The transcription factor Traffic jam orchestrates the somatic piRNA pathway in *Drosophila* ovaries

Azad Alizada[2,*], Aline Martins[1,*], Nolwenn Mouniée[1,*], Julia V. Rodriguez Suarez[2,*], Benjamin Bertin[1], Nathalie Gueguen[1], Vincent Mirouse[1], Stéphanie Maupetit-Mehouas[1], Austin J. Rivera[3], Nelson C. Lau[3], Gregory J. Hannon[2,#], Benjamin Czech Nicholson[2,#], Emilie Brasset[1,#]

[*] These authors contributed equally to this work.

[#] Corresponding authors:
GJH (greg.hannon@cruk.cam.ac.uk)
BCN (ben.nicholson@cruk.cam.ac.uk)
EB (emilie.brasset@uca.fr)

Affiliations:

[1] iGReD, Université Clermont Auvergne, CNRS, INSERM, Faculté de Médecine, 63000 Clermont-Ferrand, France

[2] Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge CB2 0RE, UK

[3] Department of Biochemistry and Cell Biology, Boston University Chobanian and Avedisian School of Medicine, Boston, MA, USA

Keywords
piRNA pathway, *Drosophila*, oogenesis, transposable elements, transcription factors

Highlights

- Traffic jam (Tj) acts as a master regulator of the somatic piRNA pathway in *Drosophila*.
- Tj directly controls the expression of the *flamenco* piRNA cluster, crucial for transposon silencing.
- Tj regulates a network of piRNA pathway genes, mirroring the gene-regulatory mechanism of A-MYB in the mouse testis.
- Cis-regulatory elements with Tj motifs are arranged in a palindromic sequence

Abstract

Transposable elements (TEs) pose a threat to genome integrity, and the piRNA pathway in animal gonads plays a crucial role in silencing TE activity. While the transcriptional regulation of the piRNA pathway components in germ cells has been documented in mice and flies, the mechanisms orchestrating the transcriptional program of the somatic piRNA pathway in *Drosophila* ovaries remains unresolved. Here, we demonstrate that Traffic jam (Tj), an orthologue of a large Maf transcription factor in mammals, is a master regulator of the piRNA pathway in ovarian somatic cells, playing a crucial role in maintaining TE silencing and genomic integrity in somatic tissues. We show that Tj directly binds to the promoters of somatic-enriched piRNA factors such as *fs(1)Yb*, *nxf2*, *panx*, and *armi*, as well as the *flamenco* piRNA cluster, a major locus for TE silencing in somatic cells. Depletion of Tj in somatic follicle cells results in a significant downregulation of these piRNA factors, a complete loss of *flam* expression and de-repression of *gypsy*-family TEs, which have gained the ability to activate in ovarian somatic cells allowing them to infect germ cells and be transmitted to future generations. We have identified an enhancer carrying Tj binding motifs located downstream of the *flam* promoter that is essential for robust and tissue-specific *flam* expression in somatic follicle cells of the adult ovary. This work uncovers a previously unappreciated layer of transcriptional regulation of the piRNA pathway, and we propose that the arms race between the host and TEs has driven the evolution of promoters in piRNA genes and clusters to respond to a unique transcription factor thereby ensuring efficient silencing of *gypsy*-family TEs.

Introduction

Transposable elements (TEs) are DNA sequences with the ability to move or replicate to new positions within the host genome[1]. Succeeding waves of TE mobilization and repression have allowed TEs to accumulate in the genomes of nearly all organisms. Approximately 20% of the *Drosophila* genome and more than 50% of the human genome is composed of TE sequences[2]. While their ability to transpose makes them powerful facilitators of genome evolution, their mobilization can also have deleterious effects, such as mutations, gene disruptions, and chromosomal rearrangements[3,4]. The resulting selective pressure has driven the co-evolution of many mechanisms within the hosts to effectively counter the disruptive TE activity[5].

