

RESEARCH ARTICLE

Genetic variation in *P*-element dysgenic sterility is associated with double-strand break repair and alternative splicing of TE transcripts

Jyoti Lama^{1,2}, Satyam Srivastav^{1,2}, Sadia Tasnim¹, Donald Hubbard¹, Savana Hadjipanteli¹, Brittany R. Smith³, Stuart J. Macdonald³, Llewellyn Green¹, Erin S. Kelleher^{1*}

1 Department of Biology and Biochemistry, University of Houston, Houston, Texas, United States of America, **2** Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York, United States of America, **3** Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas, United States of America

* eskelleher@uh.edu



OPEN ACCESS

Citation: Lama J, Srivastav S, Tasnim S, Hubbard D, Hadjipanteli S, Smith BR, et al. (2022) Genetic variation in *P*-element dysgenic sterility is associated with double-strand break repair and alternative splicing of TE transcripts. *PLoS Genet* 18(12): e1010080. <https://doi.org/10.1371/journal.pgen.1010080>

Editor: Harmit S. Malik, Fred Hutchinson Cancer Research Center, UNITED STATES

Received: February 7, 2022

Accepted: November 2, 2022

Published: December 7, 2022

Copyright: © 2022 Lama et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The RNA-seq and small RNA-seq that support the findings of this study are publicly available from SRA archive with the identifier PRJNA490147. The MSG data are publicly available from SRA archive with the identifier PRJNA893666. All other relevant data are within the manuscript and its [Supporting Information](#) files.

Funding: This research was funded by NSF Division of Environmental Biology #1457800 to

Abstract

The germline mobilization of transposable elements (TEs) by small RNA mediated silencing pathways is conserved across eukaryotes and critical for ensuring the integrity of gamete genomes. However, genomes are recurrently invaded by novel TEs through horizontal transfer. These invading TEs are not targeted by host small RNAs, and their unregulated activity can cause DNA damage in germline cells and ultimately lead to sterility. Here we use hybrid dysgenesis—a sterility syndrome of *Drosophila* caused by transposition of invading *P*-element DNA transposons—to uncover host genetic variants that modulate dysgenic sterility. Using a panel of highly recombinant inbred lines of *Drosophila melanogaster*, we identified two linked quantitative trait loci (QTL) that determine the severity of dysgenic sterility in young and old females, respectively. We show that ovaries of fertile genotypes exhibit increased expression of splicing factors that suppress the production of transposase encoding transcripts, which likely reduces the transposition rate and associated DNA damage. We also show that fertile alleles are associated with decreased sensitivity to double-stranded breaks and enhanced DNA repair, explaining their ability to withstand high germline transposition rates. Together, our work reveals a diversity of mechanisms whereby host genotype modulates the cost of an invading TE, and points to genetic variants that were likely beneficial during the *P*-element invasion.

Author summary

Transposable elements (TEs) are mobile genetic parasites that spread through host species' genomes by making additional copies of themselves in developing gametes. Transposition requires the breakage of DNA to introduce a new TE copy, which burdens the host cell to repair the damage or tolerate the mutation. To avoid these fitness costs, eukaryotic

ESK, NIH R35 GM138112 to ESK, NIH R01 OD010974 to SJM and Anthony D. Long and NIH P20 GM103638 to the University of Kansas Genome Sequencing Core. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

hosts suppress transposition in developing gametes through small-RNA mediated silencing. However, what happens when a new TE that is not recognized by existing small-RNAs invades the genome? We examined genetic variation in the degree to which *Drosophila melanogaster* oogenesis is disrupted by the transposition of a newly invading TE family. We show that while transposition results in gamete loss in some genotypes, other genotypes maintain fertility. Furthermore, we show that increased fertility likely reflects both decreased permissivity of transposition, as well as increased tolerance of DNA damage. Our observations provide a window into how host genetic variation impacts the consequences of invading TEs.

Introduction

Transposable elements (TE) are mobile DNA sequences that spread through host genomes by replicating in germline cells. Although individual TE insertions are sometimes beneficial, genomic TEs are foremost genetic parasites reviewed in [1]. Unrestricted transposition not only produces deleterious mutations, but also double-stranded breaks (DSBs) that lead to genotoxic stress in developing gametes. The mechanisms by which hosts enact silencing of resident TEs through the heritable production of regulatory small RNAs is extensively studied and broadly conserved [2,3]. However, host genomes are frequently invaded by new TE families, against which they lack small RNA-mediated “immunity” [4–7]. In the context of such novel TEs, genetic variation in the host’s ability to produce gametes could be a critical determinant of fitness. The presence and mechanisms of such host variation remain largely unstudied.

P-element DNA transposons, which invaded natural populations of *Drosophila melanogaster* around 1950, provide a unique opportunity to uncover host genetic variation in transposition-dependent sterility [8–10]. Strains isolated from natural populations prior to this invasion, referred to as M strains, do not contain genomic *P*-elements, and do not produce maternally-transmitted piRNAs that control their expression and transposition. When females from M strains are mated to males bearing genomic *P*-elements (*P*-strains), they produce dysgenic offspring that do not negatively regulate *P*-elements in germline cells [11]. A range of fertility effects result from unregulated *P*-element transposition, including the complete loss of germline cells and sterility [12]. However, naive M genotypes differ in their propensity to produce dysgenic progeny suggesting genetic variation in dysgenic sterility [8,10,13,14].

One potential source of variation in *P*-element dysgenic sterility arises from the response of germline cells to DSBs arising from transposition. In dysgenic females, primordial germ cells (PGCs) are lost beginning in the second instar larval stage, most likely due to unrepaired DSBs [15–17]. Furthermore, mutations in DNA damage response and repair proteins are known to enhance dysgenic germ cell loss [18,19]. Therefore, it is predicted that genotypes that enact more efficient DSB repair should be more tolerant of *P*-element transposition, and maintain germline cells. A related response is the production of *de novo* piRNAs, which is triggered by the activation of the DNA damage response protein checkpoint kinase 2 (CHK2) in the adult female germline [20,21]. These *de novo* piRNAs transcriptionally silence *P*-elements in a process analogous to maternally transmitted silencing, and restore fertility as dysgenic females age [20,21]. If *de novo* piRNA production activates more readily in some genotypes than others, it could lead to fertility differences in dysgenic crosses.

Another potential source of variation in dysgenic sterility lies with host co-factors of transposition, including host proteins that regulate the transcription and splicing of transposase-encoding RNA, or the activity of *P*-element transposase enzyme. In particular, differences in

splicing cofactors between germline and somatic cells ensure *P*-element transposase-encoding transcripts are only produced in germline cells [22]. However, individual germlines could differ in the production of these cofactors. Beyond transposase production, *P*-elements insert preferentially near origins of replication, a strategy that may facilitate their spread through the genome by ensuring they are replicated multiple times in a given S-phase [23]. Differences in the timing or composition of this machinery could therefore drive differences in transposition rates, and ultimately downstream germline loss.

We recently isolated natural variation in dysgenic sterility through QTL mapping, using a panel of highly recombinant inbred lines derived from M strains (*Drosophila* Synthetic Population Resource, DSPR, Population A RILs, [24]). We mapped a major effect QTL surrounding the gene *bruno*, a female germline differentiation factor [14]. Here we present results from a second QTL mapping study in an independent panel of DSPR RILs (Population B, [24]). We describe two QTL that determine differences in dysgenic sterility, one in young females only, and one in aged females only. Focusing on young dysgenic females, we further interrogated mechanisms underlying fertility differences by contrasting RNA and small RNA expression, radiation sensitivity, and *P*-element expression and splicing between fertile and sterile genotypes. Our results suggest that natural variation in dysgenic sterility arises through differences in both germline DNA repair and *P*-transposase mRNA splicing, revealing considerable complexity in host factors that modulate the fitness costs associated with transposition.

Results

QTL mapping

The DSPR RILs are all *P*-element free M-strains, which were derived from founders isolated from natural populations before the *P*-element invasion [24]. We therefore screened for alleles that influence dysgenic sterility among the panel B RIL genomes by crossing RIL females to males from the reference *P*-strain Harwich, and examining the morphology of F1 ovaries (Fig 1A). Atrophied ovaries are indicative of germline loss resulting from *P*-element activity [14,25]. Since dysgenic sterility changes across development [15], and some females exhibit age-dependent recovery from *P*-element hybrid dysgenesis through the production of *de novo* piRNAs [20], we phenotyped F1 females at two developmental time points: 3 days and 21 days post-eclosion.

Similar to our observations with the Population A RILs [14], we found continuous variation in the frequency of ovarian atrophy among dysgenic offspring of different RIL mothers, indicating genetic variation in dysgenic sterility that is unrelated to maternally deposited piRNAs (S1 and S2 Tables). Based on a combined linear model of F1 atrophy among 3 and 21 day-old females, we estimated the broad-sense heritability of dysgenic ovarian atrophy in our experiment to be ~42.5%. The effect of age on the proportion of F1 atrophy was significant but minimal ($\chi^2 = 7.03$, $df = 1$, p -value = 0.008) with 21 day-old females showing only 0.7% increase in atrophy as compared to 3 day-old females. This suggests that age-dependent recovery from dysgenic ovarian atrophy through the production of *de novo* piRNAs is not common among the genotypes we sampled.

To identify the genomic regions associated with genetic variation in dysgenic ovarian atrophy, we performed QTL mapping using the published RIL genotypes [24]. We found a large QTL peak near the 2nd chromosome centromere in both 3 and 21 day-old F1 females (Fig 1B and Tables 1, S3, and S4). However, the genomic intervals of the two QTL are non-overlapping (Fig 1C and Table 1). The major QTL in 21 day-old females (hereafter, QTL-21d) resides in the euchromatic region and is quite small (990 kb) compared to the major QTL in 3 day-old females (hereafter QTL-3d), which spans the centromere and pericentromeric regions (9.6

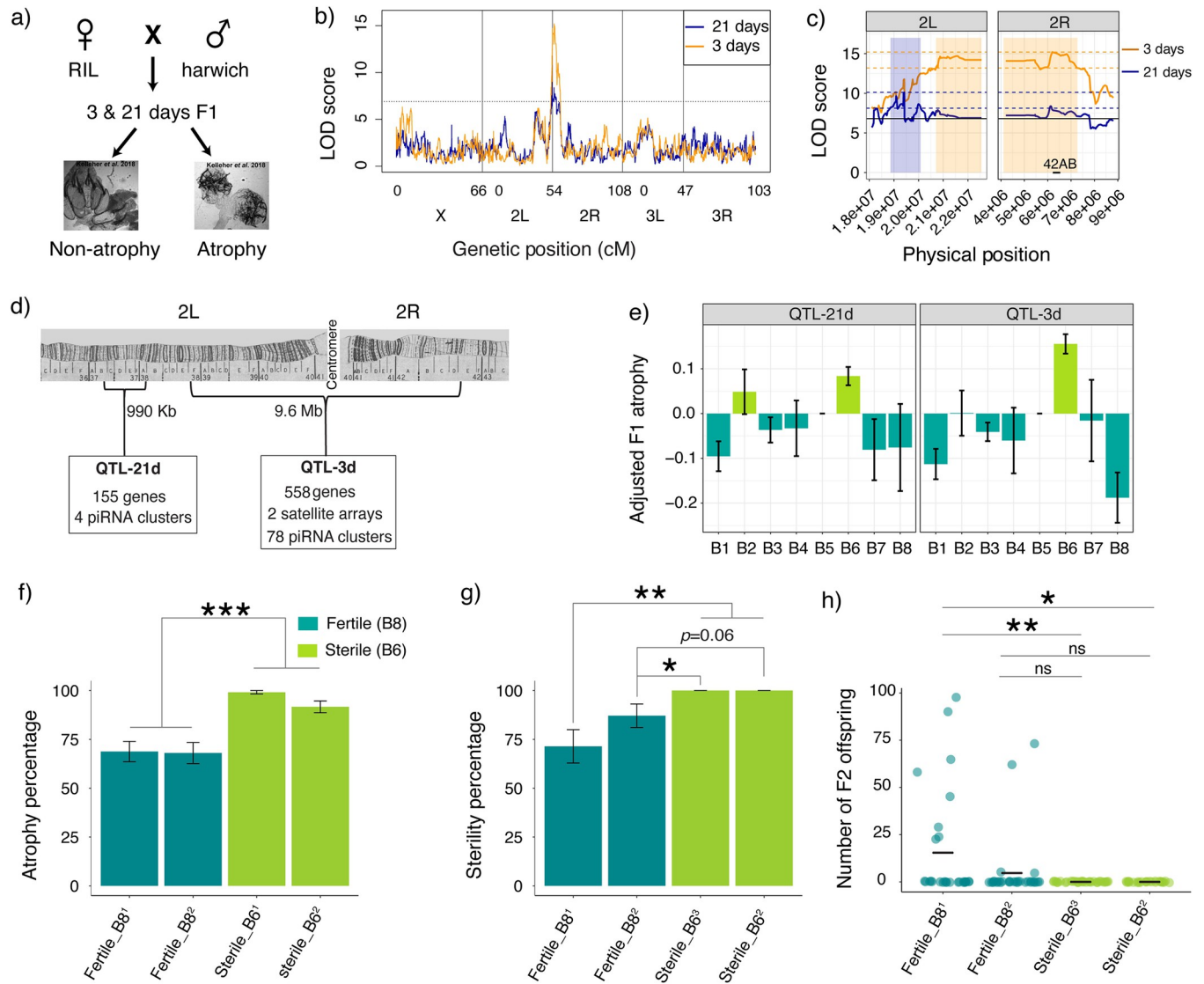


Fig 1. QTL mapping of variation in P-element induced ovarian atrophy. a) Crossing scheme to phenotype the variation in F1 ovarian atrophy among RIL offspring. Representative images of atrophied and non-atrophied ovaries are from Kelleher *et al.* [14] b) The log of odds (LOD) plot for QTL mapping of ovarian atrophy using 3 day-old (gold) and 21 day-old (blue) F1 females. The dotted line is the LOD threshold and x-axis represents the chromosomal positions. c) Zoomed-in figure of QTL mapping from 3 days (gold) and 21 days (blue). The colored boxes show the genomic interval that likely contains the causative genetic variant of each QTL, based on a $\Delta 2$ LOD drop from the peak position [26]. The pairs of dotted lines indicate the peak $\Delta 2$ LOD scores that determines the interval. The solid horizontal line is the LOD significance threshold based on 1,000 permutations of the phenotype data. d) Cytological map depicting the interval of the two QTL peaks [27,28]. e) Residual F1 atrophy (y-axis) associated with each of the eight founder alleles (x-axis) at the QTL peaks after accounting for random effects. All the QTL peaks show 2 phenotypic classes: sterile (light green) and fertile (dark green). (f-g) Percentage of (f) ovarian atrophy and (g) sterility among dysgenic female offspring from crosses between Harwich males and isogenic females carrying sterile (B6) and fertile (B8) alleles. Proportions were compared between samples using χ^2 tests of independence. h) Number of F2 offspring produced by individual dysgenic F1 females from crosses between Harwich males and isogenic fertile and sterile females. The horizontal line indicates the mean, which was compared between samples using permutation tests. Superscripts 1,2 and 3 in (f-h) denote isogenic lines that were independently generated, these sometimes differ between experiments because sterile_B61 became contaminated and was replaced with sterile_B63. Error bars in e, f and g represent the standard error. The data used to generate plot in panels b,c, and e are provided in S3 and S4 Tables and that used for plot in panels f, g and h are provided in S17 and S18 Tables respectively. * denotes $P < 0.05$, ** denotes $P < 0.01$, *** denotes $P < 0.001$.

