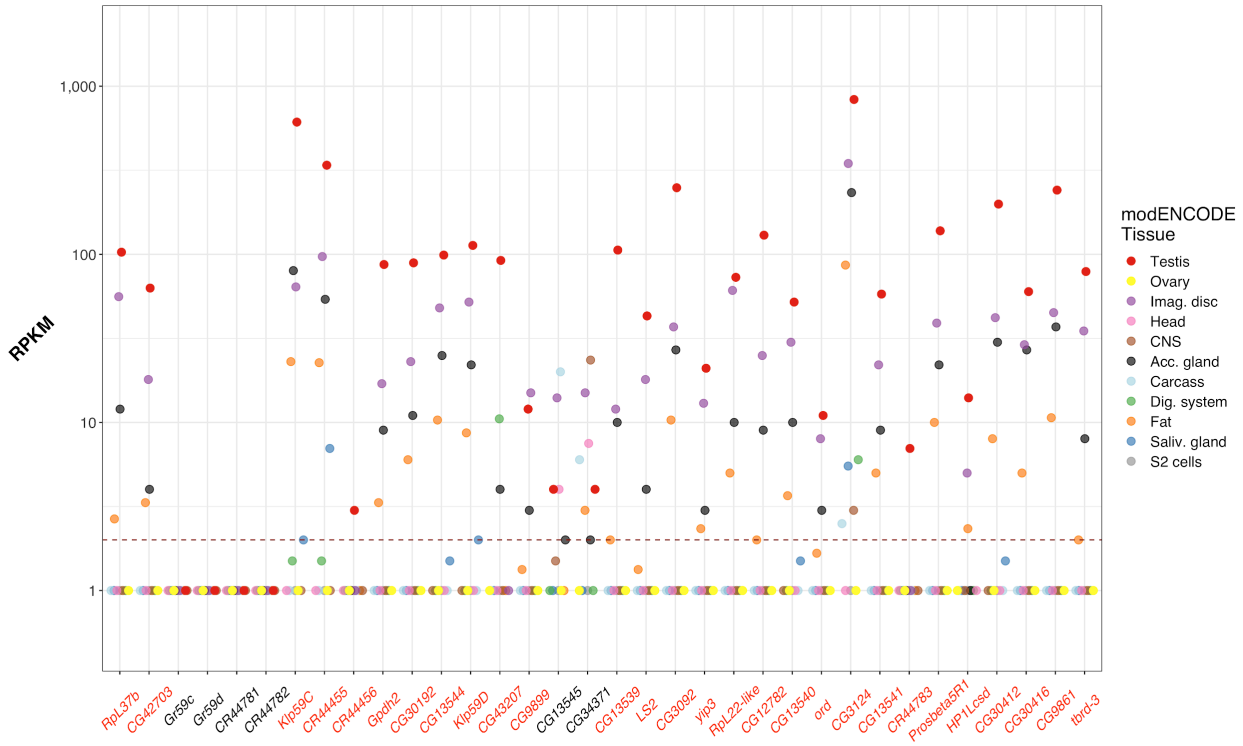


Figure 46. 59C4-59D gene cluster is highly expressed in testis, but inert in ovary. Plot of RPKM (on a $\log_{10}$ scale) of transcripts from the 59C4-59D cluster in tissues from the modENCODE anatomy RNA-Seq. Genes are plotted in order of chromosomal location. Gene names in red are enriched in testis relative to other tissues. Dotted line indicates threshold for ectopic gene expression at RPKM =2. Minimum reported RPKM is 1.

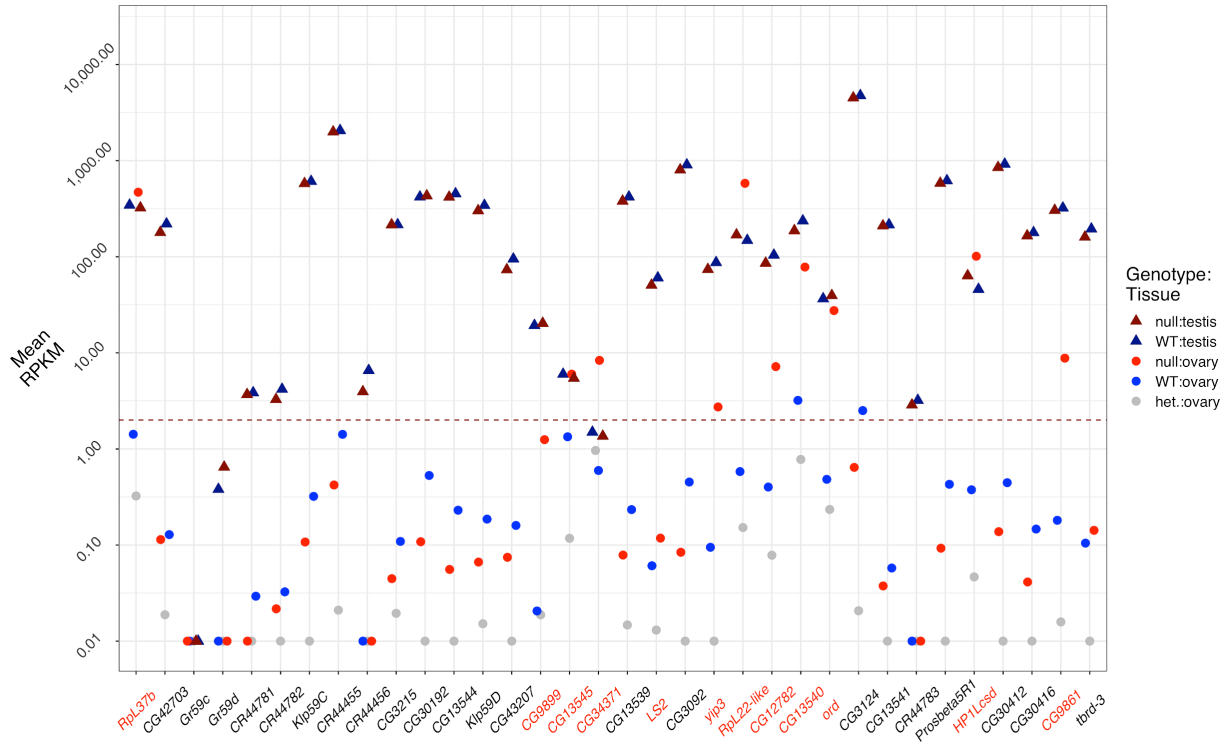


Figure 47. 59C4-59D gene cluster is de-repressed in *stwl* null ovaries.
Plot of RPKM (on a $\log_{10}$ scale) of transcripts from the 59C4-59D cluster in wild-type (blue), *stwl* null (red), and *stwl* heterozygous (gray) tissue. Genes are plotted in order of chromosomal location. Dotted line indicates threshold for ectopic gene expression at RPKM =2. Most of the genes in the cluster are highly expressed in wild-type and *stwl* null testis, and transcriptionally silent in wild-type and *stwl* heterozygous ovaries. *stwl* null ovaries exhibit a de-repression of genes indicated in red to levels similar to wild-type testis.

Symbol	Gene Name	Dist. to next gene (Kb)	Enriched tissue (modENCODE)	Expressed in WT	Upreg. in *stwl* mutant	Ectopic in *stwl* mutant
RpL37b	Ribosomal protein L37b	0.75	Testis, Imag. Disc	Testis	null ovary	null ovary, S2 cells
CG42703	uncharacterized protein	0.70	Testis, Imag. Disc	Testis	*NA*	*NA*
Gr59c	Gustatory receptor 59c	0.21	Not Expressed	None	*NA*	*NA*
Gr59d	Gustatory receptor 59d	4.81	Not Expressed	None	*NA*	*NA*
CR44781	long non-coding RNA:CR44781	2.40	Not Expressed	Testis	*NA*	*NA*
CR44782	long non-coding RNA:CR44782	9.55	Not Expressed	None	*NA*	*NA*
Klp59C	Kinesin-like protein at 59C	1.56	Testis	Testis	*NA*	*NA*
CR44455	long non-coding RNA:CR44455	0.00	Testis, Imag. Disc	Testis	*NA*	*NA*
CR44456	long non-coding RNA:CR44456	0.13	Testis	Testis	*NA*	*NA*
Gpdh2	Glycerol-3-phosphate dehydrogenase 2	8.40	Testis, Imag. Disc	Testis	*NA*	*NA*
CG30192	uncharacterized protein	0.00	Testis, Imag. Disc	Testis	*NA*	*NA*
CG13544	uncharacterized protein	1.42	Testis, Imag. Disc	Testis	*NA*	*NA*
Klp59D	Kinesin-like protein at 59D	8.25	Testis, Imag. Disc	Testis	*NA*	*NA*
CG43207	uncharacterized protein	0.54	Testis, Digestive System	Testis	*NA*	*NA*
CG9899	uncharacterized protein	0.00	Testis, Imag. Disc	Testis	null ovary	null ovary
CG13545	uncharacterized protein	4.83	Imag. Disc, Carcass	Testis	null ovary	null ovary, het. ovary
CG34371	sidestep V	0.00	Not Enriched	Testis, Ovary	null ovary, het. ovary	het. ovary
CG13539	uncharacterized protein	24.84	Testis	Testis	*NA*	*NA*
LS2	Large Subunit 2	21.09	Testis, Imag. Disc	Testis	*NA*	null ovary
CG3092	uncharacterized protein	10.12	Testis, Imag. Disc	Testis, Ovary	*NA*	*NA*
yip3	yippee interacting protein 3	1.21	Testis, Imag. Disc	Testis	null ovary	null ovary
RpL22-like	Ribosomal protein L22-like	16.38	Testis, Imag. Disc	Testis	null ovary	null ovary, S2 cells
CG12782	uncharacterized protein	4.93	Testis, Imag. Disc	Testis	null ovary	null ovary
CG13540	uncharacterized protein	0.20	Testis, Imag. Disc	Testis	null ovary	null ovary
ord	orientation disruptor	7.21	Testis, Imag. Disc	Testis	null ovary	null ovary, S2 cells, het. ovary
CG3124	uncharacterized protein	0.11	Testis, Imag. Disc	Testis, Ovary	*NA*	*NA*
CG13541	uncharacterized protein	4.86	Testis, Imag. Disc	Testis	*NA*	*NA*
CR44783	long non-coding RNA:CR44783	17.28	Testis	None	*NA*	*NA*
Prosbeta5R1	Proteasome beta5 subunit-related 1	3.20	Testis, Imag. Disc	Testis	*NA*	*NA*
HP1Lcsd	Heterochromatin Protein 1L chromoshadow domain	4.34	Testis, Imag. Disc, Fat	Testis	null ovary, S2 cells	null ovary, S2 cells, het. ovary
CG30412	uncharacterized protein	2.13	Testis, Imag. Disc	Testis	*NA*	*NA*
CG30416	uncharacterized protein	0.26	Testis, Imag. Disc, Acc. Gland	Testis	*NA*	*NA*
CG9861	uncharacterized protein	0.00	Testis, Imag. Disc	Testis	null ovary	null ovary, S2 cells
tbrd-3	testis-specifically expressed bromodomain containing protein-3	*NA*	Testis, Imag. Disc	Testis	*NA*	*NA*

Table 5. Summary of genes and expression phenotypes 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.

Loss of H3K9me3 pathway components results in ectopic expression of testis-enriched genes (Shapiro-Kulnane et al., 2015; Smolko et al., 2018). Identification of ectopically expressed genes (genes expressed in a tissue where they are normally not expressed) presents unique challenges, due to the low counts of these genes. We defined ectopic gene expression as follows: 1) mean RPKM < 2.0 in wild-type tissue, wilcoxon p value <0.1; 2) RPKM increases by at least 2-fold in mutant tissue; 3) for all genes that meet criteria 1) and 2), RPKM significantly increases in mutant, BH-adjusted wilcoxon p-value <0.25. We found that 11/34 of the genes in the 59CD-59D gene cluster meet this criteria for ectopic expression in *stwl* null ovaries (Table 5). These data indicate that *stwl* is required to maintain silencing of transcripts from the 59CD-59D gene cluster in ovaries. Further examination of differential gene expression at the 59C4-59D cluster revealed that this cluster is also up-regulated in *bam*, *sxl*, and *piwi* null ovaries, as well as *egg, wde*, and *hp1a* GLKD ovaries (Fig. 48).

A small number of testis-enriched genes are among the most upregulated genes in *stwl* null ovaries

Following the discovery of strong transcript de-repression at the 59C4-59D gene cluster in *stwl* and other GSC pathway mutant ovaries, we considered the possibility that testis-enriched genes, clustered or otherwise, are generally affected by mutations in these genes. We implemented the algorithm ClusterScan, using the function “clusterdist” to identify clusters of 2 or more testis-enriched genes, at a maximum distance of 20 kb between members of a cluster (Volpe et al., 2018). We found that the majority (2295/3144, 73%) of testis-enriched genes are clustered, according to this definition. These clusters also contain a considerable number of

"bystanders", which are genes that are not testis-enriched, but are nonetheless contained in these gene clusters. We did not identify any additional testis-enriched clusters with a de-repression effect similar to the one seen at 59C4-59D (Fig. 49).

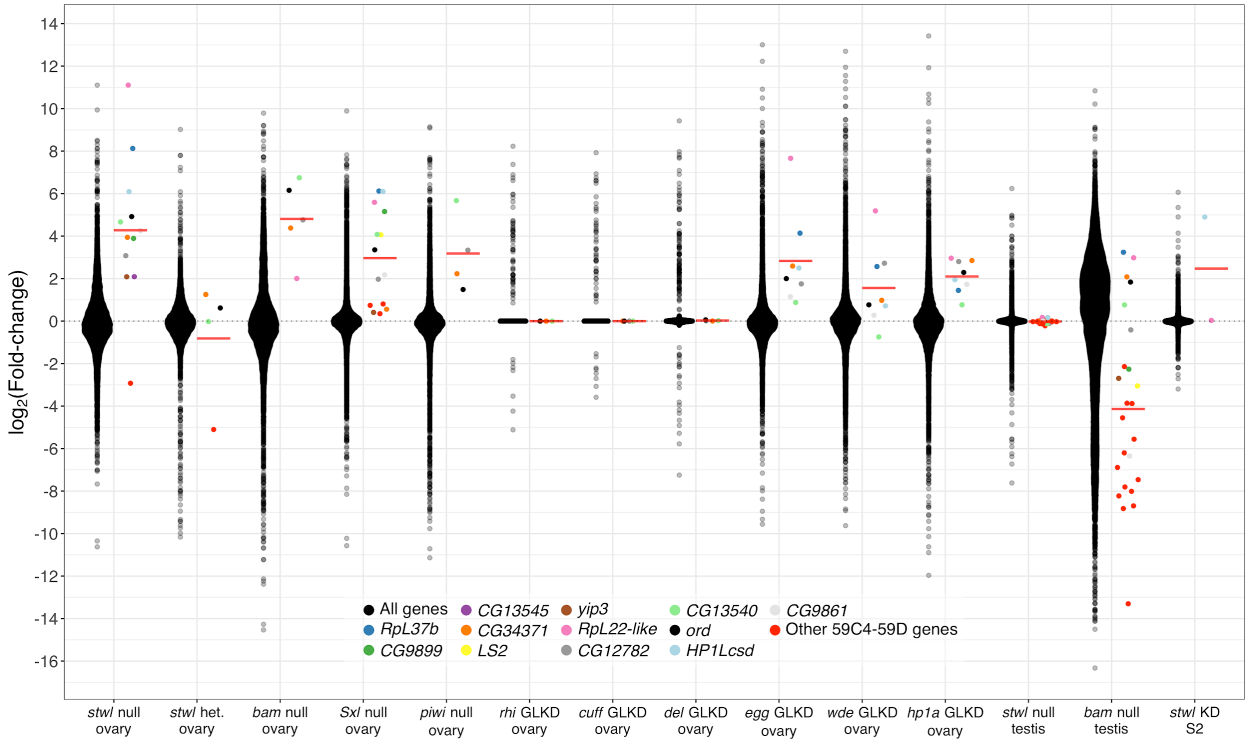


Figure 48. 59C4-59D gene cluster is de-repressed in GSC pathway mutants
$\log_2$(Fold-change) of transcript abundance from genes from the 59C4-59D cluster (multiple colors) vs all transcripts (black) for each differential expression analysis. Each data point represents the LFC of a single gene. Red crossbars show the mean LFC for all transcripts from the 59C4-59D cluster in the given differential expression experiment. Genes with low mean transcript abundance within a given experiment are not shown.

We also examined the spread of LFC values for testis-enriched genes and genes in testis-enriched clusters (including bystanders) across GSC and piRNA pathway mutant/GLKD phenotypes (Fig. 50). We confirmed that *sxl* null ovaries and *egg/wde/hp1a* GLKD exhibit up-regulation of testis-enriched genes, as previously reported (Smolko et al., 2018) (Fig. 52).

Disruption of piRNA pathway components *piwi, rhi*, *cuff*, and *del* does not result in mis-regulation of testis-enriched genes. Surprisingly, *stwl* null, *stwl* heterozygous, and *bam* null ovaries all experience a down-regulation of testis-enriched transcripts. Nonetheless, testis-enriched transcripts, including those at the 59C4-59D cluster are among the most highly up-regulated transcripts in *stwl* null ovaries (Fig. 45).

While the apparent enrichment of testis-enriched transcripts at the top of the range of LFC values in *stwl* null ovaries would normally be detected by GSEA, this phenotype is masked by the general down-regulation of testis-enriched transcripts in *stwl* null ovaries. This flaw in GSEA performance has been previously noted when attempting to perform analyses on very large and potentially complex gene sets (Hong Guini et al., 2014; Warden et al., 2013). The essential problem is that these sets include genes that are mis-regulated in either direction, presumably due to the fact that members of the same pathway may be down- or up-regulated in response to mis-expression of an upstream activator or suppressor. We performed an over-representation test for tissue-enriched genes among the top and bottom 1% and 5% of expressed genes in each RNA-Seq experiment, to identify strong biases at the tips of the LFC ranges (Fig. 53).

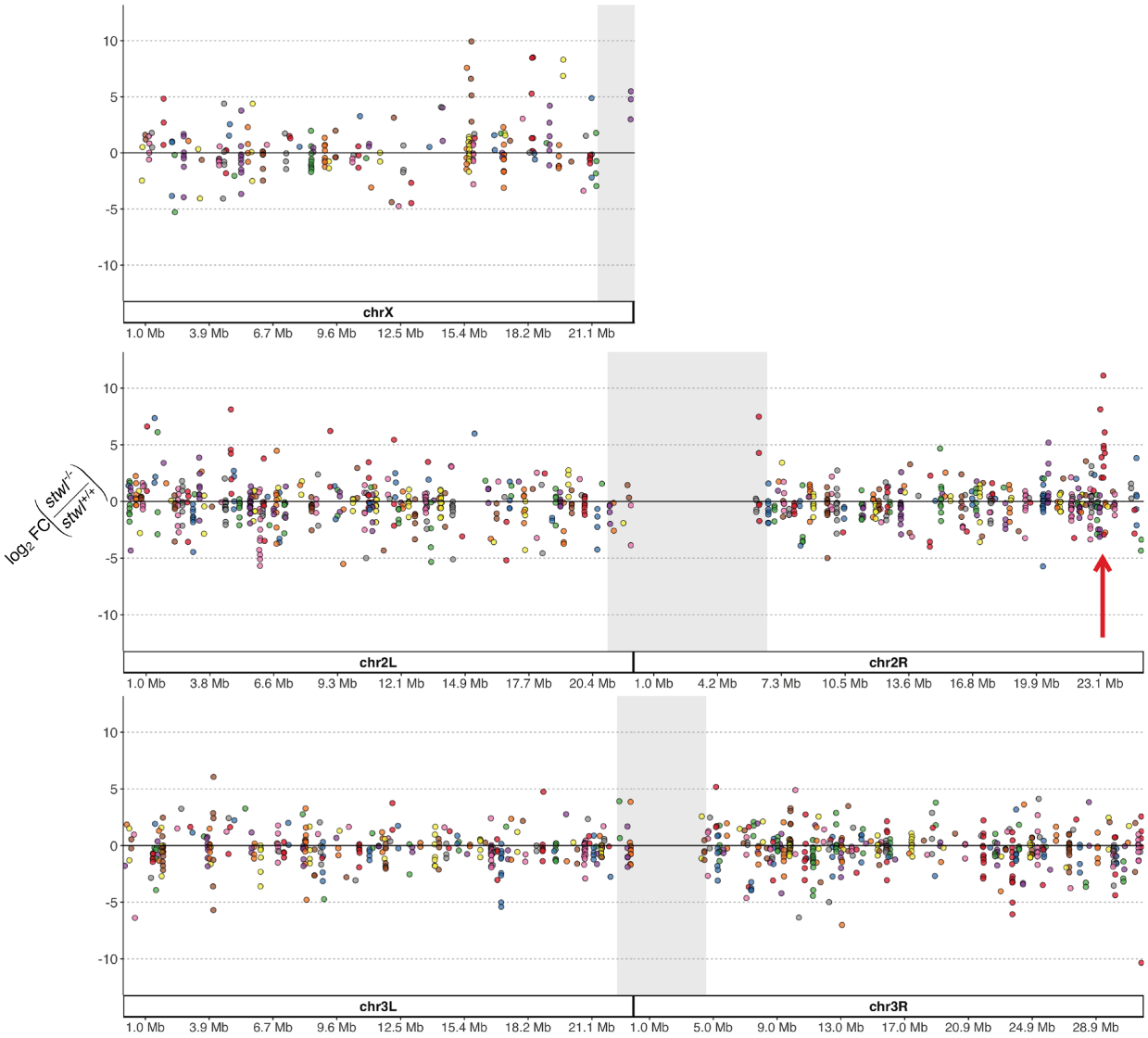


Figure 49. LFC of transcripts from testis-enriched clusters in *stwl* null ovaries
$\log_2$(Fold-change) of transcript abundance from genes in testis-enriched gene clusters, including bystanders. Clusters are defined as 2 or more testis-enriched genes within 20 kb distance from each other. Each point presents LFC of a single gene, all adjacent points of the same color are from a shared gene cluster. Red arrow at ~23 Mb on chromosome 2R points to the 59C4-59D cluster. Gray shaded rectangles indicate pericentric heterochromatin.

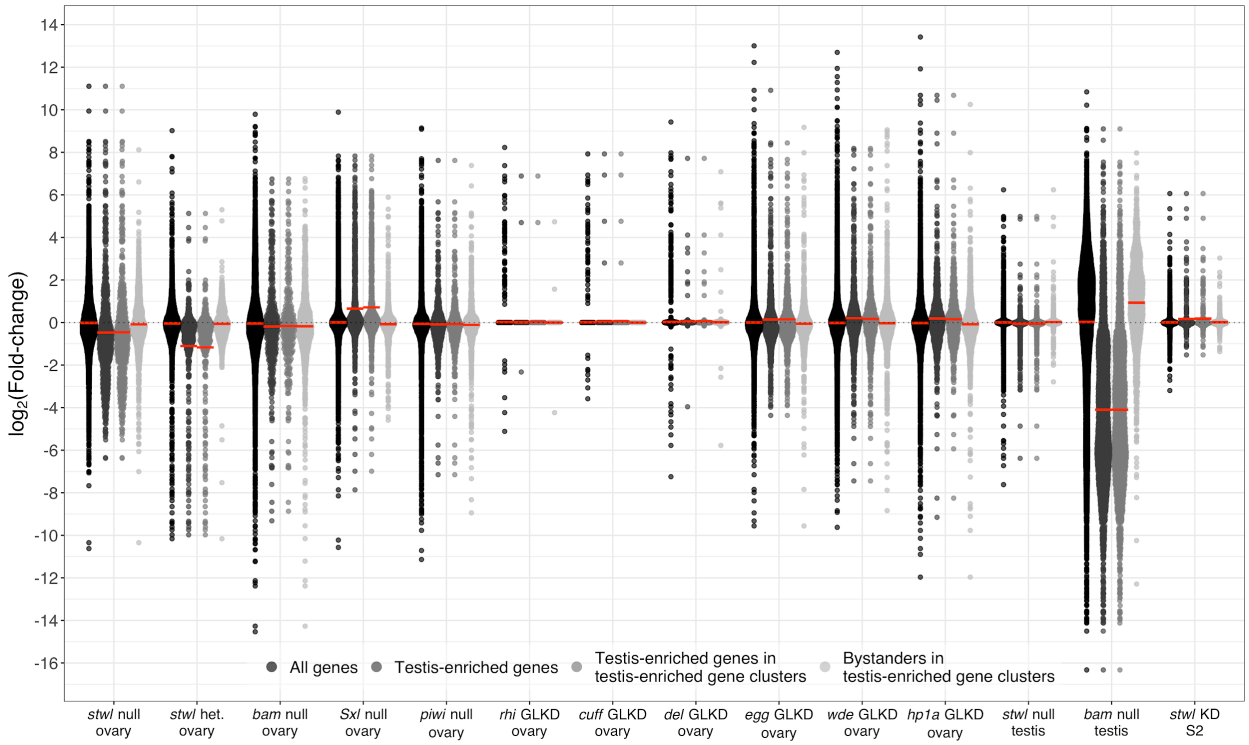


Figure 50. LFC of transcripts from testis-enriched genes and clusters in *stwl* null ovaries
$\log_2$(Fold-change) of transcript abundance from testis-enriched and clustered genes for each differential expression analysis. Each data point represents the LFC of a single gene. Red crossbars show the mean LFC for all transcripts of the indicated gene class in the given differential expression experiment; dotted line at LFC=0. Clustered, testis-enriched genes have similar mean LFC in each experiment, while the non-testis-enriched bystanders in these clusters tend to have mean LFC=0, with the exception of *bam* null testis.

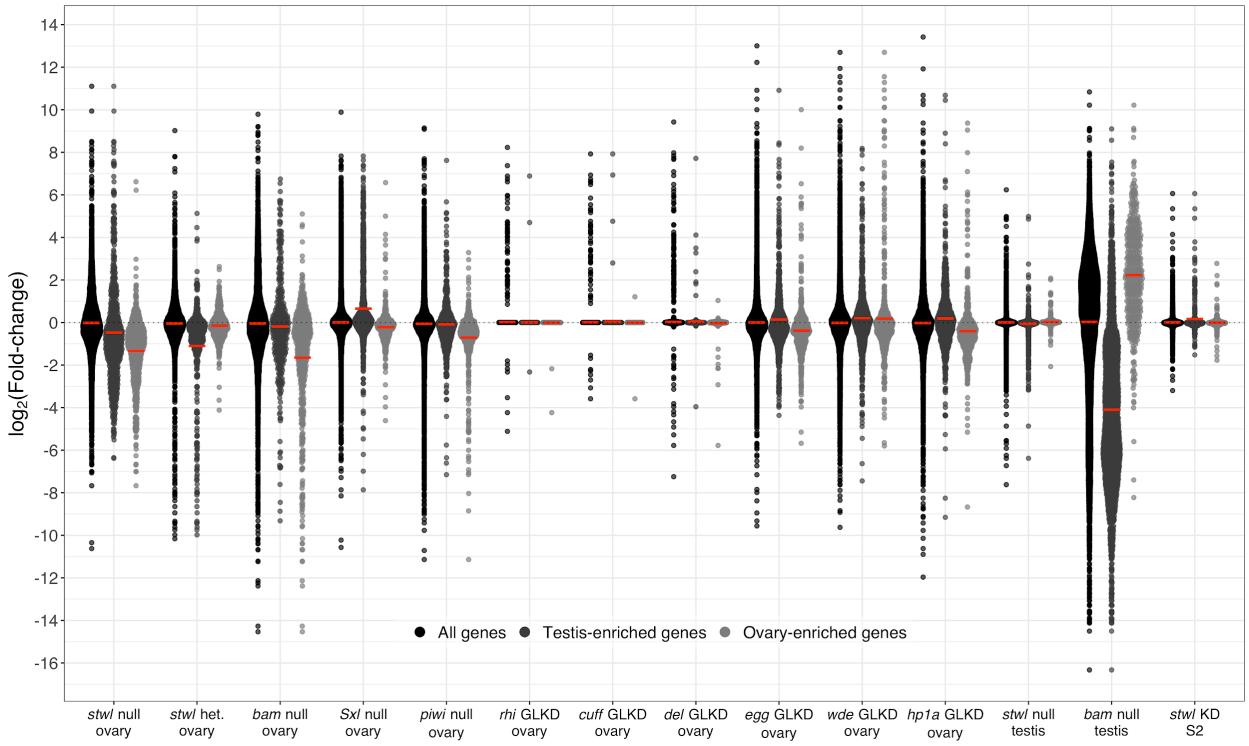
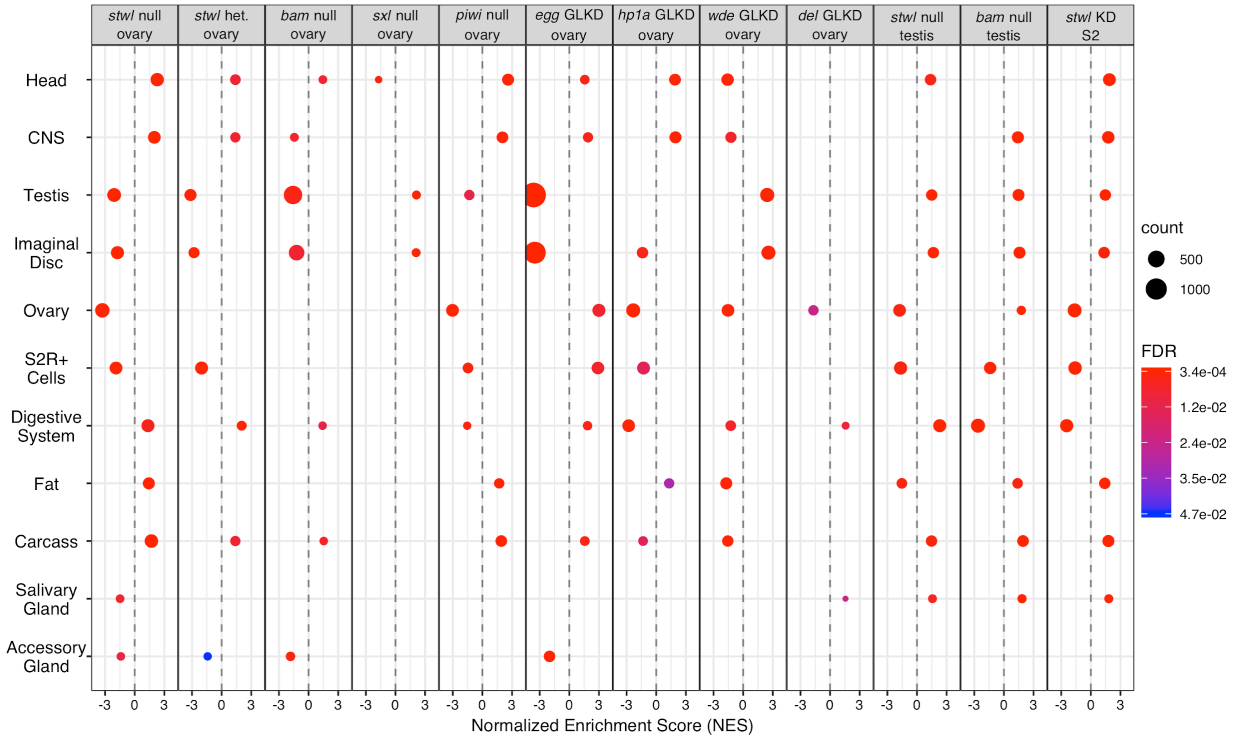


Figure 51. Testis- and ovary-enriched genes are misexpressed in mutant gonads
log_2(Fold-change) of transcript abundance from testis- and ovary-enriched genes for each differential expression analysis. Each data point represents the LFC of a single gene. Red crossbars show the mean LFC for all transcripts of the indicated gene class in the given differential expression experiment; dotted line at LFC=0.



<u>Figure 52</u>. GSEA results for tissue-enriched genes across *stwl*, piRNA pathway, and GSC regulatory mutants

Comparison of GSEA results from our *stwl* datasets with RNA-seq data from published datasets. Normalized Enrichment Score is plotted for each set of tissue-enriched genes enriched among null/WT ovaries and testes, heterozygous/WT ovaries, germline knockdown (KD)/WT ovaries, and *stwl*/*lacZ* RNAi S2 cells. Only gene sets with FDR<.05 are plotted. *rhi* and *cuff* GLKD ovaries are not shown because they were not enriched for any tissue classes. Normalized Enrichment Score (NES) reflects the strength of enrichment in the mutant or WT tissue; NES>0 and NES<0 indicate enrichment in the mutant and WT, respectively. NES is correlated to the strength of LFC of the core-enriched genes.

Our top/bottom percentile over-representation tests confirmed that testis-enriched genes are among the top 1% of most highly upregulated genes in *stwl* and *Sxl* null, as well as *egg*, *wde*, and *hp1a* GLKD ovaries; among the top 5%, we find that testis-enriched genes are overrepresented in *bam* and *piwi* null ovaries as well (Fig. 53). *Sxl* and the H3K9me3 pathway components are known silencers of testis-enriched genes in ovary; *stwl* null ovaries are the only tissues we examined where the magnitude of derepression of testis-enriched genes approached the effect sizes observed in the aforementioned mutant tissues.

Interestingly, *bam* null testes, *stwl* null ovaries, and *stwl* heterozygous ovaries showed strong downregulation of testis-enriched genes (bottom 1% LFC), an effect that was also observed to a lesser degree in *bam* null, *piwi* null, and *egg* GLKD ovaries (bottom 5% LFC); testis-enriched genes are among the most downregulated in *stwl* null testes, but the effect-size and level of enrichment is very low. We also found that ovary-enriched genes are strongly downregulated in *stwl* null ovaries; we observe the same phenotype in *bam* and *piwi* null, as well as *egg* and *hp1a* GLKD ovaries (Fig. 51, Fig. 53). These same genes are also strongly upregulated in *wde* and *hp1a* GLKD ovaries, as well as *bam* null testes.

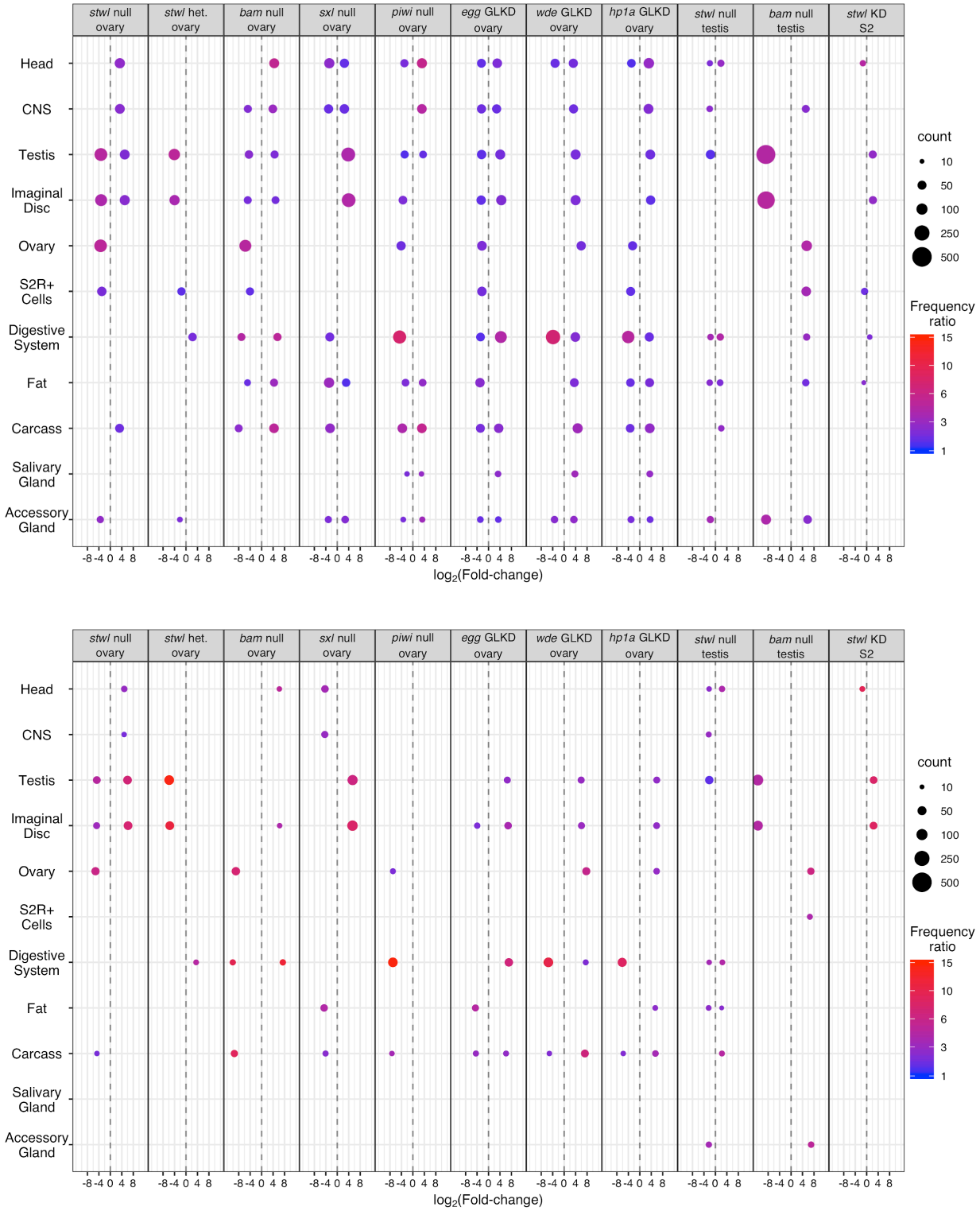
Loss of *stwl* in ovary results in upregulation of somatic-enriched genes

We also examined the impact of *stwl* loss on the expression of genes normally enriched in non-gonadal tissues (Table 3, Table 4). Mis-expression of imaginal disc-enriched genes largely mirrors misexpression of testis-enriched genes (Fig. 52, Fig. 53). Imaginal discs are larval structures responsible for the formation of adult organs and appendages, including wings, antennae, legs and gonads. *stwl* null ovaries also upregulate genes normally enriched in adult head, as well as the pharate and larval stage Central Nervous System (CNS). According to GSEA, these genes are also upregulated in *stwl* heterozygous and *piwi* null ovaries. Only *stwl*

null ovaries upregulate head- and CNS-enriched genes in the top 1% of differentially expressed genes.