In animals, the PIWI-interacting RNA (piRNA) pathway is one of the key defence mechanisms against active TEs. In *Drosophila*, the piRNA pathway is crucial for maintaining genome stability, particularly in gonadal tissues, which consist of germ cells surrounded by somatic follicle cells[6–8]. Previous work has uncovered the existence of distinct piRNA pathways in germ cells and somatic follicle cells[7–9]. Germ cells express a more sophisticated piRNA system that involves both post-transcriptional silencing of transposon transcripts via the ping-pong loop, involving Aubergine, Argonaute-3 and numerous Tudor proteins and RNA binding proteins such as Qin, Tejas, Vasa, Krimper, Spn-E and BoYb, as well as co-transcriptional silencing of nascent TEs, which relies on Piwi and its cofactors Arx, and the PICTS complex composed of Panx, Nxf2, Nxt1, and Ctp[10–16]. Furthermore, piRNAs in germ cells are predominantly generated from precursor transcripts derived from discrete genomic loci called dual-strand piRNA clusters[6,17–19]. The target repertoires of these piRNA clusters are dictated by both full-length and truncated transposon copies and their transcription and export relies on non-canonical mechanisms, involving Rhino (Rhi), Deadlock (Del), Cutoff (Cuff), Moonshiner (Moon), Nxf3, and Bootlegger (Boot)[6].

To avoid the robust silencing machinery in germ cells, some TEs such as *gypsy* or ZAM have gained the ability for exclusive expression in the somatic follicle cells, form virus-like

particles and invade germ cells such as the oocyte[20–22]. This infectious capacity enables their transposition into the germline genome and hence transmission of new TE copies to future generations. As a consequence of this host-transposon arms race, somatic follicle cells have adapted a rudimentary piRNA pathway comprised of the factors required for piRNA biogenesis (e.g., Armi, Zuc, Mino, Gasz, Daed, SoYb, Vret, Shu and the soma-specific Yb) and co-transcriptional transposon silencing (e.g., Piwi, Arx, Panx, Nxf2, Nxt1, and Ctp)[8,13,14,23–27]. In addition, somatic piRNA clusters have emerged to control TE expression in the somatic follicle cells of the ovary. In contrast to the germline-active dual-strand piRNA clusters, the soma-expressed piRNA clusters are unistrand and give rise to piRNAs in antisense orientation relative to the active TE sequence. The major somatic piRNA cluster in *Drosophila* ovaries is the *flamenco* (*flam*)[6,28]. Contrary to the dual-strand piRNA clusters, *flam* is conserved across millions of years of Drosophilid evolution[29], and transcribed by a canonical mechanism that includes a promoter, splicing, and polyadenylation[30]. Furthermore, to-date, *flam* is the only known essential piRNA cluster, with *flam* mutants showing rudimentary ovaries and complete female sterility[18,31].

Despite the importance of the somatic piRNA pathway, including the *flam* cluster and the somatic piRNA factors, the gene-regulatory circuit controlling their transcription remains to be determined. Recent studies have shown that the transcription factor (TF) Ovo plays a key role in coordinating the expression of germline-specific piRNA pathway genes in the ovaries[32]. However, it remains unclear whether an equivalent TF drives the somatic piRNA pathway in the ovarian follicle cells. Here, we uncover Traffic jam (Tj), a large Maf transcription factor, as a master regulator of the somatic piRNA pathway. Tj is critical for gonad development[33] and was previously implicated in the regulation of *piwi* expression in somatic cells[34]. In this study, we extend these findings by showing that Tj regulates the expression of many somatic piRNA pathway factors, as well as the key unistrand piRNA cluster, *flamenco*. We demonstrate that Tj directly binds to the regulatory regions of these genes, driving their soma-specific expression in ovarian follicle cells. Furthermore, we identify a specific enhancer region downstream of the *flam* TSS, which harbors Tj binding sites, that is essential for its transcriptional activity in somatic cells. Our findings shed light

on the intricate transcriptional control necessary for TE silencing in the ovarian soma, offering new insights into the regulation of the soma-specific piRNA pathway.

