LETTER

doi:10.1038/nature25018

piRNA–mediated regulation of transposon alternative splicing in the soma and germ line

Felipe Karam Teixeira[1]†, Martyna Okuniewska[1], Colin D. Malone[1]†, Rémi-Xavier Coux[1], Donald C. Rio[2] & Ruth Lehmann[1]

Transposable elements can drive genome evolution, but their enhanced activity is detrimental to the host and therefore must be tightly regulated[1]. The Piwi-interacting small RNA (piRNA) pathway is vital for the regulation of transposable elements, by inducing transcriptional silencing or post-transcriptional decay of mRNAs[2]. Here we show that piRNAs and piRNA biogenesis components regulate precursor mRNA splicing of *P*-transposable element transcripts *in vivo*, leading to the production of the non-transposase-encoding mature mRNA isoform in *Drosophila* germ cells. Unexpectedly, we show that the piRNA pathway components do not act to reduce transcript levels of the *P*-element transposon during *P–M* hybrid dysgenesis, a syndrome that affects germline development in *Drosophila*[3,4]. Instead, splicing regulation is mechanistically achieved together with piRNA-mediated changes to repressive chromatin states, and relies on the function of the Piwi–piRNA complex proteins Asterix (also known as Gtsf1)[5–7] and Panoramix (Silencio)[8,9], as well as Heterochromatin protein 1a (HP1a; encoded by *Su(var)205*). Furthermore, we show that this machinery, together with the piRNA *Flamenco* cluster[10], not only controls the accumulation of Gypsy retrotransposon transcripts[11] but also regulates the splicing of Gypsy mRNAs in cultured ovarian somatic cells, a process required for the production of infectious particles that can lead to heritable transposition events[12,13]. Our findings identify splicing regulation as a new role and essential function for the Piwi pathway in protecting the genome against transposon mobility, and provide a model system for studying the role of chromatin structure in modulating alternative splicing during development.

Hybrid dysgenesis is a syndrome that affects progeny in a non-reciprocal fashion, being normally restricted to the offspring of crosses in which males carry transposable elements but which females lack[3,14] (Fig. 1a). In *Drosophila*, the dysgenic traits triggered by the *P*-element DNA transposon[4] (Fig. 1b) are restricted to the germ line and include chromosomal rearrangements, high rates of mutation, and sterility[3]. The impairment is most prominent when hybrids are grown at higher temperatures, with adult dysgenic females being completely sterile at 29 °C[15] (Fig. 1c). Despite the severe phenotypes, little is known about the development of germ cells during *P–M* dysgenesis. To address this, we characterized germline development in the progeny obtained from reciprocal crosses between w^{1118} (*P*-element-devoid strain) and *Harwich* (*P*-element-containing strain) flies at 29 °C. In non-dysgenic progeny, germline development progressed normally throughout embryonic and larval stages (Fig. 1d, e), leading to fertile adults (Fig. 1c). Although the development of dysgenic germline cells was not disturbed during embryogenesis, germ cells decreased in number during early larval stages, leading to animals with no germ cells by late larval stages (Fig. 1d, e). These results indicate that the detrimental effects elicited by *P*-element activity are triggered early on during primordial germ cell (PGC) development in dysgenic progeny, leading to premature germ cell death.

Maternally deposited small RNAs cognate to the *P*-element are thought to provide the '*P*-cytotype' by conferring the transgenerationally inherited ability to protect developing germ cells against *P*-elements[16]. Small RNA-based transposon regulation is typically mediated by either transcriptional silencing or post-transcriptional clearance of mRNAs, both of which result in a decrease in the accumulation of transposon mRNA[2]. To understand how maternally provided small RNAs control *P*-elements in germ cells, we focused our analysis on embryonic PGCs sorted from 4- to 20-h-old embryos generated from reciprocal crosses between w^{1118} and *Harwich* strains. Surprisingly, the accumulation of *P*-element RNA as measured by quantitative reverse transcription PCR (RT–qPCR) showed no change in dysgenic PGCs when compared to non-dysgenic PGCs (open reading frame (ORF) 0–3; Fig. 1f). This indicates that *P*-cytotype small RNAs exert their function by means other than regulating *P*-element mRNA levels.

P-element activity relies on production of a functional *P*-element transposase protein, the expression of which requires precursor mRNA (pre-mRNA) splicing of three introns[17] (intervening sequence (IVS) 1–3; Fig. 1b). To analyse *P*-element RNA splicing in germ cells during hybrid dysgenesis, we designed primers that specifically anneal to spliced mRNA transcripts. The accumulation of spliced forms for the first two introns (IVS1 and IVS2) did not show changes in dysgenic PGCs when compared to non-dysgenic PGCs. By contrast, the accumulation of spliced transcripts for the third intron (IVS3) was substantially increased in dysgenic germ cells (Fig. 1f). Given that the overall accumulation of *P*-element mRNA showed no changes, our results indicate that the maternally provided *P*-cytotype can negatively regulate *P*-element IVS3 splicing and therefore inhibits the production of functional *P*-transposase in germ cells.

Analysis of publically available small RNA sequencing data from 0–2-h-old embryos laid by *Harwich* females[16] indicated that two classes of small RNAs cognate to the *P*-element are maternally transmitted: small interfering RNAs (siRNAs, 20–22-nucleotides long) and piRNAs (23–29 nucleotides long; Extended Data Fig. 1). To test the role of distinct small RNA populations on *P*-element expression, we analysed mutants uniquely affecting each small RNA biogenesis pathway in the *Harwich* background. Mutations that disrupt siRNA biogenesis components Dicer-2 (*Dcr-2*) and Argonaute 2 (*AGO2*), or mutations ablating components of the piRNA biogenesis pathway, such as the Argonautes *piwi*, *aubergine* (*aub*), and *Argonaute 3* (*AGO3*), as well as the RNA helicase *vasa* (*vas*) and *spindle E* (*spn-E*), did not affect *P*-element mRNA accumulation in adult ovaries as measured by RT–qPCR (ORF0–ORF3; Fig. 2a). However, mutations that disrupted piRNA biogenesis, and not the siRNA pathway, led to a strong and specific increase in the accumulation of IVS3-spliced mRNAs (Fig. 2a, b).

[1]Howard Hughes Medical Institute (HHMI) and Kimmel Center for Biology and Medicine of the Skirball Institute, Department of Cell Biology, New York University School of Medicine, New York, New York 10016, USA. [2]Department of Molecular and Cell Biology, Center for RNA Systems Biology, and California Institute for Quantitative Biosciences, University of California, Berkeley, California 94720, USA. †Present addresses: Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK (F.K.T.); Institute for Genomic Medicine, Columbia University Medical Center, New York, New York 10032, USA (C.D.M.).

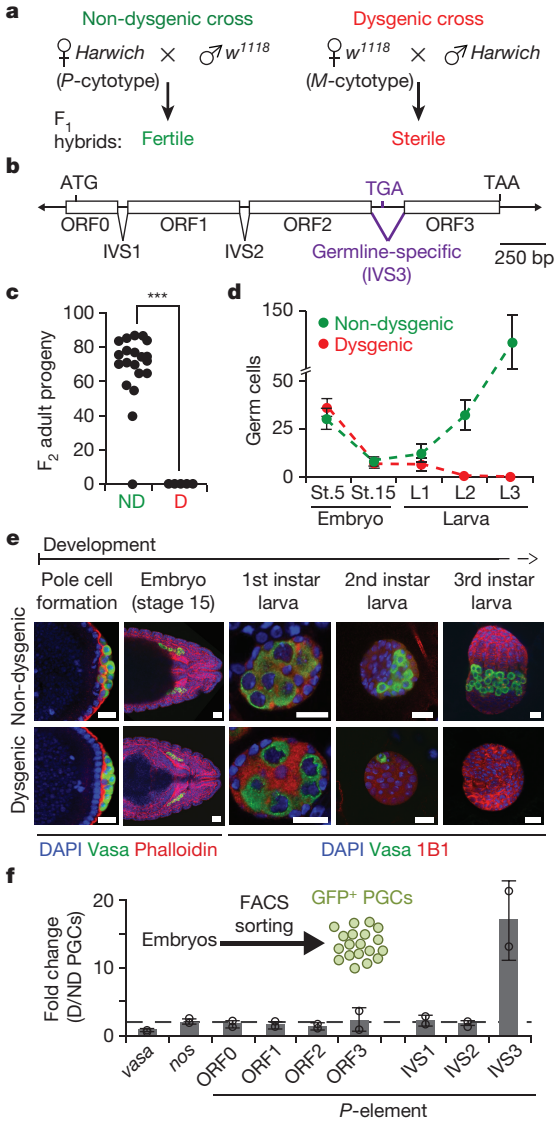


Figure 1 | *P*-element DNA transposon splicing is regulated in germ cells during hybrid dysgenesis. a, *P*–*M* hybrid dysgenesis crossing scheme. The *P*-cytotype designates animals containing *P*-elements, and the *M*-cytotype denotes those lacking them. The *Harwich* strain contains *P*-elements, whereas the w^{1118} strain does not. **b**, Diagram of *P*-element DNA transposon. Arrowheads represent terminal inverted repeats; boxes denote exons (ORF0–ORF3); inverted triangles denote introns (IVS1–IVS3). ATG, start codon; TGA and TAA, stop codons. **c**, Fertility of non-dysgenic (ND) and dysgenic (D) flies as measured by the number of F_2 adult progeny originated from single F_1 female crosses ($n = 20$ independent crosses). $^{***}P < 0.0001$ (Mann–Whitney–Wilcoxon test). **d**, Number of germ cells during embryonic and larval stages in non-dysgenic (green) and dysgenic (red) F_1 progeny. Data are mean ± s.d. Stage (st.) 5: complete germ cell count; stage 15 and later stages: germ cell count per gonad. Embryo stage 5 ($n = 28$ embryos); embryo stage 15 ($n > 30$ gonads); larva L1 ($n > 10$ gonads); larva L2 ($n = 7$ gonads); larva L3 ($n > 10$ gonads). **e**, Representative confocal images of germ-cell development in non-dysgenic and dysgenic progeny during embryonic and larval stages. Embryos and larvae were stained for Vasa (germ cells, green); DAPI (DNA, blue); phalloidin (F-actin, red; embryo stages); 1B1 (somatic cells and spectrosomes, red; larval stages). Pole cell formation is approximately 1.5 h after egg laying (AEL); embryo stage 15 is 10–12 h AEL; first instar larva is 22–48 h AEL; second instar larva is 48–72 h AEL; third instar larva is 72–120 h AEL. Scale bars, 20 μm. Experiments were repeated three or more times with similar results. **f**, RT–qPCR analysis on FACS-sorted GFP^+ PGCs from 4–20-h-old embryos generated from reciprocal crosses between w^{1118} and *Harwich* strains. Data are mean ± s.d. ($n = 2$ independent biological replicate experiments).

RNA sequencing (RNA-seq) analysis on poly(A)-selected RNAs from *aub* and *piwi* mutant adult ovaries confirmed the specific effect on IVS3 splicing (Fig. 2c–e; Extended Data Fig. 2a–e). To examine transposon expression *in tissue*, we performed RNA fluorescent *in situ* hybridization (FISH) using probes specific for the *P*-element and for the Burdock retrotransposon, a classic target of the germline piRNA pathway[6]. In mutants affecting piRNA biogenesis, increased abundance of Burdock RNA was readily observed in germline tissues, with most of the signal accumulating close to the oocyte (Fig. 2f). By contrast, we did not detect a difference in the *P*-element RNA FISH signal in piRNA biogenesis mutants compared to control. Nuclear RNA foci observed in nurse cells were of similar intensity and number regardless of the genotype, and cytoplasmic signal showed no detectable difference. Therefore, our results indicate that in germ cells, piRNAs specifically modulate IVS3 splicing. This regulation is reminiscent of the well-documented mechanism that restricts *P*-element activity to germline tissues, which involves the expression of a host-encoded RNA binding repressor protein that negatively regulates IVS3 splicing in somatic tissues[17].