<https://doi.org/10.1371/journal.pgen.1010080.g001>

Mb, Fig 1D). Therefore, there are likely at least two polymorphisms that influence tolerance near the 2nd chromosome centromere, one of which has a larger effect in young 3-day old females, with the other having a larger effect in 21 day-old females.

Table 1. QTL positions in 3 and 21-day old females. The peak position, $\Delta 2$ LOD drop confidence interval (2LOD CI), and the Bayesian Credible Interval (BCI) in dm6 [29] are provided for each analysis. The data used to identify the LOD peaks and intervals for 3 and 21-day old females can be found in S3 and S4 Tables, respectively.

Analysis	LOD Score	Peak Position	2LOD CI	BCI	% variation
3-day	15.2	2R:6,192,495	2L:20,710,000-2R:7,272,495	2L:20,820,000-2R:6,942,495	11.13
21-day	10.13	2L:19,420,000	2L:19,170,000-20,080,000	2L:19,010,000-20,000,000	9.78

<https://doi.org/10.1371/journal.pgen.1010080.t001>

We further evaluated the effect of the two linked QTL through haplotype analysis. We modeled residual F1 ovarian atrophy as a function of QTL haplotype for the 3 day and 21 day peaks, thereby disentangling synergistic (e.g. sterile 3d, sterile 21d) from opposing (e.g. sterile 3d, fertile 21d) allelic combinations (S4 Fig). We observed that the 3 day-old QTL allele is solely-determinant of ovarian atrophy in the 3 day-old offspring. However, in 21 day-old offspring only the genotypes containing fertile alleles at both QTL show decreased atrophy. This suggests that QTL-3d may determine germ cell maintenance in the larval, pupal and early adult stages, but QTL-21d may be additionally required to maintain germline cells in aging females. The presence of two QTL is further supported by the phenotypic classes we detected among founder alleles (B1-B8) for each of the QTL peaks (Fig 1E). For QTL-21d, both B2 and B6 founder alleles are associated with greatly increased dysgenic ovarian atrophy. By contrast for QTL-3d, only the B6 founder allele is associated with increased ovarian atrophy.

We next sought to determine whether reduced ovarian atrophy corresponds to restored fertility, or merely allows for the production of inviable gametes. To this end, we generated isogenic lines that carry either high-atrophy (B6) or low-atrophy (B8) alleles at both QTL loci in an otherwise identical genetic background through 6 generations of backcrossing to a marker stock (S5 Fig). Consistent with our QTL mapping, B8 alleles display less F1 ovarian atrophy (24–31%) than B6 strains when crossed with Harwich males (Fig 1F and S17 Table). Furthermore, while B6 dysgenic females produced no offspring, 13–29% of dysgenic B8 females were fertile and produced offspring (Fig 1G and S18 Table). For one B8 stock, offspring counts were significantly higher when compared to B6 (Fig 1H). In light of these observations, we refer to the low-atrophy and high-atrophy alleles hereafter as “fertile” and “sterile”. The fertility rescue conferred by these alleles would be highly beneficial in populations where dysgenic crosses are common.

Sterile and fertile alleles differ in chromatin regulation

Both the QTL regions contain large numbers of protein coding and non-coding RNA genes, piRNA clusters, and repeats, which could influence dysgenic sterility (Fig 1D). To better understand the differences between fertile and sterile genotypes, we compared their gene expression profiles in the ovaries of young 3–5 day-old females by stranded total RNA-seq. To avoid the confounding effects of germline loss under dysgenic conditions, we focused on RIL females rather than their dysgenic offspring. To account for potential background effects, we examined three pairs of RILs that carried either a sterile (B6) or fertile (B4) QTL haplotype across the QTL region (dm6 2L:19,010,000-2R:7,272,495) in otherwise similar genetic backgrounds (shared 44–47% of founder alleles outside the QTL). Please note that these lines differ from the B6 and B8 isogenic stocks we utilize in Fig 1 and later in the manuscript, which are more closely matched for genetic background. Principal component analysis (PCA) of read counts reveals two independent axes that resolve sterile and fertile gene expression profiles, which together account for 40% and 16% of variation (Fig 2A and S14 Table). One biological replicate of RIL 21188 (fertile) was an outlier, which we excluded from our downstream analysis of differentially expressed genes.

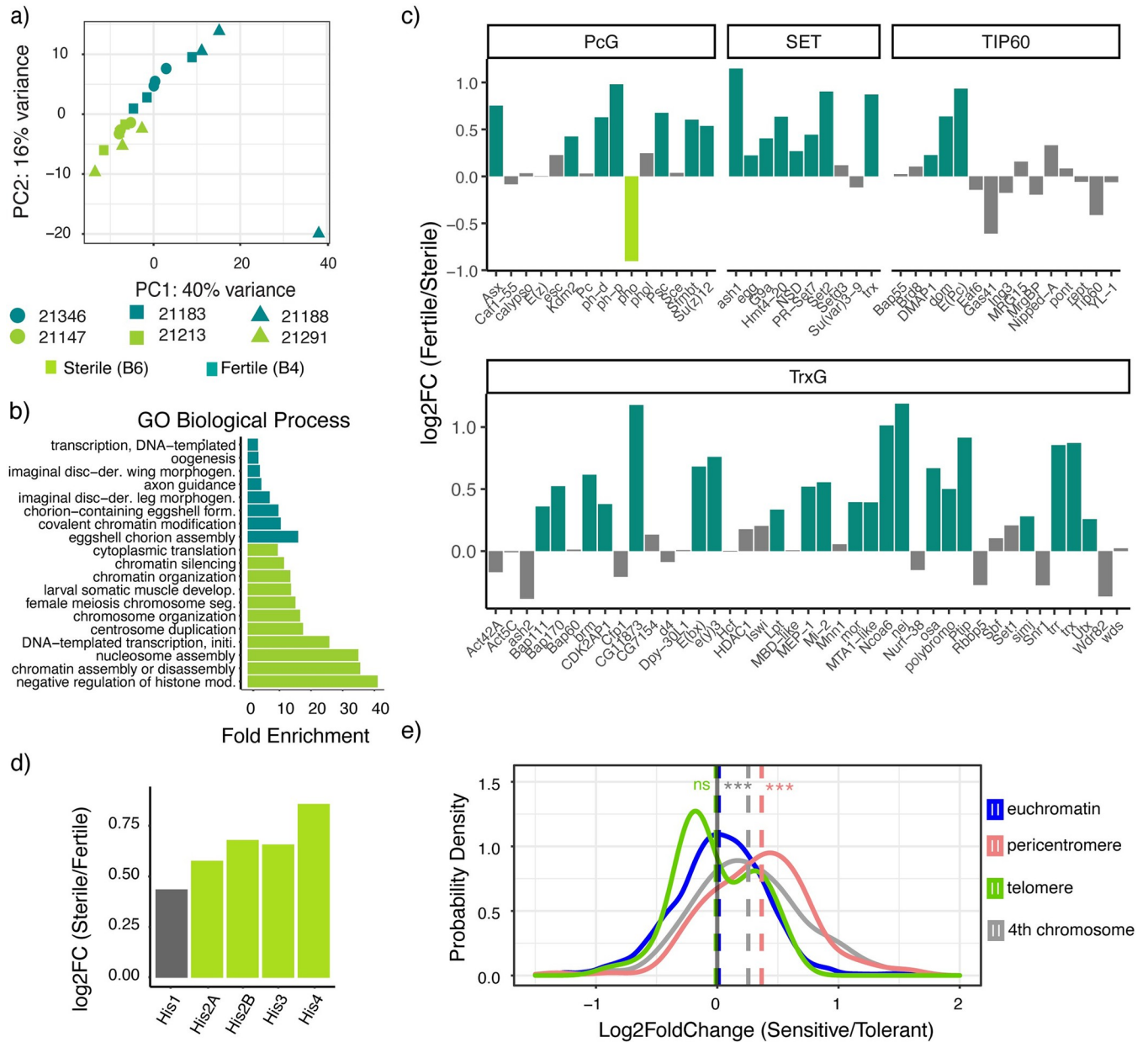


Fig 2. Fertility is associated with increased chromatin modification, whereas sterility is associated with increased expression of replication-dependent histones. **a)** PCA analysis of gene expression data for pairs of sterile (B6) and fertile (B4) RILs, that carry founder B6 and B4 haplotypes across the QTL window. Members of the same RIL pair with otherwise similar genetic backgrounds are represented by the same shape. **b)** GO terms enriched among genes upregulated in fertile and sterile genotypes. **c)** Log₂ fold differences in expression for chromatin modifiers between sterile and fertile genotypes. **d)** Log₂ fold increase in RD histone expression in sterile genotypes. **e)** Probability density plot of log₂ fold change values for all euchromatic (blue), pericentromeric (red), telomeric (green) genes and 4th chromosome (gray) between strains carrying sterile and fertile. The mean of each distribution is represented by a dotted line, and is compared between distributions with a two-sample *t*-test. The x-axis boundaries were confined from (-1.5 to 2) for a better visualization. The pericentromere-euchromatin boundaries were drawn from [29,45] and subtelomeric-euchromatin boundary coordinates from [46–48]. The data represented in panel a is provided in S14 Table and plot in panels c, d, and e in S5 Table). *** denotes $P < 0.001$.

<https://doi.org/10.1371/journal.pgen.1010080.g002>

We found a total of 530 genes differentially expressed between sterile and fertile genotypes (Benjamini-Hochberg adjusted p -value ≤ 0.05 , fold-change > 1.5 ; S5 Table). The most significantly enriched gene ontology (GO) term among genes upregulated in fertile ovaries is chomion assembly (Bonferroni corrected P value < 0.01 , Fig 2B and S7 Table). This suggests a

larger number of late-stage oocytes in fertile ovaries, as other genes that are upregulated in late oogenesis (stages 12–14) show a similar increase in expression (S6 Fig, [30]). Because atrophy results from the loss of larval PGCs and pre-meiotic adult cysts (GSCs), larger numbers of late-stage oocytes are likely unrelated to dysgenic sterility [15–17,19].

The second most significant GO term upregulated in the ovaries of fertile genotypes is covalent chromatin modification (Fig 2B). Strikingly, we discovered fertile ovaries exhibit a systematic upregulation of multiple chromatin modification complexes with key roles in oogenesis, including polycomb group, trithorax group, and the TIP60 complex, as well as many individual SET domain lysine methyltransferases (Fig 2C) [31–33]. The TIP60 complex in particular is involved in cell cycle progression and differentiation in pre-cystoblasts [32]: daughter cells of germline stem cells in which *P*-elements transpose [19,21]. Interestingly, the TIP60 complex is also involved in DSB repair [34], which could promote dysgenic germ cell survival. Similarly, polycomb-dependent gene silencing initiates in nurse cells concurrently with meiosis I: a window in which germ cell cysts with large numbers of DSBs undergo apoptosis [35].

Genes upregulated in the sterile genotypes are enriched for functions in chromatin assembly and transcription, cell division, and translation. However, a careful inspection of genes underlying these enriched terms reveals that, with the exception of translation, they are primarily explained by the increased expression of replication-dependent (RD) histone gene copies (Fig 2D). While these expression increases are modest (<2 fold), they are likely an underestimate of the true degree of histone upregulation. Histone gene expression increases dramatically in late-stage oocytes (beyond stage 10 [36]), which appear to be reduced in the ovaries of sterile genotypes (S6 Fig). Overexpression of RD histones is associated with increased sensitivity to DNA damage [37–41], and excess histones are reported to compete with DNA repair proteins for binding to damage sites [38]. Thus, while TIP60 activity might increase DSB repair in fertile ovaries promoting germ cell survival, histones might decrease repair in sterile ovaries increasing germ cell death.

Differences in ovarian chromatin modification between fertile and sterile genotypes may further be connected to histone regulation. Replication-dependent histones occur in a tandemly duplicated gene cluster that exhibits coordinated and dosage-compensated regulation in a specialized nuclear compartment known as the histone locus body (HLB, [42]). In particular, negative regulation of histone expression relies on multiple heterochromatin factors [41,43]. Consistent with reduced heterochromatin formation in the ovaries of sterile genotypes, they show higher expression of pericentromeric genes (two-sample t-test, $t_{141} = -9.32$, p -value = 2.3×10^{-16}), as well as genes on the heterochromatic 4th chromosome (two-sample t-test, $t_{53} = -4.56$, p -value = 3.0×10^{-5} , Fig 2E). The sterile B6 haplotype also exhibits increased expression of pericentromeric genes in a previously published microarray dataset from head tissue ([44] S1 Fig).

Sterile genotypes exhibit silenced piRNA loci, but no systematic TE dysregulation

In addition to gene expression, differences in the regulation of resident TEs could modulate the degree of dysgenic sterility. The *D. melanogaster* genome harbors >100 resident TE families [49,50], many of which are transpositionally active and show variable transposition rates in wild-type strains [51–53]. If sterile alleles establish weaker regulation of some resident TEs, their transposition could add to DNA damage resulting from *P*-element transposition, thereby promoting germ cell loss. Resident TEs are regulated by piRNAs, and two features of our data suggest differences in piRNA biogenesis between sterile and fertile alleles. First, QTL-3d

contains numerous piRNA clusters, including the major ovarian piRNA cluster *42AB*, which could differ in the regulation or resident TEs between sterile and fertile alleles (Fig 1D). Second, differences in chromatin regulation between sterile and fertile alleles could impact piRNA cluster expression (Fig 2E and 2F), which is dependent upon the heterochromatic histone modification, histone 3 lysine 9 trimethylation (H3K9me3) [54,55].

To look for differences in resident TE regulation, we performed small RNA-seq on the same ovarian samples from RIL females (mothers) that we used for total RNA-seq, and quantified the expression of piRNAs from clusters throughout the genome. A PCA of piRNA cluster expression reveals that sterile and fertile genotypes differ in the ovarian expression of some piRNA clusters, and are resolved by the second principal component, accounting for 22% variation in expression (Fig 3A and S15 Table). In particular, we discovered two small pericentromeric piRNA clusters located within QTL-3d that were active in fertile genotype ovaries but largely quiescent in sterile genotype ovaries (Figs 3B, 3C, 3D, S2, and S3 and S16 Table). However, the major piRNA clusters—including *42AB*—do not differ in ovarian expression in between sterile and fertile alleles, suggesting that the proposed reduction in heterochromatin formation in sterile genotypes does not majorly reduce piRNA transcription (Fig 3B and S8 Table). Furthermore, the differentially active clusters in QTL-3d seem unlikely to regulate transpositionally active resident TE families, since they are largely composed of TE fragments that are relatively divergent from the consensus (65 to 95% sequence similarity; Figs 3C, 3D, S2, and S3 and S9 Table), or are most similar to a consensus TE from other (non-*melanogaster*) *Drosophila* species. Transpositionally active TEs are generally highly similar to the consensus sequence [56], and piRNA silencing is disrupted by mismatches between the piRNA and its target [57]. Nevertheless, we cannot rule out the possibility that sterile or fertile genotypes could harbor TE insertions in this locus that are not represented in the *dm6* reference genome.