Results

The transcription factor Traffic jam controls somatic piRNA factor expression

To uncover the regulatory mechanisms underlying the expression of somatic piRNA pathway factors, we first performed differential gene expression analysis between germline and somatic cells derived from *Drosophila* ovaries. We dissected and dissociated ovaries of transgenic *D. melanogaster* flies tagged with GFP specifically in germ cells (expressed using the *vasa* promoter) and sorted the cells using FACS into germline (Vas-GFP+) and somatic (Vas-GFP-) fractions[32]. We then extracted RNAs and generated RNA-seq libraries before carrying out a differential gene expression analysis between germline and somatic cells. This approach identified piRNA pathway genes with significant enrichment in the soma of the *Drosophila* ovary, namely*, fs(1)Yb*, *nxf2*, *panx*, and *armi* (**Figure 1A, left**). These soma-enriched piRNA factors also showed expression in ovarian somatic cells (OSCs), the cell line derived from ovary somatic follicle cells[34,35]. Of note, these piRNA factors have also been reported to function in this cell type[36], validating our approach.

Next, to identify the potential TFs controlling the expression of somatic-enriched piRNA factors (*fs(1)Yb*, *nxf2*, *panx*, and *armi)*, we performed ATAC-seq on wildtype OSCs and whole ovaries and performed motif scanning to look for transcription factor motifs that were shared between promoter ATAC-seq peaks (±500 bp of the transcription start sites) of *fs(1)Yb*, *nxf2*, *panx*, and *armi* genes (**Figure 1B and Table S1**). Surprisingly, this analysis identified only a single transcription factor motif, for Traffic jam (Tj), which was shared between the promoter ATAC-seq peaks of the four soma-enriched piRNA factors. We confirmed *in vivo* Tj binding to the Tj motifs within the promoters of *fs(1)Yb*, *nxf2*, *panx*, and *armi* (**Figure 1C**) through reanalysis of publicly available GFP-Tj ChIP-seq data from whole female flies[37]. Moreover, Tj ChIP-seq and motifs were also present within promoters of other somatic piRNA factors shared between somatic and germline cells such as *piwi* but were absent in promoters of the germline-specific piRNA factors (**Figure**

1A, right and Figure 1C), suggesting that Tj could be specifically controlling the somatic piRNA pathway genes.

To validate our hypothesis that Tj regulates the soma-expressed piRNA factors, we performed siRNA-mediated *tj* knockdowns in OSCs, followed by RT-qPCR. Depletion of Tj in OSCs resulted in a corresponding downregulation of somatic piRNA factors (**Figure 1D**). Although *tj* levels were reduced by ~40% following 48 h of *tj*-siRNA treatment, we observed a similar level of downregulation for *nxf2*, *piwi*, *armi*, and *fs(1)yb* (RT-qPCR, n=3 replicates from distinct samples, $p<0.01$) (**Figure 1D**). To corroborate these results, we performed RNA-seq from 48 h and 96 h of *tj* siRNA treatments which also showed significant downregulation of the soma-enriched piRNA factors of *fs(1)Yb*, *nxf2*, *panx*, *soYb*, and *armi* (RNA-seq, n≥3 replicates from distinct samples, p.adj.<0.01) (**Figure 1E**). These effects were stronger upon 96 h of *tj* knockdown in OSCs (**Figure S1A**). The piRNA factors shared between the somatic and germline cells which showed Tj ChIP-seq binding at their promoters (i.e., *piwi*, *vret*, *mino*, and *shu*) were similarly downregulated upon *tj* knockdowns (**Figures 1E and S1A**). These results suggest that Tj is a soma-enriched TF responsible for the regulation of the soma-expressed piRNA pathway genes in *Drosophila* ovarian somatic cells, likely via direct promoter binding.