In somatic tissues, *P*-element alternative splicing regulation is mediated by the assembly of a splicing repressor complex on an exonic splicing silencer element directly upstream of IVS3[17–19]. To test whether the *P*-element IVS3 and flanking exon sequences were sufficient to trigger the piRNA-mediated splicing regulation in germ cells, we used a transgenic reporter system for IVS3 splicing in which a heterologous promoter (*Hsp83*) drives the expression of an IVS3-*lacZ-neo* fusion mRNA specifically in the germ line[20] (Fig. 3a). Using RT–qPCR, we analysed the F_1 progeny from reciprocal crosses between w^{1118} and *Harwich* flies in the presence of the *hsp83*-IVS3-*lacZ-neo* reporter (to avoid developmental defects, F_1 progeny were raised at 18 °C; see Methods). The fraction of spliced mRNAs produced from the transgenic reporter was substantially increased in dysgenic compared to non-dysgenic adult ovaries (Fig. 3b; Extended Data Fig. 3), in agreement with previously reported results[20]. Most importantly, genetic experiments confirmed that the repression of IVS3 splicing in germ cells relies on piRNA biogenesis, as the splicing repression observed with this reporter in non-dysgenic progeny was specifically abolished in adult ovaries of *aub* and *vas* mutants (Fig. 3c; Extended Data Fig. 3).

Mechanistically, piRNA-mediated splicing regulation may be achieved through direct action of piRNA complexes on target pre-mRNAs carrying the IVS3 sequence or indirectly by piRNA-mediated changes in chromatin states. Piwi-interacting proteins such as Asterix (Arx) and Panoramix (Panx) are dispensable for piRNA biogenesis but are essential for establishing Piwi-mediated chromatin changes, possibly by acting as a scaffold to recruit histone-modifying enzymes and chromatin-binding proteins to target loci[8,9]. To test the role of these chromatin regulators on *P*-element splicing, we performed germline-specific RNA interference (RNAi) knockdown experiments in the *Harwich* background. Similar to what was observed for the piRNA biogenesis components (Fig. 2), germline knockdown of Arx and Panx showed no change in the accumulation of *P*-element RNA, but a strong and specific effect on IVS3 splicing in adult ovaries (Fig. 3d). The same pattern on IVS3 splicing was observed in the germline knockdown of HP1a and Maelstrom (Mael; Fig. 3d), both of which act downstream of Piwi-mediated targeting to modulate chromatin structure[21,22]. The same genetic requirement for Panx for IVS3 splicing control was also confirmed when using the transgenic IVS3 splicing reporter (Fig. 3c; Extended Data Fig. 3), further indicating that Piwi-mediated chromatin changes at the target locus are involved in IVS3 splicing regulation. At target loci, Piwi complexes are known to mediate the deposition of the classic heterochromatin mark histone H3 lysine 9 trimethylation (H3K9me3)[22]. To assess the effect of piRNA-targeting on *P*-element chromatin marks directly, we performed H3K9me3 chromatin immunoprecipitation followed by sequencing (ChIP–seq) or quantitative PCR on adult ovaries of progeny from reciprocal crosses between w^{1118} and *Harwich* strains (to avoid developmental defects,

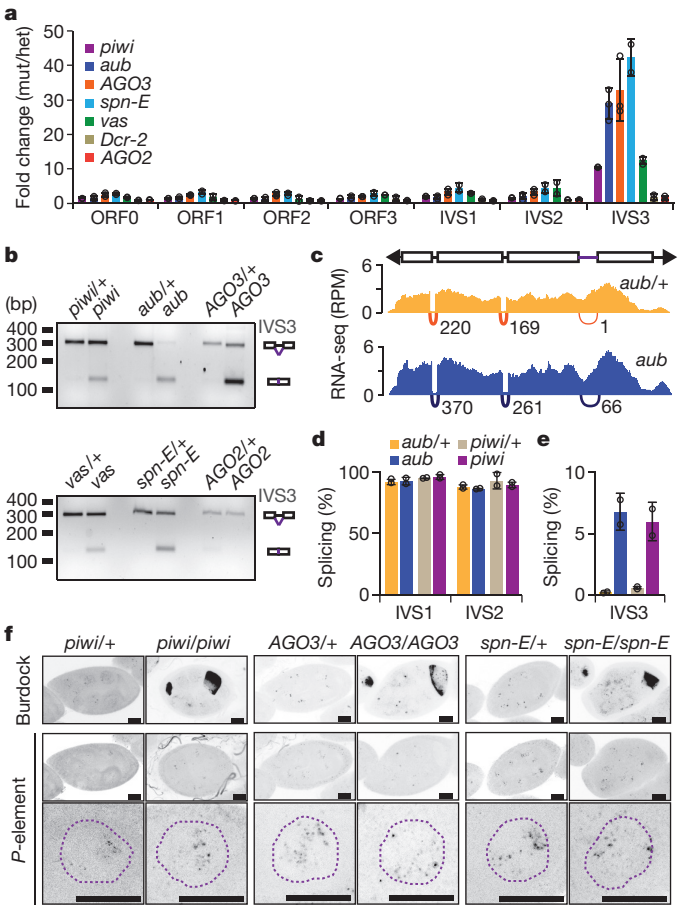


Figure 2 | piRNAs, but not siRNAs, modulate *P*-element splicing in germ cells. All analyses were performed in a *Harwich* background. **a**, RT–qPCR analysis on piRNA-biogenesis (*piwi*, *aub*, *AGO3*, *spn-E* and *vas*) and siRNA-biogenesis (*Dcr-2* and *AGO2*) mutant adult ovaries. Data are mean of fold changes in mutants in relation to respective heterozygote ± s.d. ($n \geq 2$ independent biological replicate experiments). **b**, Ethidium bromide-stained gel displaying RT–PCR reactions with primers flanking the *P*-element IVS3 intron in piRNA- and siRNA-biogenesis mutants. Size scale in base pairs (bp) is presented for each gel. Experiments were repeated twice with similar results. For gel source data, see Supplementary Fig. 1. **c**, Density plots for normalized strand-specific mRNA steady-state levels (measured by RNA-seq; RPM, reads per million) over consensus *P*-element sequence (top diagram) in *aub*/+ heterozygous (yellow, top plot) and *aub* mutant (blue, bottom plot) adult ovaries. The number and position of split-reads (represented by arcs that connect exons) observed for IVS1, IVS2 and IVS3 splicing junctions are shown below each density plot. Experiments were repeated twice with similar results. **d**, **e**, Percentage of splicing for *P*-element IVS1, IVS2 (**d**) and IVS3 (**e**). Splicing was quantified using RNA-seq analysis in *aub*/+ heterozygous (yellow), *aub* mutant (blue), *piwi*/+ heterozygous (beige), and *piwi* mutant (purple) adult ovaries. Percentage of splicing was calculated as the number of split-reads for each splicing junction normalized to the total number of reads mapping to the same junction. Data are mean ± s.d. ($n = 2$ independent biological replicate experiments). **f**, Representative confocal projections of RNA FISH signal (greyscale) showing the accumulation of sense RNA for Burdock and *P*-element transposons in heterozygous and mutant egg chambers. Bottom panels depict projections of representative nurse cell nuclei (purple dotted line) for the same genotypes. Scale bars, 20 μm. Experiments were repeated two or more times with similar results.

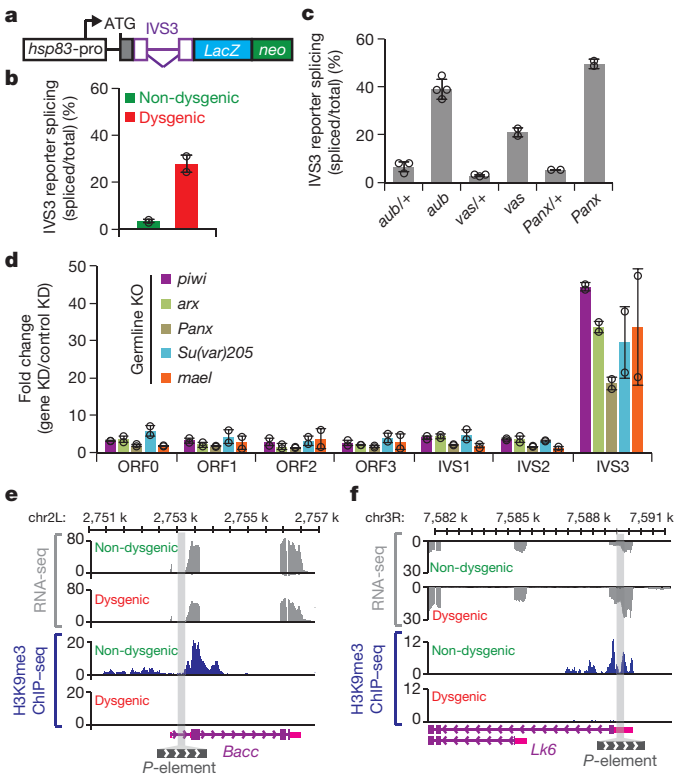


Figure 3 | piRNA-mediated chromatin changes regulate *P*-element splicing. a, Diagram of IVS3 transgenic reporter[20]. White box, germline-specific *hsp83* promoter; grey box, nuclear localization signal; purple, *P*-element fragment containing IVS3 (inverted triangle) and exonic flanking sequences (boxes); blue box, *LacZ*; green box, *neo*. **b**, RT–qPCR analysis using adult ovaries of F_1 progeny originated from reciprocal crosses between *Harwich* strain and w^{1118} flies carrying the IVS3 reporter. Bars represent the percentage of spliced IVS3 reporter transgenic transcripts, determined by the ratio of IVS3 spliced transgenic RNA (quantified using primers that specifically anneal to spliced transgenic transcripts) to total IVS3 reporter transgenic mRNA (quantified using primers that anneal within the *LacZ* coding sequence). Data are mean ± s.d. ($n = 2$ independent biological replicate experiments). **c**, RT–qPCR analysis in *aub*/+ heterozygous, *aub* mutant, *vas*/+ heterozygous, *vas* mutant, *panx*/+ heterozygous, and *panx* mutant, carrying the transgenic IVS3 reporter. All experiments were performed in a *Harwich* background ($n \geq 2$ independent biological replicate experiments). Results are represented as in **b**. **d**, RT–qPCR analysis on adult ovaries of germline knockdowns (KD) targeting piRNA pathway components involved in chromatin targeting (*piwi*, *arx*, *Panx*, *Su(var)205* and *mael*). Data are mean of fold changes in germline knockdown in relation to controls (*white* or *mCherry* germline knockdowns) ± s.d. All analyses were performed in a *Harwich* background ($n \geq 2$ independent biological replicate experiments). **e**, **f**, Genome browser view of two of the *P*-element insertions showing transcriptional activity. Normalized RNA-seq and H3K9me3 ChIP signals are presented in grey and blue, respectively. The grey bar crossing the plots represents *P*-element chromosomal insertion site. Annotation is at bottom: purple boxes, coding exons; pink boxes, untranslated regions (UTR); purple lines, introns; grey box, *P*-element insertion. View showing *P*-element insertion into *Bacc* (also known as *CG9894*) (**e**) and *Lk6* (**f**) genes.

ChIP was performed on F_1 progeny raised at 18 °C; see Methods). This analysis revealed a specific loss of global H3K9me3 levels over *P*-element insertions in dysgenic progeny when compared to non-dysgenic progeny (Extended Data Fig. 4a–d).