To directly address if fertile and sterile genotypes differ in resident TE regulation, we compared their genome-wide resident TE expression in our RNA-seq data. None of the TE families represented in the QTL-3d piRNA clusters were upregulated in sterile genotypes (Fig 3E and S10 Table). Furthermore, while some TE families are differentially expressed, there is no systematic increase in TE activity in the sterile genotypes. Rather, more TE families are upregulated in the ovaries of fertile genotypes (13 TEs) when compared to sterile (4 TEs) genotypes. Upregulation of certain TEs may be consistent with more late-stage egg-chambers in fertile genotype ovaries, since many retrotransposons accumulate transcripts in the oocyte over the course of oogenesis [58–60]. Therefore, despite the conspicuous position of QTL-3d surrounding piRNA producing-regions, as well as evidence for differential chromatin regulation that could impact piRNA biogenesis (Fig 3B and 3E), we find no evidence that fertility in dysgenic crosses is determined by resident TE silencing.

Sterile alleles increase *P*-element mRNA expression and transposase mRNA splicing

Increased dysgenic sterility associated with sterile alleles could also reflect increased *P*-element transposition, resulting from increased *P*-element mRNA expression or splicing. In particular only transcripts in which the third intron (intervening sequence, IVS3) is spliced will produce *P*-transposase, and regulation of IVS3 splicing is a key determinant of *P*-element transposition [22]. We therefore examined the abundance of different *P*-element transcripts in the ovaries of F1 dysgenic offspring of sterile (B6) and fertile (B8) isogenic females. Dysgenic offspring in these experiments were reared at 22°C to avoid germline loss [20].

Consistent with differential production of *P*-transposase, we observed differences in overall abundance of *P*-element transcripts between the F1 dysgenic offspring of sterile and fertile

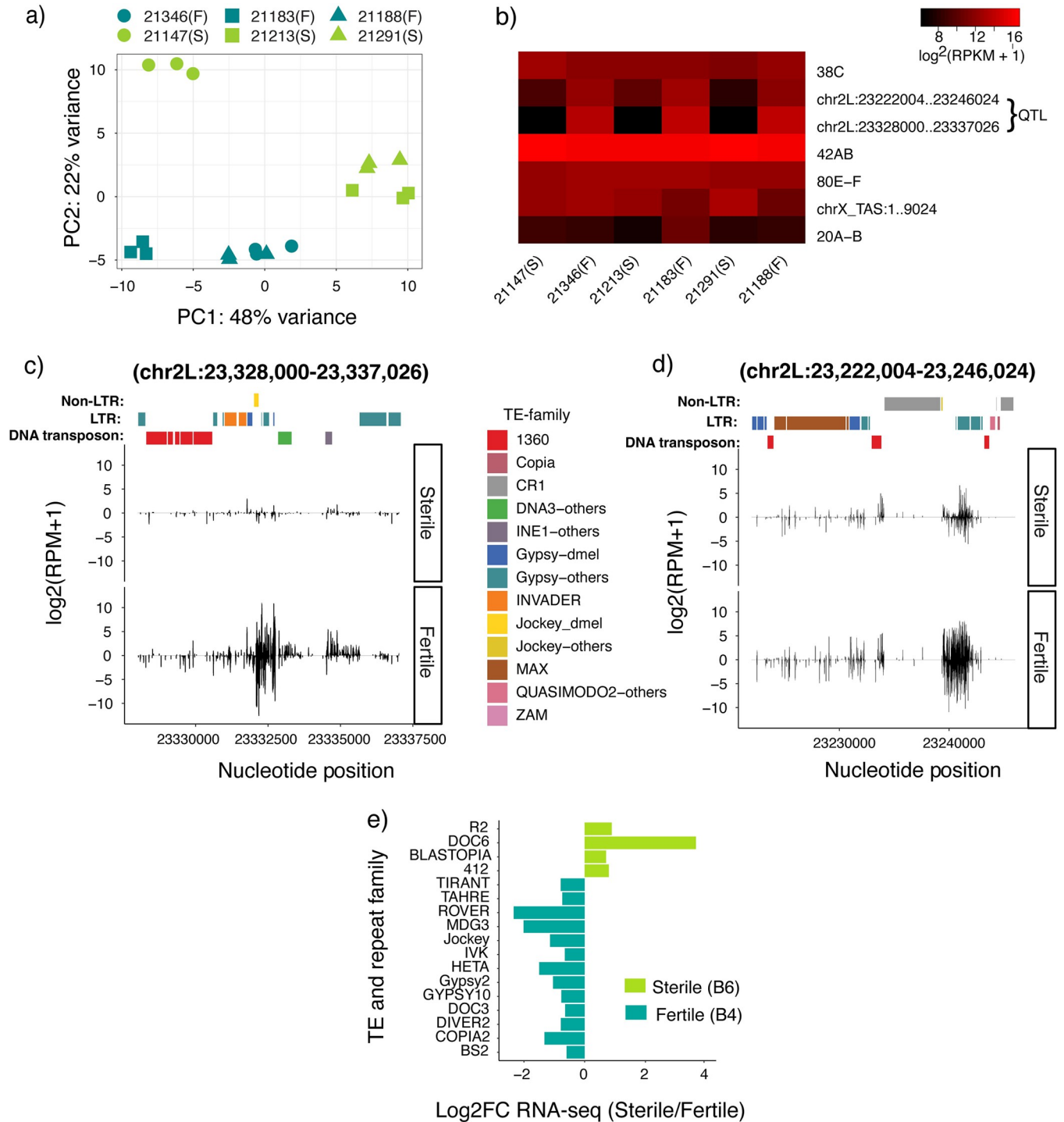


Fig 3. Dysgenic sterility is not related to differential activity of piRNA clusters or TE deregulation. **a)** PCA analysis for piRNA cluster expression data of sterile (S) and fertile (F) genotypes. Members of the same RIL pair are represented by the same shapes. **b)** Heat map showing the expression of seven major piRNA clusters [61] and the two differentially expressed QTL clusters in QTL-3d. RIL pairs are plotted adjacent to each other. **c and d)** Uniquely mapping piRNAs within two differentially active QTL-3d piRNA clusters are compared between sterile (21183) and fertile (21213) genotypes. Positive value indicates piRNAs mapped to the sense strand of the reference genome and negative value indicates those from the antisense strand. TE insertions in each cluster are presented according to family by different colors; TE-others indicate the insertion was most similar to a consensus TE from a sibling species of *D. melanogaster*. See **S2 and S3 Figs** for cluster expression in the remaining RIL pairs. For **b, c and d**, piRNA cluster expression levels are estimated by log₂ scale transformed of reads per million mapped reads [log₂(RPM+1)]. **e)** Genome-wide differences in TE family expression between sterile and fertile genotypes (fold change = 1.5, base mean >= 100, adjusted *p*-value <= 0.05), based on alignment to consensus sequences. The data used to plot panel **a** is provided in **S15 Table**, for panel **b** in **S8 Table**, for panels **c and d** in **S16 and S9 Tables**, and for panel **e** in **S10 Table**.

<https://doi.org/10.1371/journal.pgen.1010080.g003>

females (Fig 4A, $t_{10} = 13.09$, $p = 2.31 \times 10^{-15}$). On average, sterile females showed a 34% increase in *P*-element transcripts (95% CI 27–42%). Transposase-encoding (IVS3 spliced) transcripts show an even more pronounced 59% increase in expression in sterile genotypes (Fig 4A, $t_{10} = 10.27$, $p = 2.91 \times 10^{-10}$, 95% CI: 42–81%). By contrast, unspliced (IVS3 retaining) transcripts were not significantly differentially expressed between ovaries of F1 dysgenic offspring of sterile (B6) and fertile (B8) isogenic females (Fig 4A, $t_{10} = 1.68$, $p = 0.11$), although certain individual comparisons between strains were significant. To directly address whether splicing is more efficient in the ovaries of sterile dysgenic offspring, we compared the ratio of spliced to unspliced *P*-element transcripts (Fig 4B). The ratio of spliced to unspliced transcripts differed significantly between the ovaries sterile and fertile dysgenic offspring ($t_{10} = 7.45$, $p = 7.30 \times 10^{-5}$), suggesting that splicing itself is more efficient in sterile genotypes.

We also observed differences in splicing and expression between isogenic stocks carrying the same allele. In particular, the Fertile_B81 and Fertile_B82 differed in both spliced and unspliced transcripts, with a particularly pronounced 2.89-fold increase in spliced transcript expression in Fertile_B82 as compared with Fertile_B81 (Fig 4A, $t_{10} = 12.14$, $p = 9.78 \times 10^{-12}$, 95% CI: 2.61–3.17 fold). While the sample size is too small to draw any conclusions, it is notable that Fertile_B82 does not exhibit the same degree of fertility rescue as Fertile_B81, further pointing to a connection between spliced transcript production and dysgenic sterility (Fig 1G and 1H).

In germline cells, the splicing of IVS3 is known to be repressed by piRNA mediated transcriptional silencing, which is initiated by maternally transmitted piRNAs or through the production of *de novo* piRNAs in aged females [16,21]. In addition to the absence of maternally transmitted piRNAs, the splicing differences we observe here are in young (3–4 day old) dysgenic females, as opposed to aged dysgenic females, suggesting they are likely independent of the piRNA pathway. However, the splicing of IVS3 is also repressed by several host splicing factors in somatic cells, and it is proposed that some of these factors may also partially repress splicing in germline cells [62–64]. Consistent with piRNA-independent differences in splicing, we discovered that three splicing factors known to promote IVS3 retention in somatic cells, *hrp36*, *hrp38* and *P*-element somatic inhibitor (Psi) show increased expression in fertile genotypes in our ovarian RNA seq data (Fig 4C). This suggests that decreased splicing in the dysgenic offspring of fertile isogenic lines may result from increased abundance of host splicing factors.

Sterile alleles increase radiation sensitivity and accumulated mutations

Our gene expression data suggest that sterile and fertile alleles may differ in their capacity to repair germline DSBs in young (3 day) dysgenic females. Fertile alleles exhibit upregulation of the TIP60 complex (which is involved in DSB repair [34]), while sterile alleles exhibit upregulation of replication dependent histones (which may compete with DNA repair machinery [38]). Mutations in DSB repair genes are widely known to cause radiation sensitivity, which is easily quantified by measuring lethality following larval radiation exposure [65–69]. While this assay occurs in whole larvae as compared to female germlines, larvae are composed of numerous classes of mitotically dividing cells, similar to the primordial and premeiotic stages of gamete production in which *P*-element transposition occurs [16,21]. Furthermore, numerous key factors for germline DNA damage response, as well as germline *P*-element excision repair, exhibit larval radiation sensitivity phenotypes [70–75].

We therefore compared the X-ray radiation sensitivity of larvae from isogenic lines containing sterile (B6) and fertile (B8) alleles. Note these lines are the same as those in Fig 1F–1H). After exploring a range of radiation doses, we found that doses above 10 Gy showed high

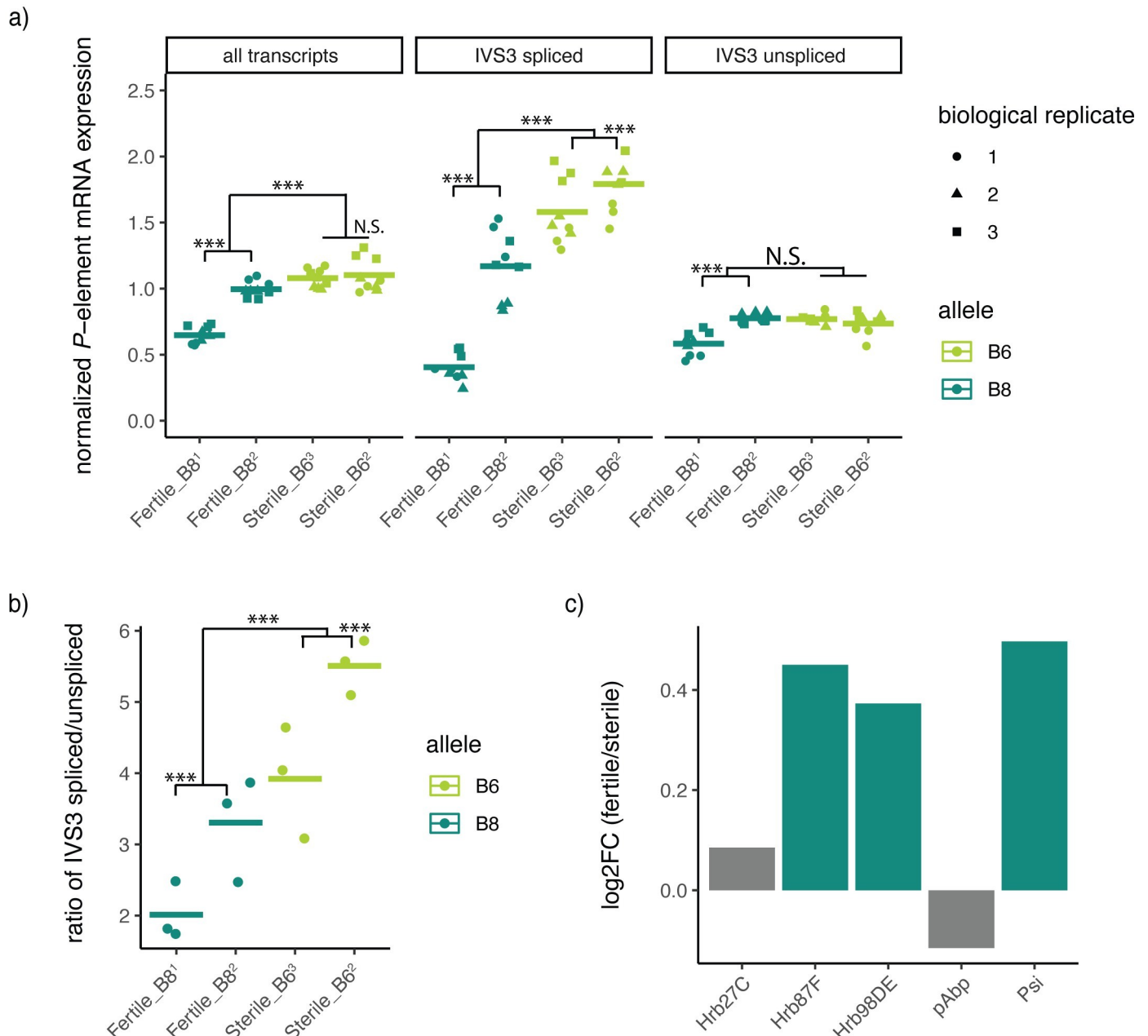


Fig 4. Decreased expression of P-transposase in fertile genotypes. A) Differential expression of P-element transcripts between fertile and sterile genotypes. Three separate qPCRs were performed, which detect all transcript isoforms, as well as IVS3 spliced and unspliced isoforms. qPCRs are normalized to rpl32. B) Ratios of IVS3 spliced to unspliced isoforms. C) Differential expression of splicing factors between sterile and fertile genotypes based on RNA-seq data. Dark green bars indicate factors that are significantly upregulated in fertile genetic backgrounds. Significant differences in qPCR data are based on linear models to detect differences between sterile and fertile genotypes, or Tukey-HSD comparisons to detect differences between genotypes containing the same allele. *** denotes $P < 0.001$.