Somatic and testis-enriched genes are ectopically expressed in *stwl* null ovaries

Custer 59C5-59D was marked not just by strong derepression of its genes, but also by ectopic expression of them (Fig. 46, Fig. 47). Due to the nature of ectopic expression (i.e. an increase in transcript abundance from a baseline of very low counts), it is often challenging to capture accurate $\log_2$(Fold-change) values, especially since GC-normalization typically requires removal of low-count genes. We therefore estimated ectopic gene expression in our *stwl* datasets as described above (before Fig. 48) and applied an overrepresentation test of tissue-enriched genes to these gene sets. We found that 3,777 genes are ectopically expressed in *stwl* null ovary; 1,716 in *stwl* heterozygous ovary, and only 376 in *stwl* null testis. We further subset the ectopically expressed *stwl* null ovary genes according to the top 5% and 1% of genes by LFC relative to wild-type. We found that testis- and imaginal disc-enriched genes are the only upregulated tissue classes among the top 1% of ectopically expressed genes in *stwl* null ovaries (Fig. 54). We also found that head- and CNS-enriched genes are consistently upregulated in all *stwl* mutant tissues, as are genes typically enriched in the adult/larval digestive system, larval/pharate/pre-pupae fat pads, larval/adult carcasses, and salivary glands.



<u>Figure 53</u>. Identification of tissue classes of genes among the 95th and 99th percentile of mis-expressed transcripts.

Top panel: Enriched tissue classes among the top and bottom **<u>5%</u>** of mis-regulated genes.

Bottom panel: Enriched tissue classes among the top and bottom **1%** of mis-regulated genes. Comparison of overrepresentation test results from our *stwl* datasets with RNA-seq data from published datasets. Significance was determined from a hypergeometric distribution comparing the gene ratio (genes enriched in tissue X among queried genes/all queried genes) to the background ratio (all genes enriched in tissue X/all genes). Frequency ratio = gene ratio/background ratio. Count, frequency ratio, and mean $\log_2$(Fold-change) is plotted for each set of tissue-enriched genes enriched among null/WT ovaries and testes, heterozygous/WT ovaries, and germline knockdown (GLKD)/WT ovaries, and *stwl*/*lacZ* RNAi S2 cells. Only gene sets with FDR<.05 are plotted.

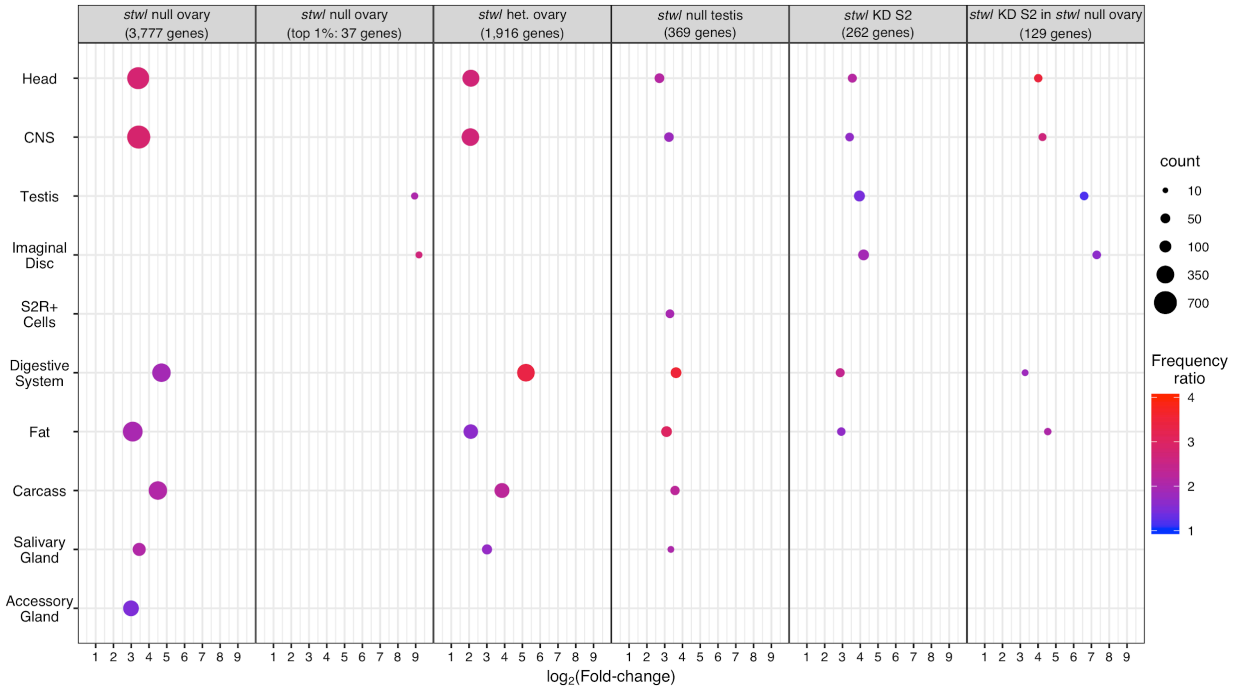


Figure 54. Identification of tissue classes of genes ectopically expressed in *stwl* mutants. Comparison of overrepresentation test results from our *stwl* datasets. Significance was determined from a hypergeometric distribution comparing the gene ratio (genes enriched in tissue X among ectopic genes/all ectopic genes) to the background ratio (all genes enriched in tissue X/all genes). Frequency ratio = gene ratio/background ratio. Count, frequency ratio, and mean $\log_2$(Fold-change) is plotted for each set of tissue-enriched genes enriched among *stwl* null/WT ovaries and testes, *stwl* heterozygous/WT ovaries, and *stwl/lacZ* dsRNA-treated S2 cells. Overrepresentation tests were also performed on the top 1% by LFC of genes ectopically expressed in *stwl* null ovary, and the subset of genes ectopically expressed in both *stwl* null ovary and *stwl* dsRNA-treated S2 cells. Only gene sets with FDR<.05 are plotted.

RNAi knockdown of *stwl* in S2 cells results in derepression of testis-enriched genes

Separating signal from noise is one of the central challenges of genomics. Even with meticulous controls, such as those implemented in our age-matched experimental design, it can be challenging to filter out genes which are differentially expressed as a consequence of the experimental condition versus those whose expression changes are incidental to the topic of investigation. Even when assayed mutant tissues are morphologically similar to wild-type, as is the case for *stwl* heterozygous ovaries and *stwl* null testes, 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 RNA extracted from S2 cells treated with *stwl* dsRNA (see methods). As a control, we treated S2 cells with similar quantities of *lacZ* dsRNA. Immunoblotting against Stwl protein confirmed that *stwl* dsRNA treatment reduced Stwl protein levels by at least 80%; RNA-Seq confirmed that *stwl* transcript was reduced by ~85% in *stwl* dsRNA treated cells (Fig. 55, Fig. 57). Read trimming, alignment, and GC-content normalization were performed as described for *stwl* ovaries and testes.

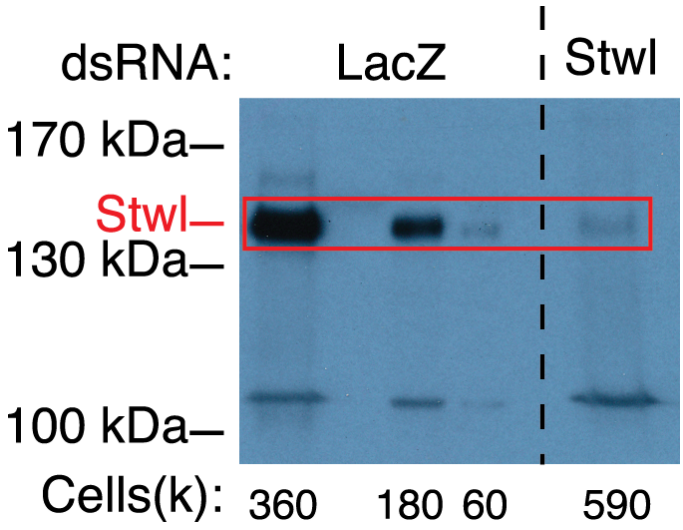


FIGURE 55. ***stwl* dsRNA treatment specifically knocks down ~130 KDa protein in S2 cells**
Western blot of S2 cells treated with either control dsRNA (*lacZ*) or dsRNA targeting the *stwl* transcript. Stwl antibody identifies a ~130 KDa signal in immunoblotting experiments. A standard curve was constructed using the signal from the *lacZ* treated cells, reflecting ~80% knockdown of Stwl signal in the *stwl* dsRNA treated cells.

Relative to loss of *stwl* in ovaries, *stwl* dsRNA treatment of S2 cells had a more subtle effect on transcript abundance (Table 2, Fig. 56). Many fewer genes are expressed in S2 cells, which are male and hematopoietic-derived (Schneider, 1972; Zhang et al., 2010). Only 6% (451/7807) of genes are differentially expressed in *stwl* dsRNA-treated S2 cells, with 61% up-regulated with a mean fold-increase of 1.9 and 39% down-regulated with a mean fold-decrease of 1.7. Repetitive elements, including satellites and transposons, are not mis-regulated or ectopically expressed in *stwl* dsRNA-treated S2 cells (Fig. 37). This is not due to an absence of repetitive transcripts from S2 cells or S2 cell libraries, as we find that many transposon transcripts are abundantly expressed in this line, as has been previously shown (Ghildiyal et al., 2008).

We were surprised to find 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. 57). 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 DESeq (Love et al., 2014) (Table 5, Fig. 48). 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 (Table 5, Fig. 58). Testis-enriched genes are up-regulated in *stwl* dsRNA-treated S2 cells as a whole according to GSEA, and are over-represented among the top 1% of both upregulated genes and ectopically expressed genes in these cells (Fig. 50, Fig. 51, Fig. 52, Fig. 53, Fig. 54, Fig. 57).. The subset of genes which are ectopically expressed in both *stwl* null and *stwl* dsRNA-treated S2 cells are highly enriched for testis, imaginal disc, head, CNS and fat transcripts (Fig. 54).

Gene Ontology (GO) analysis reveals enrichment for development-, stem-cell-regulation, brain-, and reproduction-associated terms.

We performed overrepresentation tests on upregulated genes for GO terms concerning biological processes. We found that *stwl* null and heterozygous ovaries, as well as *stwl* dsRNA-treated S2 cells, are enriched for transcripts associated with development, growth, and morphogenesis at multiple stages (Fig. 59A). A number of GO terms related to brain, CNS, eye, and head development are also associated with up-regulation in these samples, especially in null and heterozygous ovary (Fig. 59B). GO terms associated with stem cell regulation, including germ-line stem cell population maintenance, stem cell differentiation, asymmetric cell division, and stem cell division, are associated with up-regulation in *stwl* null and heterozygous ovaries, as are terms related to reproduction (Fig. 59C, Fig. 59D).

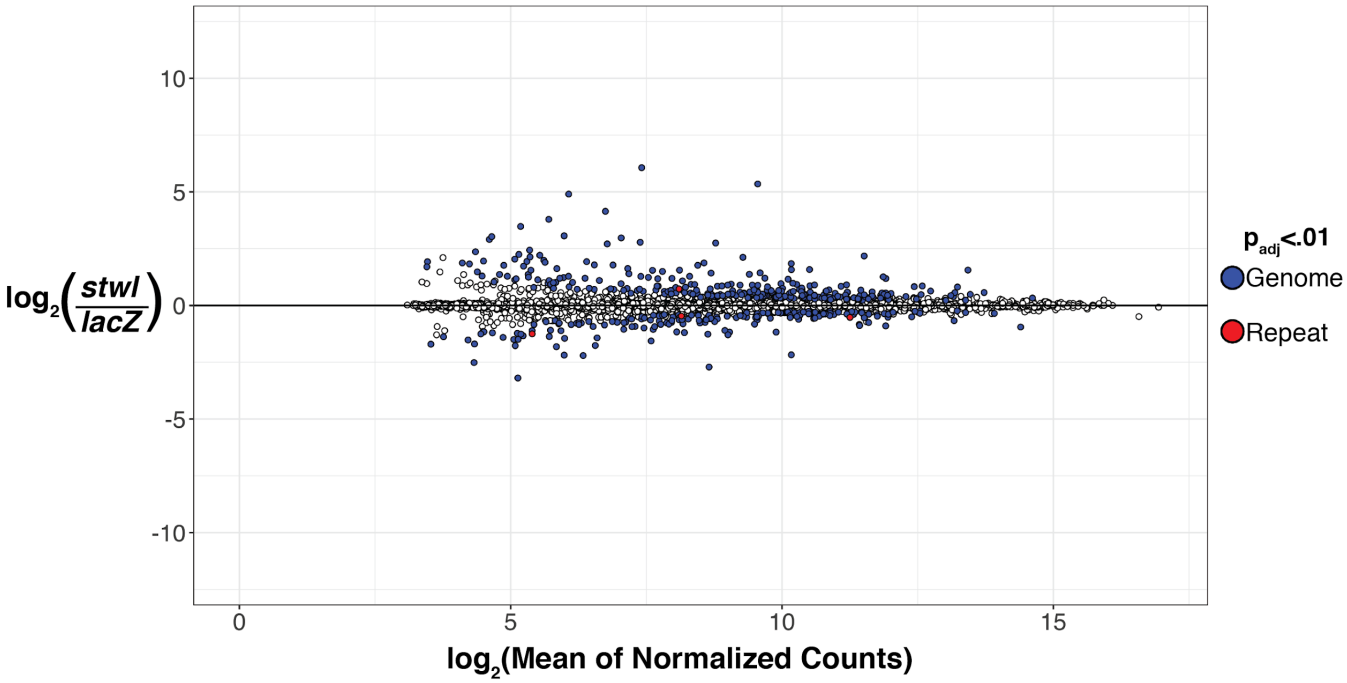


FIGURE 56. ***stwl* dsRNA minimally affects transposon expression in S2 cells.**
Fold-change for each gene is plotted against its average transcript abundance in S2 cell samples. Transcript abundance is represented by counts normalized according to GC-content and library size. The $\log_2$(Fold-change) 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 <.01) in this comparison. Red points represent entries from Repbase, blue points are from the genomic annotation.

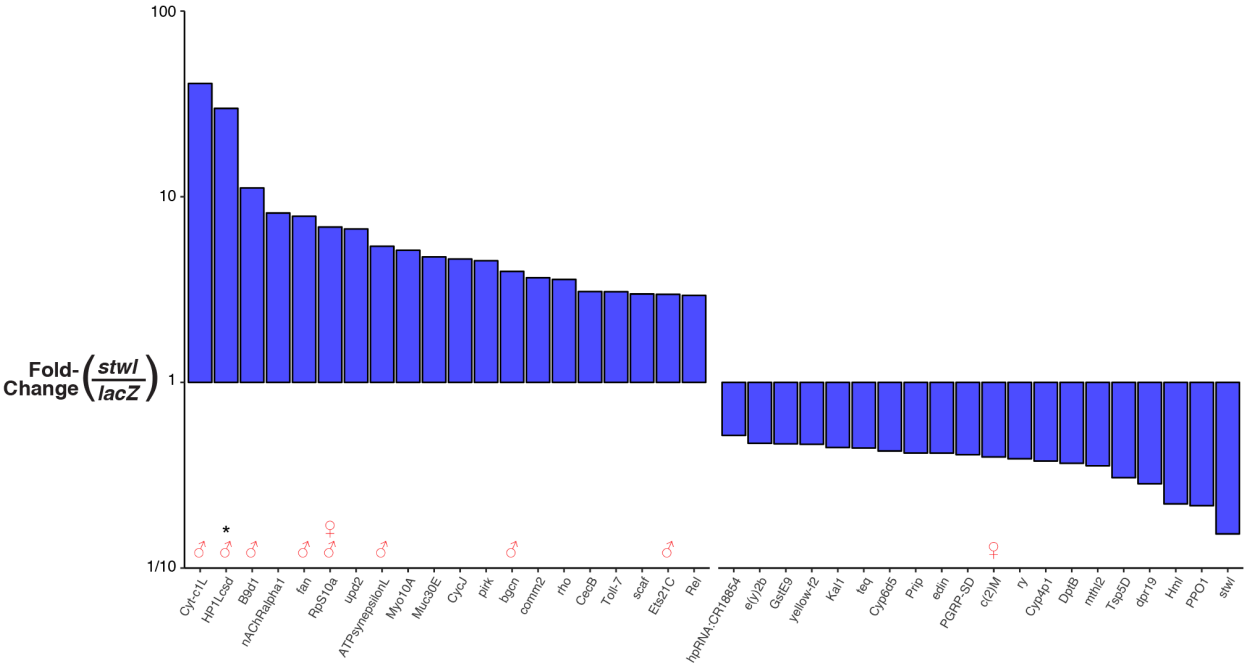


FIGURE 57. **Testis-enriched genes are upregulated in *stwl* dsRNA-treated S2 cells**
Plot of fold-change (on a $\log_{10}$ scale) of transcript abundance between S2 cells treated with *stwl* and *lacZ* dsRNA. The list excludes genes for non-coding RNA and those encoding uncharacterized proteins. Red male/female symbols identify genes with testis- and ovary-specific expression, respectively; "*" marks a gene that is part of the 59C4-59D cluster of testis-specific genes on chromosome 2R. *stwl* is the most down-regulated transcript (far-right), confirming efficacy of the *stwl* dsRNA treatment.

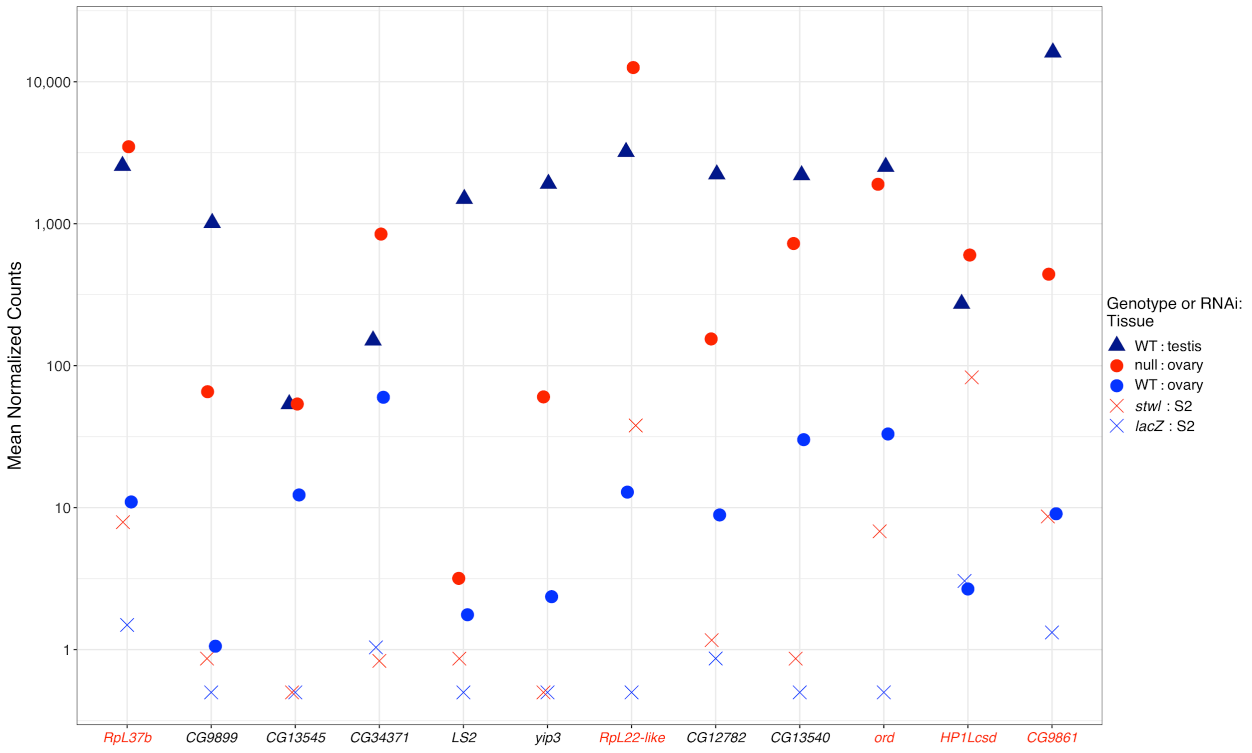
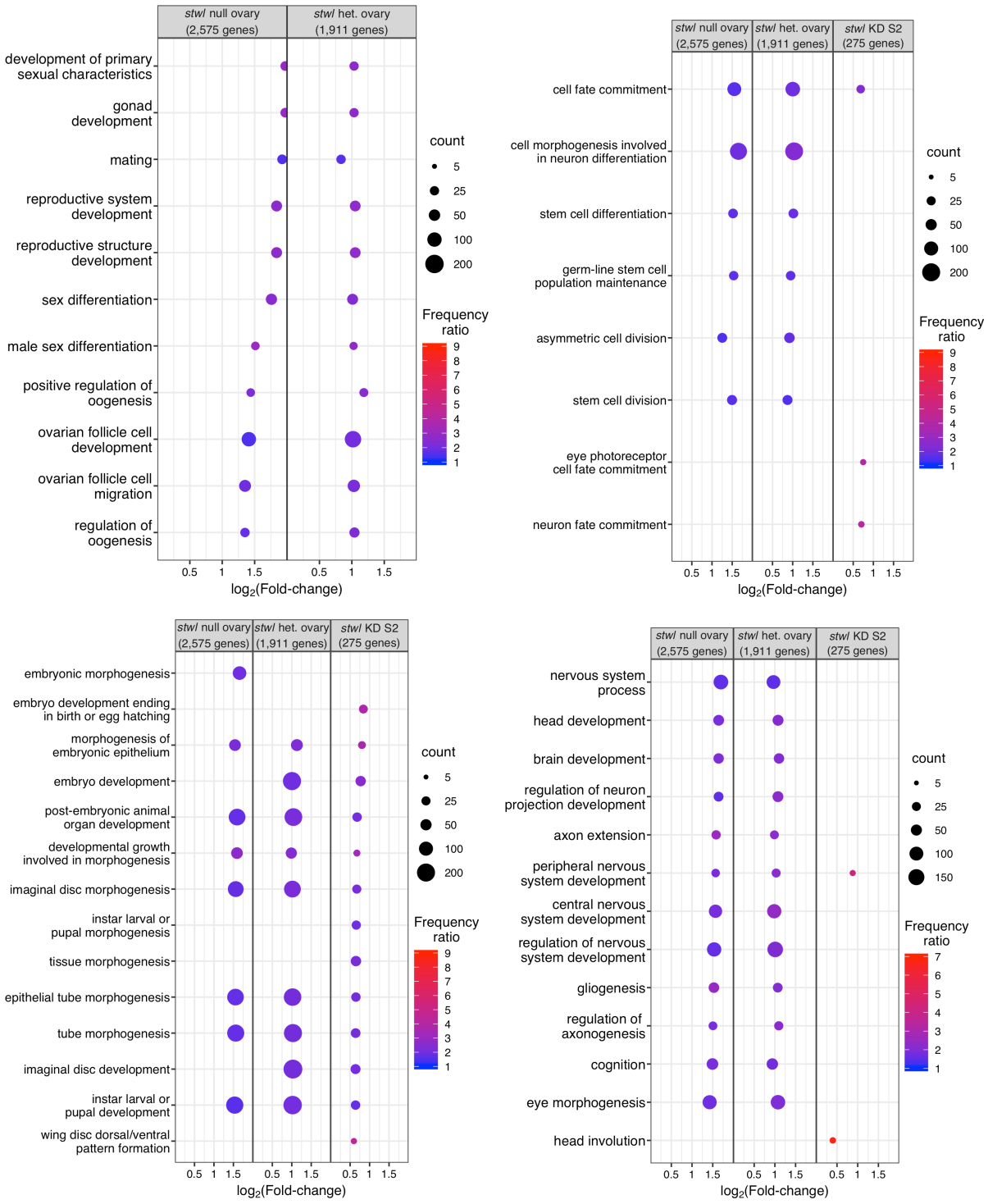


Figure 58. 59C4-59D genes are ectopically expressed in *stwl* dsRNA-treated S2 cells
Plot of library-normalized counts (on a $\log_{10}$ scale) of transcripts from the 59C4-59D in testis, ovaries and S2 cells. Genes are plotted in order of chromosomal location; only genes that showed ectopic expression and/or significant upregulation in ovaries are plotted. Gene names in red are ectopically expressed in *stwl* dsRNA-treated S2 cells. All plotted genes are highly expressed in wild-type testis, and transcriptionally silent in wild-type ovaries and S2 cells (Fig. 46, Fig. 47).



<u>Figure 59</u>. GO terms associated with up-regulation in Stwl-deficient ovaries and S2 cells. Comparison of overrepresentation test results from our *stwl* datasets. Significance was determined from a hypergeometric distribution comparing the gene ratio (genes enriched in GO term X among up-regulated genes/all up-regulated genes) to the background ratio (all genes enriched in GO term X/all genes). Frequency ratio = gene ratio/background ratio. Count,

frequency ratio, and mean $\log_2$(Fold-change) are plotted for a selection of GO-terms enriched among *stwl* null/WT ovaries, *stwl* heterozygous/WT ovaries, and *stwl/lacZ* dsRNA-treated S2 cells. Only gene sets with FDR<.05 are plotted.

The male-specific transcript of sex determination factor Phf7 is transcribed in *stwl* null ovaries

Disruption of Stwl function in females results in ectopic expression and up-regulation of non-ovarian genes in female gonads, especially transcripts typically enriched in testes. Similar phenotypes have been reported for the female sex-determination gene *Sxl*, the H3K9me3 pathway members *egg*, *hp1a*, and *wde*, and the differentiation factor *bam* (Salz et al., 2017; Shapiro-Kulnane et al., 2015; Smolko et al., 2018). The primary mechanism for maintaining female germ cell fate is the regulation of a gene called *phf7*; silencing of this gene is sufficient to rescue the tumorous germ cell phenotype of *Sxl* defective ovaries (Shapiro-Kulnane et al., 2015). In 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. 60). We found that the male-specific 5' UTR is consistently and strongly ectopically expressed in *stwl* null ovaries, regardless of age. This finding shows that *stwl* is implicated in maintaining sexual identity in ovarian germ cells.

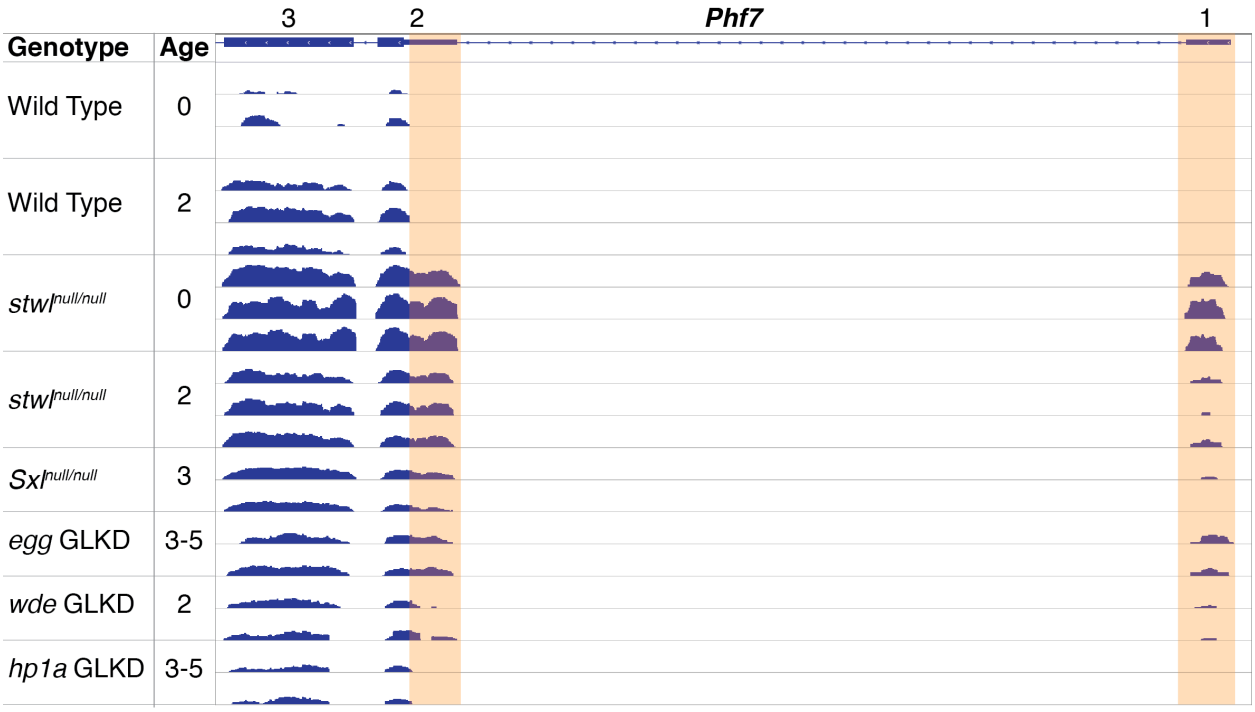


Figure 60. Male-specific variant of sex determination factor *Phf7* is transcribed in *stwl* null ovaries

Integrative Genomic Viewer (IGV) track, zoomed in on the *Phf7* gene locus. All tracks are normalized to 1x genome depth using deeptools and viewed on a log-scale (Ramírez et al., 2014).The male-specific portions of the transcript are highlighted. Transcription is initiated at exon 1 in males and exon 2 in females.

Genes ectopically expressed in both *stwl* null ovaries and *stwl*-dsRNA treated S2 cells include key sex determination and stem cell differentiation factors

Our ectopic expression analyses identified genes that are typically inert in the assayed tissue whose expression increases as a result of *stwl* mutation or depletion (Fig. 54). Interestingly, 49% (129/262) of ectopically expressed genes in *stwl*-dsRNA treated S2 cells are also ectopically expressed in *stwl* null ovaries, and these transcripts are typically enriched in head, CNS, testis and imaginal disc tissue (Fig. 54). We also found that these genes are generally upregulated in *bam, piwi,* and *Sxl* null ovaries, as well as *egg, wde,* and *hp1a* GLKD ovaries (Fig. 61, Table 6). Among these genes are *bgcn, bab1*, *esg*, *Rbp6*, *Rbp9, HP1Lcsd* and *RpL22-*

like (Fig. 61, Table 6). The protein Bgcn is a key binding partner for Bam; Bgcn binds and suppresses mRNAs associated with germline stem cell renewal, and both are required to promote differentiation of developing cystoblasts. *bab1* encodes the protein Bric-a-brac, which is involved in a myriad of functions, including pattern formation and female sex differentiation; it is regulated by sex-specific isoforms of the transcription factor Doublesex (Williams et al., 2008). Esg, or Escargot, is a transcription factor required for the maintenance of germline stem cells (non-autonomously) in testis and somatic stem cells in the midgut (Korzelius et al., 2014; Ramat et al., 2016; Streit et al., 2002; Voog et al., 2014). Rbp6 and Rbp9 are RNA-binding proteins expressed in gonads; *rbp9* mutant ovaries exhibit a tumorous phenotype similar to *bam*, *bgcn* and *Sxl* mutants (Jeong and Kim-Ha, 2004). *HP1Lcsd* and *RpL22-like* are both located in the testis-specific 59C4-59D cluster; HP1Lcsd is an HP1 homologue that likely has DNA-binding function; RpL22-like is a testis-specific ribosomal protein subunit (Levine et al., 2012).

We also found many genes that are upregulated in *stwl* null ovary and *stwl*-dsRNA treated S2 cells (Table 6). These include *chinmo*, *HP6/Umbrea*, *insc*, and *vir-1*. Chinmo is required for the establishment of male sexual identity in somatic cyst cells of the testis via positive regulation of the male isoform of Doublesex; it is a downstream effector of the Jak-STAT pathway, and is upregulated in ovarian tumor cells (Ma et al., 2014a; Shapiro-Kulnane et al., 2015). Heterochromatin Protein 6, also known as Umbrea, binds heterochromatic sequences at telomeres and centromeres (Joppich et al., 2009; Thomae et al., 2013). Insc, or Inscuteable, is required for asymmetric division and determination of neuroblasts; it is regulated in part by Escargot (Cai et al., 2003). *vir-1* is a viral response gene that is directly activated by the Jak-Stat signaling pathway (Dostert et al., 2005).

chinmo and *vir-1* are downstream effectors of the Jak-STAT signaling pathway. The canonical members of this pathway are *upd2*, *upd3*, *dome*, *hop* and *Stat92E*. We found that *upd2* and *upd3* are both upregulated in *stwl*-dsRNA treated S2 cells; *hop*, *dome*, and *Stat92E* are all upregulated in *stwl* null ovaries (Table 6). RNAi-mediated knockdown of these components rescues ovarian tumor defects caused by *Sxl* mutation (Shapiro-Kulnane et al., 2015). These data suggest that *stwl* loss results in upregulation of Jak-STAT pathway members, which results in aberrant expression of downstream effectors that cause masculinization and stem cell maintenance/differentiation defects.

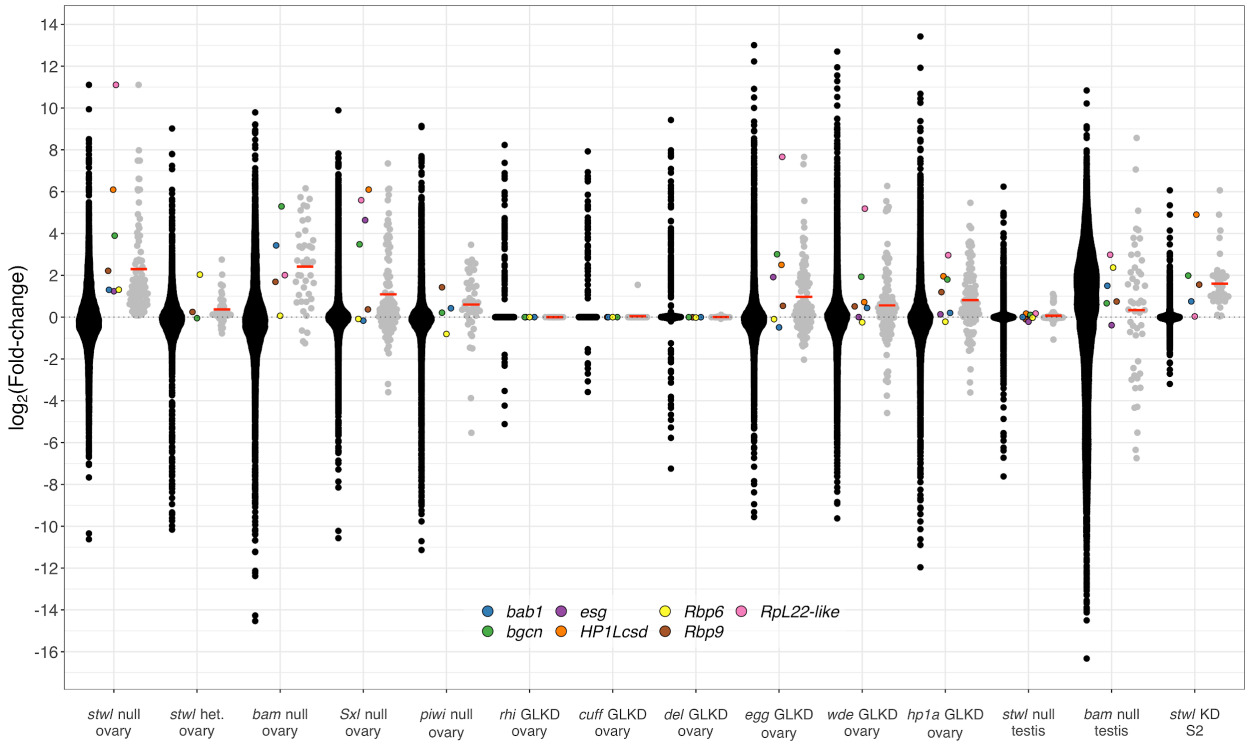


Figure 61. Genes ectopically expressed in both *stwl* null ovary and *stwl*-dsRNA treated S2 cells are highly upregulated in other GSC mutants

log_2(Fold-change) of transcript abundance from a subset of genes for each differential expression analysis. Each data point represents the LFC of a single gene. For each experiment, the LFC of each gene of interest is plotted against the LFC of all genes for that experiment (black) as well as the LFC of genes ectopically expressed in *stwl* null and *stwl*-dsRNA treated S2 cells.