To analyse the chromatin structure at individual *P*-element insertions, we used DNA sequencing (DNA-seq) data to identify all euchromatic insertions in the *Harwich* strain (Supplementary Table 1; Extended Data Fig. 5) and RNA-seq analysis to define transcriptionally active insertions (Extended Data Fig. 6). At transcriptionally active *P*-element euchromatic insertions, the spreading of H3K9me3 into the flanking genomic regions was readily observed in non-dysgenic progeny, but was completely absent in dysgenic offspring (Fig. 3e, f; Extended Data Fig. 4e–g). Similarly, a reduction in H3K9me3 modification levels was also observed over the IVS3 transgenic reporter in dysgenic progeny when compared to non-dysgenic progeny (Extended

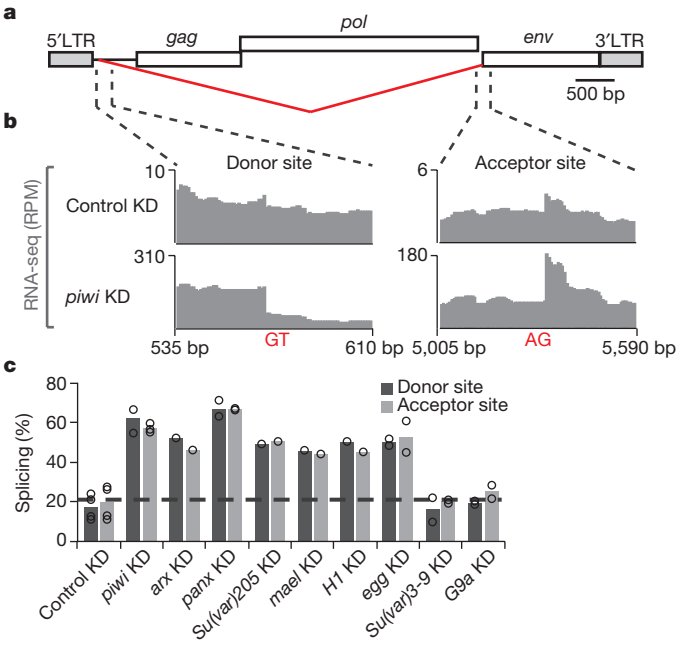


Figure 4 | piRNA-mediated chromatin targeting machinery modulates splicing of the Gypsy retrotransposon in OSCs. a, Diagram of Gypsy retrotransposon. Grey boxes, long terminal repeats (LTRs); white boxes, coding sequences; red inverted triangle, splicing that generates *env* mRNA. **b**, RNA-seq signal at the Gypsy splicing donor and acceptor sites in representative control (GFP) and *piwi* knockdown conditions. Raw datasets from ref. 9. Experiments were repeated three times with similar results. **c**, Percentage of splicing for Gypsy splicing donor and acceptor sites as determined by RNA-seq analysis performed in OSC cells knockdowns of piRNA machinery components. Bars represent the number of split-reads for the *env* splicing donor and acceptor junctions normalized to the total number of sense Gypsy reads mapping to the same junction. *SetDB1* is also known as *egg*. Data are means. With the exception *arx*, *Su(var)205*, *mael* and *H1*[7,25], experiments were repeated two or more times. Raw datasets from refs 7, 9 and 25.

Data Fig. 4h). Interestingly, euchromatic insertions with no evidence of transcriptional activity were devoid of an H3K9me3 signal in both non-dysgenic and dysgenic crosses (Extended Data Fig. 4i, j), providing further evidence for a model initially suggested in yeast[23] and more recently proposed for *Drosophila* and mammals, in which H3K9me3 deposition by piRNA complexes would require transcription of the target loci[22,24]. Mechanistically different from the well-described somatic repression, our results uncovered the existence of an unexpected piRNA-mediated, chromatin-based mechanism regulating IVS3 alternative splicing in germ cells (see also Supplementary Note 1).

To expand our analysis, we searched the literature for other cases of transposon splicing regulation. *Drosophila* Gypsy elements are retrotransposons that have retrovirus-like, infective capacity owing to their envelope (Env) protein[12,13] (Fig. 4a). These elements are expressed in somatic ovarian cells, in which they are regulated by the *flamenco* locus[10], a well-known piRNA cluster that is a soma-specific source of antisense piRNAs cognate to Gypsy[11]. Interestingly, it has been shown that mutations in *flamenco* not only elicited the accumulation of Gypsy RNA, but also modulated pre-mRNA splicing, favouring the production of the *env* mRNA and therefore germline infection[10]. To test whether the piRNA pathway, in addition to its role in regulating the accumulation of Gypsy RNA[11], is also responsible for modulating the splicing of Gypsy elements in somatic tissues, we analysed publically available RNA-seq data from poly(A)-selected RNAs extracted from *in vivo* cultures of ovarian somatic cells (OSCs)[7,9,25]. Our analysis indicates that *piwi* knockdown was sufficient to modulate *Gypsy* splicing, favouring the accumulation of *env*-encoding mRNA (Fig. 4b, c). In agreement with a chromatin-mediated regulation of alternative splicing, RNAi depletion of Arx, Panx, HP1a and Mael, as well as knockdown of the histone linker H1, was sufficient to favour Gypsy splicing, recapitulating the effect caused by Piwi depletion (Fig. 4c; Extended Data Fig. 7). Notably, this was also the case for the H3K9 methyltransferase Setdb1, but not for the H3K9 methyltransferases Su(var)3-9 and G9a, indicating specific genetic requirements (Fig. 4c; Extended Data Fig. 7). Taken together, our results indicate that the piRNA pathway, through its role in mediating changes in chromatin states, regulates the splicing of transposon pre-mRNAs in both somatic and germline tissues.

Using *P–M* hybrid dysgenesis as a model, we have uncovered splicing regulation elicited by chromatin changes as a previously unknown mechanism by which the piRNA pathway protects the genome from the detrimental effects of transposon activity. Splicing control at piRNA-target loci is likely to be mechanistically different from what has been observed for germline piRNA clusters given the low enrichment of the HP1 homologue Rhino (also known as HP1D) protein, which is required for piRNA cluster RNA processing[26], over the endogenous *P*-element insertions in the *Harwich* genome or over the transgenic IVS3 splicing reporter in non-dysgenic and dysgenic progeny (as measured by ChIP–qPCR; Extended Data Fig. 8). Because small RNA-based systems leading to chromatin mark changes at target loci are pervasive in eukaryotes[23], we expect this new type of targeted regulation to be of importance in settings far beyond the scope of the piRNA pathway and *Drosophila*. Indeed, small RNA-guided DNA methylation over the LINE retrotransposon *Karma* was recently shown to modulate alternative splicing in oil palm, disrupting nearby gene expression and ultimately affecting crop yield[27]. In this context, small RNA-based control of chromatin structure may be crucially important in genomes with a high content of intronic transposon insertions, such as the human genome, by providing a mechanism to suppress exonization of repeat elements[28]. Although the means by which piRNA-mediated changes in chromatin states could regulate alternative splicing remain to be determined, it is tempting to speculate that piRNA pathway components do so by co-transcriptionally modulating interactions between RNA polymerase II and the spliceosome[29,30].

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 March; accepted 6 November 2017.
Published online 6 December 2017.

1. Slotkin, R. K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* **8,** 272–285 (2007).
2. Czech, B. & Hannon, G. J. One loop to rule them all: the ping-pong cycle and piRNA-guided silencing. *Trends Biochem. Sci.* **41,** 324–337 (2016).
3. Kidwell, M. G., Kidwell, J. F. & Sved, J. A. Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics* **86,** 813–833 (1977).
4. Bingham, P. M., Kidwell, M. G. & Rubin, G. M. The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. *Cell* **29,** 995–1004 (1982).
5. Muerdter, F. *et al.* A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in *Drosophila*. *Mol. Cell* **50,** 736–748 (2013).
6. Dönertas, D., Sienski, G. & Brennecke, J. *Drosophila* Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. *Genes Dev.* **27,** 1693–1705 (2013).
7. Ohtani, H. *et al.* DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the *Drosophila* ovary. *Genes Dev.* **27,** 1656–1661 (2013).
8. Yu, Y. *et al.* Panoramix enforces piRNA-dependent cotranscriptional silencing. *Science* **350,** 339–342 (2015).
9. Sienski, G. *et al.* Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. *Genes Dev.* **29,** 2258–2271 (2015).
10. Pélisson, A. *et al.* Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila flamenco* gene. *EMBO J.* **13,** 4401–4411 (1994).
11. Brennecke, J. *et al.* Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128,** 1089–1103 (2007).
12. Lécher, P., Bucheton, A. & Pélisson, A. Expression of the *Drosophila* retrovirus gypsy as ultrastructurally detectable particles in the ovaries of flies carrying a permissive *flamenco* allele. *J. Gen. Virol.* **78,** 2379–2388 (1997).

13. Song, S. U., Kurkulos, M., Boeke, J. D. & Corces, V. G. Infection of the germ line by retroviral particles produced in the follicle cells: a possible mechanism for the mobilization of the gypsy retroelement of *Drosophila*. *Development* **124,** 2789–2798 (1997).
14. Malone, C. D., Lehmann, R. & Teixeira, F. K. The cellular basis of hybrid dysgenesis and Stellate regulation in *Drosophila*. *Curr. Opin. Genet. Dev.* **34,** 88–94 (2015).
15. Kidwell, M. G. & Novy, J. B. Hybrid dysgenesis in *Drosophila melanogaster*: sterility resulting from gonadal dysgenesis in the P-M system. *Genetics* **92,** 1127–1140 (1979).
16. Brennecke, J. *et al.* An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322,** 1387–1392 (2008).
17. Majumdar, S. & Rio, D. C. P transposable elements in *drosophila* and other eukaryotic organisms. *Microbiol. Spectr.* **3,** MDNA3-0004-014 (2015).
18. Siebel, C. W., Fresco, L. D. & Rio, D. C. The mechanism of somatic inhibition of *Drosophila* P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5′ splice site control U1 snRNP binding. *Genes Dev.* **6,** 1386–1401 (1992).
19. Siebel, C. W. & Rio, D. C. Regulated splicing of the *Drosophila* P transposable element third intron in vitro: somatic repression. *Science* **248,** 1200–1208 (1990).
20. Roche, S. E., Schiff, M. & Rio, D. C. P-element repressor autoregulation involves germ-line transcriptional repression and reduction of third intron splicing. *Genes Dev.* **9,** 1278–1288 (1995).
21. Wang, S. H. & Elgin, S. C. R. *Drosophila* Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proc. Natl Acad. Sci. USA* **108,** 21164–21169 (2011).
22. Sienski, G., Dönertas, D. & Brennecke, J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151,** 964–980 (2012).
23. Holoch, D. & Moazed, D. RNA-mediated epigenetic regulation of gene expression. *Nat. Rev. Genet.* **16,** 71–84 (2015).
24. Pezic, D., Manakov, S. A., Sachidanandam, R. & Aravin, A. A. piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes Dev.* **28,** 1410–1428 (2014).
25. Iwasaki, Y. W. *et al.* Piwi modulates chromatin accessibility by regulating multiple factors including histone H1 to repress transposons. *Mol. Cell* **63,** 408–419 (2016).
26. Zhang, Z. *et al.* The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell* **157,** 1353–1363 (2014).
27. Ong-Abdullah, M. *et al.* Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature* **525,** 533–537 (2015).
28. Zarnack, K. *et al.* Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* **152,** 453–466 (2013).
29. Naftelberg, S., Schor, I. E., Ast, G. & Kornblihtt, A. R. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu. Rev. Biochem.* **84,** 165–198 (2015).
30. Nojima, T. *et al.* Mammalian NET-seq reveals genome-wide nascent transcription coupled to rna processing. *Cell* **161,** 526–540 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Zamparini and J. Seifert for early discussions and assistance with developmental analysis; L. Barton for assistance with PGC sorting; T. Trcek for advice on RNA FISH; R. Sachidanandam for small RNA analysis assistance; J. Brennecke, P. Andersen and F. Roudier for advice on ChIP; A. Pelisson for early discussions; W. Theurkauf for antibody reagent; Q. Wang for assistance with JUM analysis; P. Macdonald, R. Carthew, E. Lai, P. Zamore, the Vienna *Drosophila* Resource Center, and the Bloomington *Drosophila* Stock Center for fly reagents; the NYUMC Genome Technology Center and the NYUMC Cytometry Center, supported by NIH P30CA016087, for sequencing and cell sorting support. C.D.M. was supported by a HHWF fellowship. F.K.T. was supported by EMBO and HFSP fellowships, and is funded by the Wellcome Trust Sir Henry Dale Fellowship (206257/Z/17/Z). R.L. is supported by NIH R37HD41900 and is an HHMI investigator. D.R. is supported by NIH R35-118121-02 and NIH R01-097352.

Author Contributions F.K.T. and R.L. conceived the idea. F.K.T., R.L. and D.R. designed the experiments. F.K.T. and M.O. performed the experiments, with the help of C.D.M. and R.C. F.K.T., R.L. and D.R. wrote the manuscript with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to R.L. (ruth.lehmann@med.nyu.edu) or F.K.T. (fk319@cam.ac.uk).