<https://doi.org/10.1371/journal.pgen.1010080.g004>

lethality, making it difficult to detect differences in radiation sensitivity between the genotypes (S19 Table). Therefore, we compared the response of sterile and fertile larvae to radiation doses of 0 Gy, 5 Gy and 10 Gy. We observed that fertile genotypes had significantly higher survival (53–58%) than the sterile genotypes (25–30%) at 10 Gy (Fig 5A). Given that X-ray radiation produces predominantly DSBs, these results are consistent with differences between fertile and sterile alleles in DSB repair.

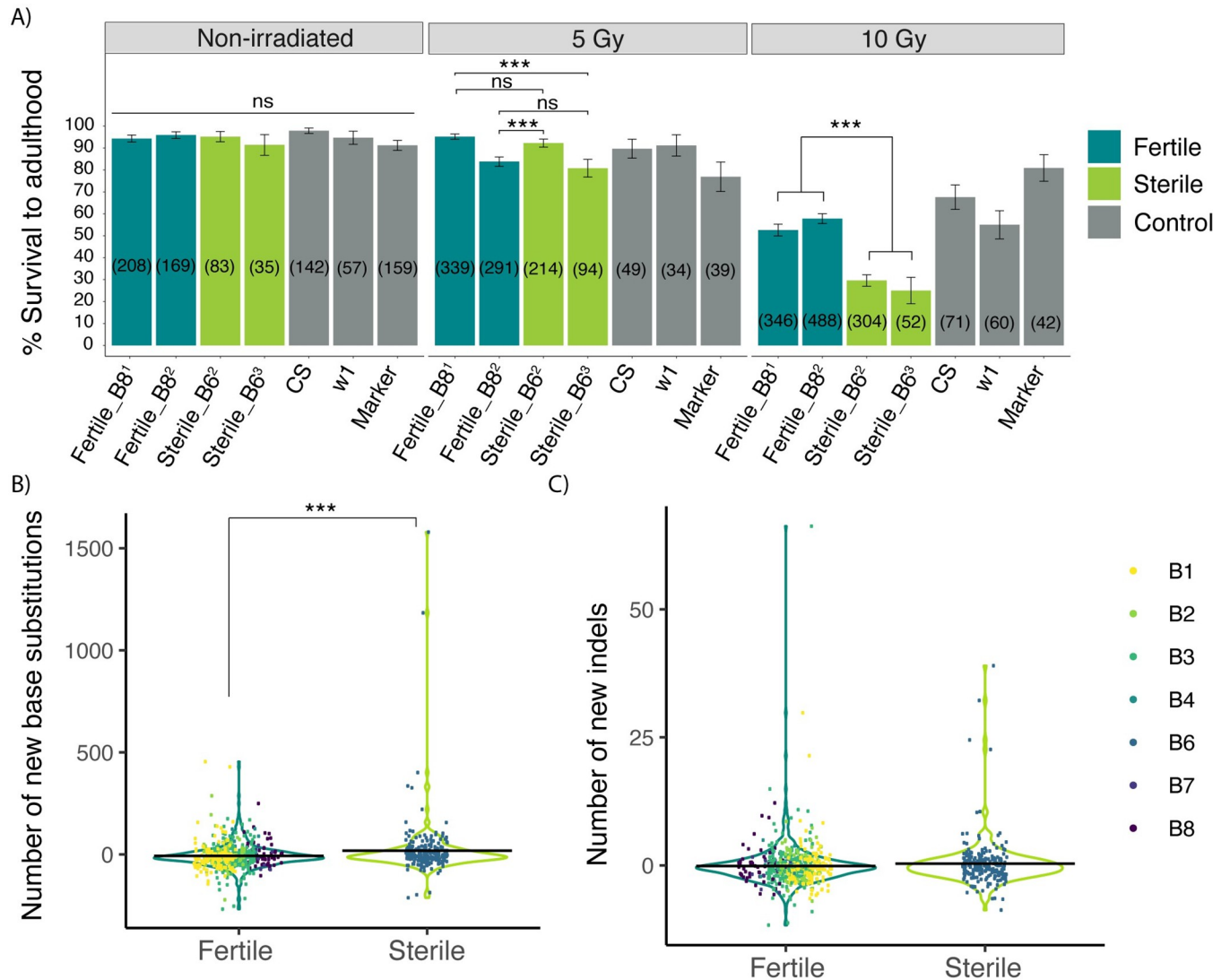


Fig 5. Sterile alleles exhibit reduced DNA repair. A) The percentage of mock treated and irradiated (5 Gy and 10 Gy) larvae that survived to adulthood for the fertile (B8), sterile (B6) and the control genotypes. CS refers to Canton-S and marker refers to the multiply marked stock *b cn* (#44229), which was used to generate isogenic lines. The numbers in the brackets refer to the sample size. The significance of comparisons between genotypes was determined by the χ^2 test-of-independence. B and C) New mutations that accumulate in RIL genomes as detected by MSG. B) New SNPs and C) new indels. An excess of new mutations was detected by a t-test comparing Sterile B6 RILs to all others. The data represented in the Fig is provided in [S19 and S20 Tables](#)) ** denotes $P < 0.01$, *** denotes $P < 0.001$.

<https://doi.org/10.1371/journal.pgen.1010080.g005>

We further looked specifically for differences in germline DSB repair by examining whether RILs carrying B6 alleles at the QTL-3d peak have accumulated more *de novo* base substitutions and small insertions or deletions. The DSPR RILs underwent 50 generations of inbreeding, and have since been maintained as isogenic lab stocks for ~175 generations, allowing ample time for new mutations to accrue as a consequence of deficient repair. To detect these new mutations, we generated multiplexed shotgun genotyping (MSG) libraries for 792 population B RILs [76]. This low coverage method (mean 2.9x) will uncover only a random subset of new mutations in each RIL, thereby underestimating the true amount of mutation accumulation. Nevertheless, we were able to detect 102,476 novel base substitutions and 5,026 novel insertions or deletions among the RIL MSG libraries.

After accounting for differences in sequencing depth and plate effects, the founder allele at QTL-3d was associated with differences in the number of new base substitutions ($-2\Delta\ln L = 15.62$, $df = 6$, $p = 0.016$). Furthermore, RILs carrying B6 alleles at the QTL-3d peak exhibit an average increase of 14.58 new base substitutions (95% CI 5.23–24.12), when compared to those carrying another founder allele ($t_{782} = 3.043$, $P = 0.0024$, **Fig 5B**). In contrast, there was no significant association between founder allele at QTL-3d and new indels ($-2\Delta\ln L = 1.37$, $df = 6$, $p = 0.97$). Given the low coverage data as well the limited potential of short-read sequencing data to identify larger structural variation [77], we cannot be conclusive about a relationship between QTL allele and indel accumulation rate. Nevertheless, the increase in base substitution supports a deficiency in germline DNA repair in association with B6 alleles for QTL-3d.

Identifying candidate genes

The QTL we map here are quite large and contain numerous candidate genes whose differential function could influence dysgenic sterility. Nevertheless, we next sought to identify candidate genes that influence dysgenic sterility for future study. We combined our own expression and mapping data with previously published polymorphism and single cell expression data to narrow candidates based on four criteria: 1) location within a QTL, 2) expression in primordial germ cells or early, pre-meiotic cysts [78,79], 3) differential expression between sterile and fertile adult ovaries, and 4) the presence of “in-phase” single nucleotide polymorphisms (SNPs) (**S11, S12 and S13 Tables**). In-phase SNPs are those where the genotypic differences between the founder alleles are consistent with their phenotype class (**Figs 1E and 6A [80]**). Of 530 differentially expressed genes, 43 are within the QTL region, representing an approximately five-fold enrichment in the QTL regions compared to the rest of the genome ($\chi^2 = 255.54$, $df = 1$, $p\text{-value} < 2.2e-16$, **Fig 6B**). Ultimately, we identified 12 and 5 differentially expressed genes and early germ cell expressed genes that also carry in-phase SNPs within the QTL-3d and 21d, respectively (**Fig 6C and 6D and S12 Table**). Furthermore, we identified 32 genes in QTL-3d and 3 genes in QTL-21d that exhibit early germ cell expression and also contain in-phase non-synonymous SNPs, which may affect the function of the encoded protein (**S13 Table**). Collectively these genes represent the strongest candidates to contain causative variants.

We next scoured our list of candidate genes for those with known functions in chromatin regulation, DSB repair, or alternative splicing, whose differential function or regulation are plausibly related to the phenotypic differences associated with sterile and fertile alleles. None of the three splicing factors we discovered are differentially expressed reside within the QTL (**Fig 4C**), suggesting their expression differences arise as a consequence of regulatory differences in *trans*. While we did not discover any transcription factors located in the QTL that are differentially expressed in fertile and sterile ovaries, we did discover three C2H2 zinc finger transcription factors, *tio* and *CG30431* (QTL-3d) and *CG17568* (QTL-21d), that are located within QTL and contain in-phase non-synonymous SNPs. Unfortunately, the genomic binding sites of these transcription factors are undetermined, so it remains unknown if they are regulators of *hrp36*, *hrp38*, or *psi* transcription.

With respect to differences in chromatin state and/or DNA repair, two genes within QTL-3d, stand out as particularly attractive candidates; *Nipped-A* and *jing*. *Nipped-A* contains a non-synonymous in-phase SNP and is expressed in both PGCs and in germline cells throughout the earliest stages of oogenesis (**S13 Table**). *Nipped-A* is a member of the TIP60 complex, which has functions in DSB repair, chromatin modification and chromatin remodeling. Additionally, we identified multiple TIP60 components upregulated in fertile ovaries (**Fig 2C**). The non-synonymous SNP that separates sterile and fertile alleles of this gene are located in the

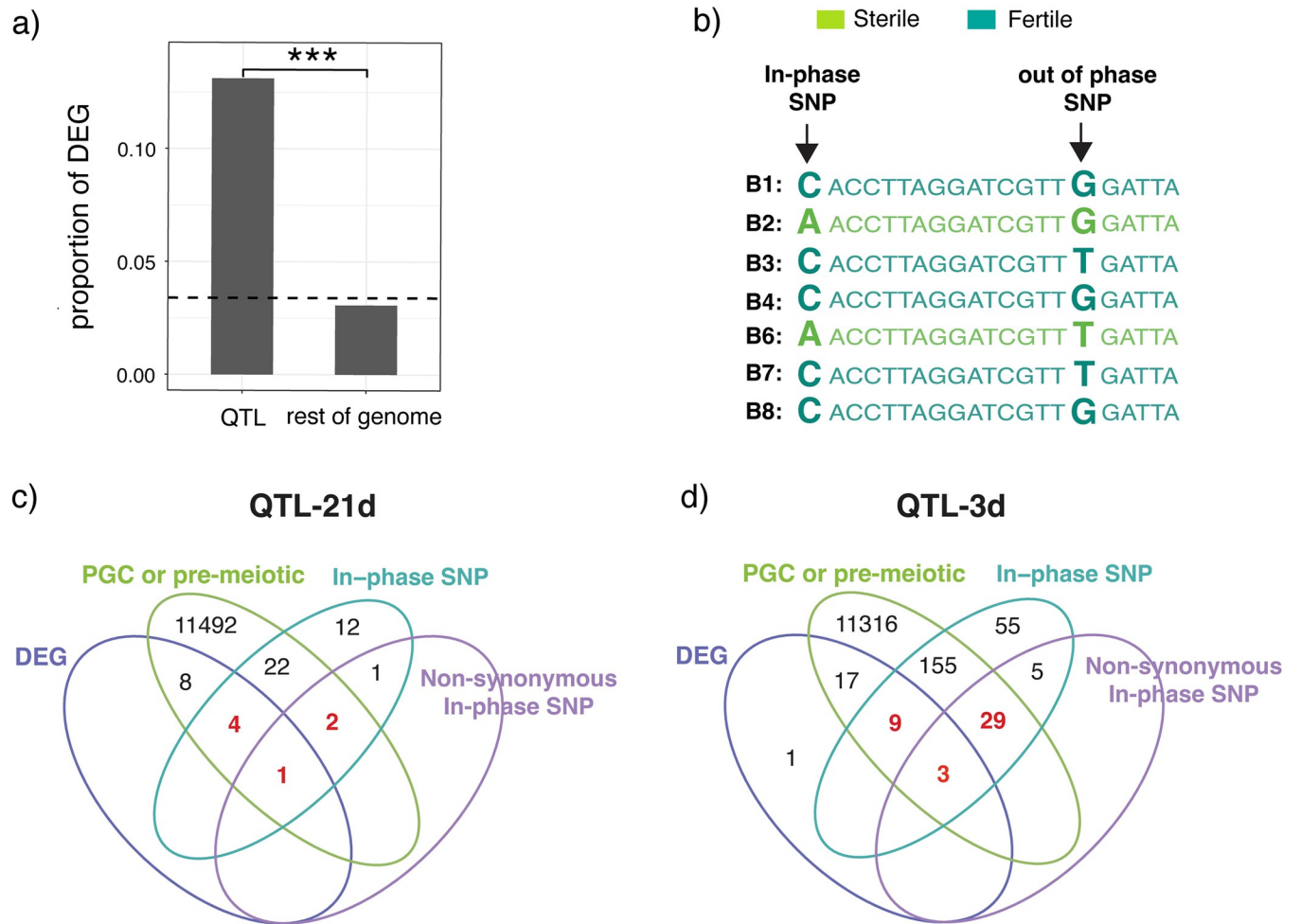


Fig 6. Differential expression and in-phase SNPs identify candidate genes. a) The proportion of genes differentially expressed (DEG) is compared inside and outside the QTL. The dotted line is the genome wide average. b) Hypothetical in-phase and out of phase SNPs are shown. Sequences of each of the B founder strains are colored based on their phenotypic classification, either fertile or sterile (Fig 1E). Bold letters indicate SNPs. c and d) Venn diagrams showing the overlap of differentially expressed genes (DEG), genes carrying in-phase synonymous and non-synonymous SNPs, and genes expressed in primordial or pre-meiotic germ cells for QTL-21d (c) and QTL-3d (d). The data for differential expression of genes for fertile and sterile genotypes is provided in S5 Table. The data on in-phase polymorphisms for each QTL peak are provided in S11 Table. List of candidate genes that have both in-phase polymorphisms and are differentially expressed, and those having non-synonymous in-phase polymorphisms are provided in S12 and S13 Tables, respectively.