Symbol	Description	Enriched tissue (modENCODE)	Upregulated in *stwl* mutant	Ectopic in *stwl* mutant
bab1	bric a brac 1, required for ovarian terminal filament organization; regulated by sexually dimorphic DSX	CNS	null ovary	null ovary, S2 cells, het. ovary
bgcn	Required for GSC differentiation	Testis, Imag. Disc	null ovary, S2 cells	null ovary, S2 cells
chinmo	Male somatic cyst identity, activated by JAK-STAT signaling	*NA*	null ovary, S2 cells, het. ovary	null ovary, het. ovary
dome	Jak/STAT signaling pathway component	*NA*	null ovary, het. ovary	*NA*
hop	Jak/STAT signaling pathway component	*NA*	null ovary	*NA*
esg	Required for maintenance of germline and intestinal stem cells	*NA*	*NA*	null ovary, S2 cells, het. ovary
HP1Lcsd	Hetochromatin Protein 1L chromoshadow domain, function unknown	Testis, Imag. Disc, Fat	null ovary, S2 cells	null ovary, S2 cells, het. ovary
HP6/Umbrea	Heterochromatin Protein, binds centromere and telomere	Testis, Imag. Disc	null ovary, S2 cells	*NA*
insc	Required for assymetric cell division	*NA*	null ovary, S2 cells, het. ovary	*NA*
Rbp6	Expressed in oocyte, testis soma; ectopic expression in testis germline causes germ cell loss	*NA*	null ovary, het. ovary	null ovary, S2 cells, het. ovary
Rbp9	Required for GSC differentiation	*NA*	null ovary, S2 cells	null ovary, S2 cells, het. ovary
Rpl22-like	Testis-specific ribosomal subunit	Testis, Imag. Disc	null ovary	null ovary, S2 cells
Stat92E	Jak/STAT signaling pathway component	*NA*	null ovary	*NA*
upd2	Jak/STAT signaling pathway component	*NA*	S2 cells	het. ovary
upd3	Jak/STAT signaling pathway component	Digestive System, S2 cells	S2 cells	het. ovary, null testis
vir-1	Activated by JAK-STAT signaling	*NA*	null ovary, S2 cells	*NA*

Table 6. Ectopically expressed and upregulated genes in *stwl* null and *stwl*-dsRNA treated S2 cells

Genes of interest whose expression increases in *stwl* null and/or *stwl*-dsRNA treated S2 cells. The discordance in ectopic expression and upregulation (i.e. ectopically expressed genes are not always upregulated) is due to the different methods for calling differential expression versus ectopic expression.

Stwl is required in germline but not follicle cells of the ovary for maintenance of fertility

RNAi knockdown experiments of *stwl* in early germ cells (*nos-Gal4*) and total germline (*MTD-Gal4*) results in severe ovarian defects and sterility (Shukla et al., 2018; Yan et al., 2014). Both of the aforementioned Gal4 lines drive expression in GSCs, where *stwl* expression is required for maintenance of the stem cell population (Akiyama, 2002; Clark and McKearin, 1996). Furthermore, overexpression of *stwl* in GSCs via *nos-Gal4* results in an expansion of GSCs (Maines et al., 2007). In addition to GSC maintenance defects, *stwl* mutant ovaries also exhibit a failure in oocyte determination that is observed immediately after eclosion (Fig. 5, Fig. 7). Oocyte determination begins shortly after the initial formation of the germline cyst, probably as early as the formation of the fusome in the first cystoblast division (Hudson and Cooley, 2014; Huynh, 2013). We therefore tested whether *stwl* loss in the female germline outside of the GSC reduces fertility. Furthermore, in light of our findings that *stwl* loss results in over-expression of somatic transcripts and has a transcriptional phenotype in non-germline S2 cells, we also tested whether *stwl* is required in ovarian somatic cells.

We performed RNAi KD by crossing *UAS-stwl-RNAi* males to females of varying *Gal4* genotypes, including the follicle cell drivers *C355, T155, GR1,* and *C306*, the germline drivers *bam* and *Mat-alpha*, and the ubiquitous drivers *tub* and *act5c*. *Gal4/UAS-stwl-RNAi* F1 females were crossed to *y w* males and progeny were counted for 13 days. Ubiquitous KD of *stwl* causes total female sterility, phenocopying the *stwl* null mutation (Fig. 4, Fig. 62). *stwl* KD in follicle cells does not impact fertility, while *stwl* KD in germline cells outside of the GSC strongly reduces fertility. Surprisingly, *stwl* KD in germline cells did not entirely phenocopy ubiquitous *stwl* loss. Individuals with *UAS-stwl-RNAi* driven by *Mat-alpha-Gal4* are initially fertile, but

become sterile by day 10 of mating. *Bam-Gal4*/*UAS-stwl-RNAi* individuals are also not completely sterile and retain low levels of fertility even after 10 days of mating.

We also identified a distinction in egg-laying behavior: females with *stwl* expression completely abolished (*stwl* null and ubiquitous Gal4s) lay no eggs whatsoever. *Bam-Gal4*/*UAS-stwl-RNAi* females lay very few eggs, which develop into a small number of adults. Surprisingly, *Mat-alpha-Gal4*/*UAS-stwl-RNAi* females lay many eggs, but almost no larvae hatch from these eggs. Previous work has identified that these females do not produce any progeny, despite having phenotypically normal ovaries (Yan et al., 2014). Further work is required to understand the phenotype of *Bam-* and *Mat-alpha-Gal4*/*UAS-stwl-RNAi* ovaries, particularly whether proper oocyte determination is maintained in these individuals over time. We have also not tested whether *stwl* is required in escort cells, which could be performed using the *Tj-* and *C587-Gal4* drivers.

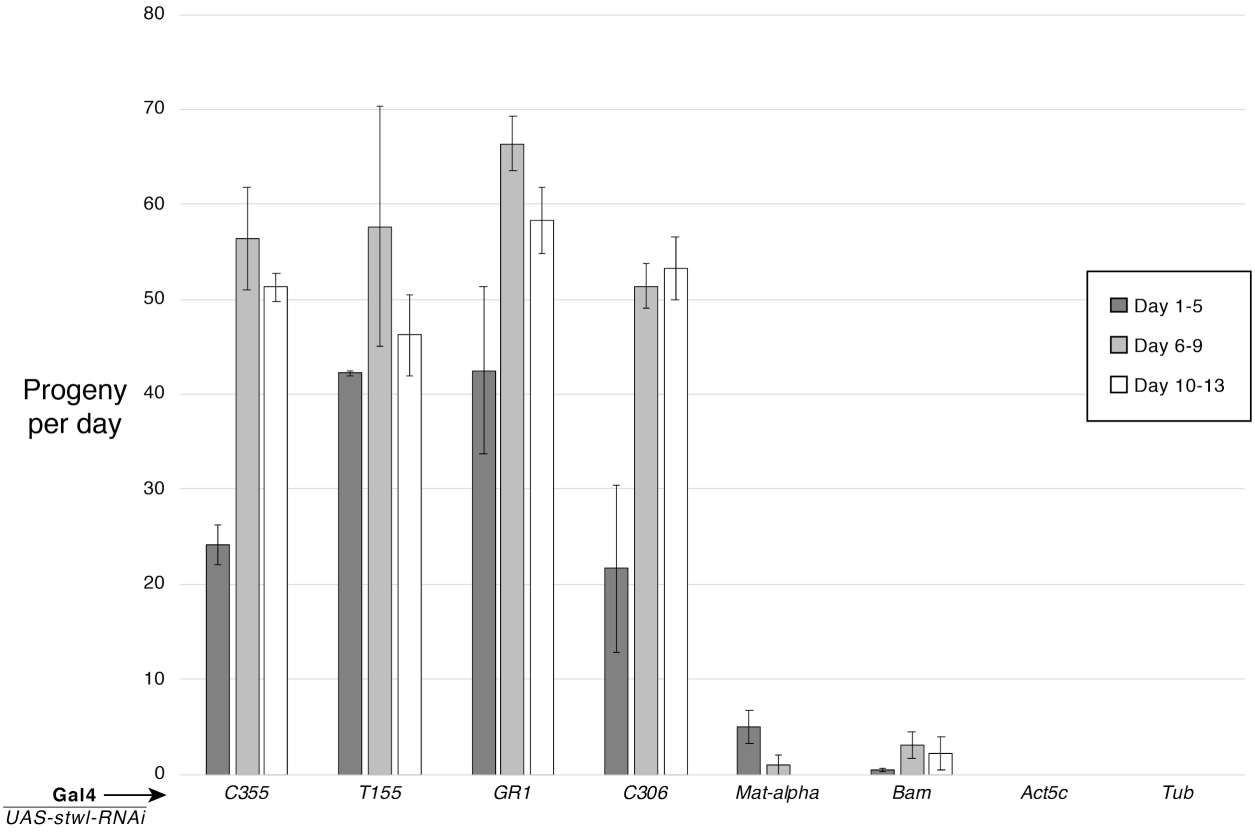


Figure 62. *stwl* expression in ovarian follicle cells is dispensable for female fertility
Age-matched Gal4/*UAS-stwl-RNAi* transheterozygous virgin females were continuously mated to 1-5 day old *y w* males to determine the effect of *stwl* KD in follicle cells and germ cells. *C355-gal4*, *T155-gal4*, 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; *Act5c-gal4* and *Tub-gal4* drive ubiquitous expression in germline and somatic cells.

Discussion (Part 2 of 2)

In part 1 of this chapter, we established that transposable elements are upregulated in *stwl* mutant ovaries. In part 2, we take a broader view at the effect of *stwl* loss on tissue-specific transcripts. First, 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. Consistent with our finding that *stwl* and *bam* null ovaries have similar transcriptional profiles, we found that *bam* null ovaries also ectopically express genes in this cluster. In fact, we found that each of *Sxl* and *piwi* null ovaries, *egg*, *wde*, and *hp1a* GLKD ovaries show strong upregulation at this gene cluster.

The 59C4-59D cluster is located with 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 (data not shown). We also do not find that Stwl is binding to this region, or overlapping with LADs in our ChIP-Seq data (Chapter 3).

A possible conclusion, and cause for concern, is that overexpression of genes at this cluster is merely a reflection of the altered cellular content of the affected tissue. In *bam* and *Sxl* null ovaries, GSC-like cells overproliferate to form tumor-like structures with stem cell-like qualities and protein expression patterns. They also tend to express transcripts associated with early gametogenesis in both sexes. Therefore, it is possible that the reported “masculinization” of the ovary as a result of *Sxl* or *bam* mutation is merely reflective of an overabundance of transcripts expressed during early stages of gametogenesis.

However, our *stwl* datasets prove that this is not the case. First of all, *stwl* null ovaries exhibit a GSC loss phenotype, not an overproliferation of such cells. Second, and most convincingly, we find that *stwl*-dsRNA treated S2 cells also ectopically express a subset of genes at this locus. S2 cells are male-derived, but our analysis nonetheless finds that these genes are almost completely silent in the control cells. We conclude that this cluster of testis-enriched genes is a consistent reporter of the "masculinization" defect associated with *stwl*, *Sxl*, *bam*, *hp1a*, *wde* and *egg* mutants. We speculate that its regulation is downstream of the masculinization factor Phf7, and possibly Bgcn, which we also find is consistently upregulated in these tissues.

In addition to the misregulated 59C4-59D cluster, we also found that testis-enriched genes are generally upregulated in *stwl*-dsRNA-treated S2 cells, as was previously observed in *Sxl* (Shapiro-Kulnane et al., 2015). We were surprised to find that in *stwl* null ovaries, testis-enriched genes show a mixed pattern of up- and down-regulation. We did confirm, however, that these genes are highly enriched among the most upregulated, ectopically expressed genes, and that genes that are ectopically expressed in both *stwl* null and *stwl*-dsRNA-treated S2 cells are biased towards testis-specificity. Additionally, we found that *stwl* null ovaries express the male-specific transcript of the master sex determination factor *Phf7*; they also highly upregulate *bgcn* transcript, which is typically expressed only in GSCs and at very low levels. *bgcn* is also upregulated in *stwl*-dsRNA-treated S2 cells, and we identify in chapter 3 that Stwl is bound to the *bgcn* locus.

Despite the fact that Bgcn has a defined and important role in the ovary, it is nonetheless largely silent throughout oogenesis. What are the consequences of ectopic *bgcn* expression outside of GSCs? Overexpression of *bam* in GSCs results in loss of GSCs, similar to *stwl* nulls;

surprisingly, *bgcn* overexpression does not result in GSC loss (Ohlstein et al., 2000). Nonetheless, non-specific expression of a *bgcn* transgene causes sterility in females, characterized by "small oocytes" in late stage egg chambers. It is reasonable to assume then, that ectopic expression of *bgcn* outside of GSCs results in sterility. It is possible that one of Stwl's functions in the female germline is to restrict expression of *bgcn* to GSCs. There is evidence that *stwl* is epistatic to *bgcn*: *bgcn stwl* double mutant ovaries form fusomes, which are not present in *bgcn* mutant ovaries (Park, Joseph, 2007).

In addition to testis-enriched genes, we also found that a variety of genes typically enriched in other tissues, including head, CNS, imaginal disc, fat, digestive system and salivary gland, are ectopically expressed in ovaries and S2 cells upon *stwl* loss. Our prevailing hypothesis is that Stwl is responsible for shutting off genes that shouldn't be expressed in the ovary, including genes associated with the male germline and somatic tissues. In chapter 3, we will explore what characteristics of Stwl make it able to perform this function.

References

1. Akiyama, T. (2002). Mutations of stonewall disrupt the maintenance of female germline stem cells in Drosophila melanogaster. Dev. Growth Differ. *44*, 97–102.

2. Brown, J.B., Boley, N., Eisman, R., May, G.E., Stoiber, M.H., Duff, M.O., Booth, B.W., Wen, J., Park, S., Suzuki, A.M., et al. (2014). Diversity and dynamics of the *Drosophila* transcriptome. Nature *512*, 393–399.

3. Cai, Y., Yu, F., Lin, S., Chia, W., and Yang, X. (2003). Apical Complex Genes Control Mitotic Spindle Geometry and Relative Size of Daughter Cells in Drosophila Neuroblast and pI Asymmetric Divisions. Cell *112*, 51–62.

4. Clark, K.A., and McKearin, D.M. (1996). The Drosophila stonewall gene encodes a putative transcription factor essential for germ cell development. Development *122*, 937–950.

5. Czech, B., Preall, J.B., McGinn, J., and Hannon, G.J. (2013). A Transcriptome-wide RNAi Screen in the Drosophila Ovary Reveals Factors of the Germline piRNA Pathway. Mol. Cell *50*, 749–761.

6. Díaz-González, J., Domínguez, A., and Albornoz, J. (2010). Genomic distribution of retrotransposons 297, 1731, copia, mdg1 and roo in the Drosophila melanogaster species subgroup. Genetica *138*, 579–586.

7. Dostert, C., Jouanguy, E., Irving, P., Troxler, L., Galiana-Arnoux, D., Hetru, C., Hoffmann, J.A., and Imler, J.-L. (2005). The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of drosophila. Nat. Immunol. *6*, 946.

8. Gan, Q., Chepelev, I., Wei, G., Tarayrah, L., Cui, K., Zhao, K., and Chen, X. (2010). Dynamic regulation of alternative splicing and chromatin structure in Drosophila gonads revealed by RNA-seq. Cell Res. *20*, 763–783.

9. Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L.W., Zapp, M.L., Weng, Z., et al. (2008). Endogenous siRNAs Derived from Transposons and mRNAs in Drosophila Somatic Cells. Science *320*, 1077–1081.

10. Hong Guini, Zhang Wenjing, Li Hongdong, Shen Xiaopei, and Guo Zheng (2014). Separate enrichment analysis of pathways for up- and downregulated genes. J. R. Soc. Interface *11*, 20130950.

11. Hudson, A.M., and Cooley, L. (2014). Methods for studying oogenesis. Methods San Diego Calif *68*, 207–217.

12. Huynh, J.-R. (2013). Fusome as a Cell-Cell Communication Channel of Drosophila Ovarian Cyst (Landes Bioscience).

13. Jeong, K., and Kim-Ha, J. (2004). Precocious expression of Drosophila Rbp9 inhibits ovarian germ cell proliferation. Mol. Cells *18*, 230–236.

14. Joppich, C., Scholz, S., Korge, G., and Schwendemann, A. (2009). Umbrea, a chromo shadow domain protein in Drosophila melanogaster heterochromatin, interacts with Hip, HP1 and HOAP. Chromosome Res. *17*, 19–36.

15. Kalmykova, A.I., Klenov, M.S., and Gvozdev, V.A. (2005). Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. Nucleic Acids Res. *33*, 2052–2059.

16. Kelleher, E.S., and Barbash, D.A. (2013). Analysis of piRNA-Mediated Silencing of Active TEs in Drosophila melanogaster Suggests Limits on the Evolution of Host Genome Defense. Mol. Biol. Evol. *30*, 1816–1829.

17. Klenov, M.S., Lavrov, S.A., Stolyarenko, A.D., Ryazansky, S.S., Aravin, A.A., Tuschl, T., and Gvozdev, V.A. (2007). Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the Drosophila melanogaster germline. Nucleic Acids Res. *35*, 5430–5438.

18. Klenov, M.S., Sokolova, O.A., Yakushev, E.Y., Stolyarenko, A.D., Mikhaleva, E.A., Lavrov, S.A., and Gvozdev, V.A. (2011). Separation of stem cell maintenance and transposon silencing functions of Piwi protein. Proc. Natl. Acad. Sci. *108*, 18760–18765.

19. Korzelius, J., Naumann, S.K., Loza-Coll, M.A., Chan, J.S., Dutta, D., Oberheim, J., Gläßer, C., Southall, T.D., Brand, A.H., Jones, D.L., et al. (2014). Escargot maintains stemness and suppresses differentiation in Drosophila intestinal stem cells. EMBO J. *33*, 2967–2982.

20. Levine, M.T., McCoy, C., Vermaak, D., Lee, Y.C.G., Hiatt, M.A., Matsen, F.A., and Malik, H.S. (2012). Phylogenomic Analysis Reveals Dynamic Evolutionary History of the Drosophila Heterochromatin Protein 1 (HP1) Gene Family. PLOS Genet. *8*, e1002729.

21. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*.

22. Ma, Q., Wawersik, M., and Matunis, E.L. (2014a). The Jak-STAT Target Chinmo Prevents Sex Transformation of Adult Stem Cells in the Drosophila Testis Niche. Dev. Cell *31*, 474–486.

23. Ma, X., Wang, S., Do, T., Song, X., Inaba, M., Nishimoto, Y., Liu, L., Gao, Y., Mao, Y., Li, H., et al. (2014b). Piwi Is Required in Multiple Cell Types to Control Germline Stem Cell Lineage Development in the Drosophila Ovary. PLoS ONE *9*.

24. Maines, J.Z., Park, J.K., Williams, M., and McKearin, D.M. (2007). Stonewalling Drosophila stem cell differentiation by epigenetic controls. Development *134*, 1471–1479.

25. Mohn, F., Sienski, G., Handler, D., and Brennecke, J. (2014). The Rhino-Deadlock-Cutoff Complex Licenses Noncanonical Transcription of Dual-Strand piRNA Clusters in Drosophila. Cell *157*, 1364–1379.

26. Morozova, T.V., Tsybul'ko, E.A., Filatov, D.A., and Pasiukova, E.G. (2004). [Regulatory elements of copia retrotransposons, controlling the level of its expression in Drosophila melanogaster testes]. Genetika *40*, 173–179.

27. Ohlstein, B., Lavoie, C.A., Vef, O., Gateff, E., and McKearin, D.M. (2000). The Drosophila cystoblast differentiation factor, benign gonial cell neoplasm, is related to DExH-box proteins and interacts genetically with bag-of-marbles. Genetics *155*, 1809–1819.

28. Park, J.K. (2007). Characterization of factors controlling germline stem cell maintenance in Drosophila melanogaster. PhD Thesis. University of Texas Southwestern Medical Center.

29. Peng, J.C., Valouev, A., Liu, N., and Lin, H. (2016). Piwi maintains germline stem cells and oogenesis in Drosophila through negative regulation of Polycomb Group proteins. Nat. Genet. *48*, 283–291.

30. Ramat, A., Audibert, A., Louvet-Vallée, S., Simon, F., Fichelson, P., and Gho, M. (2016). Escargot and Scratch regulate neural commitment by antagonizing Notch activity in Drosophila sensory organs. Development *143*, 3024–3034.

31. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

32. Salz, H.K., Dawson, E.P., and Heaney, J.D. (2017). Germ cell tumors: Insights from the Drosophila ovary and the mouse testis. Mol. Reprod. Dev. *84*, 200–211.

33. Schneider, I. (1972). Cell lines derived from late embryonic stages of Drosophila melanogaster. J. Embryol. Exp. Morphol. *27*, 353–365.

34. Shapiro-Kulnane, L., Smolko, A.E., and Salz, H.K. (2015). Maintenance of Drosophila germline stem cell sexual identity in oogenesis and tumorigenesis. Development *142*, 1073–1082.

35. Shevelyov, Y.Y., Lavrov, S.A., Mikhaylova, L.M., Nurminsky, I.D., Kulathinal, R.J., Egorova, K.S., Rozovsky, Y.M., and Nurminsky, D.I. (2009). The B-type lamin is required for somatic repression of testis-specific gene clusters. Proc. Natl. Acad. Sci. U. S. A. *106*, 3282–3287.

36. Shukla, V., Dhiman, N., Nayak, P., Dahanukar, N., Deshpande, G., and Ratnaparkhi, G.S. (2018). Stonewall and Brickwall: Two Partially Redundant Determinants Required for the Maintenance of Female Germline in Drosophila. G3 GenesGenomesGenetics *8*, 2027–2041.

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

38. Streit, A., Bernasconi, L., Sergeev, P., Cruz, A., and Steinmann-Zwicky, M. (2002). mgm 1, the earliest sex-specific germline marker in Drosophila, reflects expression of the gene esg in male stem cells. Int. J. Dev. Biol. *46*, 159–166.

39. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. *102*, 15545–15550.

40. Sun, S., and Cline, T.W. (2009). Effects of Wolbachia Infection and ovarian tumor Mutations on Sex-lethal Germline Functioning in Drosophila. Genetics *181*, 1291–1301.

41. Thomae, A.W., Schade, G.O.M., Padeken, J., Borath, M., Vetter, I., Kremmer, E., Heun, P., and Imhof, A. (2013). A Pair of Centromeric Proteins Mediates Reproductive Isolation in Drosophila Species. Dev. Cell *27*, 412–424.

42. Volpe, M., Miralto, M., Gustincich, S., and Sanges, R. (2018). ClusterScan: simple and generalistic identification of genomic clusters. Bioinforma. Oxf. Engl. *34*, 3921–3923.

43. Voog, J., Sandall, S.L., Hime, G.R., Resende, L.P.F., Loza-Coll, M., Aslanian, A., Yates, J.R., Hunter, T., Fuller, M.T., and Jones, D.L. (2014). Escargot Restricts Niche Cell to Stem Cell Conversion in the Drosophila Testis. Cell Rep. *7*, 722–734.

44. Warden, C.D., Kanaya, N., Chen, S., and Yuan, Y.-C. (2013). BD-Func: a streamlined algorithm for predicting activation and inhibition of pathways. PeerJ *1*, e159.

45. Wingett, S.W., and Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. F1000Research *7*, 1338.

46. Williams, T.M., Selegue, J.E., Werner, T., Gompel, N., Kopp, A., and Carroll, S.B. (2008). The Regulation and Evolution of a Genetic Switch Controlling Sexually Dimorphic Traits in Drosophila. Cell *134*, 610–623.

47. Yan, D., Neumüller, R.A., Buckner, M., Ayers, K., Li, H., Hu, Y., Yang-Zhou, D., Pan, L., Wang, X., Kelley, C., et al. (2014). A regulatory network of Drosophila germline stem cell self-renewal. Dev. Cell *28*, 459–473.

48. Yi, X., de Vries, H.I., Siudeja, K., Rana, A., Lemstra, W., Brunsting, J.F., Kok, R.M., Smulders, Y.M., Schaefer, M., Dijk, F., et al. (2009). Stwl Modifies Chromatin Compaction and Is Required to Maintain DNA Integrity in the Presence of Perturbed DNA Replication. Mol. Biol. Cell *20*, 983–994.

49. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS J. Integr. Biol. *16*, 284–287.

50. Zhang, Y., Malone, J.H., Powell, S.K., Periwal, V., Spana, E., MacAlpine, D.M., and Oliver, B. (2010). Expression in Aneuploid Drosophila S2 Cells. PLoS Biol. *8*.

51. Zhu, A., Ibrahim, J.G., and Love, M.I. (2018). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinforma. Oxf. Engl.

CHAPTER 3

CHIP-SEQ ANALYSIS IDENTIFIES STWL AS A LIKELY INSULATOR BINDING PROTEIN

Introduction

Stwl plays a vital role in germline stem cell maintenance and oocyte determination. We have established that loss of Stwl in the ovary results in dramatic changes in gene expression, including the ectopic expression of testis-, imaginal disc-, and somatic tissue-enriched genes, as well as transposable elements. Whether Stwl directly regulates these genes is unclear, due to the severity of the phenotype in *stwl* null ovaries. In order to more clearly understand the molecular function of Stwl, we developed ChIP-grade antibodies against it and performed ChIP-Seq experiments on S2 cells. We found that Stwl localizes to heterochromatic sequences, including telomeres, pericentric heterochromatin, and the heterochromatic 4th chromosome. At a finer scale, we also observe that Stwl binds to transcription start sites, especially at active promoters, and that Stwl-bound genes produce ovary-enriched transcripts that are upregulated in *stwl* heterozygous ovaries. Furthermore, Stwl binds to its own locus as well as those of *bgcn* and *pum*, which directly implicates it in the regulation of germline stem cell gene-expression. Finally, we find that Stwl binding sites show strong overlap with and are enriched for motifs common to insulator-binding proteins. We conclude that Stwl performs multiple essential functions of insulator-binding proteins, including the establishment of heterochromatin boundaries and the regulation of euchromatic genes involved in germline maintenance.

Methods

Production 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 (Sheffield et al., 1999). Successful assembly and transformation into chemically competent *E. coli* (One Shot® TOP10) 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 (Fig. 1); 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.

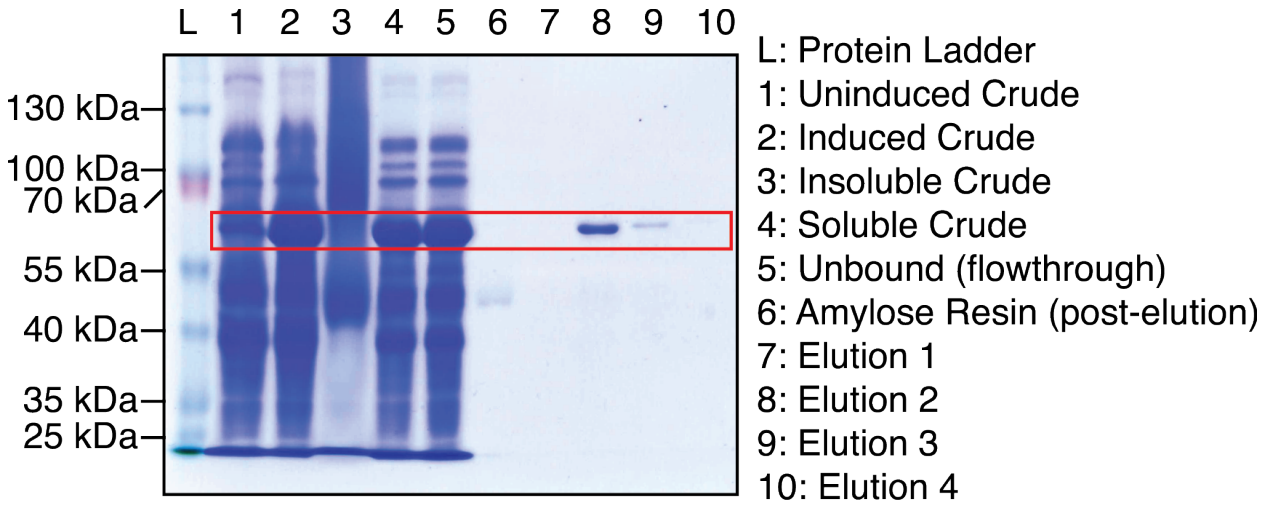


FIGURE 1. **Purification of Stwl-MBP fragment from crude lysate**
10% SDS PAGE confirming the induction of Stwl-MBP peptide via addition of IPTG and subsequent purification of the antigen from solution. "Crude" sample is lysate of bacterial cultures prior to purification. Lanes 1 and 2 are loaded with Crude extract prior to sonication, demonstrating the effective induction of the target antigen after addition of 0.2 mM IPTG. Lanes 3 and 4 are the insoluble (pellet) and soluble (supernatant) fractions of the Induced Crude lysate after sonication and centrifugation, demonstrating that the target antigen is soluble and separated from cellular debris after sonication. Lane 5 contains flow-through remaining after Soluble Crude was applied to amylose resin, indicating the resin was saturated with target antigen. Lanes 7-10 are loaded with fractions collected from 1 ml elutions from amylose resin. The first and 4th elutions (Lanes 7 and 10) contain little to no target antigen because most of the antigen was eluted in the 2nd and 3rd fractions (Lanes 8 and 9). Target antigen did not remain on Amylose Resin after elution (Lane 6).

Purified protein was submitted to Pocono Rabbit Farm & Laboratory Inc. for injection. Two guinea pigs (henceforth referred to as G76 and G77) were selected for antigen injection based on absence of background signal in pre-bleed sera (determined by probing wild type *D. melanogaster* ovaries with sera in immunofluorescence assays).

Primers used to amplify *stwl* fragment from cDNA for Gibson assembly:
Forward: TTCAGGGCGCCATGGATCCGGAATTCAAAGGTAGGAAGTCTGTGGGCTGC
Reverse: TAGTTGAGCTCGTCGACGTAGGTTAAACACCGAGTATGTCATTGTGATCC

Antibody Validation

The ENCODE and modENCODE consortia have provided helpful guidelines for antibody validation (Landt et al., 2012). We performed the following assays according to the modENCODE guidelines for antibody sera from both guinea pigs:

1. Immunoblotting.

We find that both antibodies recognize a ~135 kDa protein in wild type ovaries (Fig. 2), whole flies (Fig. 3), and S2 cells (Fig. 4). This signal is absent in null and RNA knockdown samples, as well as wild type lysates probed with pre-bleed 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 (Clark and McKearin, 1996). Both of our antibodies recognize multiple bands in immunoblots, some of which are present in pre-bleed sera while others are not (Fig. 2). Purification of antibody against Stwl-pMBP antigen resulted in loss of pre-bleed-specific bands, but other background bands remained (Fig. 5). A background band running at ~130 kDa in the G77 pre-bleed sera made it difficult to confirm absence of the presumed Stwl band in final bleed immunoblots against null flies; however, we were able to confirm Stwl-specificity of the G77 antibody in immunoblots on S2 lysates, which have significantly less background.

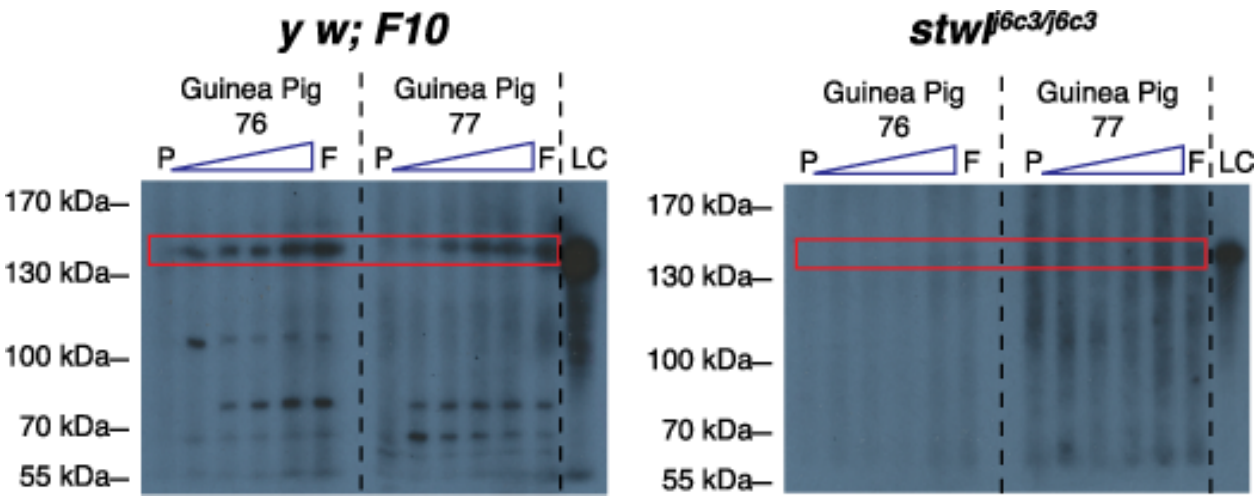


FIGURE 2. **Stwl antibodies recognize a ~130 kDA fragment in ovarian lysates**
Western blots on whole ovarian lysates from ~20 wild-type (*y w F10*) and ~20 *stwl* null (*y w; stwl$^{j6c3/j6c3}$*) individuals; ovaries were dissected from individuals aged 1-4 days. Each 6% SDS PAGE gel was loaded with a single lysate then transferred and probed with multiple antibodies as indicated. Lanes labelled "P" and "F" were probed with pre-bleeds and final-bleeds (exsanguinations) of the indicated antibody. Lanes between P and F were probed with test-bleeds collected from the indicated animal in chronological order. Test-bleeds were collected 6, 9, 10, and 12 weeks after initial immunization; final-bleeds were collected 13 weeks after initial immunization. "LC" lanes were probed with loading control (Guinea Pig α-Chromator). Test-bleeds and final-bleeds of each antibody recognize a ~130 kDa fragment in wild-type ovarian lysate that is not detected in *stwl* null ovarian lysates.

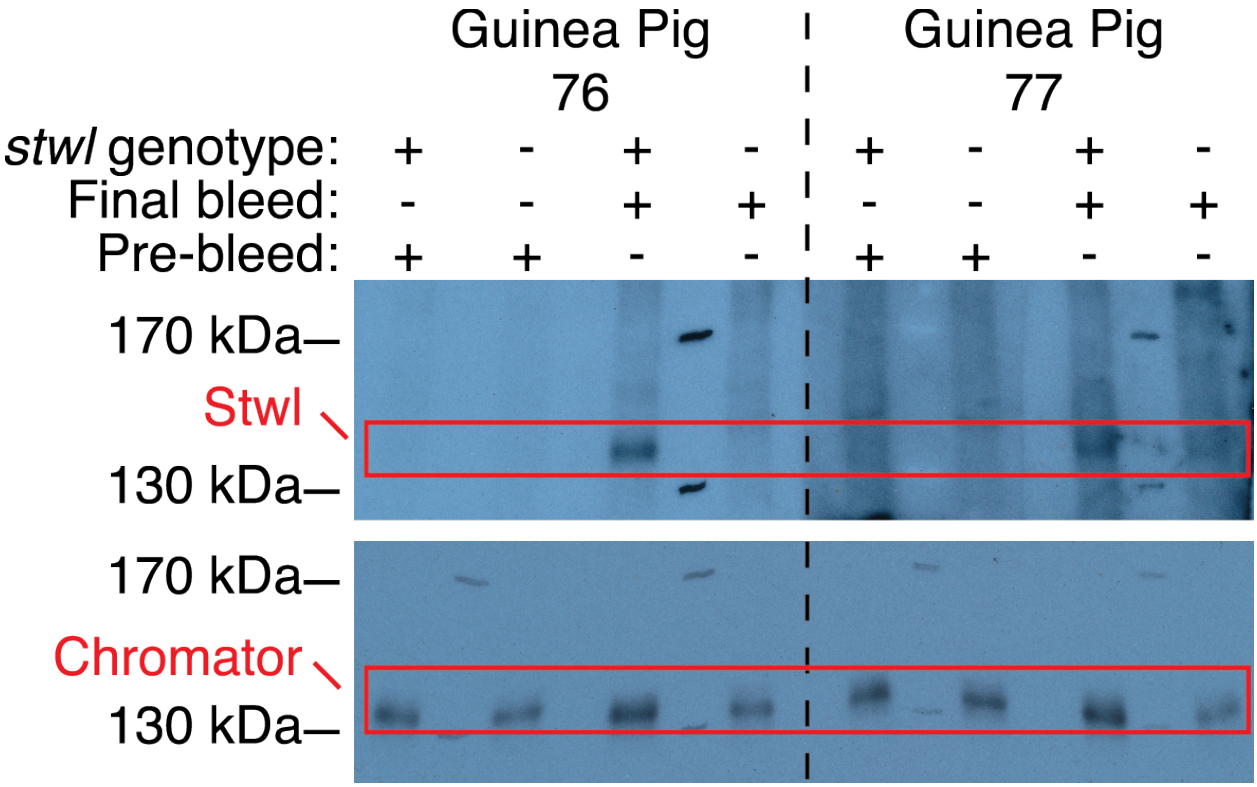


FIGURE 3. **Stwl antibodies recognize a ~130 kDA fragment in whole fly lysates**
Western blots on whole fly lysates from ~10 wild-type (*y w; F10*) and ~10 *stwl* null ($stwl^{l\Delta 95/def}$) individuals aged 1-4 days. 6% SDS PAGE gel was loaded with lysates as indicated (row labelled "*stwl* genotype"), then transferred and probed with pre-bleed or final-bleed of each antibody, as indicated. The bottom panel shows the same membrane stripped and re-probed with a loading control (Guinea Pig α-Chromator). Final-bleeds of each antibody recognize a ~130 kDa fragment in wild-type whole-fly lysate that is not detected in *stwl* null lysates.