Reviewer Information *Nature* thanks J. Brennecke, H. Malik, A. Molaro and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. Experiments were neither intentionally randomized nor intentionally ordered. Investigators were not blinded to allocation during experiments and outcome assessment.

***Drosophila* stocks, genetics and husbandry.** *Drosophila melanogaster* stocks used: w^{1118} (R. Lehmann); *Harwich* (Bloomington *Drosophila* Stock Center, BDSC# 4264); *w;;* P_{nos}*::egfp-moe::nos* 3′UTR[31], *w;; hsp83-IVS3-LacZ-neo/TM3* Sb^{20}, *w;; nos-GAL4::VP16*[32], *P{UAS-Dcr-2.D}1* w^{1118} (BDSC# 24646), w^{1118}*; piwi*^{1}*/CyO,* w^{+} (BDSC# 43637), w^{1118}*; piwi*^{2}*/CyO,* w^{+} (BDSC# 43319), w^{1118}*;* aub^{QC42} cn^{1} bw^{1}/CyO, *P{sevRas1.V12}FK1* (BDSC# 4968), aub^{HN2} cn^{1} bw^{1}*/CyO* (gift from P. Macdonald), bw^{1}*;* st^{1} *AGO3*^{T2}*/TM6B*, bw^{1}*;* st^{1} *AGO3*^{T3}*/TM6B* (gifts from P. Zamore), ru^{1} st^{1} *spn-E*^{1} e^{1} ca^{1}*/TM3* Sb^{1} Ser^{1} (BDSC# 3327), *w;; FRT82B spn-E*$^{100.37}$ e^{1}*/TM3 Sb* (R. Lehmann), vas^{PH165}*/CyO*[33], w^{1118}*; Df(2L)osp*^{29}, Adh^{UF} osp^{29} pr^{1} cn^{1}*/CyO, P{sevRas1.V12}FK1* (BDSC# 3078), w^{1118}; *CG9754*$^{[m1-9]}$/CyO (VDRC# v313502; *Panx* mutant from J. Brennecke[9]), w^{1118}; *CG9754*$^{[m3-4]}$/CyO (VDRC# v313500; *Panx* mutant from J. Brennecke[9]), w^{m4h}*;; AGO2*^{414}*/TM6b* (gift from R. Carthew), w^{1118}*;; Df(3L)BSC558/TM6c Sb* (BDSC# 25120), *Dcr-2*^{L811fsX}*, P{GMR-w.IR}; Df(2R) BSC45/CyO* (gifts from E. Lai), *aub* shRNA (Vienna *Drosophila* Resource Center, VDRC# v313413), *piwi*-dsRNA (VDRC# v101658), *arx* dsRNA (VDRC# v40408), *panx* dsRNA (VDRC# v102702), *mael* dsRNA (VDRC# v100907), *Su(var)205* shRNA (BDSC# 36792), *white* dsRNA (dsRNA control, VDRC# v30033), *mCherry* shRNA (shRNA control, BDSC# 35785).

Laboratory-strain chromosomes carrying mutations and transgenes of interest were individually transferred to the *Harwich* background through a series of backcrosses. The genotypes of the resulting *Harwich*-derived stocks, which are characterized by single chromosome permutations, are listed in Supplementary Table 2. Before further use, the ability of the *Harwich*-derived stocks to induce dysgenic ovarian phenotype in F_1 progeny was confirmed by dysgenesis testcrosses using five w^{1118} females and five males from each of the *Harwich*-derived stocks. Testcrosses were performed at 29 °C on standard medium in at least two biological replicates. Ovary phenotype in F_1 progeny was scored after dissection and results are summarized in Supplementary Table 2.

For the fertility test described in Fig. 1c, non-dysgenic and dysgenic F_1 progeny females were individually mated to two w^{1118} males. Parental flies were discarded after 3 days of mating, and number of F_2 adult progeny was assessed 12 days after that. All other crosses are described in Supplementary Table 3 and were performed in at least two biological replicates. Germline RNAi knockdowns using UAS-dsRNA lines were performed in the presence of the *P{UAS-Dcr-2.D}1* transgene, as previously described[21]. Unless stated otherwise, all crosses were performed at 29 °C on standard medium and experiments were performed with 3–5-day-old adult female flies. ChIP–seq, RNA-seq analysis, and IVS3 transgenic reporter analysis with F_1 progeny originated from reciprocal crosses between w^{1118} and *Harwich* were performed at 18 °C to minimize confounding effects derived from dysgenesis-induced developmental defects. Unlike dysgenesis-triggered developmental defects, which are only observed in flies grown in temperatures >25 °C[15], *P*-element splicing regulation is not temperature-sensitive as verified by RT–PCR, RT–qPCR, and RNA-seq analyses performed at 18 °C (Extended Data Fig. 6).

Immunofluorescence. Hand-devitellinized embryos and dissected larval ovaries were immunostained as previously described[34,35]. Samples were mounted in Vectashield media containing DAPI (Vector Laboratories). Primary antibodies: rabbit anti-Vasa serum (1:5,000, R. Lehmann) and mouse anti-1B1 (1:20, DSHB). Alexa Fluor 488- (Life Technologies) and Cy3-conjugated (Jackson Immunoresearch) secondary antibodies were used at a 1:500 dilution. Phalloidin-TRITC (Sigma) was used at 1:250 (from 20 μM stock). Fluorescent images were acquired with Plan-Apochromat 20×/NA0.8 and Plan-Apochromat 40×/NA1.4 (oil immersion) objectives on a Zeiss LSM 780 confocal microscope.

RNA FISH. RNA FISH was performed as previously described[36]. Custom Stellaris RNA FISH Probes were designed using the Stellaris RNA FISH Probe Designer (Biosearch Technologies) and labelled with CALFluor590 and Quasar670 to detect consensus *P*-element (GenBank X06779:996..3902) and Burdock (GenBank U89994) sense mRNA sequences, respectively. FISH probes are composed of 20-nucleotide-long oligonucleotide pools, as listed in Supplementary Table 4. Fluorescent images were acquired with a Plan-Apochromat 40×/NA1.4 (oil immersion) objective on a Zeiss LSM 780 confocal microscope.

FACS sorting of PGCs. GFP^{+} PGCs were isolated from 4–20-h-old embryos by FACS as previously described[37]. Reciprocal crosses to obtain non-dysgenic and dysgenic embryos were performed with stocks carrying the P_{nos}*::egfp-moe::nos* 3′UTR transgene[31], as described in Supplementary Table 3. GFP^{+} PGCs were FACS-sorted (MoFlo cell sorter) into Trizol reagent for subsequent RNA extraction.

RNA isolation and analyses. Total RNA from FACS sorted germ cells or dissected adult ovaries was isolated using Trizol reagent (Invitrogen) and quantified by Qubit (Invitrogen). Contaminating DNA was removed using RQ1 DNase (Promega). Reverse transcription was performed using random primers and SuperScript III Reverse Transcriptase (Invitrogen). Transcript levels were assessed by quantitative PCR (qPCR) using LightCycler 480 SYBR Green I Master 2X (Roche) in a Roche LightCycler 480 machine. Results were normalized to the mean value obtained for three control genes (*CG8187*, *RpL32* (also known as *Rp49* and *CG7939*), and *DCTN2-p50* (*CG8269* and *Dmn*)) with invariant expression in a range of tissues and developmental stages, as previously described[38]. For splicing analysis on gel, PCR reactions were performed in RT–qPCR-normalized cDNA to non-saturating conditions (exponential phase) using LightCycler 480 SYBR Green I Master 2X (Roche) and a Roche LightCycler 480 machine, and PCR products were separated in an ethidium bromide-stained, 2.5% agarose gel. Note that this is not a quantitative analysis, as amplification of smaller DNA fragments may be disproportionally favoured in PCR reactions targeting fragments of different length. All primers are listed in Supplementary Table 5.

Poly(A)-selected RNA-seq analysis was performed on 2.5 μg of RNA purified from adult ovaries using the NEBNext Poly(A) mRNA Magnetic Isolation Module and the NEBNext Ultra Directional RNA Library Prep Kit for Illumina. Libraries were multiplexed using the NEBNext Multiplex Oligos for Illumina and sequenced in paired-end, 100-nucleotide-long reads on an Illumina HiSeq 2500.

PCR on genomic DNA. Genomic DNA was extracted from w^{1118} and *Harwich* pools containing 20 adult female flies each, using the DNeasy Blood & Tissue kit (QIAGEN) and following manufacturer's protocol for insects. PCR was conducted on 2 ng of total DNA using Q5 Hot Start High-Fidelity DNA Polymerase (NEB). PCR products were separated in an ethidium bromide-stained, 1% agarose gel. To structurally characterize *P*-element insertions, PCR products were purified using the QIAquick PCR purification kit (QIAGEN) and sequenced using traditional Sanger sequencing. Primers are listed in Supplementary Table 5.

ChIP. ChIP was performed as previously described[39], using 100–200 dissected adult ovaries as starting material and the anti-H3K9me3 (Abcam, ab8898) or the anti-Rhino[40] (guinea pig; gift from W. Theurkauf) antibody. Before immunoprecipitation, 5% of sonicated chromatin was separated to serve as the 'input'. After reverse-crosslinking, immunoprecipitated and input DNA was extracted using QIAquick PCR Purification Kit (QIAGEN) and used for qPCR reactions as previously described or library preparation for sequencing. ChIP–qPCR results were normalized to the input and compared with signal at the *42AB* locus. Primers used for ChIP–qPCR are listed in Supplementary Table 5. For ChIP–seq, immnuoprecipitation and input libraries were prepared using NEBNext Ultra II DNA Library Prep Kit for Illumina. Libraries were multiplexed using the NEBNext Multiplex Oligos for Illumina and sequenced in paired-end, 50-nucleotide-long reads on an Illumina HiSeq 2500. ChIP experiments were performed in two biological replicates. qPCR measurements were performed in at least two technical replicates on each sample.

Small RNA analysis. Small RNA reads extracted from the NCBI Gene Expression Omnibus (GEO) database accession number GSE13081[16] were mapped to the consensus *P*-element sequence, as previously described[11].

RNA-seq analysis. Paired-end (100-nucleotide-long reads) RNA-seq data generated in this study were mapped to the *Drosophila melanogaster* genome (dm3) and the FlyBase and Repbase transposon consensus database using the piPipes package (version 1.5.0; https://github.com/bowhan/piPipes), following the RNA-seq pipeline as previously described[41]. In brief, libraries were first aligned to ribosomal RNA using Bowtie2[42]. Non-rRNA reads were then mapped to the transcriptome and transposon consensus using Bowtie2, and transcript abundance was quantified using eXpress[43]. Analyses were performed with two samples, each with two biological replicates, using the piPipes RNA-seq dual-library mode[41].

Genome-wide analysis of differentially spliced mRNAs was performed using the Junction Usage Model (JUM) package[44,45] (https://github.com/qqwang-berkeley/JUM; version 1.3.2) with default parameters and following the JUM manual. In brief, RNA-seq data were mapped to the *Drosophila* genome (dm3) using STAR[46] in two-pass mode, following instructions in the STAR manual (https://github.com/alexdobin/STAR). STAR output is then used by JUM to construct statistical models and to quantify splicing from raw counts mapped to each identified junction[44]. Each comparison was performed with two samples, each with two biological replicates. Junctions showing >2-fold changes between samples and with adjusted $P < 0.05$ were considered as differentially spliced. JUM analysis was performed using a minimal of five mapping reads as a filter for valid junctions.