<https://doi.org/10.1371/journal.pgen.1010080.g006>

HEAT2 domain, which is predicted to be essential for protein-protein interaction, and could have important implications for the function of this multi-protein complex [81–83]. *Jing* contains in-phase synonymous and non-synonymous SNPs, is upregulated in fertile ovaries, and exhibits a similar expression pattern in germline cells to that of *Nipped-A* (Fig 2E and S12 and S13 Tables). Based on a yeast-two hybrid screen *Jing* physically interacts with inverted repeat binding protein 18 (IRBP18): a DNA binding protein that comprises part of a heterodimer that binds directly to *P*-element’s transcribed inverted repeats, and facilitates repair of donor DNA after excision [84,85]. Furthermore, *irbp18* mutants exhibit larval radiation sensitivity, similar to our sterile genotypes (Fig 5A). If *Jing* determines differential activity of IRBP18 it could have a strong impact on dysgenic sterility. Beyond this function, *Jing* acts as an important cofactor of polycomb repressor complex 2, many of which showed increased expression in fertile ovaries Fig 2C, [86,87].

Discussion

Although small RNA mediated TE regulation is widely studied, little is known about genetic variation in host factors that modulate the germline transposition of invading TEs and their associated fertility effects. Here we uncovered natural variation in dysgenic sterility imposed by *P*-element DNA transposons. Our work points to two major differences between sterile and fertile genotypes, which likely explains the differential occurrence of dysgenic sterility between them (Fig 7). First, fertile alleles suppress the splicing of transposase-encoding mRNA, which likely reduces the occurrence of germline DSBs that drive germ cell loss. Second, fertile alleles are more tolerant of DSBs, perhaps due to enhanced repair, which may allow them to retain germ cells despite the genotoxic effects of transposition. We propose that these differences highlight two axes of host-TE interaction: permissivity and tolerance.

Host splicing factors determine differences in permissivity

As intracellular parasites, TEs rely on host machinery for transcription, translation and replication. Variation in host co-factors that modulate these processes could drive differences in

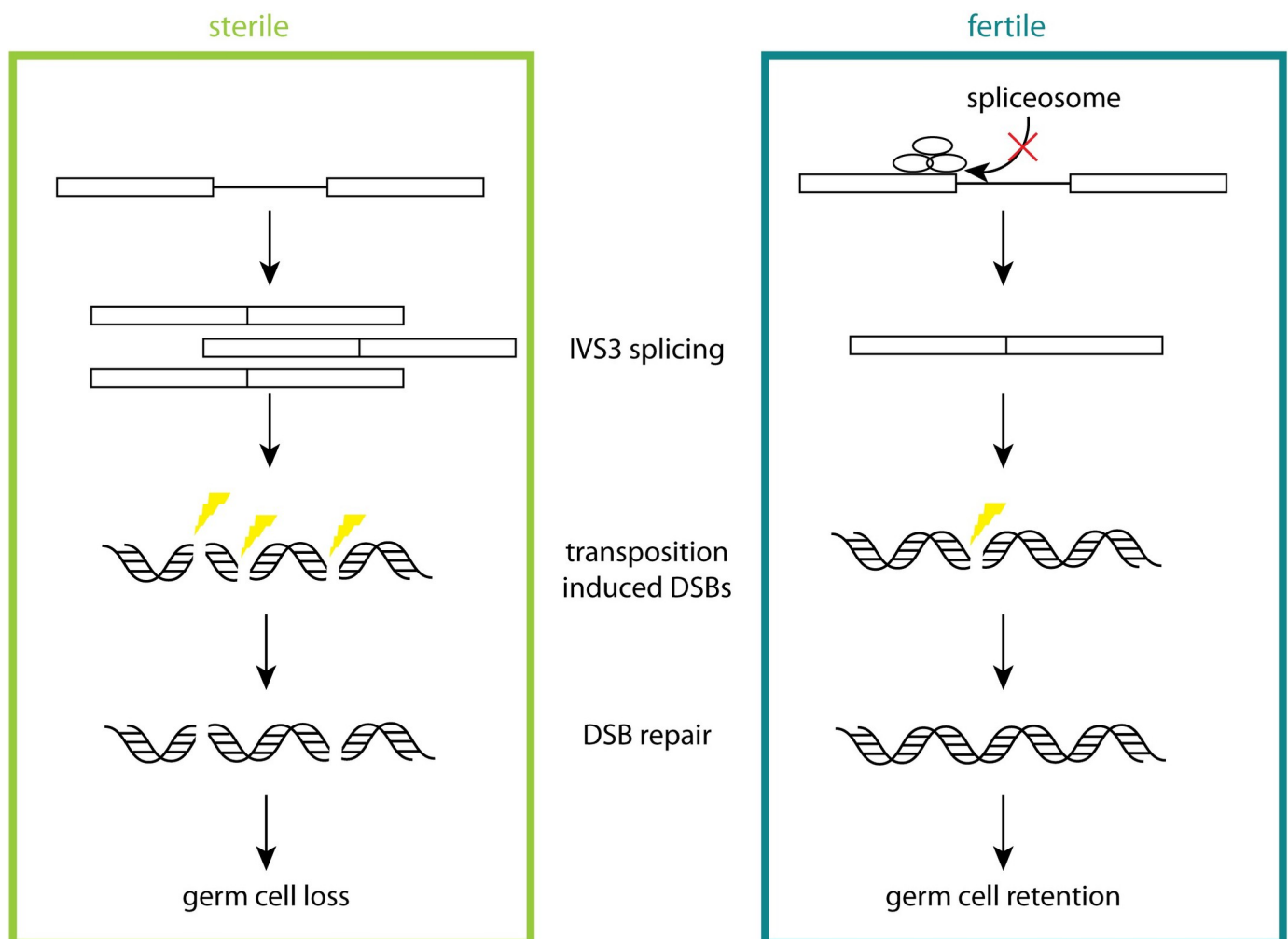


Fig 7. Schematic of phenotypic differences between sterile and fertile alleles. Differences in sterile and fertile alleles between IVS3 splicing and DNA repair are represented. In fertile alleles, host splicing suppressor expression is increased, leading to reduced production of spliced *P*-transposase encoding transcripts. As a consequence of reduced transposase production, it is predicted that fewer DSBs are produced in dysgenic females. However, it is also predicted that those breaks that are produced are repaired more efficiently.

<https://doi.org/10.1371/journal.pgen.1010080.g007>

permissivity to TE proliferation. The concept of permissivity is prevalent in virology, and refers to the degree to which an individual cell type allows a virus to replicate [88–90]. With respect to TEs, permissivity is distinguishable from repression in that host-cofactors modulating permissivity precede the introduction of a TE into the genome, and the primary function of permissivity factors is not in TE regulation. For example, it has long been known that *P*-elements do not transpose in somatic cells due to the presence of host splicing factors that prevent the correct splicing of *P*-transposase encoding mRNA [22,63,91]. While the primary function of these splicing factors is not to regulate *P*-elements, their expression renders somatic cells non-permissive.

In germline cells, maternally deposited piRNAs also regulate *P*-element transposition by promoting IVS3 retention [16]. However, even in the absence of maternally deposited piRNAs, IVS3-retaining transcripts are common, suggesting other host factors may also modulate *P*-element splicing in the germline [16,64]. Our work here reveals that there is host genetic variation in IVS3 splicing that is independent of maternally deposited piRNAs, which has potentially dramatic impacts on host fitness in dysgenic crosses (Figs 1B, 1G, 1H and 4A). While we cannot completely rule out a potential role for *de novo* piRNA production in driving these splicing differences, their occurrence in young dysgenic females and the absence of strong age effects on dysgenic sterility in our experiment points away from this explanation. Rather, the upregulation of multiple host splicing regulators in fertile genotypes suggest that the same factors that regulate IVS3 splicing in somatic cells may also modulate splicing in germline cells (Fig 4C). While we did not directly address whether these differences lead to differential transposition of *P*-transposons, these results suggest that fertile genotypes, like somatic cells, reduce permissivity through splicing regulation.

Tolerance and DSB repair. In our previous work on natural variation in dysgenic sterility, we proposed that host genotypes may differ in tolerance: the ability of germline cells to persist and divide despite the damaging effects of transposition [14]. Because hybrid dysgenesis occurs through the loss of larval primordial germ cells and adult germline stem cells [15,16], the variable expression of factors that determine stem cells maintenance and differentiation could be an important source of tolerance variation. In particular, we found that the function of *bruno*, a differentiation factor in early pre-meiotic cysts, increases dysgenic sterility [14]. Conversely, the overexpression of the stem cell factor *myc* in dysgenic larval PGCs suppresses their loss, and by association, decreases dysgenic sterility [17].

Our work here suggests that genetic variation in DNA damage response or repair provides another potential mechanism of tolerance. Since transposition results in DSBs at sites of insertion and excision, enhanced ability to detect and repair these breaks would help reduce dysgenic germ cell loss. We observed that fertile genotypes are significantly more resilient to X-ray radiation (Fig 5A), a phenotype that is widely associated with increased activity of DNA repair genes [65–69]. Indeed, the magnitude of the differences in radiation sensitivity is large, and mirrors previous comparisons between wild-type and DNA damage response mutants such as *p53* or *checkpoint kinase 1* [74,75]. We further observed that fertile genotypes exhibit fewer accumulated base substitutions (Fig 5B), suggesting heritable differences in DNA repair. While DNA damage signaling is a clear determinant of dysgenic germ cell loss [15,18,19,21], to our knowledge this is the first evidence that natural variation in DNA repair may modulate the sterility effects of transposition. However, we cannot rule out the possibility reduced DSB repair also increases permissivity, by prolonging S-phase thereby allowing more time for *P*-elements to transpose.

Something that remains puzzling about our observations regarding DNA damage and repair is that it is not intuitively obvious how deficiencies in DSB repair would lead to an accumulation of base substitutions in RILs carrying sterile alleles. However, DNA repair pathways

are interdependent, with many components impacting multiple repair processes. For example, *mei-9* is required for both nucleotide excision repair and meiotic recombination [92], and mutant alleles are known to enhance germline loss in *P*-element dysgenic males [18,93]. Beyond this, homologous repair of DSBs leads to an increased rate of base substitution, potentially due to the sensitivity of single stranded repair intermediates to other forms of DNA damage reviewed in [94]. Therefore, if the DSB repair deficiency associated with sterile alleles results in a larger number of DSBs repaired by homologous pathways, or a delay in those repairs, an increase in base substitution is predicted.

Conclusion

The degree to which innate differences among hosts govern the propagation of an invading TE, as well as its fitness effects during invasion, is an understudied aspect of TE invasion biology. Here we have uncovered two different forms of host genetic variation in dysgenic sterility, which alter the permissivity of host cells to transposition, as well as their tolerance to transposition's effects. These observations add complexity to our current understanding of how host genetic variation can modulate the fitness effects of an invading TE. The precise pathways and genetic factors whose differential function underlie these tolerance and permissivity phenotypes remain to be resolved. Similarly, the degree to which these processes reflect the pleiotropic effects of a single gene, or the combined action of multiple factors remains an important question to be addressed by future work.

Methods

Drosophila strains and husbandry

The recombinant inbred lines are described in King *et al.* [24]. Harwich (#4264) and *b cn* (#44229), were obtained from the Bloomington *Drosophila* stock center. Canton-S was obtained from Brigitte Dauwalder. All flies were maintained on standard cornmeal media.

Alleles of the second chromosome centromeric region, containing both QTL, were extracted from three recombinant inbred lines carrying B6 QTL allele (#21076, #21218, #21156) and two RILs carrying B8 QTL allele (#21077, #21154) into a common background by crossing them to multiply marked stocks *b cn* (#44229). After 7 rounds of backcrossing followed by inbreeding, the final isogenic lines (Sterile_B61, Sterile_B62, Sterile_B63 and Fertile_B81, Fertile_B82) were generated. The lines were made homozygous for the 2nd chromosome by inbreeding and selecting for wild type phenotype. The genotype of the isogenic lines were verified through PCR using five different primers within the two QTL. chr2L:19383155–19383970: AACCCCTTTTCGCTGACAATAACA, ATTATCAGCAGGA GCCGAAACTT; chr2L:21333500–21334300: AAGTGAAGCTAACAAACGTGACAAC, CGTTTGACCATCGCTTACAATAA; chr2R:2392800–2393600: AACAGGAGGTTCGAA AGCCAAATA, ATGCAGAGTCATATTCTGGGTTGG; chr2R:6203290–6204284: AATGGA GACCGTTGATTTTGGTAA, CTTTTCTGCGGCATCAGGTG; chr2R:6058000–6059000: TGGCAATTGCAATCCTTTTGGTAT, ATAACACGAACTACGACCTTTCCA.

Phenotyping. Phenotyping of ovarian atrophy was performed as described previously in Kelleher *et al.* [14]. Briefly, crosses between virgin RIL females and Harwich males were transferred to fresh food every 3–5 days. Since crosses reared at a restrictive temperature (29°C) result in complete gonadal atrophy in F1 offspring, we reared our crosses at a lower permissive temperature (25°C), which produces an intermediate phenotype that better reveals the variation in severity of dysgenesis [12,14,15,95]. F1 offspring were maintained for 3 days or 21 days, at which point their ovaries were examined using a squash prep [95]. 21 day-old females were transferred onto new food every 5 days as they aged to avoid bacterial growth. Females who

produced 1 or more chorionated egg chambers were scored as having non-atrophied ovaries, and females producing 0 egg chambers were scored as having atrophied ovaries.

Crosses and phenotyping were performed for 673 RILs across 22 experimental blocks for 3 day-old F1 females, and 552 RILs across 18 experimental blocks for 21 day-old F1 females. If fewer than 21 F1 offspring were phenotyped for the same cross, it was discarded and repeated if possible. In total, we phenotyped >20 3-day old and 21 day-old F1 female offspring for 595 RILs and 456 RILs, respectively.

QTL mapping. QTL mapping was performed as described in Kelleher *et al.* [14]. Briefly, for each developmental time point, we modeled the arcsine transformed proportion of F1 ovarian atrophy as a function of two random effects: experimental block and undergraduate experimenter. Regression models were fit using the lmer function from the lme4 package [96]. We then used the residuals as a response for QTL mapping with the DSPRqtl package [24] in R 3.02 [97]. The LOD significance threshold was determined from 1,000 permutations of the observed data, and the confidence interval around each LOD peak was identified by a difference of -2 from the LOD peak position ($\Delta 2$ -LOD) [26], or from the Bayes Confidence Interval [98]. For $\Delta 2$ -LOD intervals, we took the conservative approach of determining the longest contiguous interval where the LOD score was within 2 of the peak value. We further calculated the broad sense heritability of ovarian atrophy as in Kelleher *et al.* [14].

Estimation of founder phenotypes and QTL phasing. To estimate the phenotypic effect associated with each founder allele at the QTL peak, we considered the distribution of phenotypes from all RILs carrying the founder haplotype at the LOD peak position (genotype probability >0.95%) [24]. QTL were then phased into allelic classes by identifying the minimal number of partitions of founder haplotypes that describe phenotypic variation associated with the QTL peak, as described previously [14,24].