To experimentally verify Tj's role in regulating soma-piRNA factors, we disrupted Tj activity and examined Armi expression through immunostaining in the *Drosophila* adult ovary *in vivo*. As, Tj plays a critical role in regulating the formation and function of the germline stem cell (GSC) niche and defects in its expression leads to ovarian atrophy[38,39], we abolished Tj activity in mosaic clones of follicle cells carrying a genetic null allele of Tj, the tj^{eo2} allele[33]. This allele carries a point mutation resulting in a premature stop codon resulting in a truncated version of the Tj protein, which lacks critical DNA-binding domains, including the leucine zipper and nuclear localization signal[33]. As expected, in wild-type follicle cells, Armi, a protein essential for piRNA processing and the formation of the piRNA-induced silencing complex within Yb bodies, accumulates as a distinct dot representing the Yb body at the nuclear periphery. However, when Tj activity was abolished in mosaic clones, Armi protein was completely absent, with no detectable

accumulation (**Figure 1H**). This suggests that Tj is indeed involved in regulating *armi* expression. This finding confirms that, as observed in OSCs, Tj regulates piRNA factor expression in vivo within the follicle cells of the adult ovary.

Genome-wide GFP-Tj ChIP-seq analysis revealed 7,956 Tj binding sites, which were enriched at the promoters of the soma-expressed genes (including the somatic piRNA pathway genes), while being depleted at the promoters of the germline-enriched genes, including the germline-specific piRNA pathway genes **(Figure 1A)**. This was also evident by stronger Tj ChIP-seq signals (input normalized rpkm) at promoters of somatic-enriched genes compared to germline-enriched genes (**Figure 1F**). Specifically, Tj peaks overlapped with 714 transcription start sites (TSSs; within ±500 bp) of 419 soma-expressed genes (defined using RNA-seq $\log_2$FC<-1, adjusted p.value<0.01, n=3 from FACS-sorted vasGFP- cells), representing 62% of all soma-enriched gene TSSs (n=1,142). Interestingly, the top-scoring *de novo* motif within the Tj ChIP-seq peaks had a consensus sequence of TGCTGA (MEME e-value=1.1×10^{-20}) with its reverse complement TCAGCA often found nearby (as close as by 1 bp spacing), resulting in a palindromic motif of TGCTGA_TCAGCA which was the second top-ranked motif within Tj ChIP-seq peaks (MEME e-value=1.3×10^{-16}) (**Figure 1G**), suggesting that Tj could be binding DNA as a dimer. Overall, our results indicate that Tj specifically directs the somatic ovarian program and encompasses somatic piRNA pathway components.

Tj binds to conserved motifs in the regulatory region of the piRNA cluster *flam*

Given its role as master regulator of the somatic piRNA pathway genes, we next asked whether Tj also regulates piRNA cluster transcription. The major piRNA cluster active in follicle cells is *flamenco*[6,28]. Looking at ATAC-seq signal within 1kb upstream and 5kb downstream of the *flam* TSS in *D. melanogaster* and the orthologous region in the *D. yakuba* genome revealed three conserved accessibility peaks (ATAC-seq peaks #5, #6 and #7) in the locus (**Figure 2A**). Interestingly, the conserved ATAC-seq peaks #5 and #7 had disproportionately stronger ATAC-seq signals in OSCs when compared to the whole ovaries, suggesting their somatic enrichment (**Figure 2A**). To identify enhancer

activity across the genome, we analysed data from STARR-seq (Self-Transcribing Active Regulatory Region sequencing) performed in OSCs using the genomes of *D. melanogaster* and *D. yakuba* species[40–42]. This analysis revealed enhancer activities specifically associated with the conserved ATAC-seq peak #7 in *D. melanogaster* and its corresponding orthologous ATAC-seq peak in the *D. yakuba* ovary (**Figure 2A**). We then performed transcription factor motif scanning on the conserved ATAC-seq peak #7 to uncover potential regulators of the *flam* locus (**Figures 2A and 2C**), beyond the previously reported *cubitus interruptus,* the motif of which was found within conserved ATAC-seq peak #5[43]. Motif scanning of the conserved and STARR-seq-positive ATAC-seq peak #7 from both species revealed a set of 21 conserved transcription factor motifs that were shared between *D. melanogaster* and *D. yakuba* orthologous peak regions **(Figure 2B)**. Intriguingly, the motif for Tj was among the top hits in both *Drosophila* species. Using the ovary RNA-seq data from *D. melanogaster* and *D.yakuba*, we narrowed down the candidates to six TFs that were expressed in ovaries of both species, of which, *tj*, *crc*, *br*, and *vri* showed soma-enrichment where *tj* was the top soma-enriched TF **(Figure 2C)**. In *D. melanogaster,* Tj motifs are found between positions +430 to +439 bp and +500 to +509 bp relative to the *flam* TSS. Using Tj ChIP-seq, we confirmed Tj binding to these motifs within the conserved ATAC-seq peak #7. Our results suggests that Tj may play a regulatory role not only in the expression of somatic piRNA genes but also in the control of the piRNA cluster *flam*.