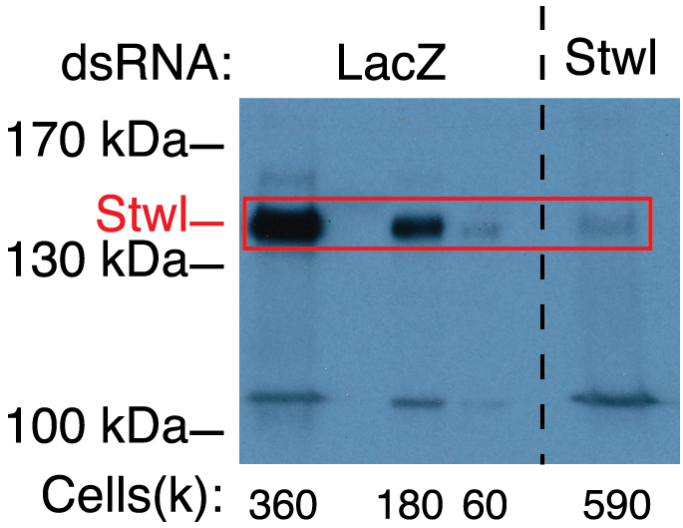


FIGURE 4. ***stwl* dsRNA treatment specifically knocks down ~130 KDa protein in S2 cells**
S2 cells were treated with either control dsRNA (*LacZ*) or dsRNA targeting the *stwl* transcript. A standard curve was constructed using the signal from the *LacZ* treated cells, indicating ~80% knockdown of Stwl signal in the *stwl* dsRNA treated cells.

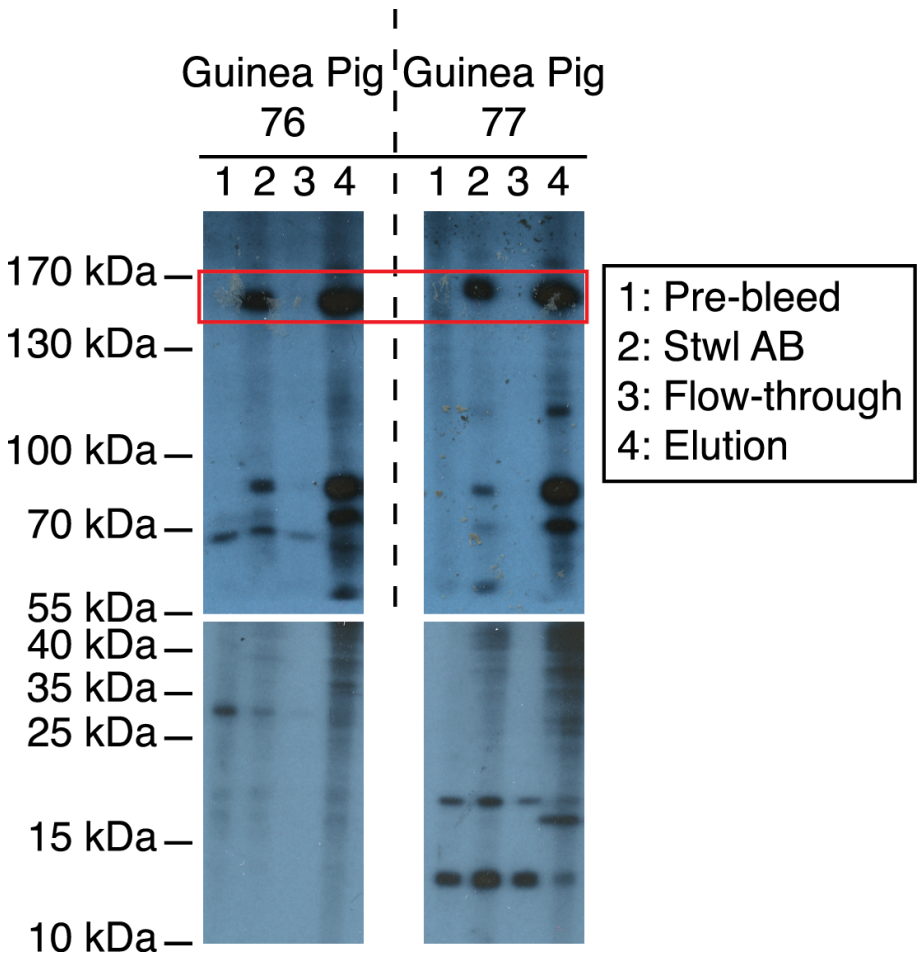


FIGURE 5. **Purification of Stwl antibodies over Stwl-MBP epitope**

Stwl-MBP target antigen was bound to CNBr-activated sepharose resin and used to purify Stwl antibodies. Each antibody was applied to columns containing Stwl-MBP-containing resin and washed repeatedly prior to elution in glycine (pH 2.8). Shown here are western blots on lysates from ~30 wild-type (*y w; F10*) ovaries dissected from individuals aged 1-4 days. 6% and 15% (top and bottom) SDS PAGE gels were loaded with wild-type ovarian lysates, then transferred and probed with pre-bleed, final-bleed, purified eluant and flow-through of each antibody, as indicated. Purification of Stwl antibodies concentrates the Stwl-specific signal at the expense of background (note the low MW bands which remain in the flowthrough).

2. Immunofluorescence.

We performed immunofluorescence experiments to confirm that the Stwl antibodies were targeting a nuclear protein and to compare to IF experiments done with other Stwl antibodies. As a first experiment, *D. melanogaster* ovaries were dissected from wild type (*y w* F10) individuals in cold 1x PBS and fixed in 4% paraformaldehyde with 0.1% Tween 20 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), and then separate incubations with primary antibody (1:100 each of antibodies G76, G77 and their respective pre-bleeds, along with 1:200 Rabbit Vasa from Santa Cruz Biotechnology, Inc.) in PBTA overnight at 4°C. 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. This initial experiment demonstrated that the Stwl antibodies failed to penetrate ovarian tissue. Stwl signal instead came mostly from epithelial sheath cells, which we interpreted as likely background noise because this signal was identical in samples probed with G76, G77 and their respective pre-bleeds.

To determine the optimum conditions for immunofluorescence imaging with the Stwl antibodies, we dissected ovaries and testes from wild-type individuals and probed them with G76 antibody in varying concentrations and fixation conditions. We varied the antibody concentration from 1:10 to 1:1000 (1:10, 1:200, 1:500, 1:1000) under standard fixation conditions (0.1% Tween 20 for 15 minutes at room temperature). We also varied fixation conditions by reducing temperature to 4°C, by increasing incubation time (from 15 to 20 minutes), by increasing Tween 20 concentration in fixative (from 0.1% to 0.2% - 0.5%), and by replacing Tween 20 in fixative solution with 0.1% or 0.2% Triton X-100. The following chart lists all of the conditions:

Fixative (+4% PFA in PBS)	Fixation time	Fixation Temp.	Antibody Concentration
0.1% Tween 20	15 minutes	25°C	1:200
0.1% Tween 20	15 minutes	25°C	1:10
0.1% Tween 20	15 minutes	25°C	1:500
0.1% Tween 20	15 minutes	25°C	1:1000
0.1% Triton X-100	10 minutes	25°C	1:100
0.2% Triton X-100	10 minutes	25°C	1:100
0.5% Tween 20	20 minutes	4°C	1:100
0.2% Tween 20	20 minutes	4°C	1:100
0.1% Tween 20	15 minutes	4°C	1:100

From these experiments, we determined that a lower concentration of antibody (1:200) and Triton X-100 instead of Tween 20 achieved better penetrance in both ovaries and testes. Subsequent immunostaining experiments were therefore conducted using 4% PFA fixative with 0.1% Triton at 4°C, followed by overnight incubation with Stwl antibodies at a concentration of 1:200 in PBTA.

The aim of the next experiment was to confirm that G76 and G77 antibodies had similar IF signals that did not appear in *stwl* mutant tissue or with pre-bleed sera. We dissected testes and ovaries from wild type (*y w* F10) and mutant (nulls or RNAi knockdown) males and females

and probed under the previously described conditions. Wild-type tissues were also probed with pre-bleeds under the same conditions. These experiments confirmed a loss of germline nuclear signal in mutant ovaries and testes, as well as an absence of this signal from wild-type tissue probed with pre-bleed sera.

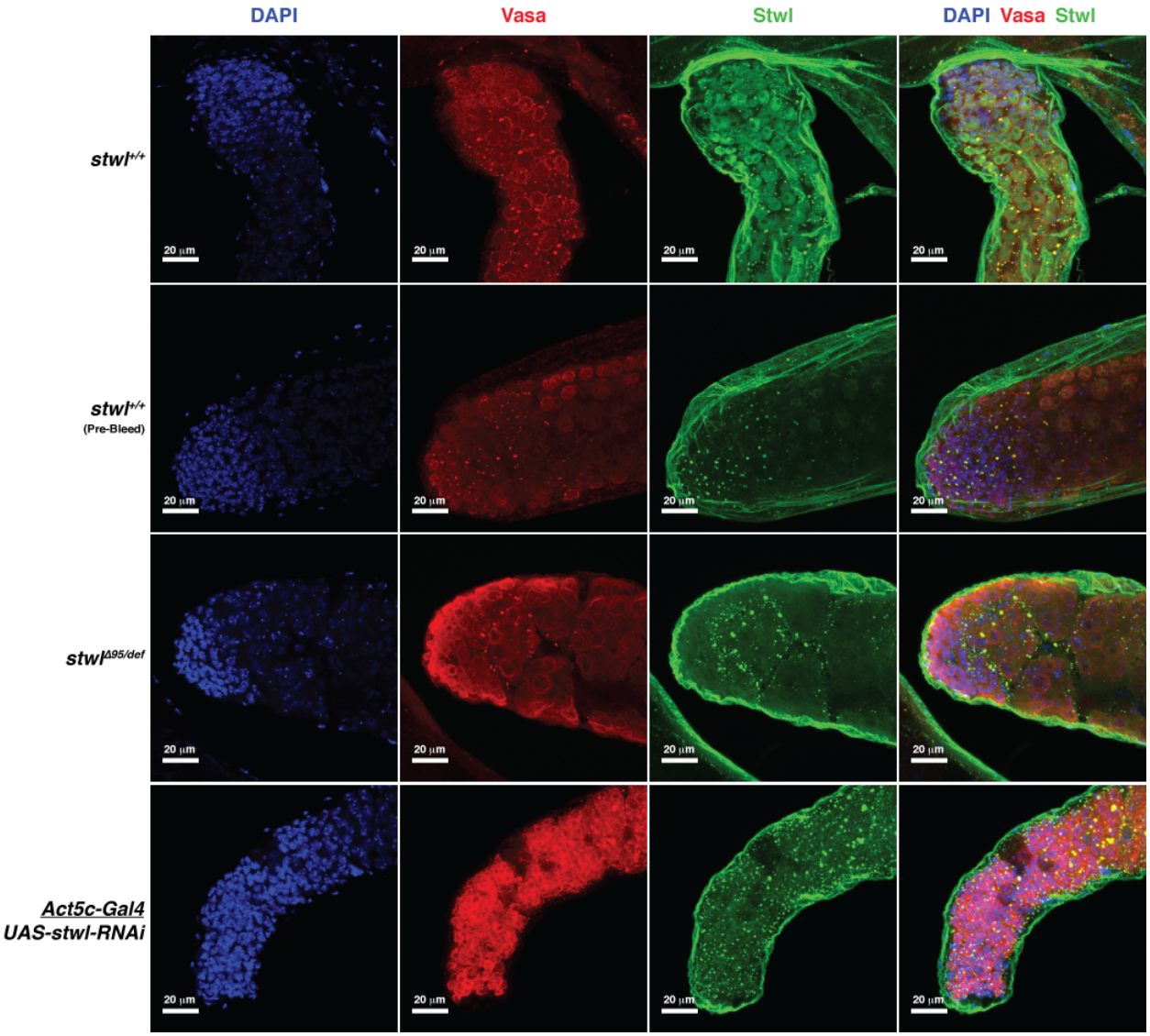


FIGURE 6a. α-Stwl GP 76 serum labels germ cell nuclei in wild-type testis
D. melanogaster testes were dissected from males of the indicated genotype 10-15 days post-eclosion. α-Vasa labels germ cells, DAPI labels cell nuclei. α-Stwl GP 76 produces a signal in wild-type germ cell nuclei. This signal is not present in testes lacking Stwl protein or in wild-type testes labeled with pre-bleed serum from the same animal. All images are maximum-

intensity projections from a z-series representing a depth of 10, 10, 10, and 5 microns respectively (top row to bottom row).

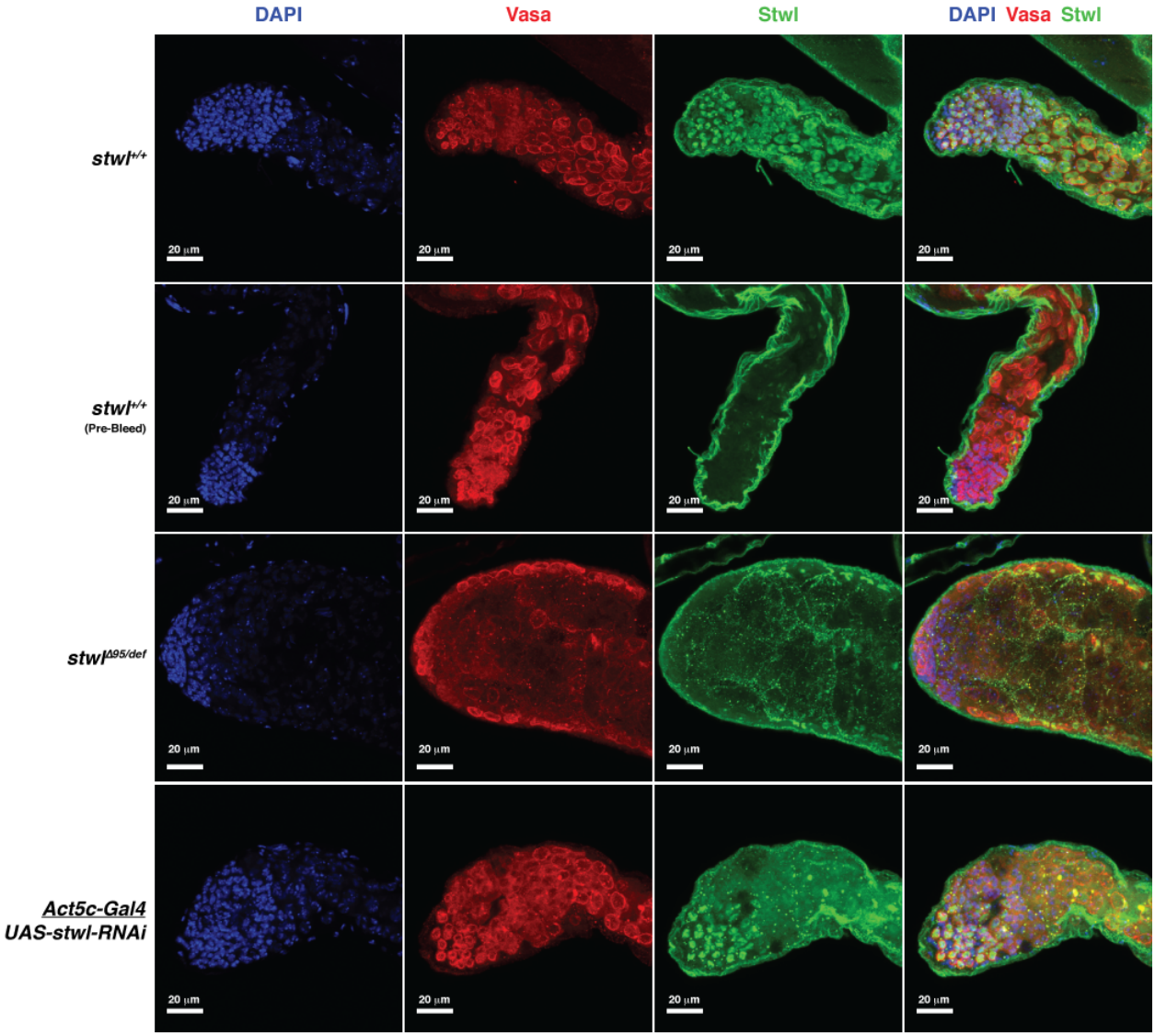


FIGURE 6b. α-Stwl GP 77 serum labels germ cell nuclei in wild-type testis

D. melanogaster testes were dissected from males of the indicated genotype 10-15 days post-eclosion. α-Vasa labels germ cells, DAPI labels cell nuclei. α-Stwl GP 77 produces a signal in wild-type germ cell nuclei. This signal is not present in *stwl* null testes or in wild-type testes labeled with pre-bleed serum from the same animal. Signal is reduced but still present in some early germ cells in *Act5c-Gal4/UAS-stwl-RNAi* testes, possibly due to incomplete reduction of Stwl protein in these animals. All images are maximum-intensity projections from a z-series representing a depth of 10, 10, 10, and 5 microns respectively (top row to bottom row).

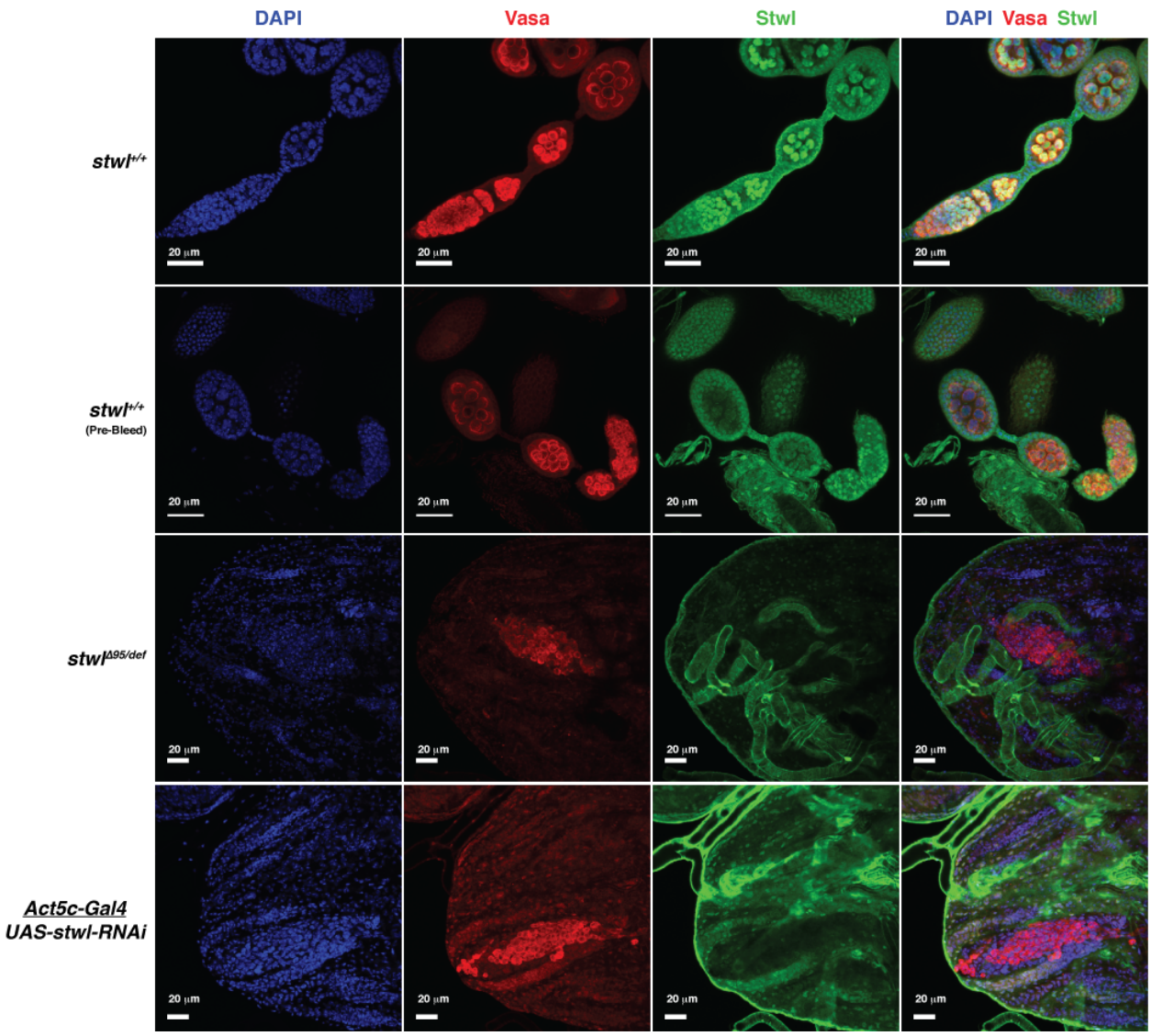


FIGURE 6c. α-Stwl GP 76 serum labels germ cell nuclei in wild-type ovary

D. melanogaster ovaries were dissected from females of the indicated genotype 10-15 days post-eclosion. α-Vasa labels germ cells, DAPI labels cell nuclei. α-Stwl GP 76 produces a signal in wild-type germ cell nuclei. This signal is not present in ovaries lacking Stwl protein or in wild-type ovaries labeled with pre-bleed serum from the same animal. Faint nuclear signal appears in follicle cell nuclei probed with pre-bleed serum; this signal is dramatically reduced relative to nuclear germline signal in ovaries probed with final bleed serum. All images are maximum-intensity projections from a z-series representing a depth of 10, 10, 10, and 5 microns respectively (top row to bottom row).

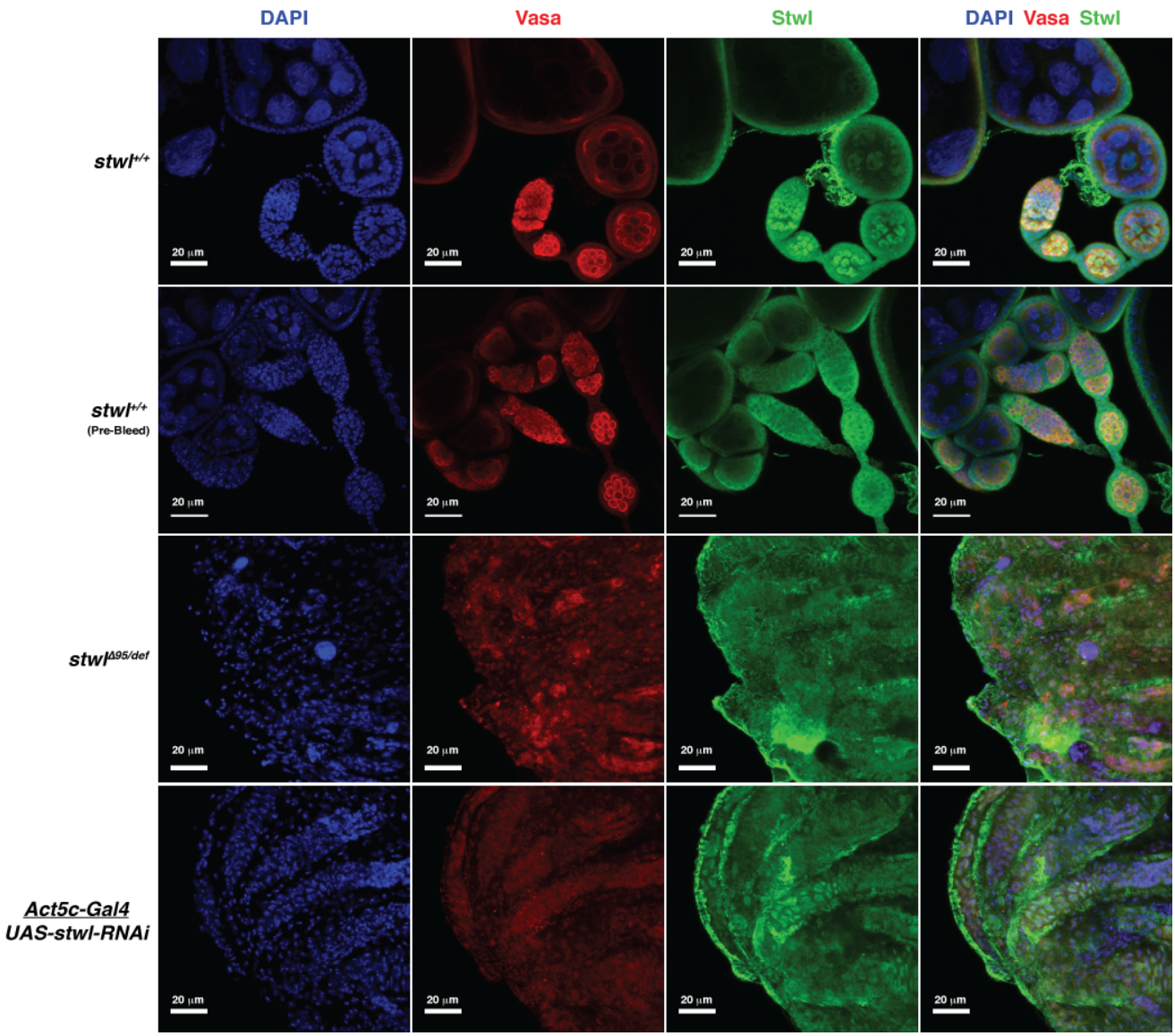
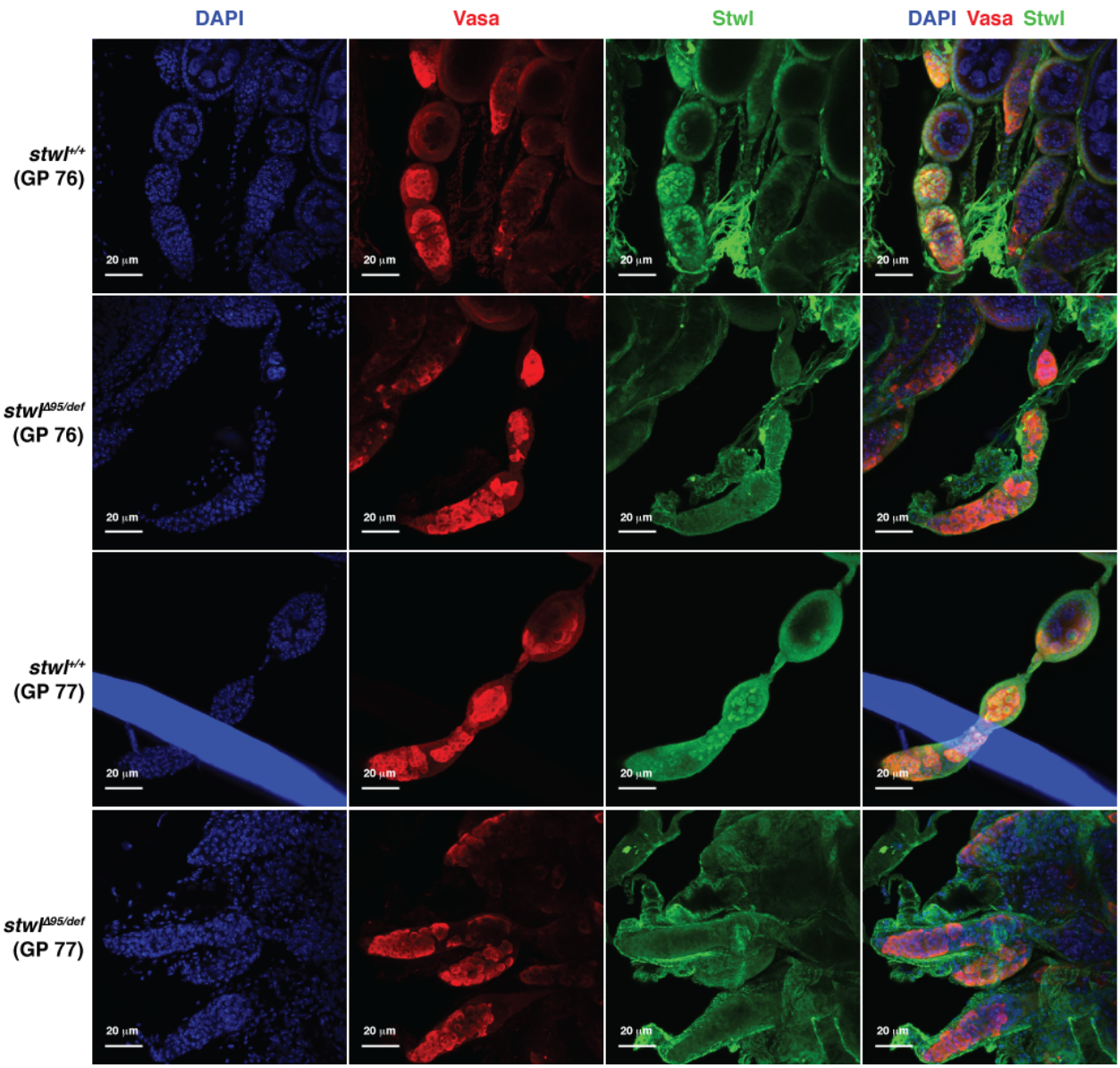


FIGURE 6d. α-Stwl GP 77 serum labels germ cell nuclei in wild-type ovaries

D. melanogaster ovaries were dissected from females of the indicated genotype 10-15 days post-eclosion. α-Vasa labels germ cells, DAPI labels cell nuclei. α-Stwl GP 77 produces a signal in wild-type germ cell nuclei. This signal is not present in *stwl* null ovaries or in wild-type ovaries labeled with pre-bleed serum from the same animal. Signal is reduced but still present in some early germ cells in *Act5c-Gal4/UAS-stwl-RNAi* testes, possibly due to incomplete reduction of Stwl protein in these animals. All images are maximum-intensity projections from a z-series representing a depth of 10, 10, 5, and 5 microns respectively (top row to bottom row).



<u>FIGURE 6E</u>. α-Stwl sera label germ cell nuclei in wild-type ovaries of newly-eclosed individuals

D. melanogaster ovaries were dissected from females of the indicated genotype <12-hours post-eclosion. α-Vasa labels germ cells, DAPI labels cell nuclei. α-Stwl sera from both animals produce a signal in wild-type germ cell nuclei. This signal is not present in *stwl* null ovaries. All images are maximum-intensity projections from a z-series representing a depth of 10 microns.

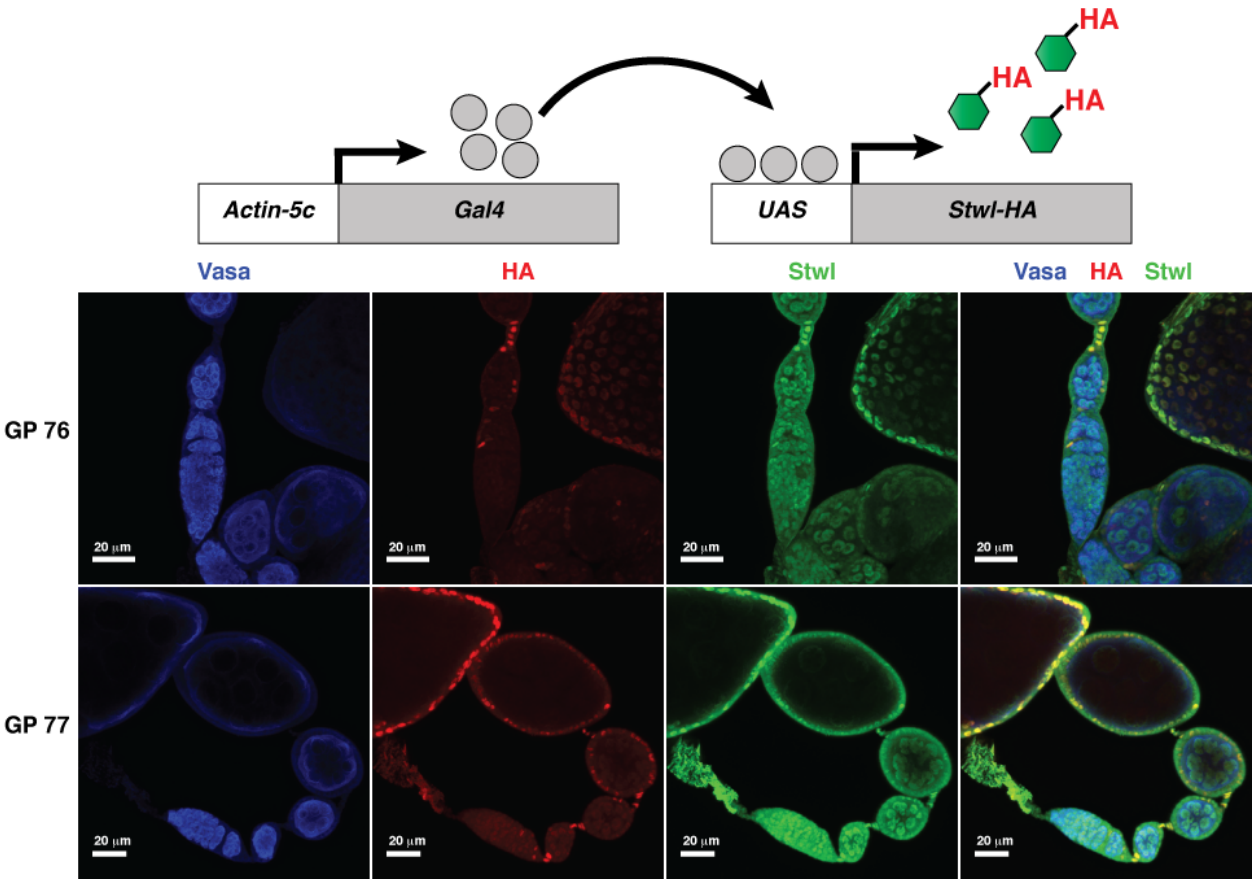


FIGURE 7. α-Stwl sera detects ectopically expressed Stwl-HA protein
D. melanogaster ovaries were dissected from *Act5c-Gal4/UAS-stwl-HA* females of the indicated genotype 0-1 days post-eclosion. Ovaries were probed with α-Vasa (germ cells), α-HA, and either GP 76 or GP 77 α-Stwl serum. α-HA signal recognizes cells in which Stwl-HA is being driven; in these examples, expression is typically 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 microns.

3. Immunoprecipitation.

Prior to performing ChIP-seq, we immunoprecipitated Stwl directly from S2 cells without formaldehyde cross-linking. Briefly, S2 cell nuclei from ~12 million cells were isolated in Farnham lysis buffer and lysed in RIPA buffer, followed by a thorough sonication. 300 ul of lysate in RIPA buffer was incubated with G76 and G77 antibodies, as well as a control antibody (Chromator) at concentrations of 1:50 and 1:100. Antibody complexes were captured using

protein-A agarose beads (Millipore: 16-125), which were then washed and eluted in 2x SDS sample buffer w/ BME. Immunoblotting of these samples against G76, G77 and Chromator antibodies demonstrated that both Stwl antibodies specifically pull down the ~130 kDa protein from input solution and that Chromator antibody does not (Fig. 8). These blots also demonstrated that G76 and G77 antibodies recognize the same protein.

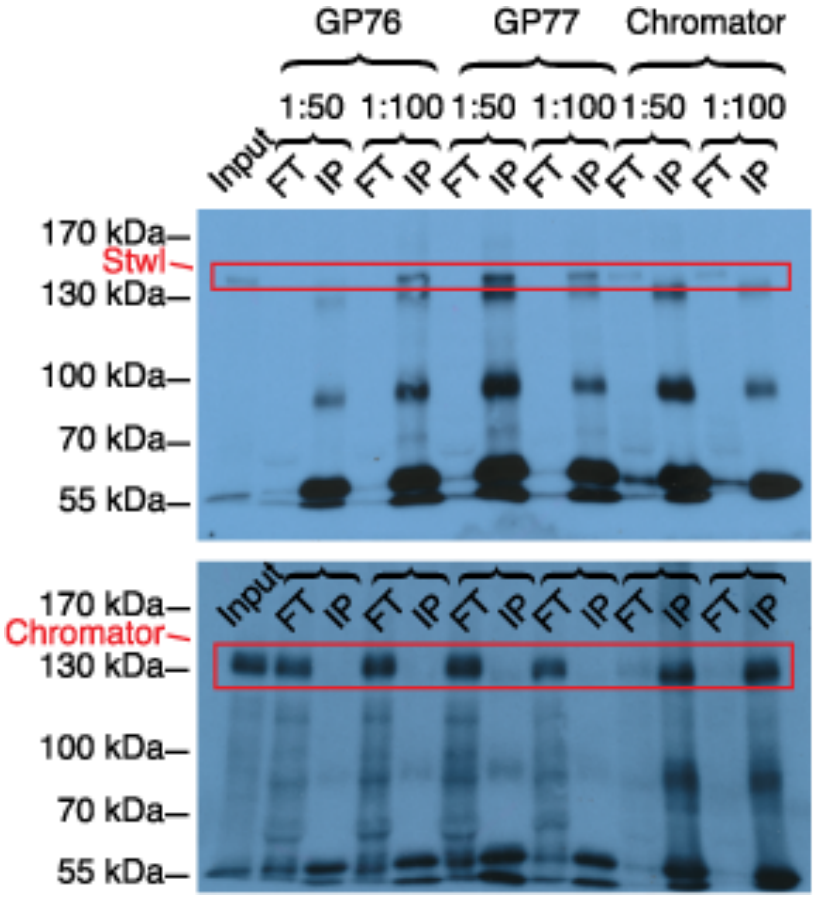


FIGURE 8. **α-Stwl sera immunoprecipitate Stwl from S2-cell lysate**
S2 cell nuclei were lysed in RIPA buffer (Input), then incubated with one of two α-Stwl serum or a control antibody (α-Chromator) at 1:50 and 1:100 dilutions. Antibody-Protein complexes were isolated with Protein-A Agarose beads. Input, flow-through (FT) and immunoprecipitated complexes (IP) were loaded on a 6% SDS PAGE gel, then transferred and 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 Figures 2, 3, and 5), as does Chromator. Both α-Stwl sera appear to IP Stwl effectively at a concentration of 1:100 (Stwl protein is eliminated from flow-through). α-Chromator antibody fails to IP Stwl (Stwl protein remains in flow-through), but successfully IPs Chromator.