Fertility assays. Virgin female offspring from dysgenic crosses between isogenic lines carrying fertile_B81/B82 (21077, 21154) and sterile_B62/B63 (21218, 21156) alleles and Harwich males were collected daily and individually placed in a vial containing two Canton-S males. Females were allowed to mate for 5 days and were transferred to a new vial for another 5 days after which the parents were discarded. The presence and total number of F2 individuals were counted from the two vials.

Selection of paired RILs with alternate QTL alleles. We identified background matched RILs containing either the B6 (sterile) or B4 (fertile) haplotypes from the start position of the QTL-21d confidence interval (2L: 19,010,000) to the end position of QTL-3d confidence interval (2R: 6,942,495) ($P > 0.9$), based on their published HMM genotypes [24]. For all possible RIL pairs (B6 and B4), we then calculated the number of 10 Kb genomic windows in which they carried the same RIL haplotype ($P > 0.9$). We selected three pairs of RILs, which carry the same founder genotype for 47% (21213 & 21183), 46% (21147 & 21346) and 44% (21291 & 21188) of genomic windows outside of the QTL.

Small RNA-seq and total RNA-seq. RILs were maintained at 25°C, and three biological replicates of 20 ovaries were dissected from 3–5 day-old females. Ovaries were homogenized in TRIzol (invitrogen) and stored at -80°C until RNA extraction. 50 µg of total RNA from each of 18 biological samples (3 biological replicates x 3 pairs) was size fractionated in a 15% denaturing polyacrylamide gel and the 18–30 nt band was excised. 2S-depleted small RNA libraries for Illumina sequencing were then constructed according to the method of Wickersheim and Blumenstiel [99]. Ovarian small RNA libraries were published previously SRP160954, [100]. Ribodepleted and stranded total RNA libraries were generated from the same ovarian samples using NuGen total RNA kit (TECAN). All 18 small RNA and total RNA libraries were sequenced on an Illumina Nextseq 500 at the University of Houston Seq-N-Edit Core.

Small-RNA analysis. Sequenced small RNAs were separated based on size into miRNAs/siRNAs (18–22nt) and piRNAs (23–30nt) [11]. Reads corresponding to contaminating rRNAs, including 2S-rRNA, were removed from each library by aligning to annotated transcripts from flybase [101]. To determine the piRNA cluster activity we first uniquely aligned the piRNAs to reference genome (dm6 [29]) using Bowtie1 (-v 1 -m 1) [102]. We then used a customized perl script to count reads that mapped to a set of previously annotated piRNA clusters from the same genotypes (497 piRNA clusters, [103]). Read counts normalized to total mapped microRNAs for each library were used to infer differential expression using DESeq2 [104]. Sliding window estimates of piRNA abundance (Fig 2C and 2D) were calculated using bedtools genomcov [105], normalizing the read counts to total mapped miRNA reads.

Total RNA analysis. Residual ribosomal RNAs (rRNAs) were identified in ribo depleted libraries based on alignment to annotated rRNAs from flybase [101], and excluded from further analysis. Retained reads aligned to the library of consensus satellite and TE sequences from repbase [106], plus additional satellite consensus sequences from Larracuent [107]. For TE expression, the total reads mapped to TE sequences were counted using unix commands (uniq -c). Remaining reads that failed to map were pseudoaligned to *D. melanogaster* transcriptome (dm6/BDGP6) using Kallisto with default parameters [108]. Differentially expressed TEs and genes were identified from a combined analysis in DESeq2 [104]. Genes and TEs with base mean ≥ 100 , Adjusted *P*-value ≤ 0.05 and whose expression pattern differed (fold change ≥ 1.5) were considered differentially expressed between the B6 and B4 QTL haplotype.

Radiation sensitivity. Third instar larvae were either mock treated or irradiated in a Rad Source RS 1800 X-ray machine set at 12.5 mA and 160 kV. To obtain 3rd instar larvae, embryos were collected for 24 hr and aged for 5 days at 25°C. The food vials containing larvae were then X-ray irradiated at doses from 5–80 Gray after which an optimal dose that clearly depicts the phenotypic difference was selected. Survival to adulthood was determined by scoring the number of empty and full pupal cases at 10 days after radiation.

Identification of novel mutations in RIL genomes. 10 females from each of 795 RILs were deposited into a well of a 96 well plate (Axygen, P-96-450R-C) on ice. DNA isolation was then executed in plates using the Gentra Puregene Cell Kit (Qiagen, 158788) using extensions of the manufacturer's protocol. Subsequently, DNA was further purified (Qiagen, QIAquick 96 PCR Purification kit, 28183), quantified via a fluorometer (ThermoFisher, Qubit dsDNA HS kit, Q32854), and diluted to 2-ng/ μ l. Each RIL sample was then subjected to the MSG (Multiplexed Shotgun Genotyping) approach developed by Andolfatto *et al.* [76], which is a form of RADseq (restriction-site associated DNA sequencing, Baird *et al.* [109]). Starting with 10-ng DNA, samples were first digested using MseI (New England Biolabs, R0525L). This enzyme has the restriction site T/TAA, cuts frequently along the genome, and unlike traditional rare-cutter RADseq strategies, yields sequencing reads spread somewhat evenly along the genome. Next, plate well-specific barcoded adapters are independently ligated onto the cut ends of each DNA sample. Fragmented, barcoded samples from a given plate are then mixed, and the 96-plex pool precipitated, purified, and size-selected to 250–300-bp via a Pippin Prep (Sage Science). Each multiplexed sample is then PCR amplified, during which a DNA plate-specific Illumina-compatible index is incorporated, and then purified. Finally, each of the independently-indexed 96-plex pools are quantified, mixed at equal concentration, and sequenced over multiple lanes of an Illumina HiSeq 2500 on “high output” mode, yielding single-end 100-bp reads (KU Genome Sequencing Core). With this MSG approach, reads from each RIL are computationally distinguished by both an Illumina index sequence (which marks the plates), and an “in line barcode” (the first 6 Read1 bases, which marks samples on any given plate).

SNP and indel variants were identified from MSG short-read data following GATK best practices for sample groups [110–112]. Briefly, Cutadapt (version 3.5; [113]) was used to demultiplex samples and trim adaptors, while alignment to the *D. melanogaster* reference genome (dm6 [29]) were performed using BWA (version 0.7.17-GCC-10.2.0; [114]). The resulting BAM files were sorted and indexed using Samtools (Li et al., 2009 [114]). Individual GVCF files were generated using HaplotypeCaller and then joint-genotyped using Genotype_GVCFs. Both SNPs and indels were extracted and filtered out following the GATK Best Practices hard filters using VariantFiltration (SNP: "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5"; indel: "QD < 2.0 || FS > 200.0", and converted to TSV in R for further analysis using the vcfR package [115].

To identify novel base substitutions that arose in RILs and not present in the founders, we filtered out all alleles with spanning deletions as well as annotated SNPs from the founder lines (<https://wfitc.bio.uci.edu/~dspr/Data/index.html>). Since founder SNPs were called in dm5, we converted their coordinates to dm6 using the NCBI Genome Remapping Service (<https://www.ncbi.nlm.nih.gov/genome/tools/remap>) before filtering. For indels, short indels were not called from the original founder sequence data. We therefore considered an indel to be novel if they were unique to a RIL and sequenced among at least 50 RILs.

Differences in the number of SNPs and indels rate was modeled using linear mixed model fitted with lmer function in the R package lme4 (Bates et al., 2013 [96]). Three models were compared, a null model (mutations ~ plate + depth), a founder allele model (mutations ~ plate + depth + founder allele), and a dysgenic sterility allele model (mutations ~ plate + depth + sterile/fertile). Founder or dysgenic sterility allele referred to the QTL-3d peak: (B1-B5, B6-B8) or sterile/fertile (B6/all other founders). Note that B5 alleles are not present among the RILs at the QTL-3d peak. The models were compared using a likelihood ratio test to determine whether founder allele or allelic class explained variation in the number of novel SNPs or indels between RILs. The effect of sterile alleles on SNP and indel number was evaluated by *t*-test.

qRT-PCR of P-element transcripts

3 biological replicates including 20 pairs of 3–5 day F1 dysgenic ovaries from crosses between fertile (B8¹, B8²) or sterile (B6², B6³) females and Harwich males were dissected and homogenized in TRIzol. Crosses were maintained at 22°C. RNA was treated with DNA-free (ThermoFisher) and reverse transcribed using oligo-dT primers and superscript IV (ThermoFisher) according to manufacturer instructions. Three different primer sets were used to amplify the 3' end of all *P*-element transcripts, IVS3 spliced transcripts and IVS3 unspliced transcripts, as well as *rpl32*. Transcripts were amplified and quantified in three technical replicates using power-SYBR green (ThermoFisher) and normalized to *rpl32* for the same sample.

Primer sequences were as follows: *rpl32*-F: 5'-CCGCTTCAAGGGACAGTATC, *rpl32*-R: 5'-GACAATCTCCTTGCGCTTCT, P-element-all F: 5'-CACCGAAATGGATGAGTTGACG, P-element-all R: 5'-TAATAAGTCCGCCGTGAGACAC, P-element IVS3 spliced F: 5'-AATA GCCAGGAATACAGAAATG, P-element IVS3 spliced R: 5'-AACATTTCTGTATTCTCTGG CTA, P-element IVS3-unspliced F: 5'-GACAAAACACAATAGACAGCACA, P-element IVS3-unspliced R: 5'-TGTGCTGTCTATTGTGTTTTGTC.

Identification of in-phase polymorphisms. The SNP data of B founders that used to infer in-phase SNPs is based on *dm3* [24]. To identify in-phase SNPs we looked for alternate SNP alleles that match the predicted phenotypic class for each of the QTL peaks. For QTL-21d we used the criteria: sterile class (B2, B6) and the fertile class (B1, B3, B4, B7, B8), whereas for QTL-3d: sterile class (B6) and the fertile class (B1, B2, B3, B4, B7, B8).

Supporting information

S1 Fig. Sterility is associated with increased expression of pericentromeric genes in the head. a) Mean expression of genes located in the pericentromere, euchromatin, telomere and the fourth chromosome from RILs carrying each of the eight B founder genotypes at the QTL-3d region. Error bars represent the standard deviation among mean expression levels of different genes. The sterile/B6 (light green) shows high pericentromeric gene expression compared to the fertile strains (dark green) (Anova; $F_{6,494} = 7.775$, $P < 5.24e-08$). The letters indicate significantly different expression levels based on Tukey-HSD comparisons between RILs with different founder alleles.

(PDF)

S2 Fig. Expression profile of QTL piRNA clusters in sterile and fertile RIL pair 2. The piRNA expression between sterile and fertile genotypes from the 21188–21291 RIL pair along the two QTL piRNA clusters: 2L:23,328,000–23,337,026 and 2L:23,222,004–23,246,024, respectively. Only uniquely mapping piRNAs are considered. The TE families at the top of each panel are represented by different colors. TE-others represent the repeat families coming from sibling species of *D. melanogaster*. Positive value indicates piRNAs mapped to the sense strand of the reference genome and negative value indicates those from the antisense strand. The piRNA cluster expression levels are estimated by log₂ scale transformed of reads per million mapped reads [$\log_2(\text{RPM}+1)$].

(PDF)

S3 Fig. Expression profile of QTL piRNA clusters in sterile and fertile RIL pair 3. The piRNA expression between sterile and fertile genotypes from the 21346–21147 RIL pair along the two QTL piRNA clusters: 2L:23,328,000–23,337,026 and 2L:23,222,004–23,246,024, respectively. Only uniquely mapping piRNAs are considered. The TE families at the top of each Fig are represented by different colors. TE-others represent the repeat families coming from sibling species of *D. melanogaster*. Positive value indicates piRNAs mapped to the sense strand of the reference genome and negative value indicates those from the antisense strand. The piRNA cluster expression levels are estimated by log₂ scale transformed of reads per million mapped reads [$\log_2(\text{RPM}+1)$].

(PDF)

S4 Fig. Sterility among 3 and 21 day females based on QTL haplotype. Four haplotypes are compared, which comprise all possible combinations of sterility alleles at 2 QTL. The allele at the 3 day QTL is indicated first and is represented by the color of the violin plot (light green = sterile, dark green = fertile). The allele at the 21 day QTL is indicated second and represented by the color of the points on the scatter plot. Y-axis is residual variation in F1 atrophy after accounting for student experimenter and block. Among 3 day old females, haplotypes containing different alleles for the 3 day old QTL are significantly different from each other (Tukey HSD $P = 0.016-0$). However, haplotypes containing alternative QTL for the 21d only do not differ from each other (Tukey HSD $P > 0.74$). This suggests phenotypic variation in 3 day old females is not influenced by their genotype at the 21 day QTL. In contrast, among 21 day old females tolerant alleles in both QTL loci are required to significantly decrease sterility below the sterile allele containing haplotypes (Tukey HSD $P = 0.01-0$).

(PDF)

S5 Fig. Crossing scheme to generate sterile and fertile alleles.

(PDF)

S6 Fig. Increased expression of genes upregulated in late-stage egg chambers in fertile ovaries. Upregulation in stage 9–10 and stage 12–14 egg chambers is from Tootle *et al.* [30]. Genes are separated into eggshell components (top) and non-eggshell components (bottom). Dark green bars indicate genes significantly upregulated in fertile genotypes whereas light green indicates genes upregulated in sterile genotypes.

(PDF)

S1 Table. Provided are the proportion of atrophy for 3-day old F1 females when recombinant inbred lines were crossed to Harwich males.

(XLSX)

S2 Table. Provided are the proportion of atrophy for 21-day old F1 females when recombinant inbred lines were crossed to Harwich males.

(XLSX)

S3 Table. Residuals from 3-day-old F1 females used for QTL mapping.

(XLSX)

S4 Table. Residuals from 21-day-old F1 females used for QTL mapping.

(XLSX)

S5 Table. Results of DESeq2 analysis of differential gene expression between sterile and fertile alleles.

(XLS)

S6 Table. List of differential expressed of Tip60 members and one of its interactors.

(XLS)

S7 Table. List of genes upregulated in sterile and fertile ovaries, as well as associated enriched GO terms.

(XLS)

S8 Table. Analysis of piRNA cluster expression and abundance of P and I element derived piRNAs in sterile and fertile ovaries.

(XLS)

S9 Table. TE composition of differentially active piRNA clusters in QTL-3d.

(XLSX)

S10 Table. Results of DESeq2 analysis of differential TE expression between sterile and fertile alleles.

(XLS)

S11 Table. In phase polymorphisms in QTL-3d and QTL-21d.

(XLSX)

S12 Table. List of candidate genes that are differentially expressed between sterile and fertile alleles and contain non-coding in-phase SNPs.

(XLSX)

S13 Table. List of candidate genes that are contain non-synonymous in-phase SNPs.

(XLSX)

S14 Table. PCA analysis of gene expression data of background-matched recombinant inbred lines.

(XLS)

S15 Table. PCA analysis of piRNA cluster expression data of background-matched recombinant inbred lines.

(XLSX)

S16 Table. Uniquely mapping read coverage of piRNAs from sterile and fertile ovaries for both plus and minus strand of differentially active piRNA clusters in QTL-3d.