Expression of *flam* in follicle cells is dependent on a regulatory region located downstream of its TSS

Previous studies using OSCs, which are derived from *Drosophila* ovarian follicle cells, have identified the major transcriptional start site, and the minimal promoter of *flam*[43]*.* To further confirm Tj's direct role in regulating *flam* expression, we explored *in vivo* the minimal sequence necessary to drive follicle cell-specific expression of *flam* in adult ovary*.* We generated reporter transgenes carrying the *tomato* reporter gene under the control of different *flam* promoter regions and followed Tomato expression in fly ovaries by immunostaining and RT-qPCR (**Figures 3A-C**). To compare the relative expression of

these reporter constructs, we have incorporated all these variants into a single commonly used genomic site at cytological position 53B2. All these constructs start at 515 bp upstream of the *flam* TSS (**Figure 3A**). With the construct containing the 6kb downstream of the TSS, a faint Tomato expression is observed in the follicle cells, along with the detection of artificial tomato piRNAs (**Figures 3B-D**). However, Tomato starts to be strongly detected in the somatic follicle cells of the ovary with the construct -515; +2086 and still be expressed at a comparable level in the -515; +1200 and to a slightly reduced extent with the -515; +718 construct. The Tj binding sites identified through ATAC-seq and ChIP-seq are located within the -515; +718 region. Overall, these results emphasize the necessity of this region, including the Tj binding site, for the soma-specific expression of *flam*.

The increased Tomato expression observed in the -515; +1200 line prompted us to explore this region further. We noted that the Tj ChIP-seq peak within ATAC-seq peak 7 spans a broad region, ranging from -350 bp to +1200 bp relative to the *flam* TSS. Additionally, a third, weaker Tj motif was identified at +921 bp (with a motif detection p-value threshold of 0.001). This result explains the enhanced Tomato expression observed when the region extends slightly beyond +718. Surprisingly almost no artificial tomato piRNAs are detected with these last constructs (**Figure 3D**). These results suggest that the first 718 bp of *flam* transcript previously described as a piRNA trigger sequence (PTS) in OSS cells[44] is not sufficient to direct artificial piRNA processing from the *tomato* reporter sequence *in vivo* (**Figure 3**). As expected, no Tomato expression is detected in the germline (**Figure 3B**).

Neither the transgene carrying the -515; +30 *flam* promoter which includes the minimal promoter nor the version comprising -515; +432 which includes the first exon of the *flam* promoter are sufficient to conduct a strong Tomato expression in the somatic follicle cells (**Figure 3B-C**). These results suggest that while the TF Ci binding site (-346 bp to the *flam* TSS) is required for minimal *flam* expression in OSCs, it is not sufficient on its own to induce strong *flam* expression in adult follicle cells. To confirm that the minimal *flam* promoter we have previously identified can efficiently drive transcription *in vivo,* we added