Cell Culture and RNAi

S2 cells were a gift from the Lis lab. 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 ug/ml of *lacZ*- or *stwl*-dsRNA for 60 minutes before addition of M3/BPYE medium containing 13% FBS (final concentration, 10% FBS, 7.5 ug/ul 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 all targeting the second exon of *stwl*, which is present in all *stwl* transcripts. Each of the dsRNA sequences was already generated and confirmed as successful for RNAi in S2 cells. S2 cells were treated with 10 ug/ml of each dsRNA.

Primers for generation of DNA template to be used for T7 RNA synthesis
T7 promoter sequence is in lower-case letters. The last 3 primer sets are to *stwl*.

LacZ_F: gaattaatacgactcactatagggagaGATATCCTGCTGATGAAGC
LacZ_R: gaattaatacgactcactatagggagaGCAGGAGCTCGTTATCGC
23913_F: taatacgactcactataggGGAAAAATCGTCCCAAGACA
23913_R: taatacgactcactataggGGAGAAGTAGTGTCGCTGCC
37657_F: taatacgactcactataggAAATGATCAAGCCCCAGATG
37657_R: taatacgactcactataggTACCACCCATAGCCGCTAAC
36036_F: taatacgactcactataggAGGTATTCCGTGTGCGGTAG
36036_R: taatacgactcactataggCCAGTGGGTGAGATTTGCTT

Chromatin Immunoprecipitation in S2 cells

Once we verified that both Stwl antibodies are Stwl specific and successfully immunoprecipitate Stwl from S2-cell lysates, we began preparation for ChIP-Seq experiments. We performed a number of pilot experiments to confirm that we could successfully recover DNA from immunoprecipitated lysates. We attempted various fixation methods, lysis buffers and sonication times and intensities before settling on the following methods.

Subsequent to dsRNA treatment for 3 days, cells were centrifuged for 5 minutes at 1000xg followed by removal of media. Cells were washed once in 1x PBS to remove cellular debris and improve cross-linking performance. They were 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, then spun and washed in 1x PBS brought to 4°C. Cells were 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 in a Bioruptor (Diagenode) water bath to shear DNA to desired size range (300-500 bp). The following settings were used: 20 seconds on, 1 minute rest, high intensity, 45 minutes total, with quick spins of lysates every 15 minutes to settle sample and re-fill bioruptor with ice-cold water.

6 ul of freshly thawed final bleed and pre-bleed sera were added to 300 ul of cell lysate (1:50 dilution) and incubated overnight at 4°C. Cell lysates were prepared with approximately 34,000 cells per ul, so that each IP experiment was performed on roughly 10,000,000 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 ul 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). RNA-Seq data processing, QC and analysis of S2 cells samples was performed as described in chapter 2.

ChIP-Seq data processing and GC-correction

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) (Andrews, 2010; Bolger et al., 2014). Reads were aligned to the unmasked *Drosophila melanogaster* genome (r6.03) using bowtie2 default settings (Langmead and Salzberg, 2012). All reads with mapping quality <20 were removed using SAMtools (Li et al., 2009). All alignment files were corrected for GC bias using the deepTools commands computeGCbias and correctGCbias (Ramírez et al., 2014). 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.

Peak calling and IDR

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 final bleeds 76 and 77 (IP), pre-bleeds 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 loosely followed peak calling standards established by the ModERN consortium (Kudron et al., 2018). We performed peak calling on all mock and IP samples using the peak calling algorithm MACS2 (Feng et al., 2012). 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 (Li et al., 2011). 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 (Fig. 9). After IDR, any peaks in our IDR 76 and IDR 77 IP peaksets that overlapped with spurious peaks from either mock were removed using BEDTools subtract (Quinlan and Hall, 2010). 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.

Repetitive DNA alignment and analysis

Limitations and challenges of identifying enriched repetitive elements from ChIP-Seq data have been well documented (Huang et al., 2013; Lin et al., 2015; Marinov et al., 2015). 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.

As described for RNA-Seq analysis, 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). Unmapped reads from this alignment were saved and aligned to the unmasked *Drosophila melanogaster* genome (r6.03) using bowtie2 default settings (Langmead and Salzberg, 2012). We used a custom perl script to count the number of reads aligning to repetitive sequences. Rather than counting reads aligned to gene bodies as we did for RNA-Seq analysis, we calculated the number of reads aligned to each 10 kb bin of the genome. We concatenated the repetitive and genomic read counts into a single file for each sample.

We analyzed count data using DESeq2 (Love et al., 2014). We imported raw counts and removed genes with mean read count <=10 across all samples. We estimated differential enrichment of genetic segments between IP and mock samples, taking into account the antibody used (GP76 or GP77) and the source S2 population (replicate 1 or 2). A genetic 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 is explained by variance between mock and IP samples (Fig. 19).

Subsequent to differential expression analysis, all $\log_2$(fold-change) estimates were transformed using apeGLM shrinkage estimator to reduce variability in LFC values among low-count genes (Zhu et al., 2018). 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 (Yu et al., 2012).

Results

Stwl binds to active promoters upstream of transcription start sites

Both Stwl antibodies 76 and 77 met all of our antibody validation criteria, including recognition of the target protein in immunoblotting and immunofluorescence experiments, successful immunoprecipitation of the target protein, and relatively low background in S2 cells (Fig. 1-8). Despite ample evidence that these antibodies specifically bind to Stwl, confirming that an antibody is "ChIP-grade" is more challenging, as it requires that the antibody is capable of binding the target protein-DNA complex after chemical crosslinking. The quantities of DNA recovered after ChIP are often quite low, and validating a successful ChIP experiment is further complicated when the binding targets are unknown. Fortunately, our ChIP-Seq experiments produced a robust set of peaks when compared to both input and mock samples (Fig. 9).

We identified 2,153 Stwl binding sites across the genome, mostly on the autosomes and the 4th chromosomes (Fig. 18). Upon closer inspection, we found that most Stwl peaks (92%) were within 1 kb of a promoter, which we defined as the region 1 kb upstream and downstream of a transcription start site (Fig. 10). We also found that Stwl appears to bind preferentially to active promoters, as defined by activity (RPKM>=2) in S2 cell RNA-Seq. In fact, among genes with TSS within 1 kb of a Stwl peak, 68% are active, compared to only 40% of genes genome-wide (Fig. 11). Plotting peak density centered at promoters revealed that Stwl peaks are highly enriched and centered ~100 bp upstream of transcription start sites (Fig. 12).

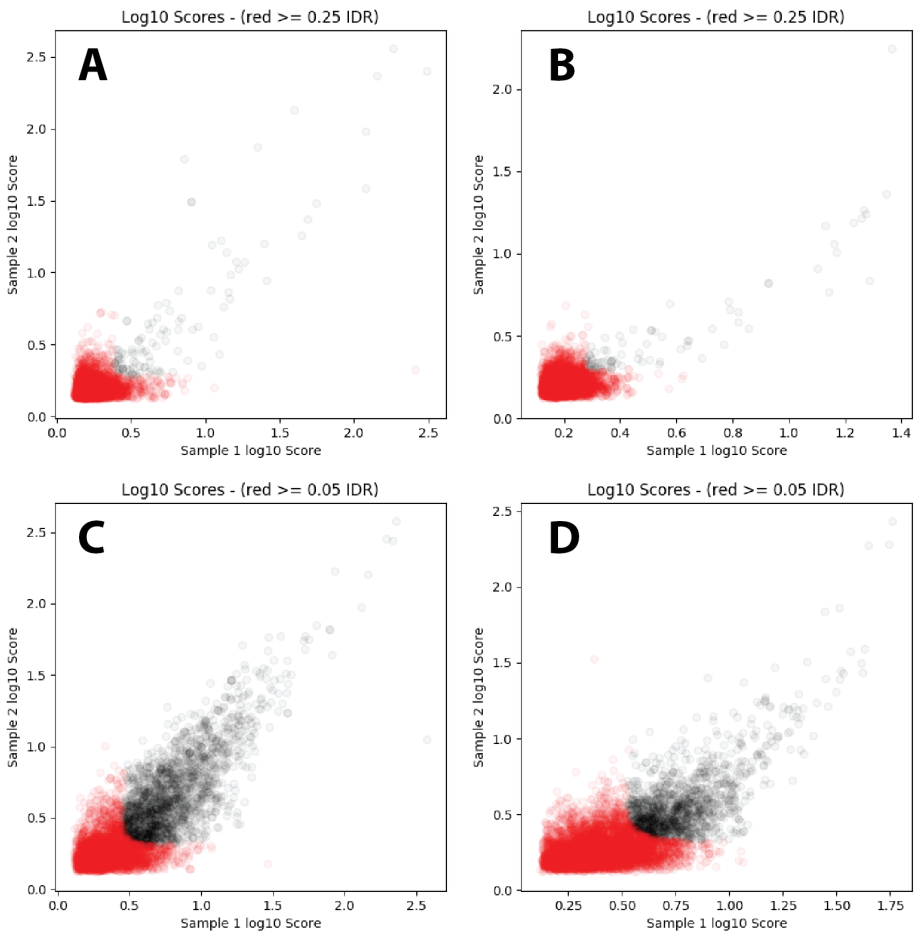


Figure 10 Stwl IP replicates create reproducible peaksets

IDR plots generated as output from IDR program. These 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 or mock. 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. score) 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. Peaks generated from mock samples are shown in A and B; peaks from Stwl IPs are in C (GP76) and D (GP77). Very few peaks were identified in mock ChIP-Seq experiments, even with a relaxed IDR threshold of 0.25.

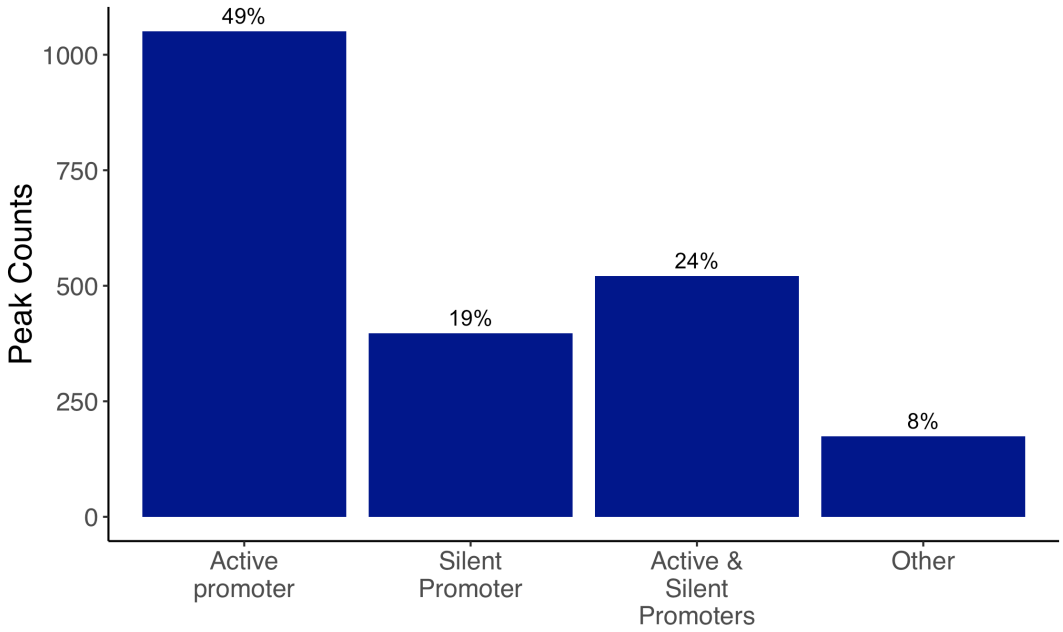


Figure 11 Stwl peaks primarily occupy promoters.
Breakdown of Stwl binding site preferences, as indicated by number of peaks within 1 kb of a promoter, and the transcriptional activity of that promoter. Promoter activity was determined from RNA-Seq data in S2 cells; active promoters have RPKM>=2.0. Stwl peaks not within 1 kb of a promoter are categorized as "Other". A majority of peaks are within 1 kb of multiple promoters; in such cases where all promoters are active or silent, the peak was categorized as such; when adjacent promoters were a mixture of active and silent, the peak was categorized as binding to active and silent promoters.

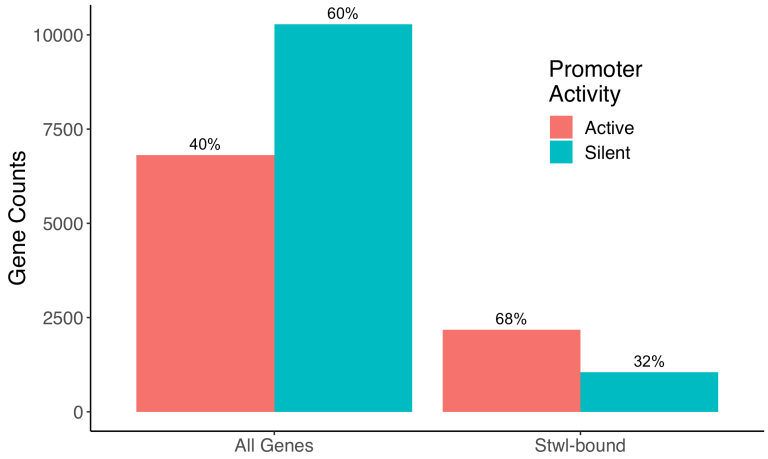


Figure 12 Majority of Stwl-bound promoters are active in S2 cells.
Due to the compact nature of the *D. melanogaster* genome, most Stwl-bound peaks occupy multiple promoters. The majority of these promoters are active (68%), in contrast to the majority of total promoters, which are silent (40%).

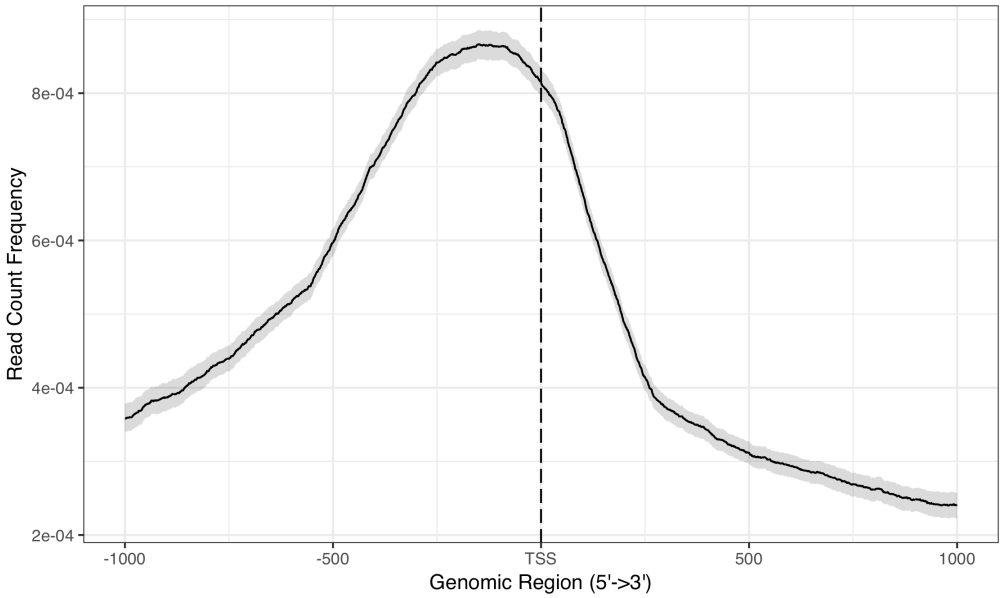


Figure 13 Stwl peaks are enriched just upstream of transcription start sites.
Binding profile of Stwl-bound promoters. This binding profile is common among known insulator-binding proteins (Nègre et al., 2010). Grey boundaries represent 95% CI around the read count frequency.

Stwl binds to active promoters upstream of transcription start sites

One of the major motivations for performing ChIP-Seq against Stwl was to identify genes that are directly regulated by Stwl. Our RNA-Seq analyses found that a large number of genes are misregulated in *stwl* null ovaries, many of which may be indirectly affected by Stwl loss, such as transcripts that are mis-expressed as a consequence of genome instability in apoptotic cells. We also identified a surprising number of differentially expressed genes in *stwl* heterozygous ovaries, which lack obvious morphological defects. Finally, we found that a number of testis-enriched genes are significantly upregulated in *stwl*-dsRNA-treated S2 cells, where the homogenous nature of the cellular population reduces biological noise associated with complex tissue. We expected to find that genes which are upregulated in *stwl*-dsRNA-treated cells also tend to be adjacent to Stwl binding sites.

To test this hypothesis, we performed overrepresentation tests on genes adjacent to Stwl binding sites (referred to as Stwl-bound). We looked for overrepresentation of genes misregulated or ectopically expressed in *stwl* mutants or *stwl*-dsRNA-treated S2 cells among Stw-bound genes. Surprisingly, Stwl-bound genes did not tend to be misregulated in *stwl*-dsRNA-treated S2 cells (Fig.14). We found that 99 genes in these cells are upregulated/ectopically expressed and Stwl-bound, out of a total of 491 upregulated/ectopically expressed genes. Among this small class of genes are *bgcn* and *pum*, which are required for GSC differentiation and maintenance, respectively.

Contrary to the lack of correlation between S2 cell RNA-Seq and ChIP-Seq data, overrepresentation analysis determined that Stwl-bound are prefertentially upregulated in *stwl* heterozygous ovaries (651 Stwl-bound and upregulated/1,911 upregulated). In addition to searching for enrichment of genes misregulated in *stwl* mutant RNA-Seq, we performed overrepresentation analysis against genes with tissue-specific enrichment. Our expectation was that testis- and imaginal disc-enriched genes would be preferentially bound by Stwl. Instead, we found that Stwl is more likely to bind near ovary- and S2-cell-enriched genes. Neither of these gene classes is upregulated or ectopically expressed in any of our *stwl* mutant tissues or *stwl*-dsRNA-treated S2 cells; rather, they tend to be downregulated in these samples (Chapter 2, Fig. 51-54). The subset of ovary- and S2-cell-enriched genes that are Stwl-bound is not the same genes that are misregulated in *stwl* mutant tissues or *stwl*-dsRNA-treated S2 cells, making it difficult to infer biological significance from this result.

We also performed overrepresentation tests on Stwl-bound genes stratified by activity in S2 cells (Fig. 14). The upregulation of Stwl-bound genes in *stwl* heterozygous ovaries is consistent between active and silent genes. Otherwise, Stwl-bound active genes are not enriched

for any tissue-specific genes, while Stwl-bound silent genes are enriched for S2- and ovary-specific genes, as noted for all Stwl-bound genes. Interestingly, Stwl-bound silent genes are also enriched for genes that are upregulated in *stwl* null ovary, and ectopically expressed in *stwl* null and *stwl* heterozygous ovary. Furthermore, these genes are also biased towards expression in head, CNS and salivary gland. Each of these tissue classes is enriched among ectopically expressed genes in all *stwl* mutant tissues (*stwl* null ovary, *stwl* heterozygous ovary, and *stwl* null testis); head- and CNS-specific genes are also highly enriched among *stwl*-dsRNA-treated S2 cells. We posit that while Stwl appears biased towards binding active promoters (as defined by S2 cell expression), many sites of binding do not correlate with expression of adjacent genes.

To further investigate the possibility that Stwl-bound silent genes are associated with Stwl function, we performed GO analysis for enriched biological process terms among Stwl-bound genes, active and silent. We found that Stwl-bound genes (as a whole) and Stwl-bound active genes are enriched for a handful of GO terms, including genes associated with spermatogenesis, oogenesis, GSC maintenance, and cell fate determination (Fig. 15). However, enrichment for these GO terms was more pronounced among Stwl-bound silent genes, and many more related terms were enriched among this class of genes than among active and total Stwl-bound genes.

It is important to exercise caution in interpreting these results. Our ChIP-Seq data was generated from S2 cells, primarily to avoid the challenges that come with generating and interpreting high-throughput data from complex tissues such as the *Drosophila* ovary. It also provided us with an opportunity to examine the effect of Stwl loss in a homogenous cell type, where reduction of Stwl has a more subtle phenotype than in animals. A limitation is that we define active and silent promoters based on the transcript abundance estimated from RNA-Seq of

S2 cells. We note that there are substantial differences in gene expression between S2 cells and ovaries (Fig. 16a). However, re-annotating promoters according to RNA levels in WT ovary did not affect the results of the overrepresentation tests, largely due to the strong overlap between Stwl-bound, ovary active and Stwl-bound S2 cell active genes (Fig. 16b).

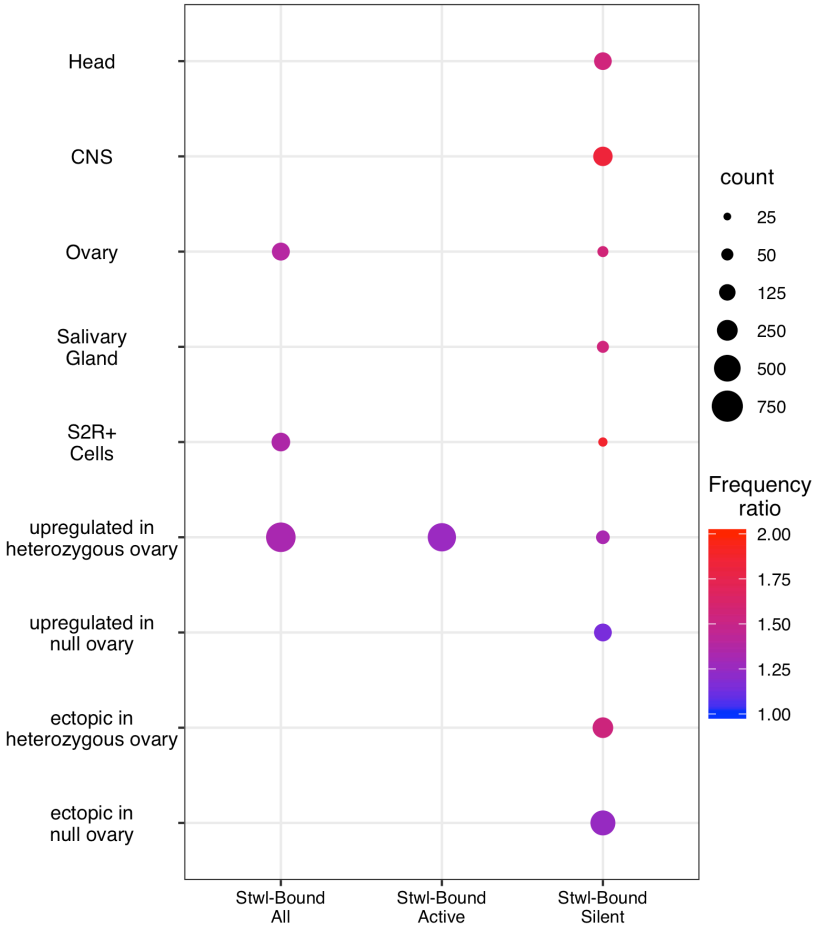
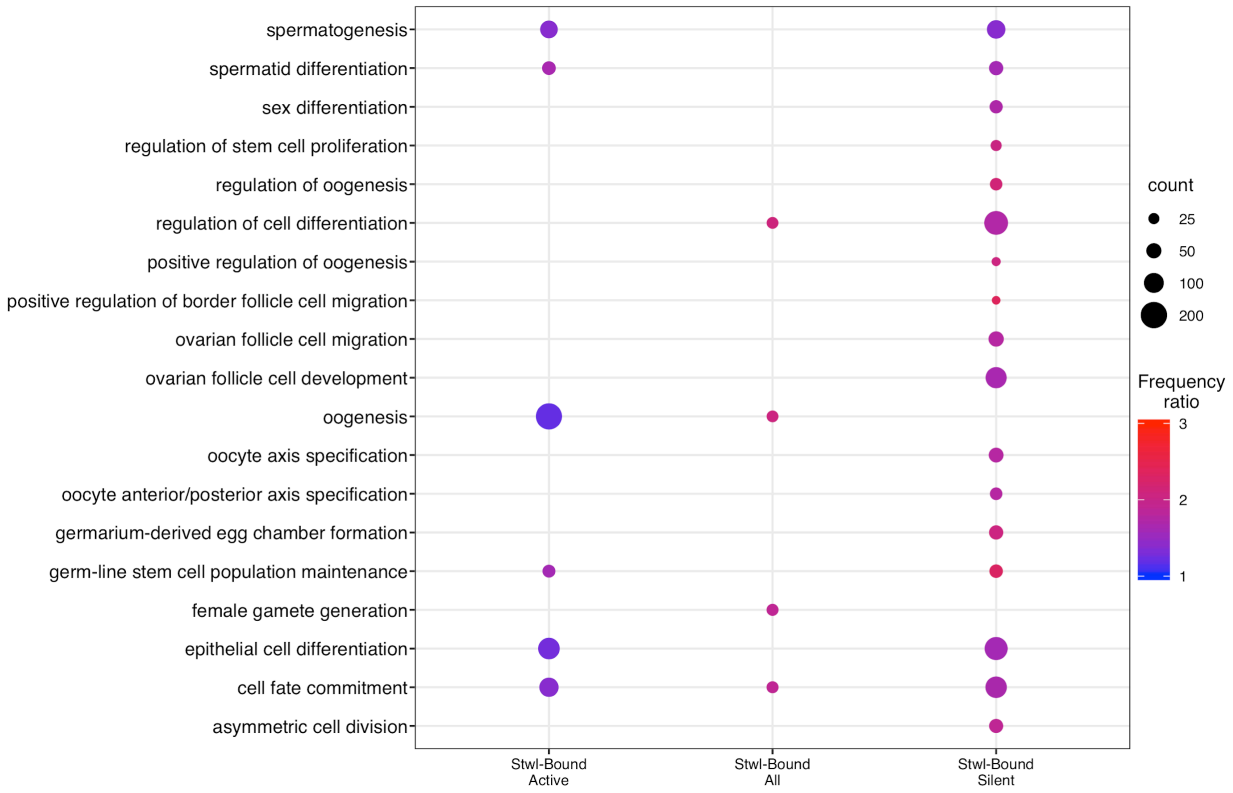


Figure 14 Stwl-bound genes are upregulated in *stwl* heterozygous ovaries.
Comparison of overrepresentation test results for Stwl-bound genes. Genes upregulated or ectopically expressed in *stwl* mutant tissue (Chapter 2 of this thesis) were included as a classification. Significance was determined from a hypergeometric distribution comparing the gene ratio (genes enriched in tissue X among queried genes/all queried genes) to the background ratio (all genes enriched in tissue X/all genes). For Stwl-bound active and silent genes, the background gene set was restricted to all active and silent genes, respectively (as determined by RPKM in S2 cells). Frequency ratio = gene ratio/background ratio. Count and frequency ratio are plotted for each set of tissue-enriched genes. Only gene sets with FDR<.05 are plotted.



<u>Figure 15</u> Transcriptionally inert genes bound by Stwl are enriched for GO terms related to Stwl function.

Comparison of overrepresentation of GO Biological Process terms for Stwl-bound genes. Significance was determined from a hypergeometric distribution comparing the gene ratio (genes enriched for term X among queried genes/all queried genes) to the background ratio (all genes enriched for term X/all genes). For Stwl-bound active and silent genes, the background gene set was restricted to all active and silent genes, respectively (as determined by RPKM in S2 cells). Frequency ratio = gene ratio/background ratio. Count and frequency ratio are plotted for each set of GO-enriched genes. Only gene sets with FDR<.05 are plotted.

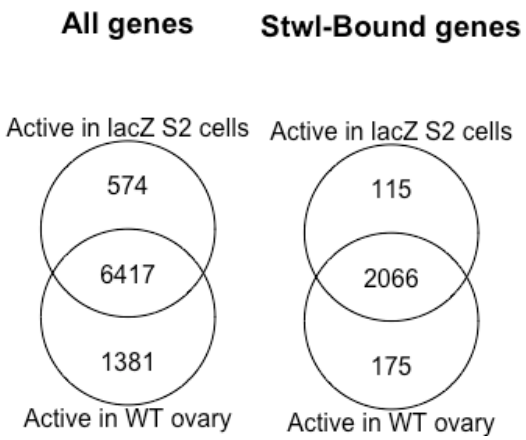


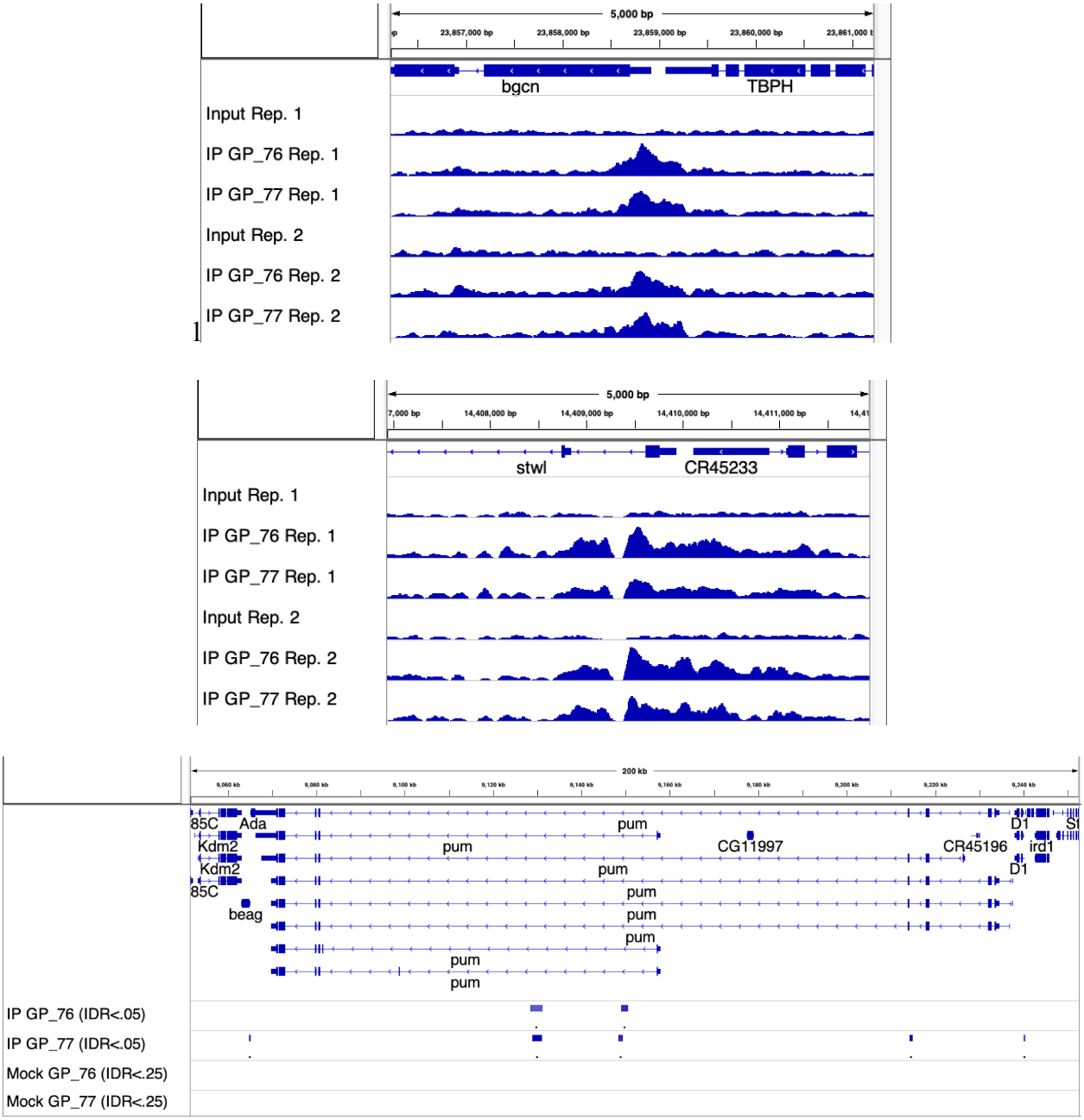
Figure 16 Stwl-bound, active genes are active in ovary and S2 cells
Overlap of genes identified as active (RPKM>=2.0) in WT ovary and *lacZ*-dsRNA-treated S2 cells. 82% of genes active in ovary are also active in S2 cells; 92% of active genes in S2 cells are also active in ovary. 88% of all Stwl-bound, active genes are active in both cell types.

Stwl binds to promoters of *bgcn*, *pum*, and *stwl*

We followed up on our overrepresentation findings by examining genes of interest that are Stwl-bound and upregulated in our RNA-seq datasets. *bgcn* is among the most upregulated genes in *stwl* null and *stwl*-dsRNA-treated S2 cells; it is ectopically expressed in both samples upon Stwl loss/reduction, and is also upregulated in *bam*, *Sxl*, *egg*, *wde*, and *hp1a* mutant ovaries (Chapter 2, Fig. 61, Table 6). We found very strong fold-enrichment of Stwl at the *bgcn* promoter (4.1-fold over input, IDR=1.3e^{-5}) (Fig. 16). We also identified a similar level of Stwl enrichment at the *stwl* promoter (4.0-fold over input, IDR=3.0e-5) (Fig. 16). The peaks at these two loci are among the top 1% of Stwl peaks when ranked according to fold-enrichment. Although not as striking, we also found that Stwl was enriched at the *pum* locus (Fig. 16). *pum* produces multiple transcripts, most of them very long; the gene body is almost 200 kb. We found multiple peaks of varying signal strength along the gene body, including 2 localized to introns,1 at the fourth coding exon, and 1 at the promoter. We also found that *pum* is significantly upregulated in both *stwl* heterozygous ovaries and *stwl*-dsRNA-treated S2 cells, although only to a moderate degree (1.6- and 1.3-fold increases, respectively).

bgcn transcript is expressed at very low levels in *drosophila* ovaries; its expression is limited largely to GSC, where it is critical for promoting asymmetrical division into cystoblast daughters (Ohlstein et al., 2000). 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 (McCarthy et al., 2018; Ohlstein et al., 2000). Despite this observation, we cannot rule that Stwl is required for suppression of *bgcn* outside of early germ cells. Experiments have not been conducted to identify the impact of *bgcn* overexpression in germline cysts, where an accumulation of Bgcn could adversely affect oocyte determination. Of note, we found that RNAi-mediated knockdown of *stwl* outside of GSCs significantly reduced female fertility (Chapter 2, Fig. 62).

The presence of a strong Stwl binding signal at the *stwl* locus suggests that Stwl is a negative self-regulator. Negative self-regulation has not previously been described as a property of GSC genes. More work needs to be done to confirm whether Stwl binding at its own locus results in repression of the *stwl* gene (see discussion).