(XLSX)

S17 Table. Incidence of ovarian atrophy among in F1 females from crosses between isogenic Fertile/Sterile lines and Harwich males.

(XLSX)

S18 Table. Fertility of F1 females from crosses between isogenic tolerant/sensitive lines and Harwich males.

(XLS)

S19 Table. Radiation sensitive of sterile and fertile alleles and controls.

(XLSX)

S20 Table. Number of novel SNPs and indels in population B RIL genomes.

(XLSX)

S21 Table. qPCR estimates of P-element derived mRNA abundance in dysgenic ovarian RNA.

(XLSX)

Author Contributions

Conceptualization: Jyoti Lama, Stuart J. Macdonald, Llewellyn Green, Erin S. Kelleher.

Data curation: Jyoti Lama, Stuart J. Macdonald, Erin S. Kelleher.

Formal analysis: Jyoti Lama, Llewellyn Green, Erin S. Kelleher.

Funding acquisition: Stuart J. Macdonald, Erin S. Kelleher.

Investigation: Jyoti Lama, Satyam Srivastav, Sadia Tasnim, Donald Hubbard, Savana Hadjipanteli, Brittany R. Smith, Erin S. Kelleher.

Methodology: Brittany R. Smith, Erin S. Kelleher.

Project administration: Erin S. Kelleher.

Supervision: Erin S. Kelleher.

Validation: Savana Hadjipanteli.

Visualization: Erin S. Kelleher.

Writing – original draft: Jyoti Lama.

Writing – review & editing: Jyoti Lama, Stuart J. Macdonald, Erin S. Kelleher.

References

1. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet.* 2017; 18: 71–86. <https://doi.org/10.1038/nrg.2016.139> PMID: 27867194
2. Senti K-A, Brennecke J. The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet.* 2010; 26: 499–509. <https://doi.org/10.1016/j.tig.2010.08.007> PMID: 20934772

3. Fultz D, Choudury SG, Slotkin RK. Silencing of active transposable elements in plants. *Curr Opin Plant Biol.* 2015; 27: 67–76. <https://doi.org/10.1016/j.pbi.2015.05.027> PMID: 26164237
4. Gilbert C, Schaack S, Pace JK, Brindley PJ, Feschotte C. A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature.* 2010; 464: 1347–1350. <https://doi.org/10.1038/nature08939> PMID: 20428170
5. Pace JKE 2nd, Gilbert C, Clark MS, Feschotte C. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proc Natl Acad Sci U S A.* 2008; 105: 17023–17028. <https://doi.org/10.1073/pnas.0806548105> PMID: 18936483
6. Thomas J, Schaack S, Pritham EJ. Pervasive horizontal transfer of rolling-circle transposons among animals. *Genome Biol Evol.* 2010; 2: 656–664. <https://doi.org/10.1093/gbe/evq050> PMID: 20693155
7. Schaack S, Gilbert C, Feschotte C. Promiscuous DNA: Horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol.* 2010; 25: 537–546. <https://doi.org/10.1016/j.tree.2010.06.001> PMID: 20591532
8. Kidwell M.G., Frydryck T., and Novy J.B. The hybrid dysgenesis potential of *Drosophila melanogaster* strains of diverse temporal and geographical natural origins. *Drosophila Information Service.* 1983; 59: 59–63.
9. Kidwell MG. Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 1983; 80: 1655–1659. <https://doi.org/10.1073/pnas.80.6.1655> PMID: 6300863
10. Anxolabéhère D, Kidwell MG, Periquet G. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile P elements. *Mol Biol Evol.* 1988; 5: 252–269. <https://doi.org/10.1093/oxfordjournals.molbev.a040491> PMID: 2838720
11. Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science.* 2008; 322: 1387–1392. <https://doi.org/10.1126/science.1165171> PMID: 19039138
12. Kidwell MG, Kidwell JF, Sved JA. Hybrid Dysgenesis in *DROSOPHILA MELANOGASTER*: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. *Genetics.* 1977; 86: 813–833. <https://doi.org/10.1093/genetics/86.4.813> PMID: 17248751
13. Ignatenko OM, Zakharenko LP, Dorogova NV, Fedorova SA. P elements and the determinants of hybrid dysgenesis have different dynamics of propagation in *Drosophila melanogaster* populations. *Genetica.* 2015; 143: 751–759. <https://doi.org/10.1007/s10709-015-9872-z> PMID: 26530414
14. Kelleher ES, Jaweria J, Akoma U, Ortega L, Tang W. QTL mapping of natural variation reveals that the developmental regulator *bruno* reduces tolerance to P-element transposition in the *Drosophila* female germline. *PLoS Biol.* 2018; 16: e2006040. <https://doi.org/10.1371/journal.pbio.2006040> PMID: 30376574
15. Dorogova NV, Bolobolova EU, Zakharenko LP. Cellular aspects of gonadal atrophy in *Drosophila* P-M hybrid dysgenesis. *Dev Biol.* 2017; 424: 105–112. <https://doi.org/10.1016/j.ydbio.2017.02.020> PMID: 28283407
16. Teixeira FK, Okuniewska M, Malone CD, Coux R-X, Rio DC, Lehmann R. piRNA-mediated regulation of transposon alternative splicing in the soma and germ line. *Nature.* 2017; 552: 268–272. <https://doi.org/10.1038/nature25018> PMID: 29211718
17. Ota R, Kobayashi S. Myc plays an important role in *Drosophila* P-M hybrid dysgenesis to eliminate germline cells with genetic damage. *Commun Biol.* 2020; 3: 185. <https://doi.org/10.1038/s42003-020-0923-3> PMID: 32322015
18. Margulies L. A high level of hybrid dysgenesis in *Drosophila*: high thermosensitivity, dependence on DNA repair, and incomplete cytotypic regulation. *Mol Gen Genet.* 1990; 220: 448–455. <https://doi.org/10.1007/BF00391752> PMID: 2160053
19. Tasnim S, Kelleher ES. p53 is required for female germline stem cell maintenance in P-element hybrid dysgenesis. *Dev Biol.* 2018; 434: 215–220. <https://doi.org/10.1016/j.ydbio.2017.12.021> PMID: 29294306
20. Khurana JS, Wang J, Xu J, Koppetsch BS, Thomson TC, Nowosielska A, et al. Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell.* 2011; 147: 1551–1563. <https://doi.org/10.1016/j.cell.2011.11.042> PMID: 22196730
21. Moon S, Cassani M, Lin YA, Wang L, Dou K, Zhang ZZ. A Robust Transposon-Endogenizing Response from Germline Stem Cells. *Dev Cell.* 2018; 47: 660–671.e3. <https://doi.org/10.1016/j.devcel.2018.10.011> PMID: 30393075
22. Laski FA, Rio DC, Rubin GM. Tissue specificity of *Drosophila* P element transposition is regulated at the level of mRNA splicing. *Cell.* 1986; 44: 7–19. [https://doi.org/10.1016/0092-8674\(86\)90480-0](https://doi.org/10.1016/0092-8674(86)90480-0) PMID: 3000622

23. Spradling AC, Bellen HJ, Hoskins RA. Drosophila P elements preferentially transpose to replication origins. *Proc Natl Acad Sci U S A*. 2011; 108: 15948–15953. <https://doi.org/10.1073/pnas.1112960108> PMID: 21896744
24. King EG, Merkes CM, McNeil CL, Hooper SR, Sen S, Broman KW, et al. Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. *Genome Res*. 2012; 22: 1558–1566. <https://doi.org/10.1101/gr.134031.111> PMID: 22496517
25. Schaefer RE, Kidwell MG, Fausto-Sterling A. Hybrid Dysgenesis in DROSOPHILA MELANOGASTER: Morphological and Cytological Studies of Ovarian Dysgenesis. *Genetics*. 1979; 92: 1141–1152. <https://doi.org/10.1093/genetics/92.4.1141> PMID: 17248944
26. Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*. 1989; 121: 185–199. <https://doi.org/10.1093/genetics/121.1.185> PMID: 2563713
27. Bridges CB. SALIVARY CHROMOSOME MAPS: With a Key to the Banding of the Chromosomes of Drosophila Melanogaster. *J Hered*. 1935; 26: 60–64.
28. Bridges PN. A NEW MAP OF THE SALIVARY GLAND 2L-CHROMOSOME: of Drosophila Melanogaster. *J Hered*. 1942; 33: 403–408.
29. Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, et al. The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome Res*. 2015; 25: 445–458. <https://doi.org/10.1101/gr.185579.114> PMID: 25589440
30. Tootle TL, Williams D, Hubb A, Frederick R, Spradling A. Drosophila eggshell production: identification of new genes and coordination by Pxt. *PLoS One*. 2011; 6: e19943. <https://doi.org/10.1371/journal.pone.0019943> PMID: 21637834
31. Navarro-Costa P, McCarthy A, Prudêncio P, Greer C, Guilgur LG, Becker JD, et al. Early programming of the oocyte epigenome temporally controls late prophase I transcription and chromatin remodelling. *Nat Commun*. 2016; 7: 12331. <https://doi.org/10.1038/ncomms12331> PMID: 27507044
32. McCarthy A, Deiluo A, Martin ET, Upadhyay M, Rangan P. Tip60 complex promotes expression of a differentiation factor to regulate germline differentiation in female Drosophila. *Mol Biol Cell*. 2018; 29: 2933–2945. <https://doi.org/10.1091/mbc.E18-06-0385> PMID: 30230973
33. DeLuca SZ, Ghildiyal M, Pang L-Y, Spradling AC. Differentiating Drosophila female germ cells initiate Polycomb silencing by regulating PRC2-interacting proteins. *Elife*. 2020; 9. <https://doi.org/10.7554/eLife.56922> PMID: 32773039
34. Kusch T, Florens L, Macdonald WH, Swanson SK, Glaser RL, Yates JR 3rd, et al. Acetylation by Tip60 is required for selective histone variant exchange at DNA lesions. *Science*. 2004; 306: 2084–2087. <https://doi.org/10.1126/science.1103455> PMID: 15528408
35. Shim HJ, Lee E-M, Nguyen LD, Shim J, Song Y-H. High-dose irradiation induces cell cycle arrest, apoptosis, and developmental defects during Drosophila oogenesis. *PLoS One*. 2014; 9: e89009. <https://doi.org/10.1371/journal.pone.0089009> PMID: 24551207
36. Ruddell A, Jacobs-Lorena M. Biphasic pattern of histone gene expression during Drosophila oogenesis. *Proc Natl Acad Sci U S A*. 1985; 82: 3316–3319. <https://doi.org/10.1073/pnas.82.10.3316> PMID: 3923476
37. Gunjan A, Verreault A. A Rad53 kinase-dependent surveillance mechanism that regulates histone protein levels in *S. cerevisiae*. *Cell*. 2003; 115. [https://doi.org/10.1016/s0092-8674\(03\)00896-1](https://doi.org/10.1016/s0092-8674(03)00896-1) PMID: 14651846
38. Liang D, Burkhart SL, Singh RK, Kabbaj M-HM, Gunjan A. Histone dosage regulates DNA damage sensitivity in a checkpoint-independent manner by the homologous recombination pathway. *Nucleic Acids Res*. 2012; 40: 9604. <https://doi.org/10.1093/nar/gks722> PMID: 22850743
39. Murga M, Jaco I, Fan Y, Soria R, Martinez-Pastor B, Cuadrado M, et al. Global chromatin compaction limits the strength of the DNA damage response. *J Cell Biol*. 2007; 178: 1101. <https://doi.org/10.1083/jcb.200704140> PMID: 17893239
40. Landais S, D'Alterio C, Jones DL. Persistent replicative stress alters polycomb phenotypes and tissue homeostasis in Drosophila melanogaster. *Cell Rep*. 2014; 7. <https://doi.org/10.1016/j.celrep.2014.03.042> PMID: 24746823
41. Ozawa N, Furuhashi H, Masuko K, Numao E, Makino T, Yano T, et al. Organ identity specification factor WGE localizes to the histone locus body and regulates histone expression to ensure genomic stability in Drosophila. *Genes Cells*. 2016; 21. <https://doi.org/10.1111/gtc.12354> PMID: 27145109
42. McKay DJ, Klusza S, Penke TJR, Meers MP, Curry KP, McDaniel SL, et al. Interrogating the function of metazoan histones using engineered gene clusters. *Dev Cell*. 2015; 32: 373–386. <https://doi.org/10.1016/j.devcel.2014.12.025> PMID: 25669886

43. Ner SS, Harrington MJ, Grigliatti TA. A Role for the *Drosophila* SU(VAR)3-9 Protein in Chromatin Organization at the Histone Gene Cluster and in Suppression of Position-Effect Variegation. *Genetics*. 2002; 162: 1763–1774. <https://doi.org/10.1093/genetics/162.4.1763> PMID: 12524347
44. King EG, Sanderson BJ, McNeil CL, Long AD, Macdonald SJ. Genetic dissection of the *Drosophila melanogaster* female head transcriptome reveals widespread allelic heterogeneity. *PLoS Genet*. 2014; 10: e1004322. <https://doi.org/10.1371/journal.pgen.1004322> PMID: 24810915
45. Riddle NC, Minoda A, Kharchenko PV, Alekseyenko AA, Schwartz YB, Tolstorukov MY, et al. Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila* heterochromatin. *Genome Res*. 2011; 21: 147. <https://doi.org/10.1101/gr.110098.110> PMID: 21177972
46. Karpen GH, Spradling AC. Analysis of Subtelomeric Heterochromatin in the *Drosophila* Minichromosome Dp1187 by Single P Element Insertional Mutagenesis. *Genetics*. 1992; 132: 737. <https://doi.org/10.1093/genetics/132.3.737> PMID: 1334894
47. Yin H, Lin H. An epigenetic activation role of Piwi and a Piwi-associated piRNA in *Drosophila melanogaster*. *Nature*. 2007; 450: 304–308. <https://doi.org/10.1038/nature06263> PMID: 17952056
48. Walter MF, Jang C, Kasravi B, Donath J, Mechler BM, Mason JM, et al. DNA organization and polymorphism of a wild-type *Drosophila* telomere region. *Chromosoma*. 1995; 104: 229–241. <https://doi.org/10.1007/BF00352254> PMID: 8565699
49. Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, et al. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol*. 2002; 3: 1–20. <https://doi.org/10.1186/gb-2002-3-12-research0084> PMID: 12537573
50. Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, et al. Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput Biol*. 2005; 1: 166–175. <https://doi.org/10.1371/journal.pcbi.0010022> PMID: 16110336
51. Nuzhdin SV, Mackay TF. Direct determination of retrotransposon transposition rates in *Drosophila melanogaster*. *Genet Res*. 1994; 63: 139–144. <https://doi.org/10.1017/s0016672300032249> PMID: 8026740
52. Nuzhdin SV, Mackay TF. The genomic rate of transposable element movement in *Drosophila melanogaster*. *Mol Biol Evol*. 1995; 12: 180–181. <https://doi.org/10.1093/oxfordjournals.molbev.a040188> PMID: 7877494
53. Nuzhdin SV, Pasyukova EG, Morozova EA, Flavell AJ. Quantitative genetic analysis of copia retrotransposon activity in inbred *Drosophila melanogaster* lines. *Genetics*. 1998; 150: 755–766. <https://doi.org/10.1093/genetics/150.2.755> PMID: 9755206
54. Mohn F, Sienski G, Handler D, Brennecke J. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell*. 2014; 157: 1364–1379. <https://doi.org/10.1016/j.cell.2014.04.031> PMID: 24906153
55. Le Thomas A, Stuwe E, Li S, Du J, Marinov G, Rozhkov N, et al. Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev*. 2014; 28: 1667–1680. <https://doi.org/10.1101/gad.245514.114> PMID: 25085419
56. Bergman CM, Bensasson D. Recent LTR retrotransposon insertion contrasts with waves of non-LTR insertion since speciation in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 2007; 104: 11340–11345. <https://doi.org/10.1073/pnas.0702552104> PMID: 17592135
57. Post C, Clark JP, Sytnikova YA, Chirn G-W, Lau NC. The capacity of target silencing by *Drosophila*-PIWI and piRNAs. *RNA*. 2014. pp. 1977–1986. <https://doi.org/10.1261/rna.046300.114> PMID: 25336588
58. Chambeyron S, Popkova A, Payen-Groschêne G, Brun C, Laouini D, Pelisson A, et al. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proc Natl Acad Sci U S A*. 2008; 105: 14964–14969. <https://doi.org/10.1073/pnas.0805943105> PMID: 18809914
59. Tiwari B, Kurtz P, Jones AE, Wylie A, Amatruda JF, Boggupalli DP, et al. Retrotransposons Mimic Germ Plasm Determinants to Promote Transgenerational Inheritance. *Curr Biol*. 2017; 27: 3010–3016.e3. <https://doi.org/10.1016/j.cub.2017.08.036> PMID: 28966088
60. Akkouche A, Rebollo R, Burlet N, Esnault C, Martinez S, Viginier B, et al. tirant, a newly discovered active endogenous retrovirus in *Drosophila simulans*. *J Virol*. 2012; 86: 3675–3681. <https://doi.org/10.1128/JVI.07146-11> PMID: 22278247
61. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*. 2007; 128: 1089–1103. <https://doi.org/10.1016/j.cell.2007.01.043> PMID: 17346786
62. Adams MD, Tarnig RS, Rio DC. The alternative splicing factor PSI regulates P-element third intron splicing in vivo. *Genes Dev*. 1997; 11: 129–138. <https://doi.org/10.1101/gad.11.1.129> PMID: 9000056