UAS sequences in front of the -515; +432 *flam* promoter (**Figure 4A**). Combined with a follicle cell-specific *traffic jam* (*tj*)-Gal4 driver we showed a strong Tomato expression in somatic follicle cells demonstrating that this minimal promoter is functional (**Figures 4B-C**). Using CRISPR-Cas9 we have deleted this region (**Figure 4A**) and then confirm that this minimal promoter region is required for *flam* expression (**Figures 4D-F**). Indeed, these *flam* mutants, named *flam*Δ*prom*, are female sterile with ovaries which are fully atrophied and exhibited loss of *flam* piRNAs whereas other piRNA cluster known to be expressed in the germline were unaltered (**Figure 4D-F**). These results suggest that the sequence between -515 and +432 around the *flam* TSS contains the minimal promoter of *flam* required to conduct expression but does not contain any regulatory sequences. Furthermore, no Tomato expression is detected in fly ovaries carrying the transgene +770; +2086 in front of the tomato reporter sequence excluding the presence of an alternative TSS in this region (**Figures S2H-I**). Overall, these results showed that the sequence between +432 and +1200 downstream of the *flam* TSS and including the three Tj binding sites is an enhancer sequence required and sufficient to drive efficient *flam* expression specifically in the somatic follicle cells of adult ovary.

Finally, we have also tested the influence of the upstream sequence of the *flam* TSS. We replaced the minimal *flam* promoter by a larger *flam* promoter encompassing the 1624 bp that separate the *flam* TSS and the promoter of the closest gene *DIP1* (**Figure 3D**). The addition of this sequence does not allow to increase Tomato expression in the somatic follicle cells of the -1626; +718 and -1626; +2086 transgenic lines (**Figures 3A-C**). These findings suggest the absence of any additional enhancers in the upstream region of *flam* TSS.

Tj is the master transcription factor for follicle cell-specific expression of *flam*

To explore whether Tj, in addition to certain piRNA pathway genes, is involved in the direct regulation of the *flam* piRNA cluster *in vivo*, we abolished Tj activity and monitored *flam* expression by smFISH in mosaic clones of follicle cells (**Figures 5A and S3**). We observed a complete absence of *flam* signal in *tj*-deficient mutant cells, whereas in

neighboring control cells expressing Tj, *flam* RNA was detected, accumulating as a distinct dot at the nuclear periphery consistent with previous descriptions[45,46]. Furthermore, sequencing of small RNAs from ovaries, where Tj was knocked down in follicle cells using a *tj*-Gal4 driver (*tj* SKD), showed a complete loss of piRNAs from the *flam* locus. (**Figure 5B**). All together these findings demonstrate that Tj regulates the soma-specific expression of *flam*.

To investigate whether somatic transcription factors beyond Tj might be involved in the follicle cell-specific expression of *flam* in vivo, we carried out a Translating Ribosome Affinity Purification (TRAP) experiment. This method utilizes the UAS/Gal4 system and allows the identification of cell-specific messenger RNA (mRNA) translation[47]. In adult Drosophila ovaries, we applied TRAP to the GFP-labeled follicle cell population using the *tj*-Gal4 driver (UAS-GFP::Rpl10AGFP; *tj*-Gal4) and to the germ cell population with *nos*-Gal4 driver (UAS-GFP::Rpl10AGFP; *nos*-Gal4).Transcriptional profiling of purified mRNA followed by bioinformatic analyses led to identify around 900 differentially enriched genes in somatic compared to germline cells (FC>2, pvalue <0.05). In support of this, genes enriched in positive Tj cells are associated for specific gene ontology biological processes related to the development of follicle somatic cells (**Figure S4**). Among them, thirty DNA binding factors have been identified including Tj (**Table S2**). We have also performed an RNAi screen targeting each candidate obtained by TRAP, using *tj*-Gal4 as a driver in somatic follicle cells, and retained only those that reproduced the *flam* mutant phenotype[31,48], characterized by atrophied ovaries. This resulted in *tj*, *Stat92E*, *schurri*, *tango*, *chd64* and *mamo* as factors whose depletion leads to atrophic ovaries. To ask whether these factors might be involved in *flam* piRNA precursor transcription, we investigated whether a loss of function of these factors has an impact on *flam* expression by smFISH (**Figure S5**). Because the ovaries are strongly atrophied in a SKD of *Stat92E*, *tango*, *chd64*, *mamo* and *schnurri*, we transiently depleted these factors in the adult follicle cells (*tj*-aSKD), using the thermosensitive *tj*-Gal80ts-Gal4 inhibitor[49] that was activated by a 18 °C-to-28 °C shift during adult ovaries to allow the *tj*-Gal4-driven expression of the RNAi later in oogenesis. In contrast to what we observed in *tj*-deficient cells, no impact on *flam* expression was detected in somatic follicle cells following

depletion of *schnurri*, *tango*, *chd64*, and *mamo* using *tj*-Gal80ts-Gal4 driver, as *flam* RNAs are detected and accumulate in a dot at the nuclear periphery, similar to the pattern observed in wild-type cells. While we cannot entirely rule out the contribution of multiple transcription factors, Tj appears to be the master TF driving the soma-specific expression of *flam*.