Splicing of *P*-element and Gypsy transposons was defined manually according to the method used in the JUM package[44]. To do so, RNA-seq data were initially analysed on the Galaxy web-based platform (https://usegalaxy.org)[47]. After the removal of Illumina adaptors and other low quality sequences using the 'Trim Galore!' package, stranded RNA-seq data were mapped to consensus *P*-element (GenBank X06779:996..3902) and Gypsy (GenBank M12927) sequences using TopHat[48], allowing for up to 3 mismatches. To calculate the percentage of splicing,

TopHat output was used to define the number of split-reads mapping to a junction (N_1), as well as the number of non-split-reads fully and uniquely contained within the interval surrounding the same exon–intron junction (N_2). This interval is centred of the exon–intron boundary and its size is defined based on the length of RNA-seq, in which size is equal to: (2 × length of RNA-seq reads) − 4. Percentage of 'splicing in' is then defined by the $(N_1/(N_1+N_2))$ ratio multiplied by 100. Fold changes are calculated based on percentage of splicing values. Quality of split-read alignments was visually inspected using IGV browser (version 2.3.77)[49], and low-quality alignments containing <3 nucleotides overlapping one side of the splice junction were discarded. For *P*-element splicing analysis, biological replicates generated in this study were individually analysed and used to calculate average of percentage of splicing and standard deviation. For Gypsy analysis, raw RNA-seq data sets from OSCs were retrieved from the GEO database, with accession numbers GSE47006[7], GSE74097[9] and GSE81434[25]. Given the distance between Gypsy *env* donor and acceptor sites (4,982 bp), the percentage of splicing for each exon–intron boundary was individually calculated as described above. Data obtained for the same tested conditions were pooled whenever possible.

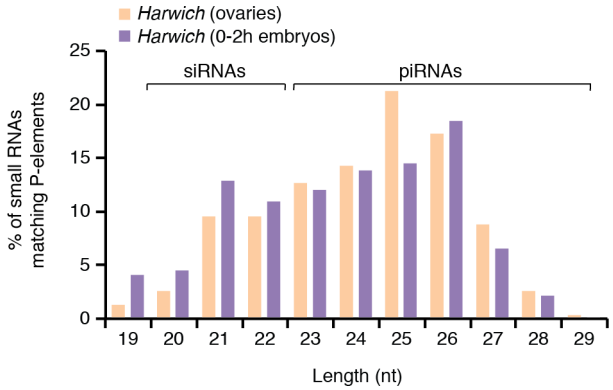
Analysis of transcriptionally active *P*-element insertions was performed on the Galaxy web-based platform using RNA-seq data from non-dysgenic and dysgenic progeny and previously described method to identify non-reference transposon insertions using split-reads[50]. In brief, paired-end RNA-seq data (100-nucleotide-long reads) were mapped to the *Drosophila* genome (dm3) using Bowtie2. Next, unmapped reads were aligned to the *P*-element consensus sequence using Bowtie2, with sensitive-local settings to obtain soft clip alignments. Reads with <20 nucleotide mapping to *P*-element extremities were removed, and the remaining reads were then treated to remove all *P*-element-derived sequences and mapped to the *Drosophila* genome (dm3) using USCS genome Browser BLAT Search (http://genome.ucsc.edu/cgi-bin/hgBlat). The number of RNA-seq split-reads mapping to a given locus in relation to total number of split-reads (excluding singletons) was used to determine the relative expression of individual insertions.

ChIP–seq analysis. ChIP–seq analysis was performed on the Galaxy web-based platform following previously described methods[41]. In brief, Illumina adaptors and other low quality sequences were removed using the 'Trim Galore!' package. ChIP–seq input and immunoprecipitation filtered data were mapped to the *Drosophila* genome (dm3) and transposon consensus sequences using Bowtie2, using the -k = 1 parameter. For the *Drosophila* genome, peak calling and signal normalization (Poisson *P* value [ppois] correction) was performed using MACS2[51]. Heat maps were generated using SeqPlots, using default parameters (http://przemol.github.io/seqplots/). For transposon consensus sequences, input-corrected H3K9me3 occupancy was normalized to the total number of genome-mapped aligned reads and represented as reads per million (RPM).

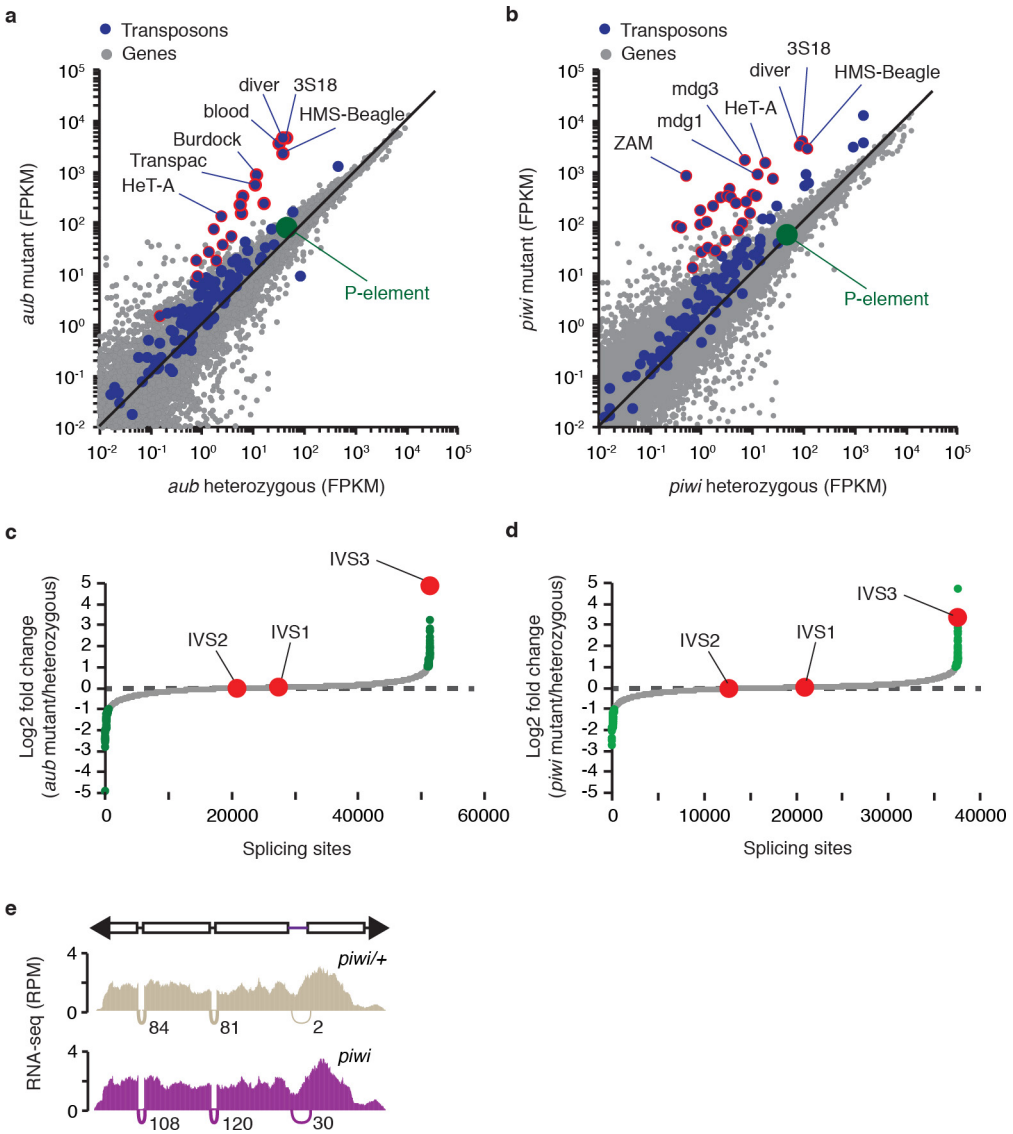
DNA-seq analysis. Identification of *P*-element insertions in the *Harwich* strain was performed using the input data from ChIP–seq and the Genome-seq pipeline provided in the piPipes package[41], as previously described[52]. In brief, genomic sequencing data were first mapped to the *Drosophila* genome (dm3) using BWA[53]. Then, using the consensus *P*-element sequence (GenBank X06779:996..3902) as a template, TEMP[54] was used to locate *P*-element insertions in the genome. This analysis was individually performed in four independent repeats, provided by two biological replicates of each progeny originated from reciprocal crosses between w^{1118} (*P*-element-devoid strain) and *Harwich*. Shared insertions identified in all four experiments were used to reconstruct the set of *P*-element euchromatic insertion existing in the *Harwich* strain (Supplementary Table 1). Euchromatic coordinates were obtained from ref. 22.

Data availability. Sequencing data generated during the current study are available in the NCBI Gene Expression Omnibus (GEO) repository under accession number GSE103582.

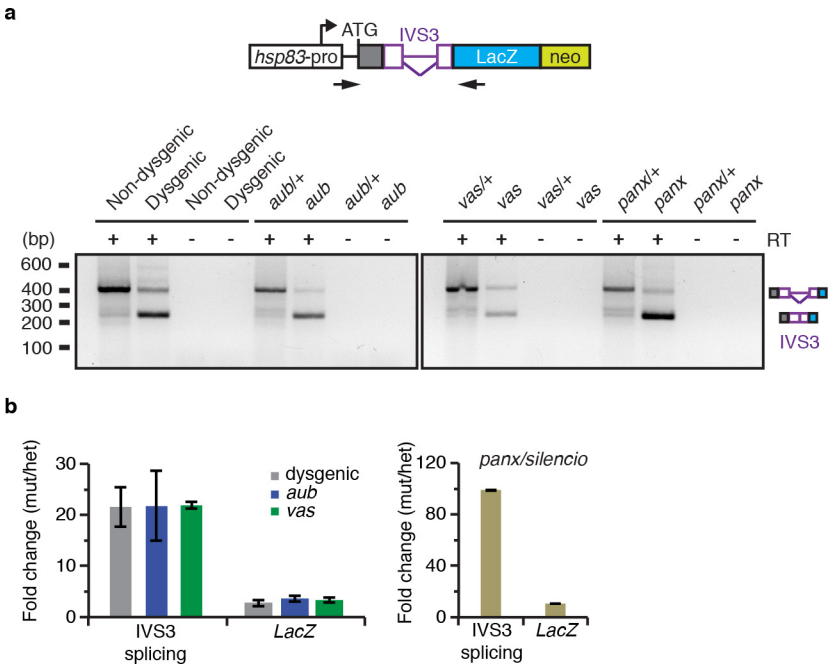
31. Sano, H., Renault, A. D. & Lehmann, R. Control of lateral migration and germ cell elimination by the *Drosophila melanogaster* lipid phosphate phosphatases Wunen and Wunen 2. *J. Cell Biol.* **171,** 675–683 (2005).
32. Van Doren, M., Williamson, A. L. & Lehmann, R. Regulation of zygotic gene expression in *Drosophila* primordial germ cells. *Curr. Biol.* **8,** 243–246 (1998).
33. Styhler, S., Nakamura, A., Swan, A., Suter, B. & Lasko, P. vasa is required for GURKEN accumulation in the oocyte, and is involved in oocyte differentiation and germline cyst development. *Development* **125,** 1569–1578 (1998).
34. Seifert, J. R. K. & Lehmann, R. Drosophila primordial germ cell migration requires epithelial remodeling of the endoderm. *Development* **139,** 2101–2106 (2012).
35. Maimon, I. & Gilboa, L. Dissection and staining of *Drosophila* larval ovaries. *J. Vis. Exp.* **13,** 2537 (2011).
36. Trcek, T. *et al.* *Drosophila* germ granules are structured and contain homotypic mRNA clusters. *Nat. Commun.* **6,** 7962 (2015).
37. Ricardo, S. & Lehmann, R. An ABC transporter controls export of a *Drosophila* germ cell attractant. *Science* **323,** 943–946 (2009).
38. Teixeira, F. K. *et al.* ATP synthase promotes germ cell differentiation independent of oxidative phosphorylation. *Nat. Cell Biol.* **17,** 689–696 (2015).
39. Rozhkov, N. V., Hammell, M. & Hannon, G. J. Multiple roles for Piwi in silencing *Drosophila* transposons. *Genes Dev.* **27,** 400–412 (2013).
40. Klattenhoff, C. *et al.* The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* **138,** 1137–1149 (2009).
41. Han, B. W., Wang, W., Zamore, P. D. & Weng, Z. piPipes: a set of pipelines for piRNA and transposon analysis via small RNA-seq, RNA-seq, degradome- and CAGE-seq, ChIP-seq and genomic DNA sequencing. *Bioinformatics* **31,** 593–595 (2015).
42. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).
43. Roberts, A. & Pachter, L. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat. Methods* **10,** 71–73 (2013).
44. Wang, Q. & Rio, D. The Junction Usage Model (JUM): A method for comprehensive annotation-free differential analysis of tissue-specific global alternative pre-mRNA splicing patterns. Preprint at bioRxiv https://doi.org/10.1101/116863 (2017).
45. Wang, Q. *et al.* The PSI-U1 snRNP interaction regulates male mating behavior in *Drosophila*. *Proc. Natl Acad. Sci. USA* **113,** 5269–5274 (2016).
46. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
47. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* **44** (W1), W3–W10 (2016).
48. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).
49. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29,** 24–26 (2011).
50. Quadrana, L. *et al.* The *Arabidopsis thaliana* mobilome and its impact at the species level. *eLife* **5,** 6919 (2016).
51. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
52. Khurana, J. S. *et al.* Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell* **147,** 1551–1563 (2011).
53. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
54. Zhuang, J., Wang, J., Theurkauf, W. & Weng, Z. TEMP: a computational method for analyzing transposable element polymorphism in populations. *Nucleic Acids Res.* **42,** 6826–6838 (2014).