63. Horan L, Yasuhara JC, Kohlstaedt LA, Rio DC. Biochemical identification of new proteins involved in splicing repression at the *Drosophila* P-element exonic splicing silencer. *Genes Dev.* 2015; 29: 2298–2311. <https://doi.org/10.1101/gad.268847.115> PMID: 26545814
64. Ghanim GE, Rio DC, Teixeira FK. Mechanism and regulation of P element transposition. *Open Biol.* 2020; 10: 200244. <https://doi.org/10.1098/rsob.200244> PMID: 33352068
65. Koval L, Proshkina E, Shaposhnikov M, Moskalev A. The role of DNA repair genes in radiation-induced adaptive response in *Drosophila melanogaster* is differential and conditional. *Biogerontology.* 2019; 21: 45–56. <https://doi.org/10.1007/s10522-019-09842-1> PMID: 31624983
66. Uri A, Martha K, Veronika B-I, Anna B, Trudi S. An essential role for *Drosophila* hus1 in somatic and meiotic DNA damage responses. *J Cell Sci.* 2007; 120: 1042. <https://doi.org/10.1242/jcs.03414> PMID: 17327271
67. Ruike T, Takeuchi R, Takata K-I, Oshige M, Kasai N, Shimanouchi K, et al. Characterization of a second proliferating cell nuclear antigen (PCNA2) from *Drosophila melanogaster*. *FEBS J.* 2006; 273: 5062–5073. <https://doi.org/10.1111/j.1742-4658.2006.05504.x> PMID: 17087725
68. Staeva-Vieira E, Yoo S, Lehmann R. An essential role of DmRad51/SpnA in DNA repair and meiotic checkpoint control. *EMBO J.* 2003; 22: 5863–5874. <https://doi.org/10.1093/emboj/cdg564> PMID: 14592983
69. Sterpone S, Cozzi R. Influence of XRCC1 Genetic Polymorphisms on Ionizing Radiation-Induced DNA Damage and Repair. *J Nucleic Acids.* 2010; 2010. <https://doi.org/10.4061/2010/780369> PMID: 20798883
70. Kooistra R, Pastink A, Zonneveld JB, Lohman PH, Eeken JC. The *Drosophila melanogaster* DmRAD54 gene plays a crucial role in double-strand break repair after P-element excision and acts synergistically with Ku70 in the repair of X-ray damage. *Mol Cell Biol.* 1999; 19: 6269–6275. <https://doi.org/10.1128/MCB.19.9.6269> PMID: 10454573
71. Xu J, Xin S, Du W. *Drosophila* Chk2 is required for DNA damage-mediated cell cycle arrest and apoptosis. *FEBS Lett.* 2001; 508: 394–398. [https://doi.org/10.1016/S0014-5793\(01\)03103-9](https://doi.org/10.1016/S0014-5793(01)03103-9) PMID: 11728459
72. Adams MD, McVey M, Sekelsky JJ. *Drosophila* BLM in double-strand break repair by synthesis-dependent strand annealing. *Science.* 2003; 299: 265–267. <https://doi.org/10.1126/science.1077198> PMID: 12522255
73. McVey M, Radut D, Sekelsky JJ. End-joining repair of double-strand breaks in *Drosophila melanogaster* is largely DNA ligase IV independent. *Genetics.* 2004; 168: 2067–2076. <https://doi.org/10.1534/genetics.104.033902> PMID: 15611176
74. Wichmann A, Jaklevic B, Su TT. Ionizing radiation induces caspase-dependent but Chk2- and p53-independent cell death in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 2006; 103: 9952–9957.
75. Gladstone M, Frederick B, Zheng D, Edwards A, Yoon P, Stickel S, et al. A translation inhibitor identified in a *Drosophila* screen enhances the effect of ionizing radiation and taxol in mammalian models of cancer. *Dis Model Mech.* 2012; 5: 342–350. <https://doi.org/10.1242/dmm.008722> PMID: 22344740
76. Andolfatto P, Davison D, Erezylmaz D, Hu TT, Mast J, Sunayama-Morita T, et al. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res.* 2011; 21: 610–617. <https://doi.org/10.1101/gr.115402.110> PMID: 21233398
77. Mahmoud M, Gobet N, Cruz-Dávalos DI, Mounier N, Dessimoz C, Sedlazeck FJ. Structural variant calling: the long and the short of it. *Genome Biol.* 2019; 20: 246. <https://doi.org/10.1186/s13059-019-1828-7> PMID: 31747936
78. Slaidina M, Banisch TU, Gupta S, Lehmann R. A single-cell atlas of the developing *Drosophila* ovary identifies follicle stem cell progenitors. *Genes Dev.* 2020; 34: 239–249. <https://doi.org/10.1101/gad.330464.119> PMID: 31919193
79. Slaidina M, Gupta S, Banisch TU, Lehmann R. A single-cell atlas reveals unanticipated cell type complexity in *Drosophila* ovaries. *Genome Res.* 2021; 31: 1938–1951. <https://doi.org/10.1101/gr.274340.120> PMID: 34389661
80. Long AD, Macdonald SJ, King EG. Dissecting Complex Traits Using the *Drosophila* Synthetic Population Resource. *Trends Genet.* 2014; 30: 488. <https://doi.org/10.1016/j.tig.2014.07.009> PMID: 25175100
81. Knutson BA, Hahn S. Domains of Tra1 important for activator recruitment and transcription coactivator functions of SAGA and NuA4 complexes. *Mol Cell Biol.* 2011; 31: 818–831. <https://doi.org/10.1128/MCB.00687-10> PMID: 21149579
82. Perry P, Sauer S, Billon N, Richardson WD, Spivakov M, Warnes G, et al. A dynamic switch in the replication timing of key regulator genes in embryonic stem cells upon neural induction. *Cell Cycle.* 2004; 3: 1645–1650. PMID: 15611653

83. Murr R, Vaissière T, Sawan C, Shukla V, Herceg Z. Orchestration of chromatin-based processes: mind the TRRAP. *Oncogene*. 2007; 26: 5358–5372. <https://doi.org/10.1038/sj.onc.1210605> PMID: 17694078
84. Francis MJ, Roche S, Cho MJ, Beall E, Min B, Panganiban RP, et al. *Drosophila* IRBP bZIP heterodimer binds P-element DNA and affects hybrid dysgenesis. *Proceedings of the National Academy of Sciences*. 2016; 113: 13003–13008.
85. Shokri L, Inukai S, Hafner A, Weinand K, Hens K, Vedenko A, et al. A Comprehensive *Drosophila melanogaster* Transcription Factor Interactome. *Cell Rep*. 2019; 27: 955–970.e7. <https://doi.org/10.1016/j.celrep.2019.03.071> PMID: 30995488
86. Herz H-M, Mohan M, Garrett AS, Miller C, Casto D, Zhang Y, et al. Polycomb repressive complex 2-dependent and -independent functions of Jarid2 in transcriptional regulation in *Drosophila*. *Mol Cell Biol*. 2012; 32: 1683–1693. <https://doi.org/10.1128/MCB.06503-11> PMID: 22354997
87. Kang H, McElroy KA, Jung YL, Alekseyenko AA, Zee BM, Park PJ, et al. Sex comb on midleg (Scm) is a functional link between PcG-repressive complexes in *Drosophila*. *Genes Dev*. 2015; 29: 1136–1150. <https://doi.org/10.1101/gad.260562.115> PMID: 26063573
88. Blight KJ, McKeating JA, Rice CM. Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication. *J Virol*. 2002; 76: 13001–13014. <https://doi.org/10.1128/jvi.76.24.13001-13014.2002> PMID: 12438626
89. Seal S, Dharmarajan G, Khan I. Evolution of pathogen tolerance and emerging infections: A missing experimental paradigm. *Elife*. 2021; 10. <https://doi.org/10.7554/eLife.68874> PMID: 34544548
90. Lee JY, Wing PAC, Gala DS, Noerenberg M, Järvelin AI, Titlow J, et al. Absolute quantitation of individual SARS-CoV-2 RNA molecules provides a new paradigm for infection dynamics and variant differences. *Elife*. 2022; 11. <https://doi.org/10.7554/eLife.74153> PMID: 35049501
91. Siebel CW, Admon A, Rio DC. Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. *Genes Dev*. 1995; 9: 269–283. <https://doi.org/10.1101/gad.9.3.269> PMID: 7867926
92. Yildiz Ö, Majumder S, Kramer B, Sekelsky JJ. *Drosophila* MUS312 Interacts with the Nucleotide Excision Repair Endonuclease MEI-9 to Generate Meiotic Crossovers. *Mol Cell*. 2002; 10: 1503–1509. [https://doi.org/10.1016/s1097-2765\(02\)00782-7](https://doi.org/10.1016/s1097-2765(02)00782-7) PMID: 12504024
93. Boyd JB, Golino MD, Setlow RB. The mei-9 alpha mutant of *Drosophila melanogaster* increases mutagen sensitivity and decreases excision repair. *Genetics*. 1976; 84: 527–544. <https://doi.org/10.1093/genetics/84.3.527> PMID: 826452
94. Rodgers K, McVey M. Error-Prone Repair of DNA Double-Strand Breaks. *J Cell Physiol*. 2016; 231: 15–24. <https://doi.org/10.1002/jcp.25053> PMID: 26033759
95. Srivastav SP, Kelleher ES. Paternal Induction of Hybrid Dysgenesis in *Drosophila melanogaster* Is Weakly Correlated with Both P-Element and *hobo* Element Dosage. *G3: Genes|Genomes|Genetics*. 2017; g3.117.040634.
96. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw*. 2015; 67: 1–48.
97. Team, TRDC. The R Project for Statistical Computing. <http://www.r-project.org>. [cited 3 Mar 2021]. Available: <https://ci.nii.ac.jp/naid/10027310073>
98. Manichaikul A, Dupuis J, Sen S, Broman KW. Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. *Genetics*. 2006; 174: 481–489. <https://doi.org/10.1534/genetics.106.061549> PMID: 16783000
99. Wickersheim ML, Blumenstiel JP. Terminator oligo blocking efficiently eliminates rRNA from *Drosophila* small RNA sequencing libraries. *Biotechniques*. 2013; 55: 269–272. <https://doi.org/10.2144/000114102> PMID: 24215643
100. Zhang S, Kelleher ES. piRNA-mediated silencing of an invading transposable element evolves rapidly through abundant beneficial de novo mutations. *bioRxiv*. 2019. p. 611350. <https://doi.org/10.1101/611350>
101. Gramates LS, Marygold SJ, Santos G dos, Urbano J-M, Antonazzo G, Matthews BB, et al. FlyBase at 25: looking to the future. *Nucleic Acids Res*. 2017; 45: D663–D671. <https://doi.org/10.1093/nar/gkw1016> PMID: 27799470
102. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10: R25. <https://doi.org/10.1186/gb-2009-10-3-r25> PMID: 19261174
103. Zhang S, Pointer B, Kelleher ES. Rapid evolution of piRNA-mediated silencing of an invading transposable element was driven by abundant de novo mutations. *Genome Res*. 2020; 30: 566–575. <https://doi.org/10.1101/gr.251546.119> PMID: 32238416

104. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15: 550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
105. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics.* 2014; 47: 11.12.1–34. <https://doi.org/10.1002/0471250953.bi1112s47> PMID: 25199790
106. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA.* 2015; 6: 11. <https://doi.org/10.1186/s13100-015-0041-9> PMID: 26045719
107. Larracuente AM. The organization and evolution of the Responder satellite in species of the *Drosophila melanogaster* group: dynamic evolution of a target of meiotic drive. *BMC Evol Biol.* 2014; 14: 233. <https://doi.org/10.1186/s12862-014-0233-9> PMID: 25424548
108. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016; 34: 525–527. <https://doi.org/10.1038/nbt.3519> PMID: 27043002
109. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One.* 2008; 3: e3376. <https://doi.org/10.1371/journal.pone.0003376> PMID: 18852878
110. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics.* 2013; 43: 11.10.1–11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43> PMID: 25431634
111. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011; 43: 491–498. <https://doi.org/10.1038/ng.806> PMID: 21478889
112. Samuk K, Manzano-Winkler B, Ritz KR, Noor MAF. Natural Selection Shapes Variation in Genome-wide Recombination Rate in *Drosophila pseudoobscura*. *Curr Biol.* 2020; 30: 1517–1528.e6. <https://doi.org/10.1016/j.cub.2020.03.053> PMID: 32275873
113. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal.* 2011; 17: 10–12.
114. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009; 25: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
115. Knaus BJ, Grünwald NJ. vcfr: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour.* 2017; 17: 44–53. <https://doi.org/10.1111/1755-0998.12549> PMID: 27401132