<u>Figure 16</u> Stwl binds to key GSC maintenance and differentiation genes, including itself. Read coverage across *bgcn* (top) and *stwl* (middle) promoters, and peak calls at the *pum* (bottom) gene body. *bgcn* and *stwl* reads produced highly enriched, highly reproducible peaks spanning the promoter and 5' UTR. At the *pum* locus, we show peak calls after IDR filtering for both antibody IPs (GP 76 and GP 77) and their respective mocks. The final peak calls for all sites was the union of peaks in IP, subtracting any sites that intersect with enriched regions in the mock datasets. Tracks of read coverage and peak signal were printed from Integrative Genomics Viewer (IGV) (Robinson et al., 2011).

modENCODE Transcription Factor	Description	fold-enrichment	query peaks	overlap with Stwl peaks	% of Stwl peaks	% of Stwl coverage	% of query peaks	% of query coverage
Stwl	A protein of some interest	11.1	6869	1639	76.1	18.6	23.9	25.5
Pzg	Negative regulator of JAK-Stat signalling	10.9	5808	1577	73.2	19.7	27.2	25.0
Dp	Required for axis specification in oogenesis; represses genes during quiescence in G0	10.0	4908	1365	63.4	20.5	27.8	22.8
BEAF-32	Insulator binding protein; contains a BESS motif	9.5	5036	1345	62.5	23.6	26.7	21.8
Ets97D	Required in germline for female fertility	9.2	5202	1256	58.3	14.8	24.1	21.4
E(bx)	Negative regulator of JAK-Stat signalling; member of NURF remodeling complex	8.8	8110	1685	78.3	23.9	20.8	20.7
Lhr	Hybrid incompatibility gene; required for regulation of repetitive transcripts; Contains MADF, BESS domains; binds to HMR, which is a putative insulator binding protein	8.6	4905	1093	50.8	12.1	22.3	20.4
ZIPIC	Insulator binding protein, interacts with CP190	8.4	2815	587	27.3	6.4	20.9	19.3
Myb	Contains a SANT/Myb domain, which has similarity to MADF domain; transcriptional repressor	8.2	5533	1193	55.4	14.1	21.6	18.5
CTCF	Insulator binding protein	6.5	5098	860	39.9	11.1	16.9	15.0
CoRest	Contains a SANT/Myb domain, which has similarity to MADF domain; transcriptional repressor	6.1	8185	1214	56.4	14.2	14.8	13.8
su(Hw)	Insulator binding protein; binds the gypsy insulator; required for female fertility; mutant males exhibit a sperm exhaustion phenotype	5.9	5366	818	38.0	11.4	15.2	13.7
Hmr	Hybrid incompatibility gene; required for regulation of repetitive transcripts; Contains MADF, BESS domains; Insulator binding protein	5.2	7528	1016	47.2	11.4	13.5	11.8
tj	Required for germline stem cell formation	4.4	8370	1041	48.4	13.3	12.4	9.6

Table 1 Notable ModERN transcription factors with binding profiles similar to Stwl. We compared our Stwl ChIP-Seq peakset to all peaksets compiled by the ModERN group (including their independent analysis of Stwl). This table represents the 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. We also show the number of peaks in the listed peakset, the number of these peaks that overlap with the Stwl peakset, the % of Stwl peaks are in this overlap (from a total of 2153 peaks), the % of Stwl sites (bp) are covered in the listed peakset, the % of overlap peaks in the listed peakset, and the % sites in the listed peakset covered by Stwl.

Stwl binding peak profiles are similar to profiles of known insulator binding proteins

The ModERN consortium recently made available a database of ChIP-Seq experiments conducted in *D. melanogaster* (Kudron et al., 2018). Most of the IPs were performed against GFP on whole fly embryos or larvae expressing a GFP-fusion construct of a particular TF. We compared our Stwl ChIP-Seq profile to the peaks generated from all 475 ModERN experiments

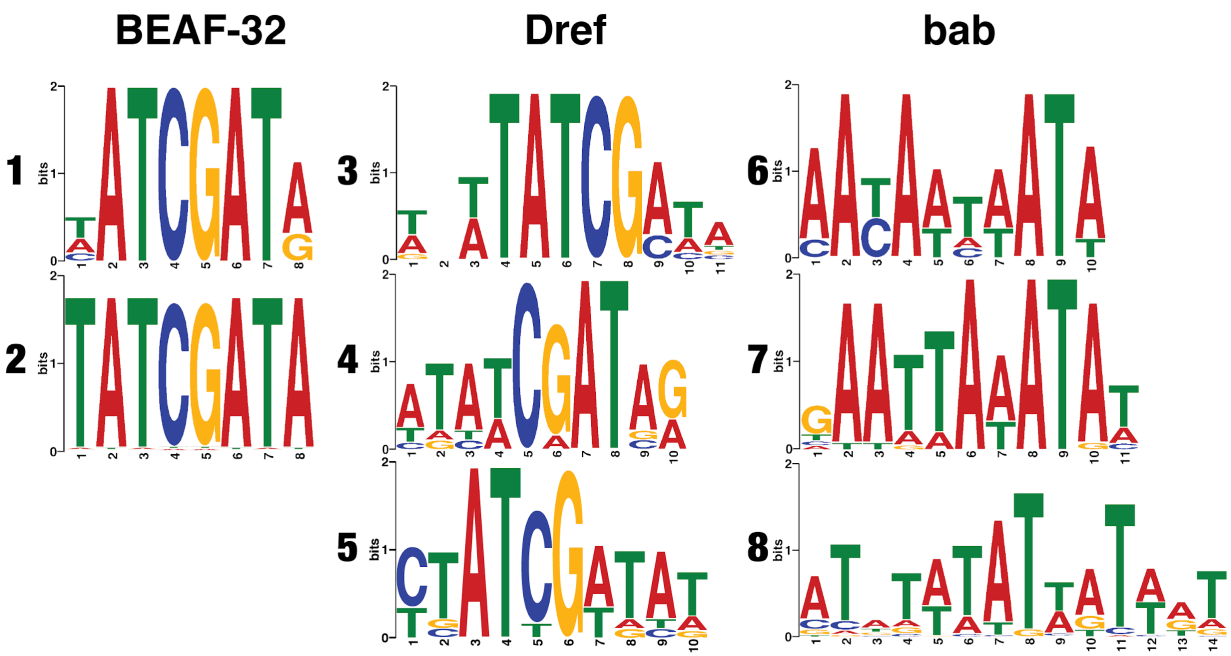
using the Genomic Association Tester (GAT) program (Heger et al., 2013). Briefly, GAT simulates a null distribution of peaks based on the size of each peakset, then estimates the number of overlaps expected by chance and compares this to the number of observed overlaps. GAT found that a majority (453/475) of ModERN peak profiles share similarity with stwl binding sites. This is to be expected, since most TFs are enriched at gene bodies. We therefore examined the most similar binding profiles according to fold-enrichment and % overlap.

According to fold-enrichment, ChIP-Seq against Stwl-GFP had the most similar binding profile to our Stwl peakset (Table 1). 76% of Stwl peaks were also identified in the Stwl-GFP ChIP. This experiment was performed in embryos 0-10 hour embryos with 2 biological replicates. This dataset can be used in future to validate certain results observed in S2 cell ChIP-Seq. 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), ZIPIC, Pita, Hmr, and Lhr.

Insulators are genomic regions that, when appropriately bound, prevent interaction between enhancers and their target promoters and prevent the spreading of chromatin modifications. BEAF-32 and CTCF, in addition to CP190, bind to Class I insulator elements (Nègre et al., 2010). These elements primarily serve to establish chromatin boundaries and regulate differentially expressed promoters. Differentially expressed promoters are promoter pairs where one promoter is active and the other is silent. Class I insulators are often found between these differentially expressed promoters, establishing boundaries to prevent the spreading of active/repressed chromatin marks to one or the other. Class 2 insulators are bound by Su(Hw) and are highly enriched for *gypsy* and *gypsy*-like elements. ZIPIC and Pita are recently identified insulator binding protein with confirmed enhancer blocking properties

(Maksimenko et al., 2015). Hmr and Lhr are physically interacting speciation factors that are necessary for TE silencing; Hmr was recently characterized as an insulator binding protein that binds to pericentromeric heterochromatin and gypsy insulators (Gerland et al., 2017; Maheshwari et al., 2008; Satyaki et al., 2014; Thomae et al., 2013). With few exceptions, these TFs typically bind ~ 100 bp upstream of transcription start sites, similar to Stwl's binding profile at promoters (Fig. 13).

We utilized the Meme Suite to identify enriched binding motifs in S2 cell Stwl ChIP-Seq (Bailey et al., 2009). Binding motifs are short stretches of sequence that are overrepresented and centrally enriched among ChIP-Seq peaks. Identification of conserved motifs can help to identify TF function, especially if those motifs are conserved with other known TF. We found that Stwl peaks are enriched for DNA sequence motifs that are common to BEAF-32, Dref and bab (Fig. 17). Dref is a known insulator binding protein that is additionally required for telomere maintenance (Tue et al., 2017). bab, which we identified as ectopically expressed in *stwl* null and *stwl*-dsRNA-treated cells, plays an important role in female sex differentiation (Williams et al., 2008). The occurrence of insulator motifs in Stwl ChIP-Seq provides strong evidence that Stwl binds to insulators. Previous work demonstrating that Stwl is necessary for limiting the expansion of heterochromatin in Su(var) experiments further support this finding (Maines et al., 2007; Yi et al., 2009).



Motif	TF Match	E-value	Matches	Union	Intersect	% Matches	% Union	% Intersect
1	BEAF-32	2.40E-10	277	538	260	20	39	19
2	BEAF-32	3.90E-08	521	538	260	38	39	19
3	Dref	1.10E-09	498	749	286	36	54	21
4	Dref	1.20E-08	513	749	286	37	54	21
5	Dref	1.60E-04	519	749	286	38	54	21
6	bab1	9.70E-06	301	1022	141	22	74	10
7	bab1	5.90E-04	854	1022	141	62	74	10
8	bab1	1.50E-03	473	1022	141	34	74	10

Figure 17 Stwl peaks are enriched for motifs common to the insulator binding proteins BEAF-32 and Dref, and the sex differentiation factor bab.

Enriched motifs identified in Stwl peaks using Meme Suite (Bailey et al., 2009). For each motif, we include the transcription factor that motif matches with, the confidence that the given motif is centrally enriched in Stwl peaks (e-value), and the number of Stwl peaks that contain the given motif. Union and Intersect columns indicate the number of Stwl peaks that contain the union or intersect of all motifs matching one of the three identified transcription factors. The % columns identify the percent of the given motifs found in Stwl peaks.

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 co-localizes with HP1 at heterochromatin-like structures nuclear periphery (Maines et al., 2007; Yi et al., 2009). We performed additional analyses to further establish that Stwl localizes to heterochromatic regions. First, we estimated Stwl coverage across 100-kb genomic bins using our ChIP-Seq peakset (Fig. 18). We found that Stwl is highly enriched across the dot chromosome (chromosome 4), which is a highly repetitive and mostly heterochromatic chromosome (Riddle and Elgin, 2018). We also saw enrichment 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 (Gerland et al., 2017). Hmr localization at chromosome 2 has also been verified via immunofluorescence (Thomae et al., 2013).

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 stretches of DNA, we re-analyzed our ChIP-Seq data and instead calculated differential enrichment of reads in IP samples relative to mock. We counted reads aligning to repetitive DNA, similar to our approach for RNA-Seq data. In order to normalize our ChIP-Seq libraries for differences in sequencing depth, we also counted the number of reads aligning to 1-kb bins in the genome. We performed differential enrichment analysis using DESeq, identifying genomic bins and repetitive DNA that are significantly enriched in IP samples, but not variable between source animal (IP 76 vs IP 77) or biological replicate. Principal Components Analysis

(PCA) of the count data showed that our method mostly produces variance between mock and IP samples, not within mock and IP groups (Fig. 19).

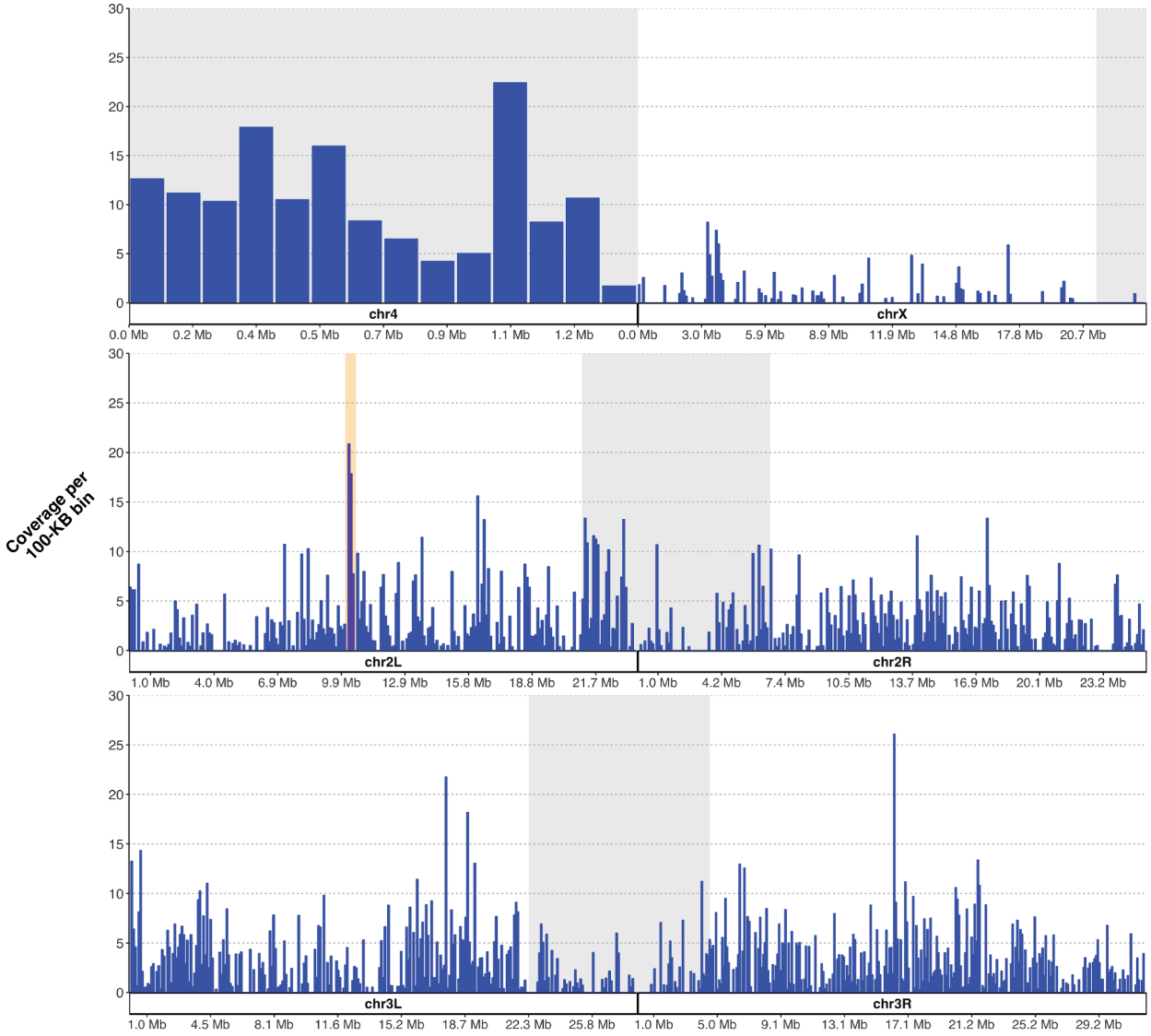


Figure 18 Stwl is enriched at chromosome 4, pericentromeric heterochromatin of chromosome 2, and cytological region 31.
Plot of % coverage of Stwl peaks per 100-kb bin of the genome. % coverage is calculated as the sum of the length of all Stwl-bound regions within a 100-kb bin, divided by 100 kb. Shaded grey areas represent constitutive heterochromatin (pericentromeric regions and chromosome 4). Cytological region 31 on chromosome 2L is highlighted.

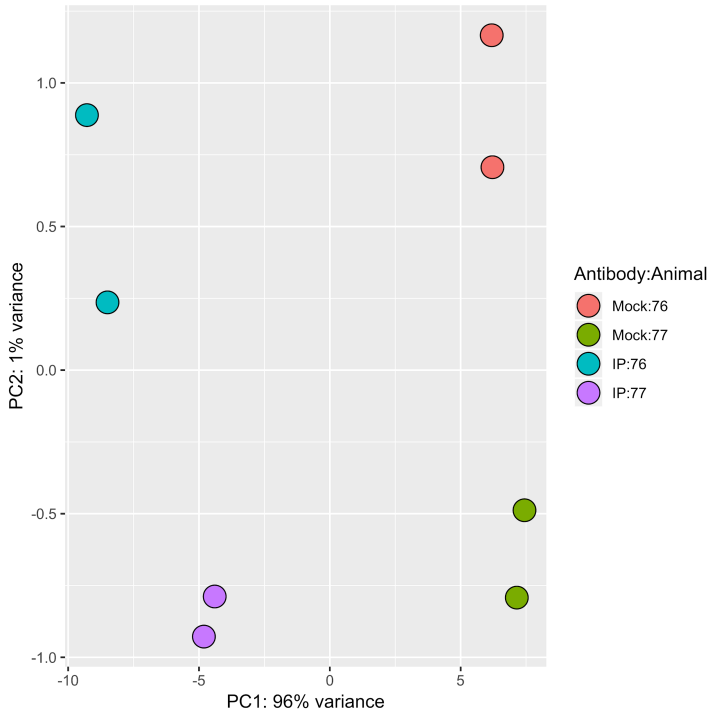
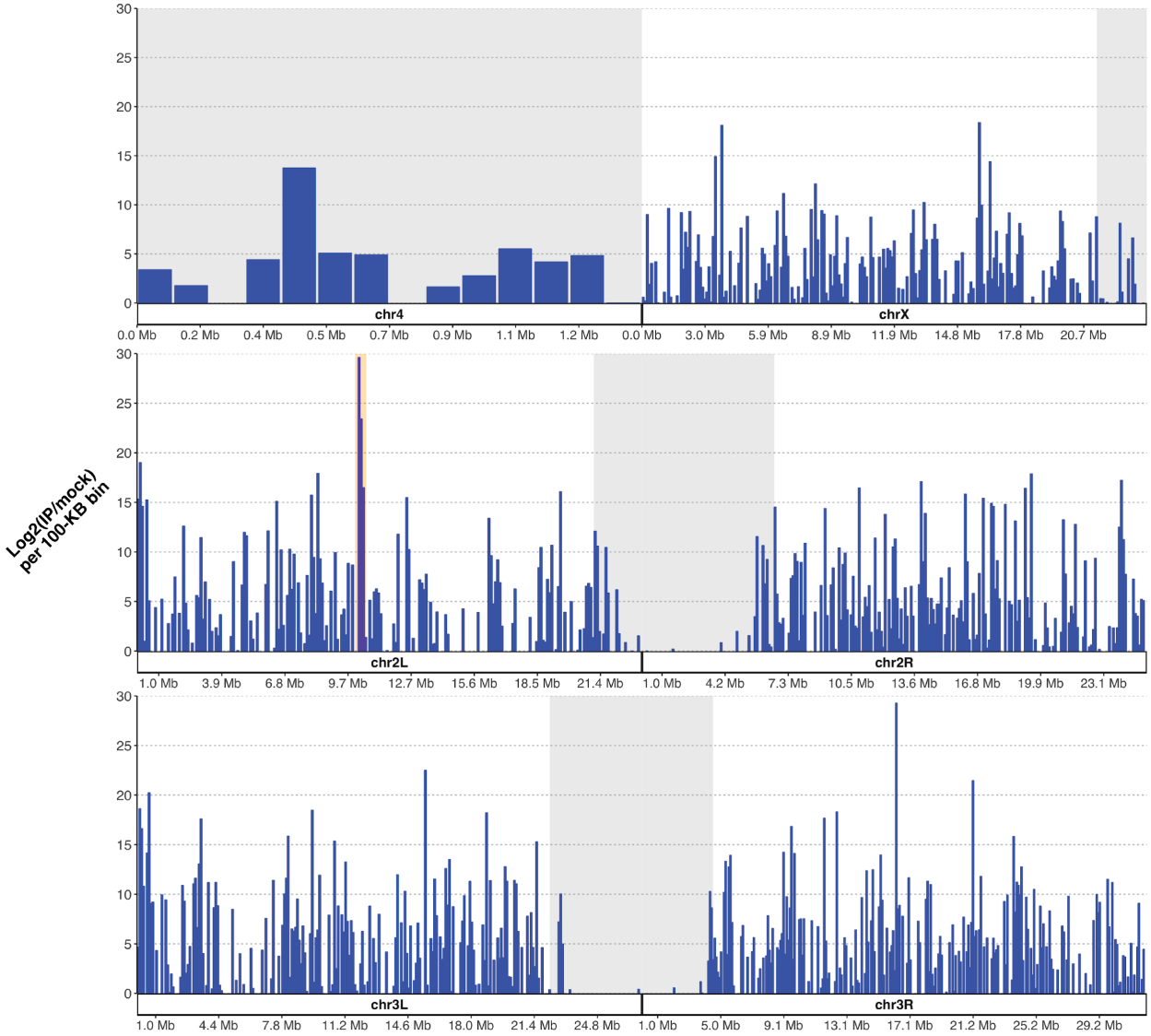
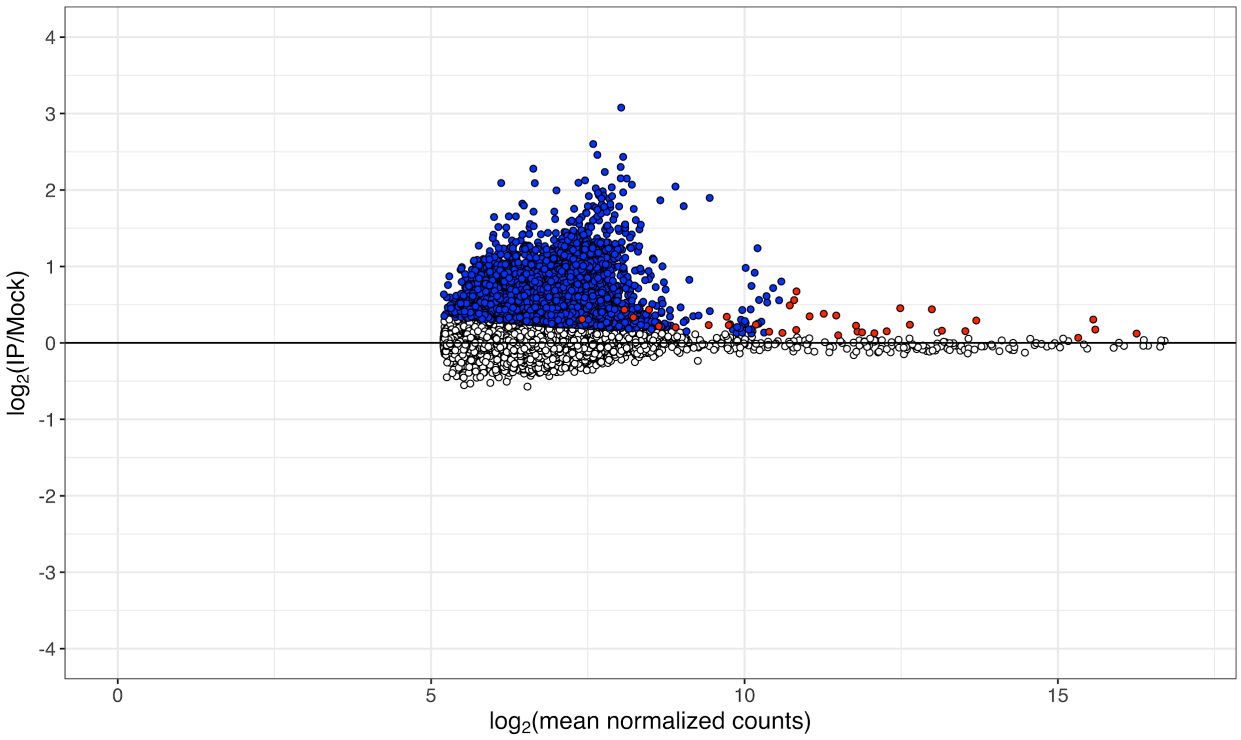


Figure 19 PCA of α-Stwl ChIP-Seq read counts separate mock from IP
Principal components analysis for read counts generated from alignment to genomic bins and repeat index. Experiments labeled as mock were performed with pre-bleed antibodies, IP with final-bleed antibodies. Antibodies were generated from two different animals (referred to as 76 and 77) using the same epitope. DNA for all experiments, including input (not shown) was immunoprecipitated and isolated from two pools of S2 cells. The majority of the variance in the data is contained in PC1 and is explained by differences between mock and IP condition, not by differences in the source animal or replicate pools.

As a further validation of our approach, we plotted $\log_2$(fold-change) of IP/mock in 100-kb bins against chromosomal location, similar to our plotting of peak coverage in Figure 18. While there are notable differences, including a reduction in signal at chromosome 4 and an increase at the X, we still find strong enrichment of Stwl-bound reads at cytological region 31 on 2L, and at pericentromeric heterochromatin boundaries of chromosome 2. We suspect that signal at 4 was lost due to the removal of repetitive DNA reads for alignment to our repeat index.

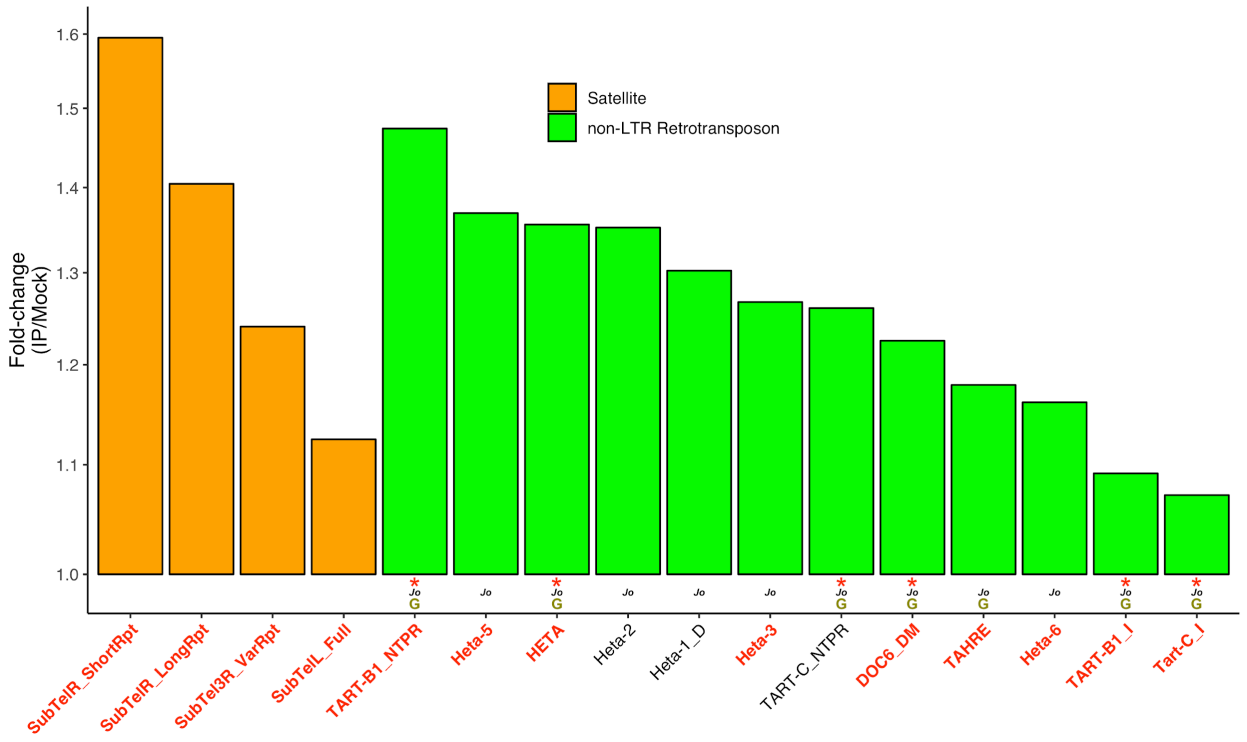


<u>Figure 20</u> LFC of read counts across chromosome arms is similar to peak coverage signal
Average $\log_2$(Fold-change) of IP/mock is plotted for each 100-kb genetic bin. The $\log_2$(Fold-change) values (LFC) were "shrunk" to minimize the variance associated with low-count genes. Cytological region 31 on chromosome 2L is highlighted.



<u>Figure 21</u> α-Stwl ChIP is enriched for TE reads relative to mock ChIP
log_2(fold-change) for each gene is plotted against its average read count across all samples (mock and IP). Read count is represented by counts normalized according to library size. The log_2(fold-change) values (LFC) were "shrunk" to minimize the variance associated with low-count genes. Filled points (blue and red) identify genetic features which are differentially enriched (adjusted p-value <.01) in this comparison. Red points represent entries from Repbase, blue points are from 1-kb genomic bins.

Differential enrichment analysis identified repetitive DNAs enriched in Stwl IP samples relative to mock (Fig. 21, Fig. 22). 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 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 Stwl-bound peaks was 2.0. We are therefore cautiously optimistic that our Stwl ChIP-Seq has identified binding sites at repetitive DNA.

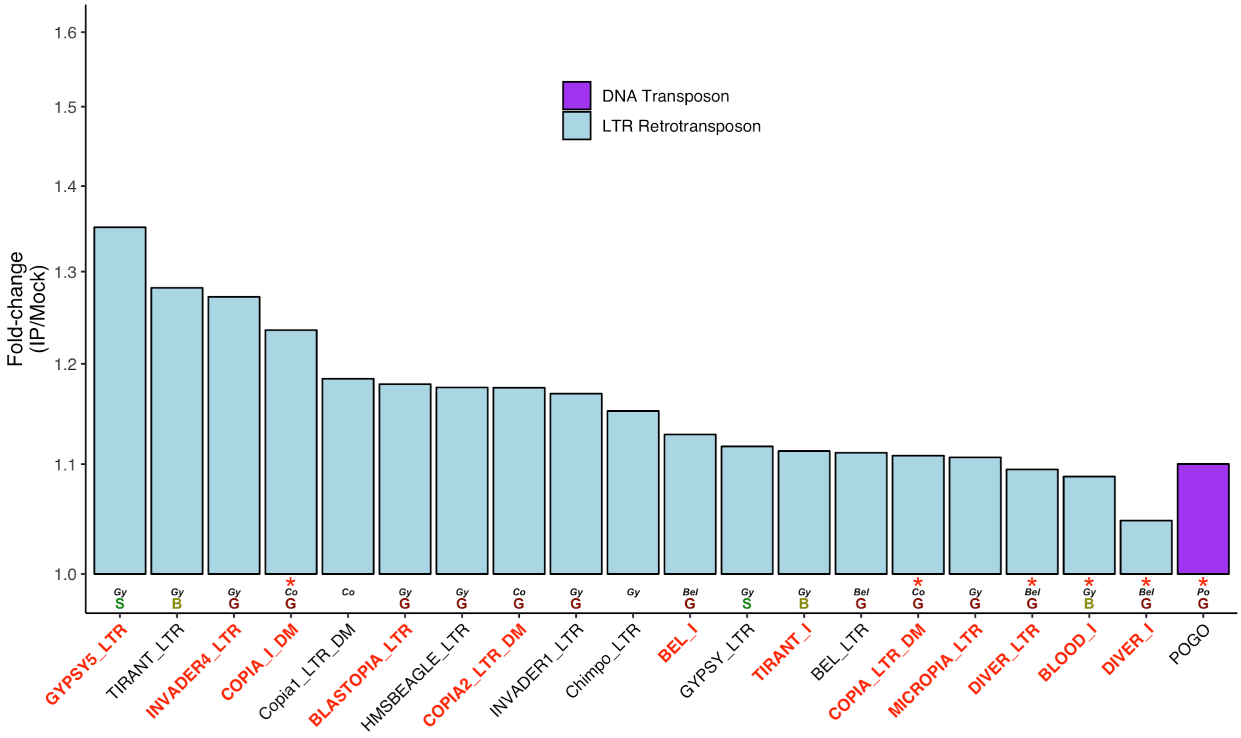


<u>Figure 22</u> Telomere-associated sequences are enriched (at low fold-change) in Stwl IP
Fold-change of read count abundance of satellite and non-LTR retrotransposons for IP/mock comparison. Y-axis is in $\log_2$-scale. All significantly enriched satellites and non-LTR retrotransposons (adjusted p <.01) are plotted. Red asterisks indicate elements that are actively transposing in *D. melanogaster* (Kelleher and Barbash, 2013). "G","B","S" indicates whether transposons are typically expressed in germline cells, somatic cells, or both. All of the bound non-LTR retrotransposons are *Jockey* elements (indicated by *Jo*). Repetitive elements whose names are in red are upregulated in *stwl* null or *stwl* heterozygous ovaries.

We were surprised to find that telomere-associated sequences are consistently enriched in Stwl IP (Fig. 22). 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, *HetA*, *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 motifs with Dref (Table 1, Fig. 17). Each of these factors localizes to and is necessary for telomere maintenance (Andreyeva et al., 2005; Silva-Sousa et al., 2013). Lastly, we find that a majority of Stwl-bound telomeric sequences are

also upregulated in either *stwl* null or heterozygous ovaries (gene names in red in Fig. 22). These findings suggest that Stwl localizes to telomeres, although more data is required to validate this hypothesis.

In addition to telomeric sequences, we also found that Stwl ChIP-Seq is enriched for LTR retrotransposons (Fig. 23). We find that these enriched TEs are members of the *copia*, *gypsy*, and *bel* superfamilies. *Copia* elements are among the most highly upregulated transcripts in *stwl* null ovaries, and we find that they are also upregulated in *stwl* heterozygous ovaries and *stwl* null testes (Chapter 2, Fig. 22). Despite the apparent enrichment of *gypsy* LTR sequences in our Stwl ChIP, we did not detect enrichment of internal *gypsy* sequences. The *gypsy* insulator is located at the 5' end of the internal *gypsy* sequence; we also did not detect motifs indicative of *gypsy* insulator binding in our motif enrichment analysis. It is possible that these LTR retrotransposons are enriched in Stwl IP data by virtue of being commonly located at pericentromeric heterochromatin.



<u>Figure 23</u> Stwl binds to *Gypsy*, *Bel*, and *Copia* LTR-retrotransposons.
Fold-change of read count abundance of DNA transposons (purple) and LTR retrotransposons (light blue) for IP/Mock comparison. Y-axis is in $\log_2$-scale. All significantly enriched DNA transposons and LTR retrotransposons (adjusted p <.01) are plotted. Red asterisks indicate elements that are actively transposing in *D. melanogaster* (Kelleher and Barbash, 2013). "G", "B", "S" indicates whether transposons are typically expressed in germline cells, somatic cells, or both. TE super-families are reported as follows: *Bel=Bel-Pao*, *Co=Ty1-Copia*, *Gy=Gypsy*, *Po=Pogo*. Repetitive elements whose names are in red are upregulated in *stwl* null or *stwl* heterozygous ovaries.

Discussion

In chapter 2, we established that *stwl* mutant tissues exhibit a wide array of phenotypes. In addition to GSC loss and complete sterility due to defects in oocyte determination, we found that *stwl* null ovaries overexpress transposable elements and *stwl* heterozygous ovaries overexpress satellite transcripts, that a number of somatic-enriched genes are ectopically expressed in *stwl* mutant tissues, and that testis-specific genes are upregulated in S2 cells upon *stwl* KD. We also showed a link between *stwl* and sex determination, by demonstrating that the male-specific variant of the sex determination factor *Phf7* is transcribed in *stwl* null ovaries, and that many ectopically expressed genes in *stwl* null ovaries and *stwl*-dsRNA-treated S2 cells are members of the Jak/STAT signaling pathway, which is required for male germline sex determination. Finally, we found that Stwl is required outside of germline stem cells in females to maintain fertility. With so many distinct phenotypes associated with loss of a single gene product, it is difficult to disentangle direct effects of *stwl* mutation from indirect effects. For example, does an overexpression of TEs in *stwl* mutants mean that Stwl is involved in TE defense or is this phenotype a result of genomic instability in apoptotic cells? Is upregulation at *Phf7* and *bgcn* a result of Stwl binding to these loci in WT tissue or the terminal result of a cascade of misregulated genes in *stwl* mutants? Lastly, what is the molecular basis of these distinct phenotypes?

In order to address these questions and others, we performed a ChIP-Seq experiment on S2 cells against Stwl, using a pair of antibodies we developed against a Stwl epitope. Our analyses indicate that Stwl is likely binding to insulator elements. Most Stwl peaks are located just upstream of promoters, with a preference towards active 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). Lastly, we found that *stwl* peaks accumulate at pericentromeric heterochromatin boundaries.

These data, in conjunction with published literature regarding Stwl's role in heterochromatin maintenance, strongly suggest that Stwl is an insulator binding protein. However, more work needs to be done to determine how Stwl specifically functions in insulator binding. It is unclear whether Stwl functions by blocking enhancer-promoter interactions, or by establishing boundaries to prevent the spread of chromatin modifications, or both. For starters, we need to examine whether Stwl is preferentially binding at differentially expressed promoters; we know that 25% of Stwl peaks are within 1 kb of 2 or more promoters that are a mixture of active or inactive, but have not determined how many of these peaks specifically sit between differentially expressed promoter pairs. Insulators are critical for establishing boundaries between differentially expressed promoter pairs.

In order to thoroughly characterize Stwl's insulator functions, we will need to determine where Stwl peaks sit relative to other insulator sites, and in what combination. While we know that Stwl peaks overlap significantly with BEAF-32, what subset of Stwl peaks are bound by a combination of BEAF-32 and other known insulators, such as CTCF? Perhaps most critically, we hope to examine the chromatin state of Stwl-bound peaks and examine how they change upon reduction or loss of Stwl. We speculate that Stwl-bound sites are located adjacent to regions of repressed chromatin, and that loss of Stwl would result in spreading of these repressed chromatin marks to neighboring loci.

We were surprised to find that Stwl binds prominently to its own promoter. It is tempting to suggest that Stwl negatively self-regulates; that is, as Stwl becomes more abundant in a given

cell type, that cell experiences a reduction in the production of *stwl* transcript. However, it is unclear whether Stwl binding alone is sufficient for establishment of insulating boundaries and/or recruitment of transcriptional silencing machinery. According to IF and EM assays, Stwl co-localizes with Hp1a at nucleolar like structures, but we did not identify significant enrichment between Stwl and Hp1a ChIP-Seq profiles (Yi et al., 2009). Therefore, it is as yet unclear what Stwl is doing at the *stwl* promoter. Of note, we have been told through personal communications that Bam may also self-regulate; if true, this suggests that self-regulation is a shared feature of GSC regulatory genes.