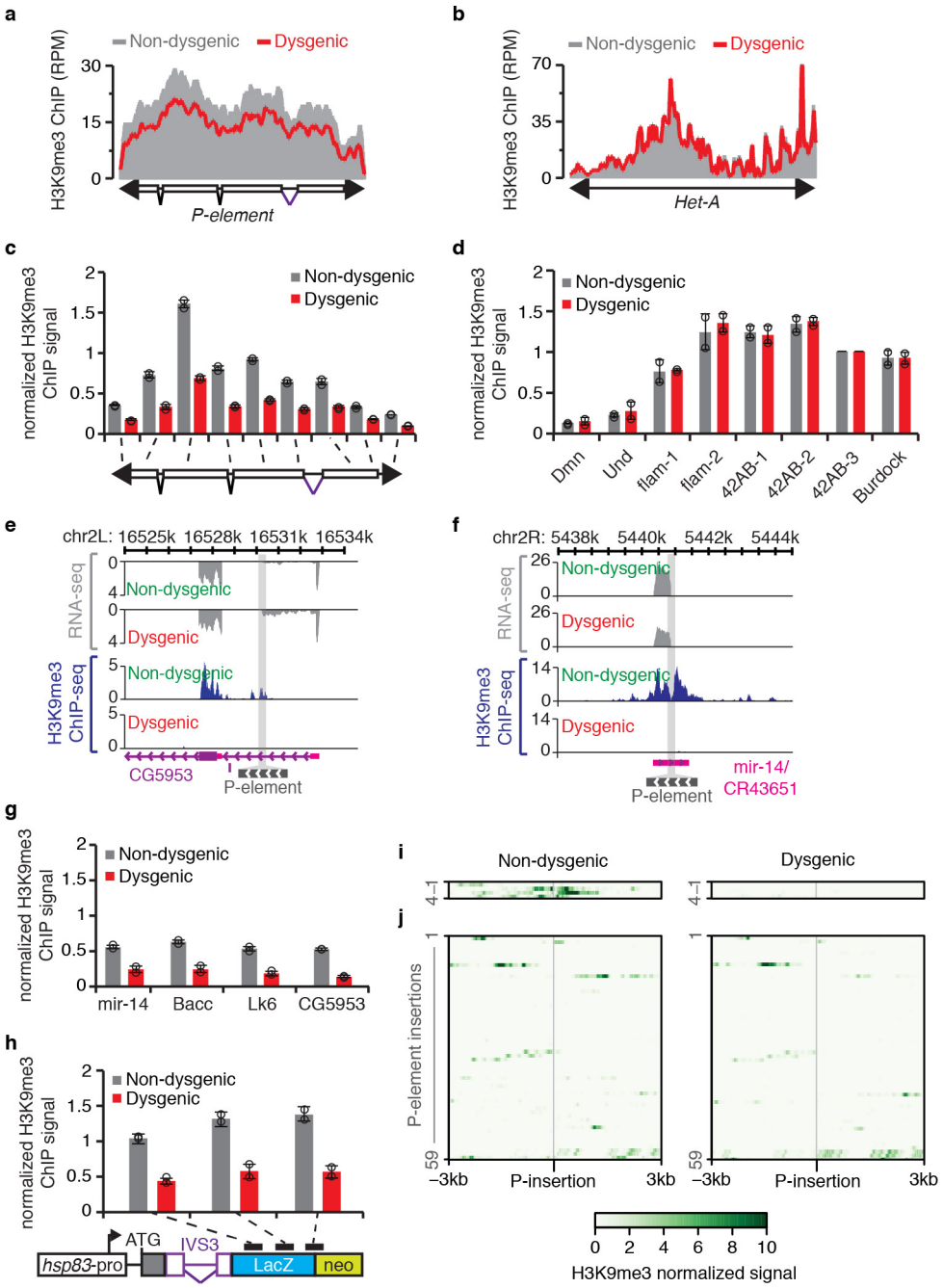
Extended Data Figure 1 | Size distribution of small RNAs matching *P*-elements in small RNA libraries. Libraries were obtained from *Harwich* ovaries (orange) or from 0–2-h-old embryos laid by *Harwich* females (purple). Analysis was performed on data from ref. 16. siRNA (20–22 nucleotides long) and piRNA (23–29 nucleotides long) populations are indicated.



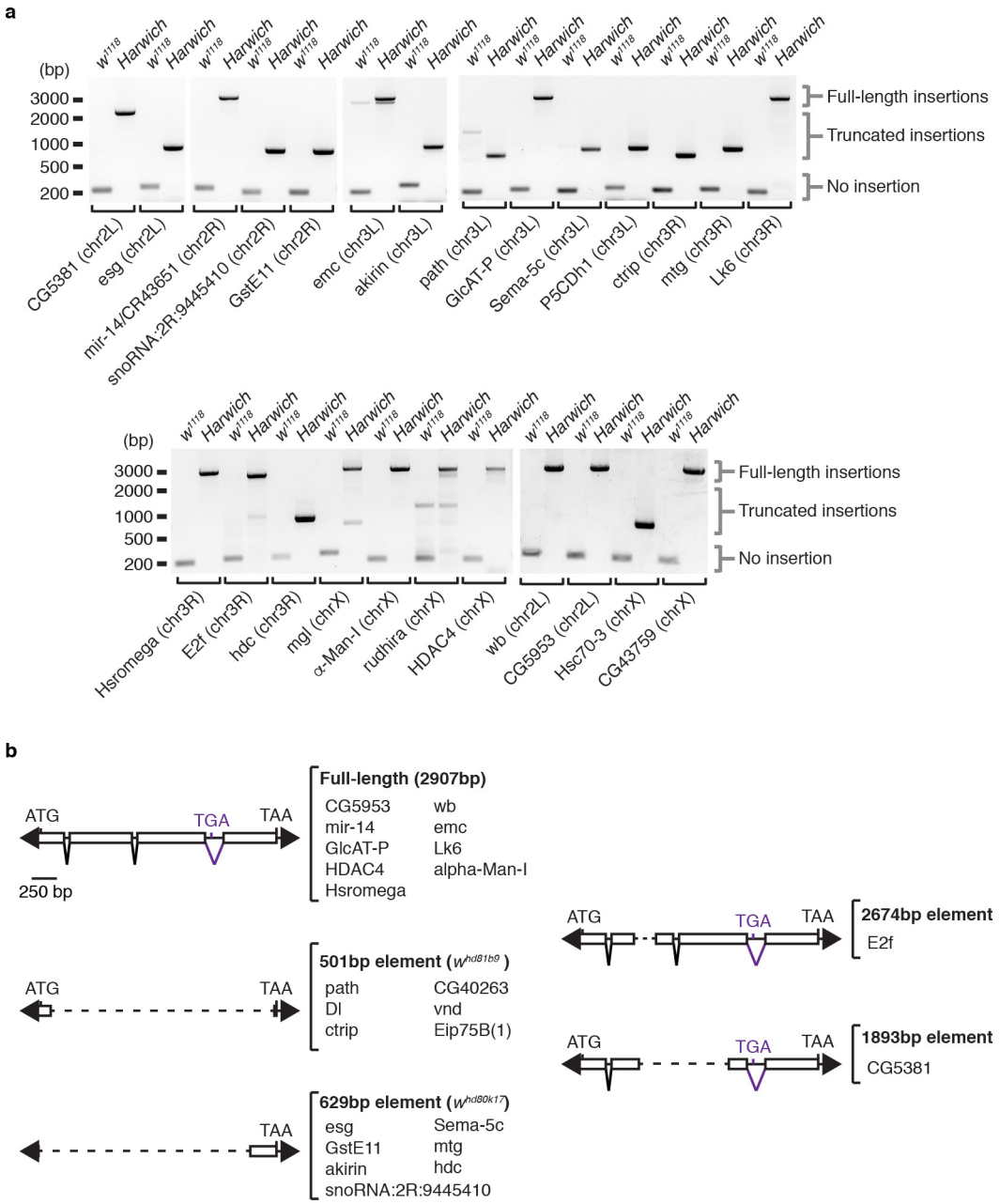
Extended Data Figure 2 | *P*-element mRNA steady-state levels do not change in piRNA mutants in comparison to respective heterozygous. **a**, **b**, Scatterplots showing the expression of genes (grey dots) and transposons (blue dots), as measured by RNA-seq analysis (expressed in fragments per kilobase per million fragments (FPKM), $\log_{10}$), in *aub*/+ heterozygous versus *aub* mutant adult ovaries (**a**) and *piwi*/+ heterozygous versus *piwi* mutant adult ovaries (**b**) comparisons. *P*-element expression is shown in green. Transposons with >10-fold change in RNA abundance are outlined in red. Experiments were repeated twice with similar results. **c**, **d**, Genome-wide analysis of splicing changes in *aub*/+ heterozygous versus *aub* mutant adult ovaries (**c**) and *piwi*/+ heterozygous versus *piwi* mutant adult ovaries (**d**) comparisons. Quantification of splicing changes was performed using RNA-seq data and the JUM method[44,45]. Results are expressed as $\log_2$ fold changes in splicing (mutant/heterozygous). Grey dots represent individual splice junctions identified, sorted by fold change values. Green dots represent splice junctions with statistically significant changes in heterozygous versus mutant comparisons (adjusted $P < 0.05$). Fold changes for *P*-element splice junctions (IVS1, IVS2 and IVS3) are presented in red. Note that approximately 70% of genic splice junctions showing statistically significant changes in mutant comparisons (green dots) are located in the same chromosome as the inducing mutations (second chromosome), and may be due to genetic background differences. **e**, Density plots for normalized strand-specific mRNA steady-state levels (measured by RNA-seq) over consensus *P*-element sequence (top diagram) in *piwi*/+ heterozygous (beige, top plot) and *piwi* mutant (purple, bottom plot) adult ovaries. The number and position of split-reads (represented by arcs that connect exons) observed for IVS1, IVS2 and IVS3 splicing junctions are shown below each density plot. Experiments were repeated twice with similar results.