We are intrigued by the correlation of Stwl ChIP-Seq with RNA-Seq of *stwl* heterozygous ovaries. We were hesitant to draw conclusions from *stwl* heterozygous ovaries because of the lack of any obvious phenotype in this tissue. However, the observation that Stwl-bound peaks tend to bind near promoters of genes upregulated in *stwl* heterozygous ovaries suggests that these genes may be directly regulated by Stwl. Again, we wish to establish a more direct link between Stwl binding and activation/silencing of the bound region. It is important to determine whether this experiment should be done in embryos, ovaries, or S2 cells; fortunately, the availability of two ChIP grade antibodies gives us the option to perform experiments in any of these cell types.

How does the ChIP-Seq data inform our view on Stwl's broader role in GSC maintenance and oocyte determination? First and foremost, we found that Stwl binds strongly at *bgcn* and *pum*, both of which are upregulated in S2 cells after *stwl* KD and in *stwl* mutant ovaries, as discussed in chapter 2. Furthermore, we find that Stwl-bound genes, particularly those that are normally silent in S2 cells, are enriched for GO terms having to do with gametogenesis, GSC population maintenance and oocyte axis specification, all of which are known broad biological

functions of Stwl. We did not find Stwl peaks at the *Phf7* locus, or at other genes we identified in our RNA-Seq dataset as being linked to masculinization of ovaries and S2 cells. We did find that Stwl peaksets share similarities with a number of known regulators of the Jak/STAT signaling pathway, which is linked to male sex determination in testis. It is possible that Stwl is regulating an as yet unidentified intermediary gene, which in turn regulates transcription of the Jak/STAT pathway in ovaries.

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 these genes in a given cell type. We are especially interested in finding whether Stwl-binding sites are consistent between S2 cells and ovary, or whether these sites are altered depending on the accessibility of chromatin in a given cell type.

References

1. Andreyeva, E.N., Belyaeva, E.S., Semeshin, V.F., Pokholkova, G.V., and Zhimulev, I.F. (2005). Three distinct chromatin domains in telomere ends of polytene chromosomes in Drosophila melanogaster Tel mutants. J. Cell Sci. *118*, 5465–5477.

2. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202-208.

3. Clark, K.A., and McKearin, D.M. (1996). The Drosophila stonewall gene encodes a putative transcription factor essential for germ cell development. Development *122*, 937–950.

4. Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat. Protoc. *7*.

5. Gerland, T.A., Sun, B., Smialowski, P., Lukacs, A., Thomae, A.W., and Imhof, A. (2017). The Drosophila speciation factor HMR localizes to genomic insulator sites. PLOS ONE *12*, e0171798.

6. Heger, A., Webber, C., Goodson, M., Ponting, C.P., and Lunter, G. (2013). GAT: a simulation framework for testing the association of genomic intervals. Bioinformatics *29*, 2046–2048.

7. Huang, X.A., Yin, H., Sweeney, S., Raha, D., Snyder, M., and Lin, H. (2013). A Major Epigenetic Programming Mechanism Guided by piRNAs. Dev. Cell *24*, 502–516.

8. Kelleher, E.S., and Barbash, D.A. (2013). Analysis of piRNA-Mediated Silencing of Active TEs in Drosophila melanogaster Suggests Limits on the Evolution of Host Genome Defense. Mol. Biol. Evol. *30*, 1816–1829.

9. Kudron, M.M., Victorsen, A., Gevirtzman, L., Hillier, L.W., Fisher, W.W., Vafeados, D., Kirkey, M., Hammonds, A.S., Gersch, J., Ammouri, H., et al. (2018). The ModERN Resource: Genome-Wide Binding Profiles for Hundreds of Drosophila and Caenorhabditis elegans Transcription Factors. Genetics *208*, 937–949.

10. Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., et al. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res *22*, 1813–1831.

11. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. *25*, 2078–2079.

12. Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. *5*, 1752–1779.

13. Lin, H., Chen, M., Kundaje, A., Valouev, A., Yin, H., Liu, N., Neuenkirchen, N., Zhong, M., and Snyder, M. (2015). Reassessment of Piwi Binding to the Genome and Piwi Impact on RNA Polymerase II Distribution. Dev. Cell *32*, 772–774.

14. Maheshwari, S., Wang, J., and Barbash, D.A. (2008). Recurrent Positive Selection of the Drosophila Hybrid Incompatibility Gene Hmr. Mol. Biol. Evol. *25*, 2421–2430.

15. Maines, J.Z., Park, J.K., Williams, M., and McKearin, D.M. (2007). Stonewalling Drosophila stem cell differentiation by epigenetic controls. Development *134*, 1471–1479.

16. Maksimenko, O., Bartkuhn, M., Stakhov, V., Herold, M., Zolotarev, N., Jox, T., Buxa, M.K., Kirsch, R., Bonchuk, A., Fedotova, A., et al. (2015). Two new insulator proteins, Pita and ZIPIC, target CP190 to chromatin. Genome Res. *25*, 89–99.

17. Marinov, G.K., Wang, J., Handler, D., Wold, B.J., Weng, Z., Hannon, G.J., Aravin, A.A., Zamore, P.D., Brennecke, J., and Toth, K.F. (2015). Pitfalls of mapping high-throughput sequencing data to repetitive sequences: Piwi's genomic targets still not identified. Dev. Cell *32*, 765–771.

18. McCarthy, A., Deiulio, A., Martin, E.T., Upadhyay, M., and Rangan, P. (2018). Tip60 complex promotes expression of a differentiation factor to regulate germline differentiation in female *Drosophila*. Mol. Biol. Cell *29*, 2933–2945.

19. Nègre, N., Brown, C.D., Shah, P.K., Kheradpour, P., Morrison, C.A., Henikoff, J.G., Feng, X., Ahmad, K., Russell, S., White, R.A.H., et al. (2010). A Comprehensive Map of Insulator Elements for the Drosophila Genome. PLOS Genet. *6*, e1000814.

20. Ohlstein, B., Lavoie, C.A., Vef, O., Gateff, E., and McKearin, D.M. (2000). The Drosophila cystoblast differentiation factor, benign gonial cell neoplasm, is related to DExH-box proteins and interacts genetically with bag-of-marbles. Genetics *155*, 1809–1819.

21. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing

genomic features. Bioinformatics *26*, 841–842.

22. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

23. Riddle, N.C., and Elgin, S.C.R. (2018). The Drosophila Dot Chromosome: Where Genes Flourish Amidst Repeats. Genetics *210*, 757–772.

24. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative Genomics Viewer. Nat. Biotechnol. *29*, 24–26.

25. Satyaki, P.R.V., Cuykendall, T.N., Wei, K.H.-C., Brideau, N.J., Kwak, H., Aruna, S., Ferree, P.M., Ji, S., and Barbash, D.A. (2014). The Hmr and Lhr Hybrid Incompatibility Genes Suppress a Broad Range of Heterochromatic Repeats. PLoS Genet. *10*.

26. Silva-Sousa, R., Varela, M.D., and Casacuberta, E. (2013). The Putzig partners DREF, TRF2 and KEN are involved in the regulation of the Drosophila telomere retrotransposons, HeT-A and TART. Mob. DNA *4*, 18.

27. Thomae, A.W., Schade, G.O.M., Padeken, J., Borath, M., Vetter, I., Kremmer, E., Heun, P., and Imhof, A. (2013). A Pair of Centromeric Proteins Mediates Reproductive Isolation in Drosophila Species. Dev. Cell *27*, 412–424.

28. Tue, N.T., Yoshioka, Y., Mizoguchi, M., Yoshida, H., Zurita, M., and Yamaguchi, M. (2017). DREF plays multiple roles during Drosophila development. Biochim. Biophys. Acta Gene Regul. Mech. *1860*, 705–712.

29. Williams, T.M., Selegue, J.E., Werner, T., Gompel, N., Kopp, A., and Carroll, S.B. (2008). The Regulation and Evolution of a Genetic Switch Controlling Sexually Dimorphic Traits in Drosophila. Cell *134*, 610–623.

30. Yi, X., de Vries, H.I., Siudeja, K., Rana, A., Lemstra, W., Brunsting, J.F., Kok, R.M., Smulders, Y.M., Schaefer, M., Dijk, F., et al. (2009). Stwl Modifies Chromatin Compaction and Is Required to Maintain DNA Integrity in the Presence of Perturbed DNA Replication. Mol. Biol. Cell *20*, 983–994.

31. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS J. Integr. Biol. *16*, 284–287.

32. Zhu, A., Ibrahim, J.G., and Love, M.I. (2018). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinforma. Oxf. Engl.

CHAPTER 4

SPERM EXHAUSTION ASSAYS DEMONSTRATE THAT STWL IS REQUIRED IN TESTIS FOR SUSTAINED FERTILITY

Introduction

Previous studies failed to identify a role for *stwl* in male fertility(Clark, 1996; Clark and McKearin, 1996; Maines et al., 2007). While a Stwl antibody shows Stwl localization in germ cells of male third-instar larvae and adult testis, *stwl* mutant males are fertile and have no obvious defects in testis morphology. However, male fertility defects may be subtle, such as declining fertility over time. To test whether *stwl* is required for male fertility, we performed a series of sperm exhaustion assays on *stwl*--deficient males. A sperm exhaustion assay differs from a conventional fertility assay in that it determines a male fly's ability to consistently generate progeny after successive matings. Since *stwl* transcript is abundant in embryos and nurse cells, it is likely maternally deposited; this means that *stwl* null males could still have some baseline level of Stwl protein during development that is initially sufficient for production of motile sperm. If Stwl is required for continued sperm production, then depletion of sperm by repeated matings would render *stwl* null males sterile.

Methods

We performed a series of sperm exhaustion assays to identify subtle defects in fertility of *stwl* mutant males. All crosses were performed at 25° C. For all assays, males were collected as virgins and aged for 2,3, or 10 days prior to the start of the assay. *y w* F10 females were also collected as virgins and aged for 2 days. Subsequent to initial collection for virgining, all flies in the mating assay were handled without any other anesthesia that could impact courtship or egg-laying behavior; flies were either transferred via flips or mouth aspiration. Each male was paired

with 2 2-day old *y w* F10 virgin females for approximately 24 hours and subsequently transferred to a new vial with another 2 2-day old *y w* F10 virgin females, for 5 to 12 days. Following the single day of mating, females were allowed to egg-lay for 15 days (flipped to new vials every 3-5 days).

Total progeny were counted and recorded from each mating; progeny from each vial were recorded on the 3rd and 6th days post-eclosion. In rare instances, 1 or both females escaped or died during the 15-day egg-laying period; the date of escape/death was recorded in each case. To account for this, all progeny counts were normalized to a full 30-day egg-laying period (e.g. if one female died after 10 days of egg-laying and the other survived all 15 days, normalization would be 30*(total progeny count)/25). Each experiment was replicated 10-20 times (i.e. 10-20 males were assayed for each genotype). Males which did not survive the full length of the mating assay were not included in these analyses.

In the first assay, we compared *stwl* heterozygotes to *stwl* trans-heterozygotes. $y^1\ w^*$; $P^{\{w[+mC]=lacW\}}stwl^{j6C3}$, $l(3)j6C3^{j6C3}/TM3,\ Sb^1$ and y^1; $P^{y[+mDint2]}w^{[BR.E.BR]=SUPor-P}\ stwl^{KG07971}\ ry^{506}/TM3,\ Sb^1\ Ser^1$ males were used as heterozygous controls; $stwl^{j6c3}$ females were crossed to $stwl^{KG07971}$ males to produce $stwl^{j6c3}/stwl^{KG07971}$ trans-heterozygous males. These males were assayed after collected as virgins and aged for 10 days. In follow-up experiments, $stwl^{j6C3}$ homozygous males were collected from an outbred *y w;* $stwl^{j6C3}/TM6b$ stock; *y w* F10 males were used as a wild-type control. All males were collected as virgins and aged for 2 days, then mated to 2-day old *y w* F10 virgin females for 5-10 days. For the recovery assay, we performed the same experiment as described with 2-day old *y w;* $stwl^{j6C3}$ and *y w* F10 males, except that males were mated for 6 days, isolated for a 3-day "recovery period", and then mated for another 6 days. Testes from these males were then dissected and immunolabeled for Vasa and Hts-1B1, as

described below. For the Gal4/UAS sperm exhaustion assay, we crossed Gal4 females (*Actin5c-Gal4: y^1 w^*; P{Act5C-GAL4}17bFO1/TM6B, Tb^1* and *nos-Gal4: w^{1118}; P{GAL4::VP16-nos.UTR}CG6325^{MVD1}*) to *UAS-stwl-RNAi* (*$y^{[1]}$ $sc^{[*]}$ $v^{[1]}$; P{y[+t7.7] v[+t1.8]=TRiP.GL00337}attP2*) and *y w* F10 males [Bloomington Stock #s 3954, 4937 and 35415, respectively]. We collected males as virgins from each of these crosses (*Actin5c-Gal4*/*UAS-stwl-RNAi*, *Actin5c-Gal4*/+, *nos-Gal4*/*UAS-stwl-RNAi, nos-Gal4*/+) and aged them for 3 days prior to assaying them for sperm exhaustion defects, as described above.

Results

We first assayed 10-day old *stwl* trans-heterozygous mutant males (*stwl^{j6c3}*/*stwl*^{KG0797}) against *stwl* heterozygous controls (*stwl^{j6c3}*/*TM3, Sb^1* and *stwl*^{KG0797}/*TM3, Sb^1, Ser^1*) for their ability to produce offspring after multiple successive matings. We found that *stwl* heterozygotes maintain fertility (100% of males produced at least some progeny at conclusion of assay) over 7 days of mating , although there is an obvious decline in fecundity of both genotypes over the course of the experiment (fig. 1). In contrast, the trans-heterozygotes are significantly less fertile than the heterozygote controls, even during the first day of mating when the total progeny number is highest. The sterility phenotype is exacerbated over time; on day 2 the mutant males are 2% as fertile as the controls, and 80% of tested males are completely sterile. By the 7th day of mating, 90% of males are completely sterile, and total fertility is reduced to less than 1% of control.

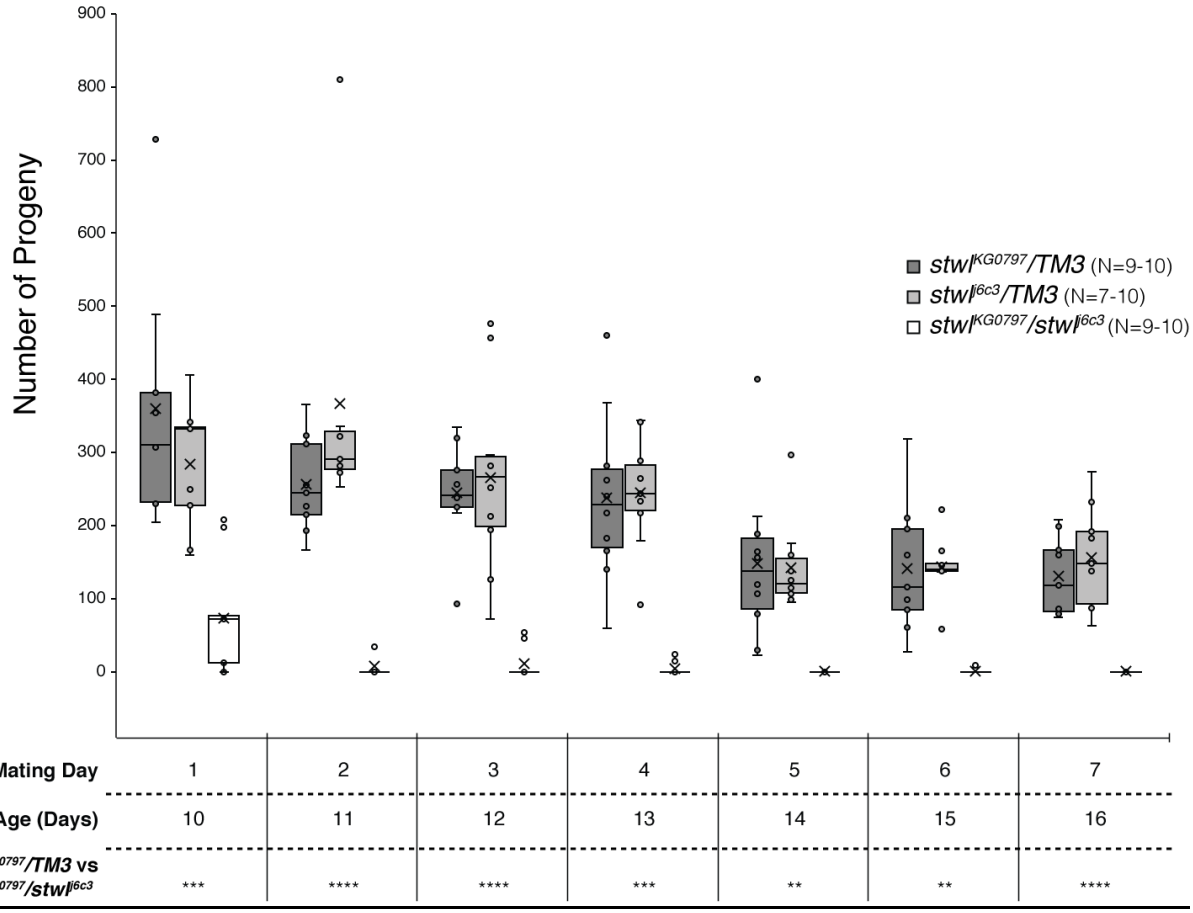


FIGURE 1. **10-day old *stwl* trans-heterozygotes exhibit sperm exhaustion defects**
Fecundity of aged males having either one functional *stwl*$^{+}$ allele (from the *TM3* balancer) or none. The Y-axis shows the total number of progeny that each male produced from a single day of mating with two virgin *y w F10* females. The X-axis portrays the mating day and the age of the flies on that day (males were aged for 10 days prior to start of assay, hence age on mating day 1 is 10 days). Total number of males for each genotype is displayed next to the legend (at a few time points, egg-laying females were lost, which is why N is variable across days). Statistical analysis compares *stwl*^{j6c3}/*stwl*^{KG0797} fertility for each mating day against the indicated control genotype (*stwl*^{KG0797}/*TM3, Sb*^{1}*, Ser*^{1}). Two-tailed p-values from a Student's t-test (two-sample assuming unequal variances) are presented as: p<.05 =*, p<.01=**, p<.001=***, p<.0001=****, NS=not significant (p>.05).

We conducted a follow-up experiment utilizing outbred *stwl*^{j6c3} to test whether genetic background is a variable in male fertility. The *stwl*^{j6c3} stock was outcrossed to an inbred lab strain (*y w* F10) for 8 generations to remove recessive lethal mutations, which made it possible to then assay *stwl*^{j6c3} homozygotes. We compared fertility of these homozygous males to *stwl*$^{+}$ *y w* F10

males. This time we began the assay with 2-day old males to ascertain whether the sterility phenotype is age-dependent. The first trial for this experiment was cut short after 5 days of mating due to mold contamination in the fly food (Figure 2). This trial revealed to us that *stwl*^{j6c3} homozygous flies have reduced viability; this may be due to recessive deleterious mutations linked to the *stwl*^{j6c3} allele that were not removed in the 8 generations of outcrossing. We therefore expanded the number of *stwl*^{j6c3} individuals for our trial from 10 to 30, and only analyzed data from males who survived the entirety of the mating assay.

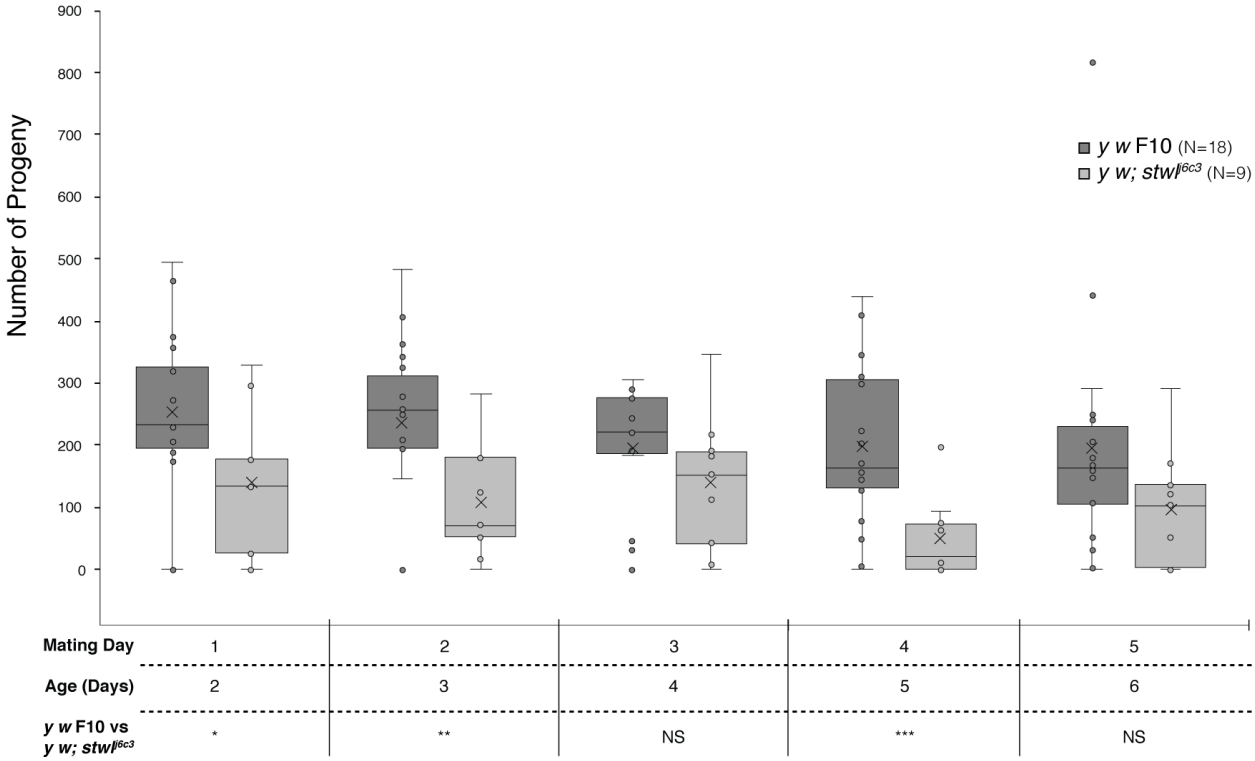


FIGURE 2. **Aborted trial for 2-day old males**

Fecundity of young, background-matched males having either two functional *stwl* alleles (*y w* F10) or none (*y w; stwl^{j6c3}*). The Y-axis shows the total number of progeny that each male produced from a single day of mating with two virgin *y w F10* females. The X-axis portrays the mating day and the age of the flies on that day (males were aged for 2 days prior to start of assay, hence age on mating day 1 is 2 days). Total number of males for each genotype is displayed next to the legend. Statistical analysis compares *y w; stwl^{j6c3}* fertility for each mating day against *y w* F10. Two-tailed p-values from a Student's t-test (two-sample assuming unequal variances) are presented as: p<.05 =*, p<.01=**, p<.001=***, p<.0001=****, NS=not significant (p>.05).

We found that young, background-matched *y w; stwl*^{j6c3} males have a sperm exhaustion phenotype similar to the one observed in older, trans-heterozygous males. By the 4th day of mating, *y w; stwl*^{j6c3} progeny number is 8% of the progeny number of the control group; at the conclusion of the assay, it is reduced to 3% of the control group (Figure 3). A decrease in progeny number (66% of control) is seen in the first day of mating, though this decline is not as extreme as the one observed in older males (23% of control). A greater proportion of young *y w; stwl*^{j6c3} mutants are initially fertile compared to the older trans-heterozygotes (94% vs 78%), and the decline in the proportion of fertile males is much slower.

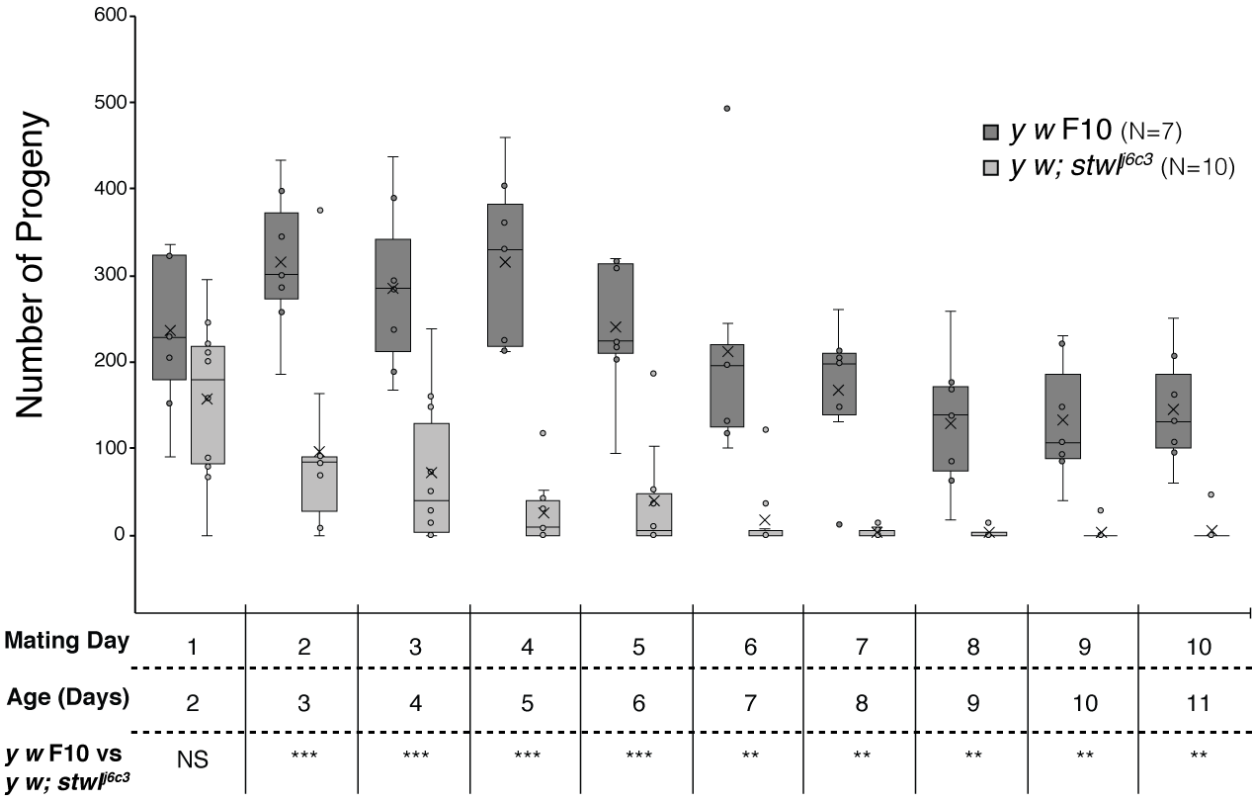


FIGURE 3. ***stwl^{j6c3}* homozygotes exhibit sperm exhaustion defects**

Fecundity of young, background-matched males having either two functional *stwl* alleles (*y w* F10) or none (*y w; stwl^{j6c3}*). The Y-axis shows the total number of progeny that each male produced from a single day of mating with two virgin *y w F10* females. The X-axis portrays the mating day and the age of the flies on that day (males were aged for 2 days prior to start of assay, hence age on mating day 1 is 2 days). Total number of males for each genotype is displayed next to the legend. Statistical analysis compares *y w; stwl^{j6c3}* fertility for each mating day against *y w* F10. Two-tailed p-values from a Student's t-test (two-sample assuming unequal variances) are presented as: p<.05 =*, p<.01=**, p<.001=***, p<.0001=****, NS=not significant (p>.05).

Our findings indicate that *stwl* mutant males are incapable of continuously producing progeny after successive matings. We posited that these mutant males may be able to regain fertility if given time to recover; were this to happen, it would suggest that *stwl* mutant males have a decreased rate of spermiogenesis, rather than a total loss of sperm production over time. We therefore performed a "sperm recovery assay", wherein we mated *y w* F10 and *y w; stwl^{j6c3}* homozygous males to *y w* F10 females for 6 days, transferred them to empty vials for 3 days, and resumed mating for another 6 days (Fig. 4). We found that neither genotype experiences a significant increase in fertility after the 3-day recovery period (day 6 vs day 10, p=0.59, 0.99 for *y w* F10 and *y w; stwl^{j6c3}*, respectively). Mutant males are 2% less fecund on the 1st day post-recovery compared to the fecundity prior to the rest-period, and none of the males that were sterile prior to the rest period regained fertility afterwards. We also counted male and female progeny to look for biases in sex ratio as a result of *stwl* loss; sex ratio is not altered in progeny of *stwl^{j6c3}* males (Fig. 5).

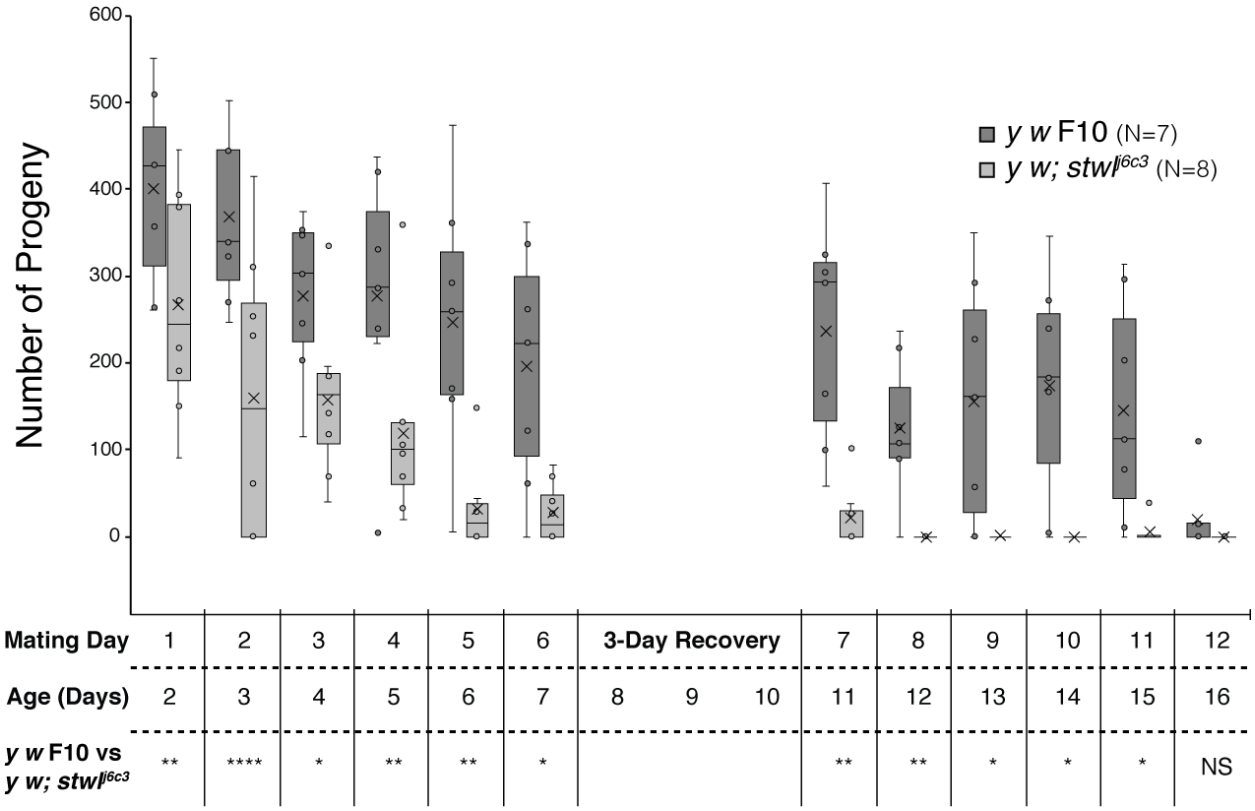


FIGURE 4. **Fertility of y w; *stwl*^{j6c3} males is not restored after a brief recovery period**
Fecundity of young, background-matched males having either two functional *stwl* alleles (*y w* F10) or none (*y w; stwl*^{j6c3}). The Y-axis shows the total number of progeny that each male produced from a single day of mating with two virgin *y w F10* females. The X-axis portrays the mating day and the age of the flies on that day (males were aged for 2 days prior to start of assay, hence age on mating day 1 is 2 days). During days 8, 9, and 10 males were isolated and not placed with females to copulate with. Total number of males for each genotype is displayed next to the legend. Statistical analysis compares *y w; stwl*^{j6c3} fertility for each mating day against *y w* F10. Two-tailed p-values from a Student's t-test (two-sample assuming unequal variances) are presented as: p<.05 =*, p<.01=**, p<.001=***, p<.0001=****, NS=not significant (p>.05).

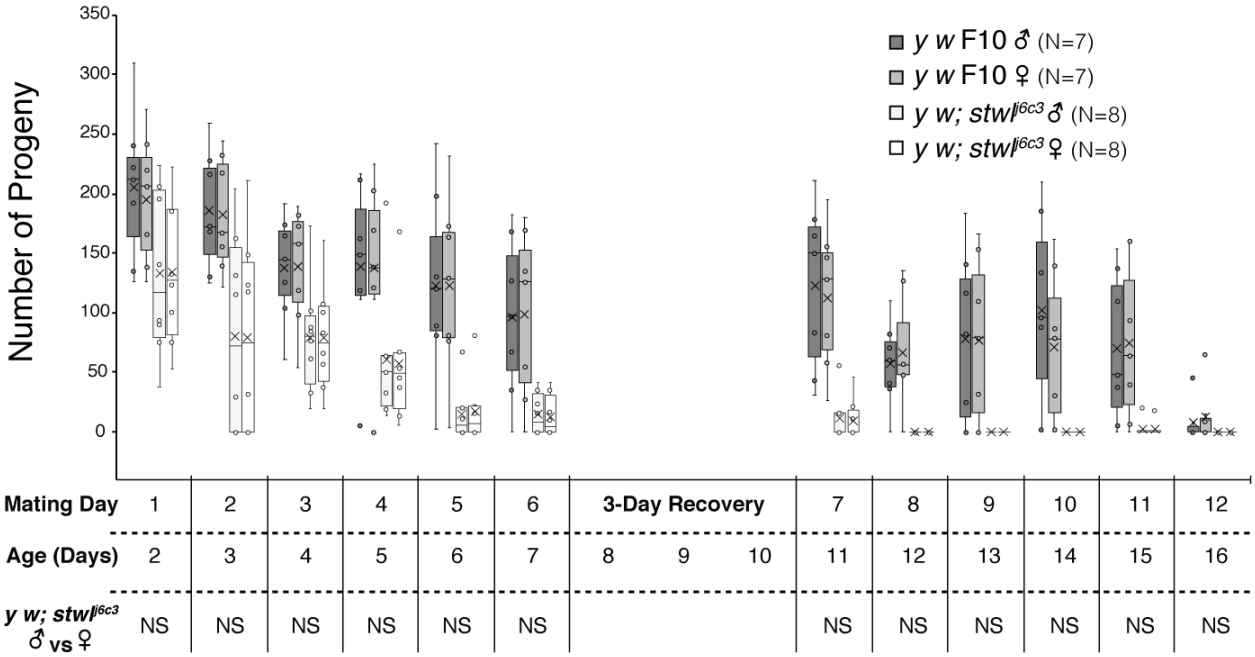


FIGURE 5. **Loss of *stwl* in males does not cause a distortion in sex ratio**

Fecundity of young, background-matched males having either two functional *stwl* alleles (*y w* F10) or none (*y w; stwl^{j6c3}*). The Y-axis shows the total number of male or female progeny that each male produced from a single day of mating with two virgin *y w F10* females. The X-axis portrays the mating day and the age of the flies on that day (males were aged for 2 days prior to start of assay, hence age on mating day 1 is 2 days). During days 8, 9, and 10 males were isolated and not placed with females to copulate with. Total number of males for each genotype is displayed next to the legend. Statistical analysis compares male and female progeny number from *y w; stwl^{j6c3}* for each mating day. Two-tailed p-values from a Student's t-test (two-sample assuming unequal variances) are presented as: p<.05 =*, p<.01=**, p<.001=***, p<.0001=****, NS=not significant (p>.05).

To better understand the spatio-temporal requirement for *stwl* in the testis, we utilized the *UAS/Gal4* system to target *stwl* transcript in germline and somatic cells of the testis. We knocked down *stwl* using a *UAS-RNAi* construct that targets all *stwl* transcripts and phenocopies *stwl* mutants when expressed in ovarian germ cells (Chapter 2, Fig.4). We drove expression of *UAS-stwl-RNAi* using *Act5c-Gal4* and *nos-Gal4*, which drive ubiquitous and early germline-restricted expression, respectively. We confirmed *nos-Gal4* expression by examining GFP localization in *nos-Gal4/UAS-GFP* testis, where it is largely restricted to germline stem cells and spermatogonia (Fig.6). Knockdown of *stwl* transcripts in ovaries with each of these drivers results in loss of ovarian germline; we therefore looked for a similar phenotype in testis. We confirmed knockdown of *stwl* transcript in testes via RT-PCR (Fig. 7). Surprisingly, *nos-Gal4* driven *stwl-RNAi* does not reduce fecundity in a sperm exhaustion assay (Fig. 8). 95% of *nos-Gal4/UAS-stwl-RNAi* males remain fertile after 10 days of mating, and these males are typically as fecund as their control counterparts. The only exceptions are on the first and final days of the assay. On day 1, *nos-Gal4/+* males are surprisingly less fecund than their knockdown counterparts; on the 10th day, these same flies significantly increase progeny production. The *Act5c-Gal4/UAS-stwl-RNAi* males effectively phenocopy the *stwl* null flies assayed previously; they are initially fertile (94% fertility on day 1, 60% as fecund as control groups), and they become less fecund steadily over time; after 8 days of mating, these males are completely sterile. This experiment also rules out the possibility that recessive, deleterious mutations were contributing to decreased fertility of test subjects; *Act5c-Gal4/UAS-stwl-RNAi* flies do not have a heightened mortality rate compared to controls.