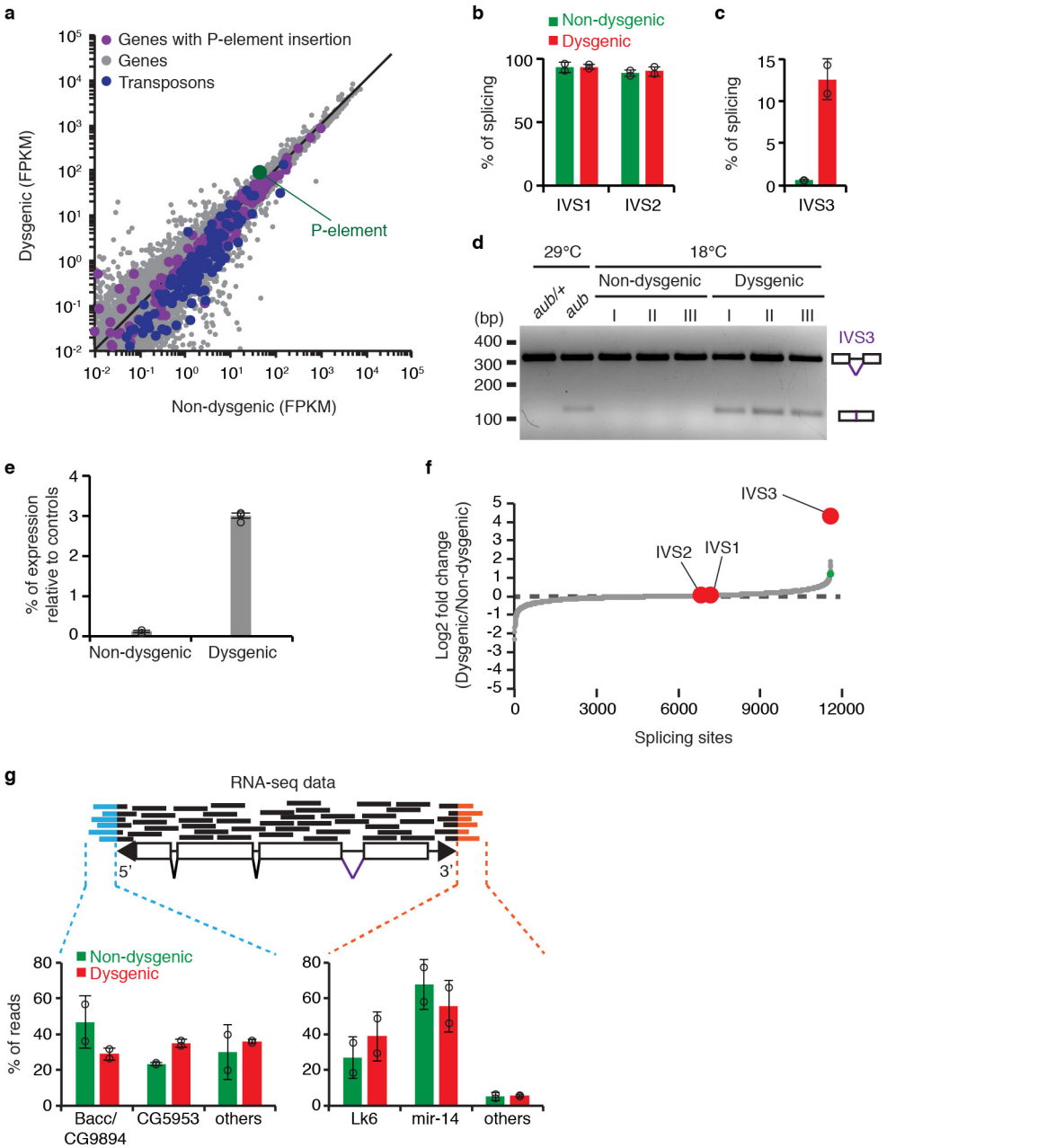
Extended Data Figure 3 | *P*-cytotype and piRNA pathway components involved in piRNA biogenesis and targeting regulate the splicing of the IVS3 transgenic reporter *in vivo*. **a**, Ethidium bromide-stained gel displaying RT–PCR reactions with primers flanking the transgenic reporter IVS3 intron. Analysis was performed with adult ovaries of non-dysgenics and dysgenic flies grown at 18 °C, or with adult ovaries of heterozygous and mutants for the piRNA components *aub*, *vas* and *Panx*. Mutant analyses were performed in a *Harwich* background, at 29 °C. Size scale in base pairs is presented for each gel. Control reactions omitting reverse transcriptase (RT−) are also presented. Diagram of IVS3 transgenic reporter (as in Fig. 3a) and primers (arrows) used in RT–PCR reactions are depicted in the top of the ethidium bromide-stained gel. Experiments were repeated at least twice with similar results. For gel source data, see Supplementary Fig. 1. **b**, RT–qPCR analysis using adult ovaries of F_1 progeny carrying the IVS3 reporter, probing spliced (IVS3 splicing, quantified using primers that specifically anneal to spliced transgenic transcripts) and total (*LacZ*, quantified using primers that anneal within the *LacZ* coding sequence) IVS3 reporter transgenic mRNA levels. F_1 progeny originated either from reciprocal crosses between *Harwich* and w^{1118} strain flies, or in *aub*/+ heterozygous, *aub* mutant, *vas*/+ heterozygous, *vas* mutant, *panx*/+ heterozygous, and *panx* mutant. Results are presented as mean fold changes in the mutants (or in dysgenic) in relation to the respective heterozygous siblings (or non-dysgenic) ± s.d. ($n \geq 2$ independent biological replicate experiments).



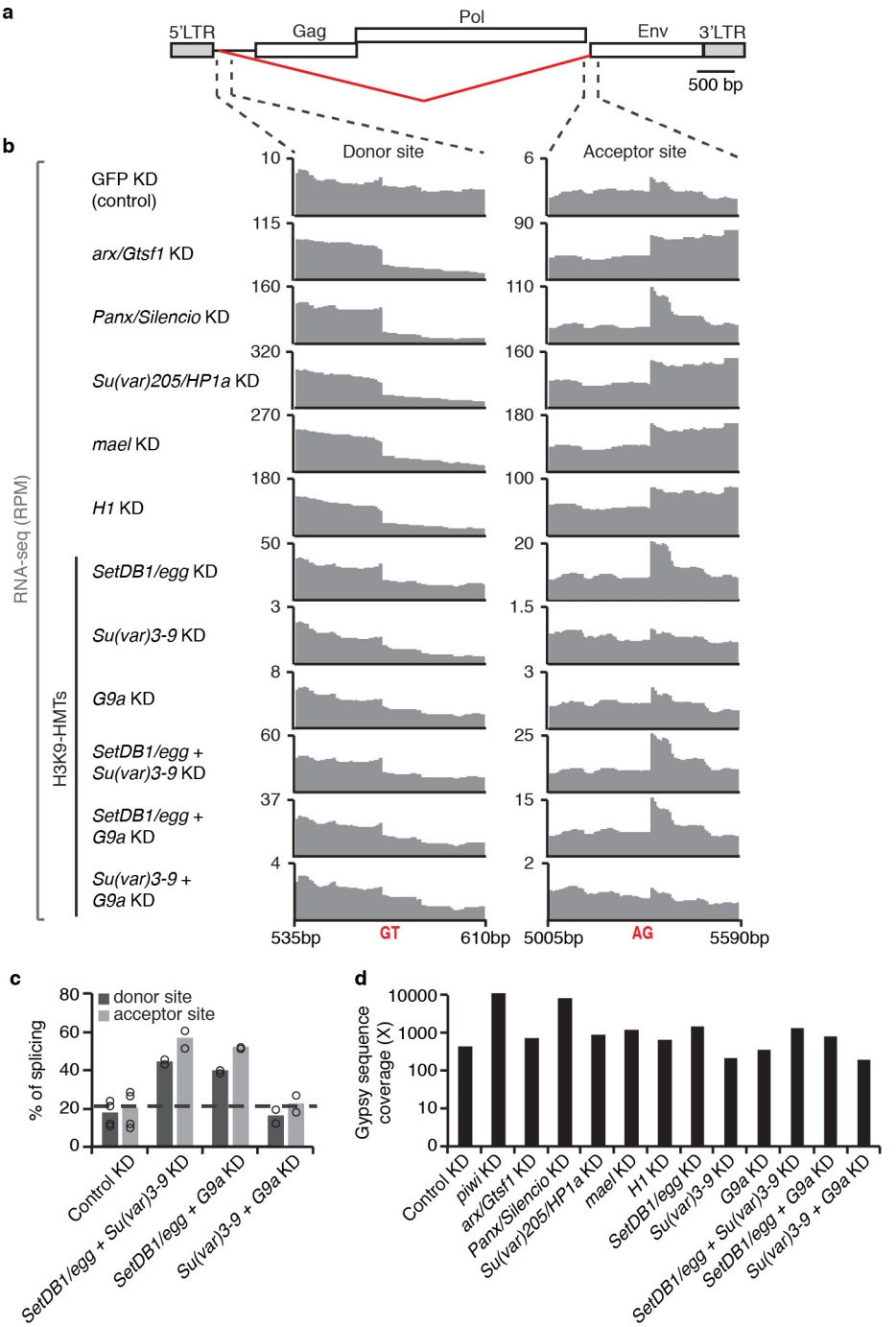
Extended Data Figure 4 | Loss of H3K9me3 in dysgenic progeny is restricted to transcriptionally active *P*-elements. a, **b**, Density plots for normalized H3K9me3 ChIP–seq signals over consensus *P*-element (**a**) and Het-A (**b**) sequences in non-dysgenic (grey) and dysgenic (red) progeny. Experiments were performed with adult ovaries of flies grown at 18 °C. Experiments were repeated twice with similar results. **c**, **d**, H3K9me3 ChIP–qPCR analysis on *P*-element (**c**) and controls (**d**) in non-dysgenic (grey) and dysgenic (red) progeny. Bars represent means of H3K9me3 signal, normalized to a control genomic region in the *42AB* locus (*42AB-3*) ± s.d. ($n = 2$ independent biological replicate experiments). **d**, *und* and *Dmn* are negative controls. **e**, **f**, Genome browser view of two of the *P*-element insertions showing transcriptional activity. Normalized RNA-seq and H3K9me3 ChIP signals are presented in grey and blue, respectively. Grey bar crossing the plots represents *P*-element insertion site. Chromosome coordinates are at the top. Annotation is at the bottom: purple boxes, coding exons; pink boxes, untranslated regions; purple lines, introns; grey box, *P*-element insertion. View showing *P*-element insertion into *CG5953* (**e**) and *CR43651* (**f**) genes. Experiments were repeated twice with similar results. **g**, **h**, ChIP–qPCR analysis on sequences flanking transcriptionally active *P*-element insertions (**g**) and on the IVS3 transgenic reporter (**h**) in non-dysgenic (grey) and dysgenic (red) progeny. Bars represent means of H3K9me3 signal, normalized to a control genomic region in the *42AB* locus (*42AB-3*), ± s.d. ($n = 2$ independent biological replicate experiments). **h**, Diagram of IVS3 transgenic reporter is presented below the graph. Black rectangles indicate regions probed by ChIP–qPCR. **i**, **j**, Heat maps depicting normalized H3K9me3 levels obtained by ChIP–seq analysis within the 6-kb flanking *P*-element insertions in non-dysgenic (left) and dysgenic (right) progeny. Experiments were repeated twice with similar results. **i**, 4 *P*-element insertions with evidence of transcriptional activity. **j**, 59 *P*-element insertions with no evidence of transcriptional activity. All experiments were performed with ovaries of adult progeny from non-dysgenic and dysgenic crosses grown at 18 °C.



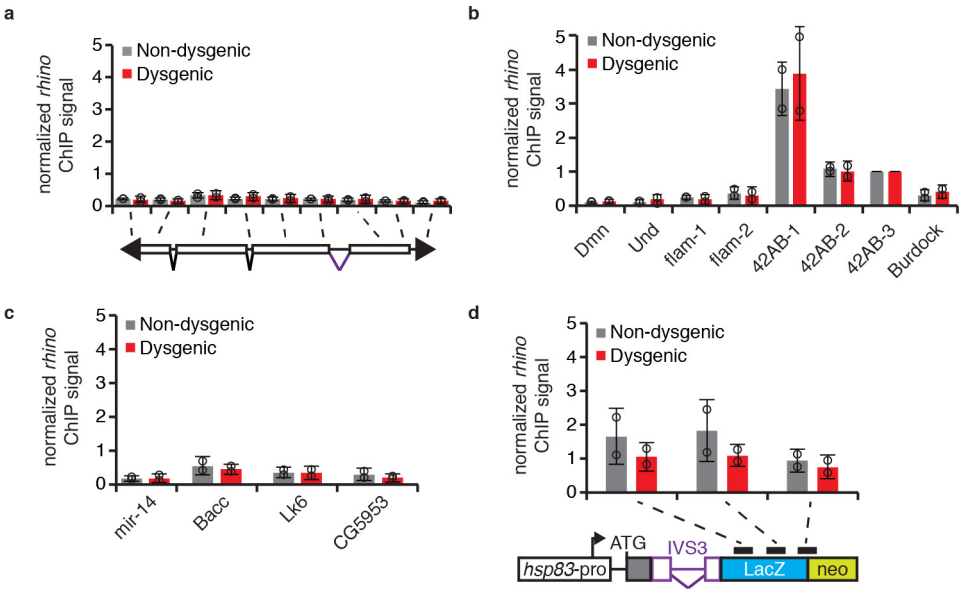
Extended Data Figure 5 | Characterization of euchromatic *P*-element insertion in *Harwich* strain. a, Ethidium bromide-stained gels displaying genomic PCR reactions with 25 primer sets flanking *P*-element insertions uncovered by DNA-seq analysis. PCR reactions were performed with genomic DNA extract from w^{1118} and *Harwich* pools containing 20 adult females each. Size scale in base pairs is presented on the right side of each gel. Amplicon sizes representing absence (no insertion) or presence (truncated or full-length elements) of *P*-element insertion are indicated on the left of each gel. Targeted insertion and chromosome localization are displayed at the bottom of each gel. Experiments were repeated twice with similar results. For gel source data, see Supplementary Fig. 1. **b**, Schematic representation of five structurally different elements regrouping 24 *P*-element insertions characterized by DNA sequencing in the *Harwich* strain. Elements size, as well as a list of respective insertions is indicated on the right of each diagram. Arrowheads represent terminal inverted repeats; boxes, exons; inverted triangles, introns. Dashed lines represent internal deletions.



Extended Data Figure 6 | Analysis of *P*-element expression and splicing in adult ovaries of non-dysgenic and dysgenic progeny grown at 18 °C. a, Scatterplot showing the expression of genes (grey dots) and transposons (blue dots), as measured by RNA-seq analysis (log_{10}), in adult ovaries of non-dysgenic versus dysgenic progeny grown at 18 °C. *P*-element expression is shown in green. Genes containing a *P*-element insertion in *Harwich* strain are depicted in purple. **b**, **c**, Percentage of splicing for *P*-element IVS1, IVS2 (**b**), and IVS3 (**c**) splicing junctions as determined by RNA-seq analysis in non-dysgenic (green) and dysgenic (red) adult ovaries. Bars represent percentage of splicing, calculated as the number of split-reads for each splicing junction normalized to the total number of reads mapping to the same junction. Data are mean ± s.d. ($n = 2$ independent biological replicate experiments). **d**, Ethidium bromide-stained gel displaying RT–PCR reactions with primers flanking the *P*-element IVS3 intron in adult ovaries of non-dysgenic and dysgenic progeny grown at 18 °C, as well as *aub*/+ heterozygous and *aub* mutant grown at 29 °C. Size scale in base pairs is presented. As shown, experiments were repeated three times with similar results. For gel source data, see Supplementary Fig. 1. **e**, RT–qPCR analysis testing accumulation of IVS3-spliced mRNA on non-dysgenic and dysgenic progeny (ovaries) grown at 18 °C. Results are expressed as mean of percentage of expression relative to controls ± s.d. ($n = 3$ independent biological replicate experiments). **f**, Genome-wide analysis of splicing changes in in adult ovaries of non-dysgenic versus dysgenic progeny grown at 18 °C. Quantification of splicing changes was performed using RNA-seq data and the JUM method[44,45]. Results are expressed as log_2 fold changes in splicing (dysgenic/non-dysgenic). Grey dots represent individual splice junctions identified, sorted by fold change values. Green dots represent splice junctions with statistically significant changes in heterozygous versus mutant comparisons (adjusted $P < 0.05$). Fold changes for *P*-element splice junctions (IVS1, IVS2 and IVS3) are presented in red. **g**, Analysis of RNA-seq data obtained from non-dysgenic (green) and dysgenic (red) progeny (adult ovaries) to identify transcriptionally active *P*-element insertions. Reads partially mapping to the *P*-element extremities were subsequently mapped to the *Drosophila* genome. Results are expressed as percentage of reads mapping to a given genomic location in relation to total reads. Data are mean ± s.d. ($n = 2$ independent biological replicate experiments). All experiments were performed with ovaries of adult progeny from non-dysgenic and dysgenic crosses grown at 18 °C.



Extended Data Figure 7 | Analysis of Gypsy splicing in OSCs. **a**, Diagram of Gypsy retrotransposon, as in Fig. 4a. **b**, RNA-seq signal at the Gypsy splicing donor and acceptor sites in representative control (GFP knockdown) and knockdowns of *arx*, *Panx*, *Su(var)205*, *mael*, *H1*, *SetDB1*, *Su(var)3-9* and *G9a*. Data for double knockdowns of *SetDB1* + *Su(var)3-9*, *SetDB1* + *G9a*, and *Su(var.)3-9* + *G9a* are also shown. With the exception of *arx*, *Su(var)205*, *mael* and *H1*[7,25], experiments were repeated twice with similar results. **c**, Percentage of splicing for Gypsy donor and acceptor splicing sites as determined by RNA-seq analysis performed in OSCs double knockdowns of H3K9 methyltransferases. Bars represent the number of split-reads for the *env* donor and acceptor splicing junctions normalized to the total number of sense Gypsy reads mapping to the same junction. Results are represented as means. Experiments were repeated twice with similar results. **d**, Coverage of Gypsy consensus sequence by RNA-seq data. Coverage was calculated as: (number of reads matching Gypsy consensus sequence × read length in nucleotides)/length of consensus Gypsy sequence in nucleotides (7,469 nucleotides). Raw datasets are from refs 7, 9 and 25.



Extended Data Figure 8 | *rhino* ChIP–qPCR analysis in adult ovaries of non-dysgenic and dysgenic progeny grown at 18 °C. *rhino* ChIP–qPCR analysis on *P*-element (**a**), controls (**b**), sequences flanking transcriptionally active *P*-insertions in the *Harwich* strain (**c**), and transgenic IVS3 splicing reporter (**d**) in non-dysgenic (grey) and dysgenic (red) progeny. Bars represent means of Rhino signal, normalized to a control genomic region in the *42AB* locus (*42AB-3*), ± s.d. ($n = 2$ independent biological replicate experiments). **b**, *Dmn*, *und* and *flam* are negative controls; *42AB* is a positive control. **d**, Diagram of transgenic IVS3 splicing reporter is presented below the graph. Black rectangles indicate regions probed by ChIP–qPCR. All experiments were performed with ovaries of adult progeny from non-dysgenic and dysgenic crosses grown at 18 °C.

nature research

Corresponding author(s): Ruth Lehmann, Felipe K Teixeira

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

No statistical methods were used to predetermine sample size. Information about sample collection and replication is presented in the Methods section.

2. Data exclusions

Describe any data exclusions.

There are no data exclusion in this manuscript.

3. Replication

Describe whether the experimental findings were reliably reproduced.

Experiments were performed in at least four replicates (at least two technical replicates for at least each of the two biological replicates) and successfully reproduced.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Experiments were neither intentionally randomized nor intentionally ordered.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

With the exception of NGS procedures, investigators were not blinded to allocation during experiments and outcome assessment.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	All used software and codes were previously published by others and are publicly available at GitHub or Galaxy platform, as following: piPipes package (version 1.5.0; https://github.com/bowhan/piPipes); Junction Usage Model (JUM) package (version 1.3.2; https://github.com/qqwang-berkeley/JUM); Galaxy (www.usegalaxy.org).

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	There are no restrictions on availability.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	Antibodies used have been previously validated and published by others: rabbit anti-Vasa serum (R. Lehmann, NYU/HHMI), mouse anti-1B1 (DSHB), rabbit anti-H3K9me3 (Abcam, ab8898), guinea pig anti-rhino (W. Theurkauf, UMass).

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used.
b. Describe the method of cell line authentication used.	No eukaryotic cell lines were used.
c. Report whether the cell lines were tested for mycoplasma contamination.	No eukaryotic cell lines were used.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No eukaryotic cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	All experiments were performed with Drosophila melanogaster. Standard laboratory procedures have been applied throughout the study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	n/a.

nature research

Corresponding author(s): Ruth Lehmann, Felipe K Teixeira

☐ Initial submission ☐ Revised version ☒ Final submission

ChIP-seq Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data deposition

1. For all ChIP-seq data:

☒ a. Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ b. Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

2. Provide all necessary reviewer access links.
The entry may remain private before publication.

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE103582

3. Provide a list of all files available in the database submission.

ChIPseq_input_non_dysgenic_rep1.fastq.gz
ChIPseq_input_non_dysgenic_rep2.fastq.gz
ChIPseq_input_dysgenic_rep1.fastq.gz
ChIPseq_input_dysgenic_rep2.fastq.gz
ChIPseq_antiH3K9me3_non_dysgenic_rep1.fastq.gz
ChIPseq_antiH3K9me3_non_dysgenic_rep2.fastq.gz
ChIPseq_antiH3K9me3_dysgenic_rep1.fastq.gz
ChIPseq_antiH3K9me3_dysgenic_rep2.fastq.gz
ChIPseq_input_non_dysgenic_rep1.bedgraph
ChIPseq_antiH3K9me3_non_dysgenic_rep1.bedgraph
ChIPseq_input_non_dysgenic_rep2.bedgraph
ChIPseq_antiH3K9me3_non_dysgenic_rep2.bedgraph
ChIPseq_input_dysgenic_rep1.bedgraph
ChIPseq_antiH3K9me3_dysgenic_rep1.bedgraph
ChIPseq_input_dysgenic_rep2.bedgraph
ChIPseq_antiH3K9me3_dysgenic_rep2.bedgraph

4. If available, provide a link to an anonymized genome browser session (e.g. UCSC).

Bedgraph files are compatible with UCSC genome browser.

Methodological details

5. Describe the experimental replicates.

Two samples (non-dysgenic and dysgenic F1 progeny) were analyzed in biological replicates (4 samples total).

6. Describe the sequencing depth for each experiment.

ChIPseq_input_non_dysgenic_rep1.fastq.gz: 59936919 total reads; 57635794 reads mapped to dm3.
ChIPseq_input_non_dysgenic_rep2.fastq.gz: 53291036 total reads; 51259422 reads mapped to dm3.
ChIPseq_input_dysgenic_rep1.fastq.gz: 57825933 total reads; 55685779 reads mapped to dm3.
ChIPseq_input_dysgenic_rep2.fastq.gz: 52409949 total reads; 50455595 reads mapped to dm3.
ChIPseq_antiH3K9me3_non_dysgenic_rep1.fastq.gz: 37136316 total reads; 20506013 reads mapped to dm3.
ChIPseq_antiH3K9me3_non_dysgenic_rep2.fastq.gz: 48394679 total reads; 35108397 reads mapped to dm3.
ChIPseq_antiH3K9me3_dysgenic_rep1.fastq.gz: 33028992 total reads; 22921696 reads mapped to dm3.
ChIPseq_antiH3K9me3_dysgenic_rep2.fastq.gz: 40733662 total reads; 31479022 reads mapped to dm3.

7. Describe the antibodies used for the ChIP-seq experiments.	anti-H3K9me3 (Abcam, ab8898).
8. Describe the peak calling parameters.	Peak calling and signal normalization (Poisson p-value [ppois] correction) was performed using MACS2 (Zhang et al, 2008 - Genome Biol) with standard parameters.
9. Describe the methods used to ensure data quality.	Prior to sequencing, IP and input samples were tested by qPCR. ChIP-seq data was visually inspected in the genome browser to confirm H3K9me3 signal enrichment in heterochromatic regions and transposable elements.
10. Describe the software used to collect and analyze the ChIP-seq data.	ChIP-seq analysis was performed on the Galaxy web-based platform following previously described methods (Han et al, 2015 - Bioinformatics). Briefly, Illumina adaptors and other low quality sequences were removed using the 'Trim Galore!' package. ChIP-seq input and IP filtered data were mapped to the Drosophila genome (dm3) and transposon consensus sequences using Bowtie2 (Langmead et al, 2012 - Nat Methods), using the -k = 1 parameter. For the Drosophila genome, peak calling and signal normalization (Poisson p-value [ppois] correction) was performed using MACS2 (Zhang et al, 2008 - Genome Biology). Heatmaps were generated using SeqPlots, using default parameters (http://przemol.github.io/seqplots/). For transposon consensus sequences, input-corrected H3K9me3 occupancy was normalized to the total number of genomemapped aligned reads and represented as reads per million (RPM).