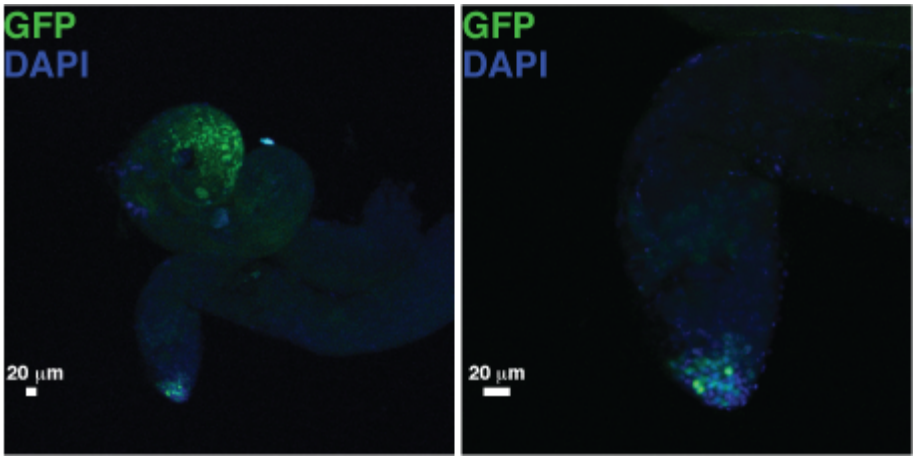


FIGURE 6. ***nanos-Gal4*** **driver is germline-specific**
Testes dissected from *nanos-Gal4*/*UAS-GFP-NLS* confirmed the driver is largely restricted to germline stem cells and spermatogonia, based on proximity to and placement at the apical end of the testis.

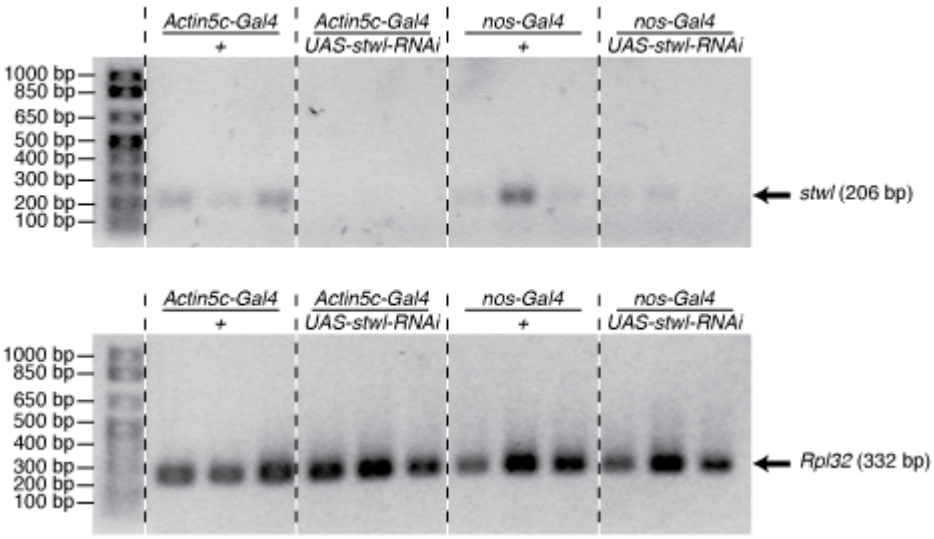


FIGURE 7. **RT-PCR shows reduction of *stwl* transcript upon induction of *UAS-stwl-RNAi***
RT-PCR against *stwl* transcripts; RNA was extracted in triplicate from ~20 pairs of testes of the indicated genotype. Each lane was loaded with PCR product generated from cDNA derived from one of these replicate pools. *Rpl32* was assayed as a loading control.

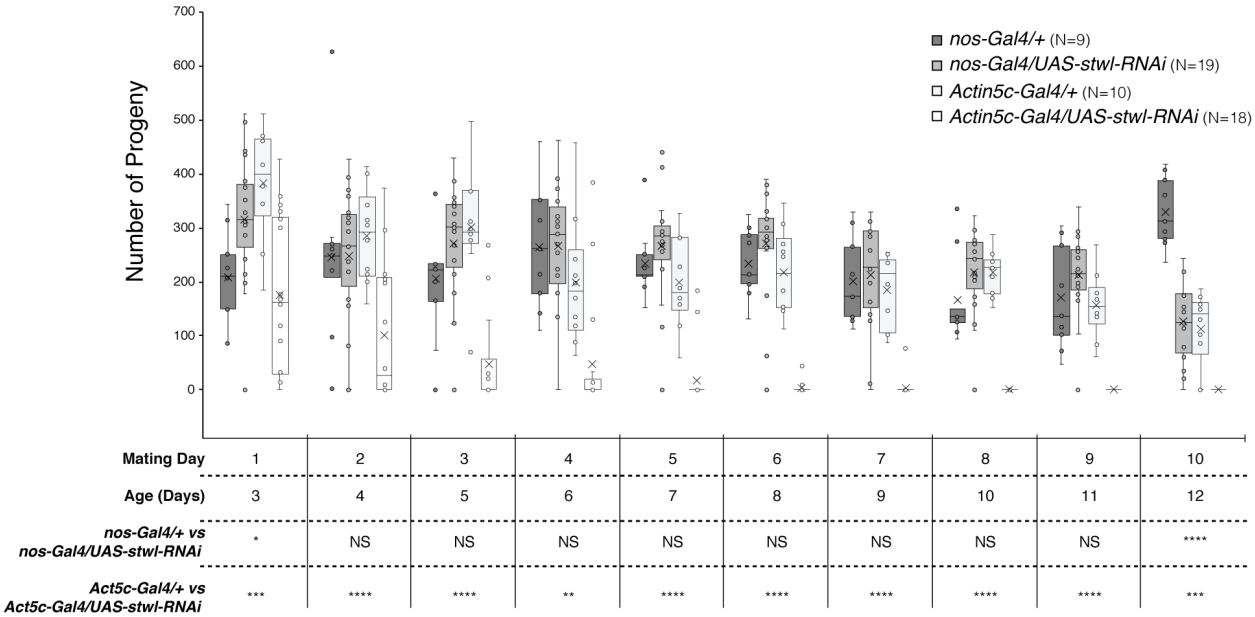


FIGURE 8. ***stwl* is not required in male GSCs for prolonged fertility**

Fecundity of males expressing *UAS-stwl-RNAi* construct driven either by *nos-Gal4* or *Actin5c-Gal4*, as well as negative control males lacking the RNAi construct. The Y-axis shows the total number of progeny that each male produced from a single day of mating with two virgin *y w F10* females. The X-axis portrays the mating day and the age of the flies on that day (males were aged for 3 days prior to start of assay, hence age on mating day 1 is 3 days). Total number of males for each genotype is displayed next to the legend. Statistical analysis compares fertility of each *Gal4*/*UAS-stwl-RNAi* genotype against its respective *Gal4/+* control. Two-tailed p-values from a Student's t-test (two-sample assuming unequal variances) are presented as: p<.05 =*, p<.01=**, p<.001=***, p<.0001=****, NS=not significant (p>.05).

In females, *stwl* mutant ovaries are obviously and dramatically altered: they are significantly smaller, lacking in developed egg chamber, display a reduction in GSCs over time and are often entirely devoid of germ cells, especially at an advanced age (10+ days). To understand the cause of male fertility reduction, we looked for morphological defects in the male testis. We found that *y w; stwl*^{j6c3} males still produced motile, bundled sperm, even after being continuously mated for 10 days (Fig. 9). We then performed immunofluorescence microscopy using antibodies against Vasa, Hts-1B1, and Fas3. Vasa is exclusive to germ cell cytoplasm; Hts-1B1 is a component of spectrosomes and branched fusomes (spectrosomes present as spheres exclusive to GSCs and gonialblasts while fusomes are branched structures within spermatogonia and spermatocytes); Fas3 is a component of the hub, which is the GSC niche in the *Drosophila* testis. We did not identify any defects in *y w; stwl*^{j6c3} testes that could explain the sperm exhaustion defect (Fig. 10-12).

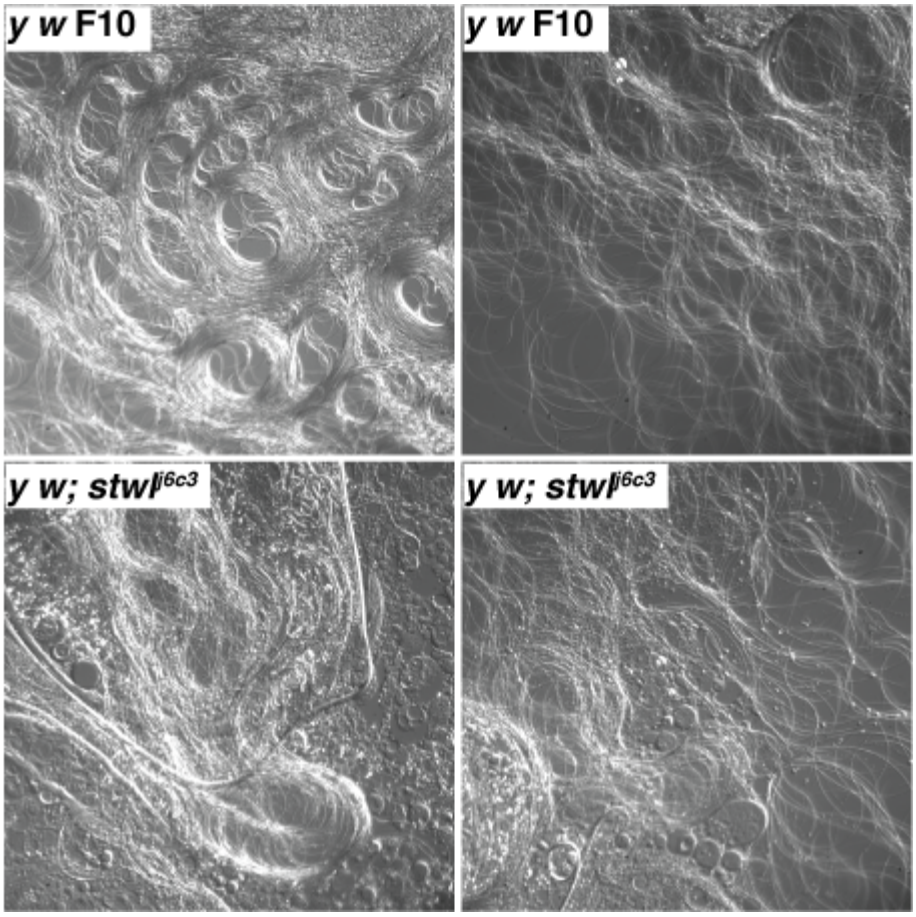
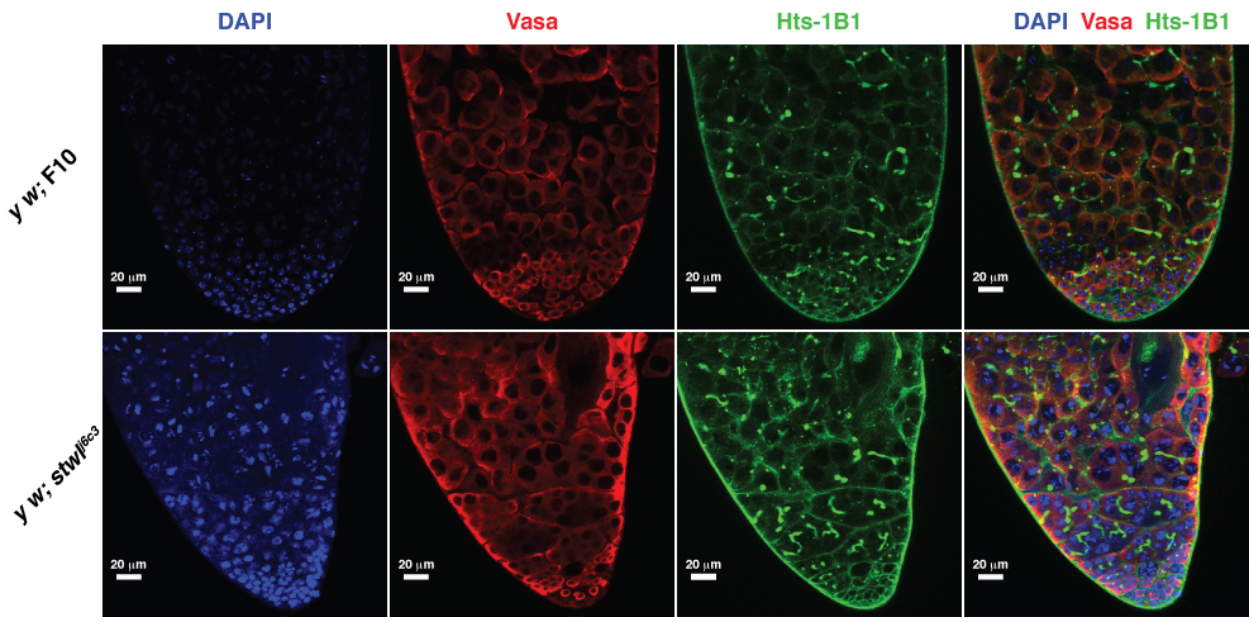


FIGURE 9. *stwl*^{j6c3} null males produce motile, bundled sperm
Testes were dissected from 10-day old, mated males of the indicated genotype. They were ruptured and subsequently imaged using DIC microscopy.



<u>FIGURE 10</u>. Young *stwl*^{j6c3} males have normal-looking testes

Testes were dissected from 2-day old, mated males of the indicated genotype. Mutant testes appear to have healthy populations of GSCs and gonialblasts (spherical spectrosome), and spermatogonia and spermatocytes (branched fusomes). All images are maximum-intensity projections from a z-series representing a depth of 10 microns.

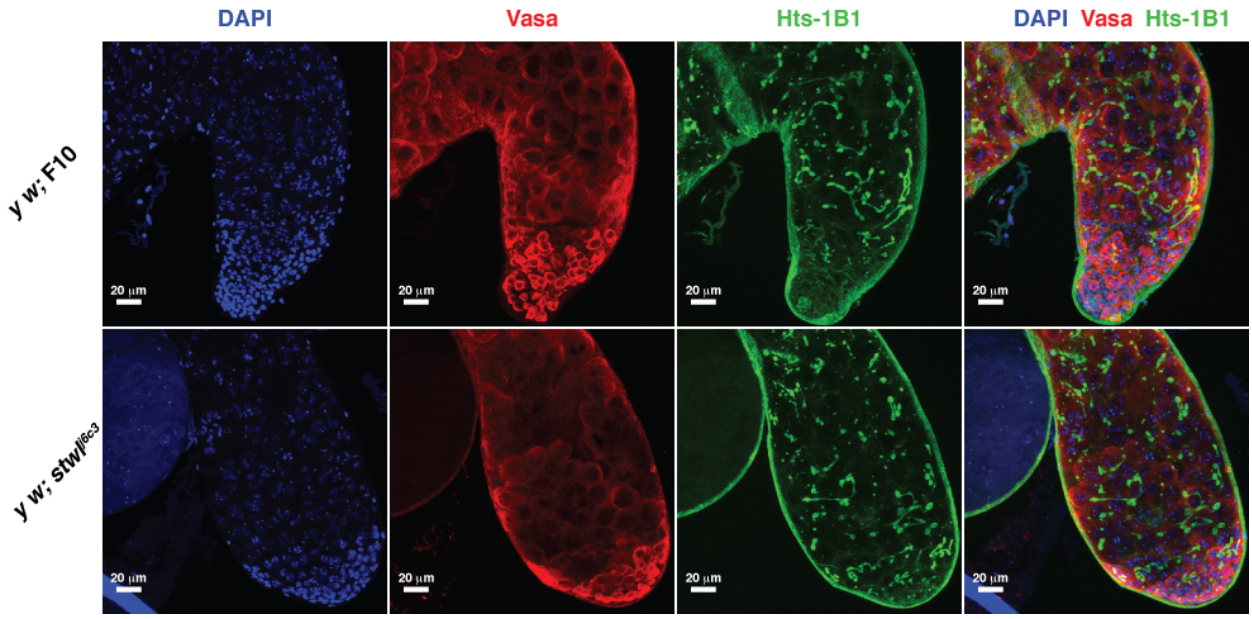


FIGURE 11. Aged *stwl*^{j6c3} males have normal testes

Testes were dissected from 15-day old, mated males of the indicated genotype. Mutant testes appear to have healthy populations of GSCs and gonialblasts (spherical spectrosome), and spermatogonia and spermatocytes (branched fusomes). All images are maximum-intensity projections from a z-series representing a depth of 10 microns.

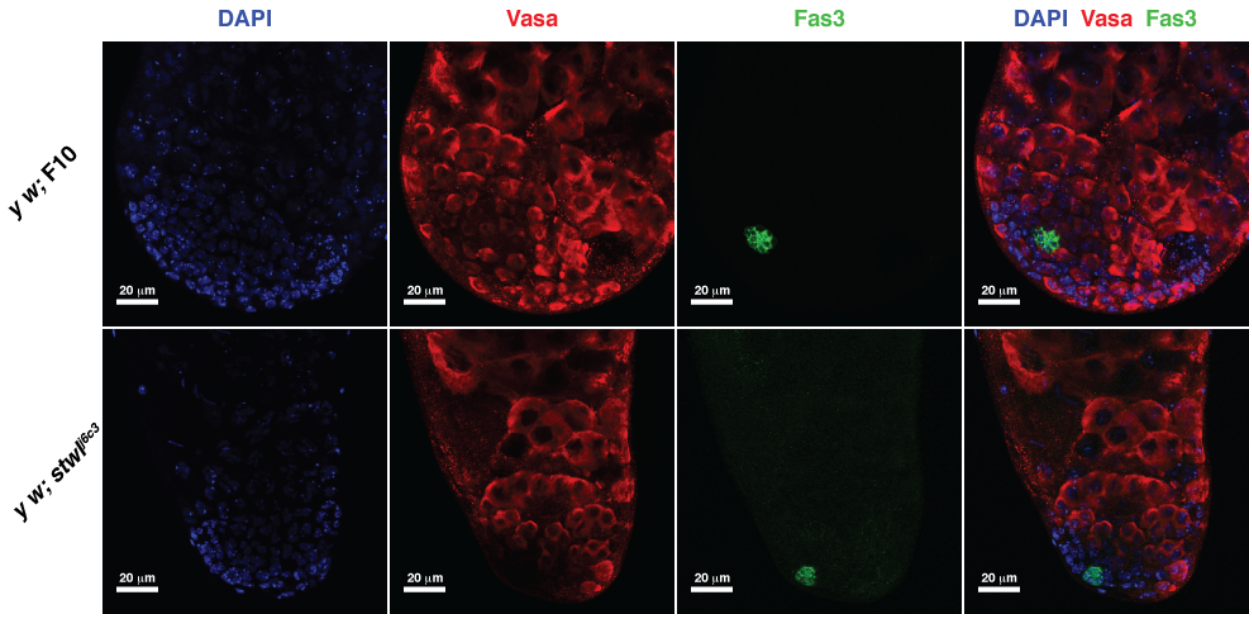


FIGURE 12. Aged *stwl*^{j6c3} males have normal-looking testes

Testes were dissected from 10-day old, mated males of the indicated genotype. No obvious perturbations to the hub (green) are evident in *stwl* mutants.

Discussion

Previous studies had shown that *stwl* is expressed in testis, and that Stwl protein localizes to germ cells in 3rd instar and adult testis, but a null phenotype was not identified in males. We report here that Stwl plays an important role in male reproduction. *stwl* null males are initially fertile, although they are less fecund than their wild type counterparts even on the first mating day. The *stwl* null males experience a dramatic reduction in fertility with each successive day of mating, typically become sterile after ~5 days of mating. We found that 10-day old *stwl* null males are also initially fertile, although they experience an even more rapid decline in fertility subsequent to the initial mating. This suggests that these males have an initial pool of potentially progeny-producing sperm that is quickly depleted. We also found that *stwl* null males do not experience an increase in fertility after a prolonged pause in mating; this rules out the possibility that they are still producing sperm at a decreased rate that is undetectable after repeated matings.

Lastly, we showed that the requirement for *stwl* is not during early stages of spermatogenesis, as *nos-Gal4/UAS-stwl-RNAi* flies did not exhibit a sperm exhaustion defect, while *Actin5c-Gal4/UAS-stwl-RNAi* males were mostly sterile after 5 days of mating. While the *Actin5c-Gal4* was constructed and described as a ubiquitous promoter, it has been reported by other labs to be specific to somatic cells in male gonads(Dorus et al., 2006; Ito et al., 1997; White-Cooper, 2012). We found that *Actin5c-Gal4/UAS-stwl-RNAi* testes exhibit a loss of Stwl signal in germ cells when probed with Stwl antibodies (see section on Antibody Validation). Therefore, we cannot confidently say that the sperm exhaustion defect was caused specifically by loss of *stwl* in somatic cells. Because *nos-Gal4* is restricted to early germ cells (GSCs and spermatogonia), we cannot conclude that *stwl* is not required in late-stage male germ cells. In order to pinpoint the requirement for Stwl in male fertility, we would like to do the following: 1)

Perform sperm exhaustion experiments with Gal4 constructs targeting somatic cells (*Upd-Gal4* in the Hub*, ptc-Gal4* and *eya-Gal4* in cyst cells) and a broader range of germ cells (*bam-Gal4*, *vasa-Gal4*), 2) Rescue sustained fertility by expressing *UAS-stwl-HA* transgene in *stwl* null background, 3) Confirm loss/expression of Stwl protein in these cells in affected males.

A recent study by Duan and Geyer hints at a possible mechanism for the sperm exhaustion defect in *stwl* null testes (Duan and Geyer, 2018). They found a similar male fertility defect in mutants for *su(Hw)*, a known insulator-binding protein. Specifically, they identified a sperm exhaustion phenotype wherein *su(Hw)* null males become ~50% sterile after 10 days of mating. They found that premeiotic cells (GSCs, spermatogonia, spermatocytes) are unaffected in these mutants, but that the testes lose Individualization Complexes (IC) over time. ICs are actin-rich structures that promote individualization of sperm nuclei after meiosis. Furthermore, Duan and Geyer found that the sterility/IC loss defects were phenocopied in flies expressing RNAi against *su(Hw)* in somatic cyst cells, but not when RNAi was directed in germ cells. It is plausible, given the data we have presented, that we would find a similar cellular requirement for *stwl* in cyst cells and for the production of Individualization Complexes. The potential link between Stwl and Su(Hw) in spermiogenesis is worth noting in light of our finding that Stwl is an insulator-binding protein (Chapter 3).

References

1. Clark, K. (1996). The Drosophila Stonewall gene encodes a putative transcription factor that regulates oocyte differentiation and germ cell viability. PhD Thesis. University of Texas Southwestern Medical Center.

2. Clark, K.A., and McKearin, D.M. (1996). The Drosophila stonewall gene encodes a putative transcription factor essential for germ cell development. Development *122*, 937–950.

3. Dorus, S., Busby, S.A., Gerike, U., Shabanowitz, J., Hunt, D.F., and Karr, T.L. (2006). Genomic and functional evolution of the Drosophila melanogaster sperm proteome. Nat. Genet. *38*, 1440–1445.

4. Duan, T., and Geyer, P.K. (2018). Spermiogenesis and Male Fertility Require the Function of Suppressor of Hairy-Wing in Somatic Cyst Cells of Drosophila. Genetics *209*, 757–772.

5. Ito, K., Awano, W., Suzuki, K., Hiromi, Y., and Yamamoto, D. (1997). The Drosophila mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurones and glial cells. Dev. Camb. Engl. *124*, 761–771.

6. Maines, J.Z., Park, J.K., Williams, M., and McKearin, D.M. (2007). Stonewalling Drosophila stem cell differentiation by epigenetic controls. Development *134*, 1471–1479.

7. White-Cooper, H. (2012). Tissue, cell type and stage-specific ectopic gene expression and RNAi induction in the Drosophila testis. Spermatogenesis *2*, 11–22.

LOOKING BACK, LOOKING AHEAD

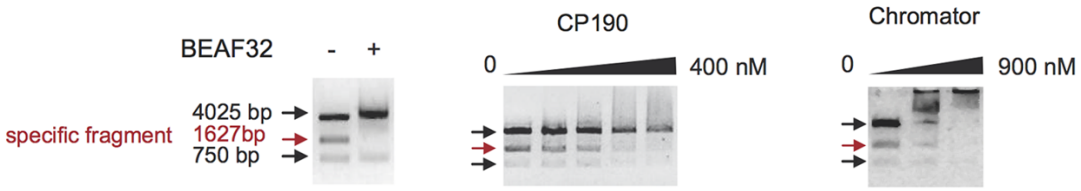
In this text, we have described a number of key features of Stwl: 1) Stwl associates with insulators in the genome, 2) Stwl associates with heterochromatin (telomeres and pericentromeres), 3) *stwl* null ovaries exhibit a "masculinization" phenotype, 4) Stwl is required for sustained male fertility. These findings present a wide range of distinct (and sometimes overlapping) biological mechanisms that Stwl is involved in. More work is needed to disentangle the complex phenotypes associated with *stwl* loss, and to understand the specific biological mechanisms that cause them.

Does Stwl directly bind insulators?

Our ChIP-Seq data (and data from an independent experiment conducted in embryos) demonstrate that Stwl protein is associating with genomic insulators (Kudron et al., 2018). Immunolabeling of Stwl in ovaries shows that it is associating with insulator complexes (Rohrbaugh et al., 2013).The next step is to better understand the role that Stwl plays in establishing and/or maintaining genome architecture.

An important follow-up experiment is to determine whether Stwl directly binds insulator sequences or is directed there by complexing with other insulator-binding proteins. There are a number of angles that need to be explored here. First, we would determine which sets of insulator-binding proteins most frequently associates with at known insulator sequences. For example, HMR most commonly occupies genomic regions that are also occupied by CP190, Mod(mdg4), Su(Hw), and BEAF-32, but not CTCF (Gerland et al., 2017). This would help us determine which proteins are potentially complexing with Stwl, and whether it or one of these other proteins is responsible for directly binding Stwl-occupied sites. This information could be more directly inferred using an electric mobility shift assay (EMSA).

EMSA (sometimes known as a gel shift assay) is used to determine whether a purified protein is capable of binding a naked piece of DNA of a given sequence. We would do this as follows: Identify a number of loci (1-5) that are occupied by Stwl and other insulator-binding proteins, especially those loci that contain Stwl binding motifs as identified in chapter 3. Clone each sequence into a plasmid and digest the plasmid such that it produces 3 or more fragments of lengths that can be easily separated via gel electrophoresis. We would then express and purify fusion proteins (e.g. Stwl-MBP, BEAF32-MBP) from *E. coli* and incubate each protein with the mixture of DNA fragments to determine whether the protein binds to all fragments (non-specific binding), no fragments (does not bind DNA), or the insulator fragment (sequence-specific binding) (Figure 1).



<u>Figure 1</u>. EMSA assay from Vogelmann et al. 2014 shows BEAF32 binds insulator fragments while CP190 and Chromator non-specifically bind DNA

Here the authors cloned a 1627 bp insulator fragment into a plasmid and digested it with restriction enzymes to produce one specific fragment and two sequences formed from the plasmid backbone. Incubation with purified MBP-BEAF32 caused an upshift of the specified fragment, but not the background fragments. Incubation with His-CP190 and His-Chromator at varying concentrations resulted in an upshift of all fragments, indicating that they do not specifically bind this insulator sequence (Vogelmann et al., 2014).

To make detection easier, we can also incubate the DNA-protein mixture with antibodies to the target proteins in order to create a larger antibody-protein-DNA complex that can be clearly observed via SDS-PAGE (supershift assay). The design of this study also would enable us to test full-length Stwl along with C-terminal (BESS domain fragment) and N-terminal (MADF domain fragment) peptides. If Stwl directly binds insulator sequences, we suspect that the MADF domain (which binds DNA) would be required to perform this function, whereas if Stwl merely complexes with other proteins at insulator sites, we expect that the protein-binding BESS domain would be required to associate to DNA (when incubated with insulator-binding proteins that directly bind DNA).

What is Stwl's function at insulators and heterochromatin?

We find that Stwl occupies heterochromatic loci, specifically telomeres and pericentromeres, and chromosome 4 (chapter 1). We have also shown that *stwl* null and heterozygous ovaries exhibit TE and satellite derepression (chapter 2). We speculate that Stwl is required for establishing heterochromatin boundaries, and that the derepression of repetitive elements in *stwl* mutants is caused by the spreading of euchromatic histone modifications and transcription machinery to typically silent loci. In order to test this hypothesis, we would incorporate modENCODE data of active/repressed chromatin states and identify overlaps with Stwl-occupied sites (Consortium et al., 2010). We would especially want to see if Stwl peaks tend to occur at or within heterochromatin boundaries. As a follow-up experiment, we would also examine changes in chromatin activity (H3K9me3, H3K27me3, RNA-PolII) in *stwl*-mutant tissues. We would perform ChIP-Seq using antibodies against the aforementioned proteins in wild-type and *stwl* heterozygous ovaries. If Stwl is required for maintaining heterochromatin

boundaries, we expect to see a decrease of repressed chromatin marks at or near Stwl-occupied loci.

With regard to Stwl function at insulator sites, we previously speculated that Stwl may be blocking enhancer-promoter interactions. To further explore this possibility, we would examine Stwl occupancy at differentially expressed promoters. If Stwl blocks enhancer-promoter interactions, we would expect to find an abundance of Stwl-occupied sites to be adjacent to differentially expressed promoters. We would compare this data to ChIP-Seq datasets for other proteins that occupy promoters, including insulator-binding proteins.

Insulator-binding proteins, such as CP190, can also mediate long-range interactions (Saha et al., 2019; Vogelmann et al., 2014). We have not determined whether Stwl is involved in establishing such interactions, which can organize domains such that regions on different chromosomes physically interact. If Stwl is involved in the formation of long-range interactions, it is possible that it could promote tethering of euchromatic insulator sites to heterochromatic regions. In order to explore this possibility, we would examine available Hi-C data to determine whether Stwl-occupied sites are preferentially enriched for sites involved in long-range interactions (Li et al., 2017; Schwartz and Cavalli, 2017; Sexton et al., 2012).

How does Stwl silence male transcripts in ovaries?

In chapter 2, we outlined the various molecular phenotypes associated with *stwl* mutants and *stwl* dsRNA-treated S2 cells. Most striking was the ectopic expression of testis-enriched transcripts in these tissues. We did not find that Stwl-bound sites were enriched at these genes, suggesting that this phenotype may be a downstream reporter of a more direct *stwl* loss phenotype. 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 (Insco et al., 2012; Ohlstein et al., 2000). *bgcn* is a strong candidate for Stwl regulation: its promoter is bound by Stwl, and it is highly upregulated in *stwl* null ovaries and *stwl* dsRNA-treated S2 cells. Furthermore, the expression of *bgcn* transcript in ovaries is anti-correlated to Stwl and females expressing a *hs-bgcn* transgene are sterile (Ohlstein et al., 2000). We propose that loss of Stwl in the developing germline cyst results in ectopic expression of Bgcn in these cells that eventually leads to female sterility (Figure 2).

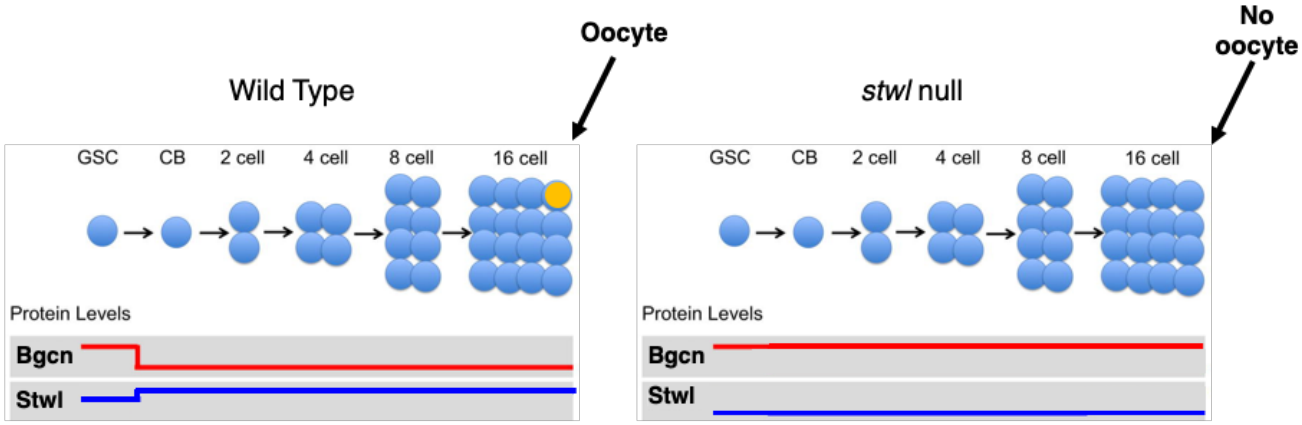


Figure 2. Proposed model for Stwl as an essential regulator of Bgcn

In wild-type ovaries, Bgcn expression is restricted to GSCs and cystoblasts (Ohlstein et al., 2000). Stwl expression is very low in these same cells and increases substantially in the developing germline cyst. We propose that loss of Stwl results in an increase of Bgcn outside of GSCs and cystoblasts, causing expression of male transcripts that ultimately disrupts formation of the oocyte in the 16-cell cyst. Figure adapted from (Salz et al., 2017).

Our first approach in confirming this model would be to examine Bgcn protein expression in wild-type and *stwl* null ovaries (Figure 2). We would then attempt to rescue oocyte formation in *stwl* null ovaries by reducing *bgcn* expression in germline cysts (*bam-gal4*; *UAS-*

bgcn-RNAi). Similarly, we would test whether ectopic expression of *stwl* (*UAS-stwl-HA*) in a given cell type results in a corresponding reduction of Bgcn protein or transcript.

References

1. Consortium, T. modENCODE, Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., et al. (2010). Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. Science *330*, 1787–1797.

2. Gerland, T.A., Sun, B., Smialowski, P., Lukacs, A., Thomae, A.W., and Imhof, A. (2017). The Drosophila speciation factor HMR localizes to genomic insulator sites. PLOS ONE *12*, e0171798.

3. Insco, M.L., Bailey, A.S., Kim, J., Olivares, G.H., Wapinski, O.L., Tam, C.H., and Fuller, M.T. (2012). A self-limiting switch based on translational control regulates the transition from proliferation to differentiation in an adult stem cell lineage. Cell Stem Cell *11*.

4. Kudron, M.M., Victorsen, A., Gevirtzman, L., Hillier, L.W., Fisher, W.W., Vafeados, D., Kirkey, M., Hammonds, A.S., Gersch, J., Ammouri, H., et al. (2018). The ModERN Resource: Genome-Wide Binding Profiles for Hundreds of Drosophila and Caenorhabditis elegans Transcription Factors. Genetics *208*, 937–949.

5. Li, Q., Tjong, H., Li, X., Gong, K., Zhou, X.J., Chiolo, I., and Alber, F. (2017). The three-dimensional genome organization of Drosophila melanogaster through data integration. Genome Biol. *18*.

6. Ohlstein, B., Lavoie, C.A., Vef, O., Gateff, E., and McKearin, D.M. (2000). The Drosophila cystoblast differentiation factor, benign gonial cell neoplasm, is related to DExH-box proteins and interacts genetically with bag-of-marbles. Genetics *155*, 1809–1819.

7. Rohrbaugh, M., Clore, A., Davis, J., Johnson, S., Jones, B., Jones, K., Kim, J., Kithuka, B., Lunsford, K., Mitchell, J., et al. (2013). Identification and Characterization of Proteins Involved in Nuclear Organization Using Drosophila GFP Protein Trap Lines. PLoS ONE *8*.

8. Saha, P., Sowpati, D.T., and Mishra, R.K. (2019). Epigenomic and genomic landscape of Drosophila melanogaster heterochromatic genes. Genomics *111*, 177–185.

9. Salz, H.K., Dawson, E.P., and Heaney, J.D. (2017). Germ cell tumors: Insights from the Drosophila ovary and the mouse testis. Mol. Reprod. Dev. *84*, 200–211.

10. Schwartz, Y.B., and Cavalli, G. (2017). Three-Dimensional Genome Organization and Function in Drosophila. Genetics *205*, 5–24.

11. Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472.

12. Vogelmann, J., Gall, A.L., Dejardin, S., Allemand, F., Gamot, A., Labesse, G., Cuvier, O., Nègre, N., Cohen-Gonsaud, M., Margeat, E., et al. (2014). Chromatin Insulator Factors Involved in Long-Range DNA Interactions and Their Role in the Folding of the Drosophila Genome. PLOS Genet. *10*, e1004544.