Tj is required for TE silencing in follicle cells by regulating *flam* and piRNA factor expression independently

Since *flam* and the somatic piRNA pathway factors control *gypsy*-family TEs in OSCs, we next asked whether Tj depletion resulted in transposon de-repression. By RT-qPCR, we observed ~4-fold and ~20-fold upregulation of *gypsy* TE expression upon 48 h and 96 h of *tj* siRNA knockdowns in OSCs, respectively (**Figures S1B and S1C**). Using mRNA-seq, we further confirmed >6-fold and >24-fold upregulation of *gypsy* and other *gypsy*-family TEs upon 48 h and 96 h of *tj* siRNA knockdowns in OSCs, respectively **(Figures 3C and S1D)**. While overall our results suggest that Tj may directly regulate *flam* expression, since Tj also controls some piRNA pathway components, the observed upregulation of transposable elements in *tj* knockdown alone is not enough to support this conclusion.

Tj is known to be a transcriptional factor required for gonad development[33] and it was proposed to drive the expression of Piwi in follicle cells[34], a finding further supported by our results from OSCs. To rule out the possibility that Tj might affect *flam* expression indirectly via Piwi, we have investigated *flam* transcription by smFISH in follicle cells on Piwi SKD ovaries. In contrast to *tj* mutant follicle cells where *flam* RNA is lost, the absence of Piwi had no effect on *flam* transcription and *flam* RNA nuclear localization (**Figures 6B-C**). We observed *flam* transcripts as a dot in all somatic follicle cells in *piwi* SKD, which was similar to the control *w* SKD. Overall, these results suggest that Tj regulates *flam* expression directly independently of Piwi. The function of Piwi is independent of its slicer activity and Piwi is not required for piRNA biogenesis[50], the loss of *flam* small RNAs in the absence of Traffic jam (**Figures 5A-B**) reinforce the conclusion that the loss of Tj

directly affects *flam* transcription, independent of its impact on Piwi expression. In conclusion, our findings strongly support the key role of Tj in directly regulating the transcription of *flam* piRNA precursors in the somatic follicle cells of the adult ovary.

Discussion

Our study identifies Tj as a pivotal transcriptional regulator of the somatic piRNA pathway in *Drosophila* ovaries. The activity of Tj behind the regulation of the somatic piRNA pathway must be crucial for suppression of *gypsy*-family TEs in somatic ovary follicle cells to prevent TE mobilization and invasion into the neighbouring germ cells. By coordinating the expression of both piRNA factors and piRNA clusters, such as *flam*, Tj ensures effective TE silencing in somatic cells, safeguarding the genomic integrity of the germ cells. Our discovery adds a key layer to our understanding of TE repression mechanisms in *Drosophila*, particularly in non-germline tissues.

A striking outcome of our work is the identification of Tj as a direct regulator of several core components of the somatic piRNA pathway, including *fs(1)Yb*, *nxf2*, *panx*, and *armi*. This regulation is mediated by conserved binding motifs within the promoter regions of these genes, positioning Tj as a master regulator of somatic piRNA pathway expression. Previous studies have demonstrated Tj's involvement in gonadal development and its regulation of piwi expression[33,34], however, our findings expand on these roles by demonstrating Tj's broader control over piRNA pathway components in somatic cells as well as that of the major piRNA cluster *flam* to control *gypsy*-family TEs that have the capability to infect germ cells.

The expression of the germline piRNA pathway components, such as *vas*, *aub*, *ago3* has been recently shown to be under the control of the germline TF Ovo[32]. Interestingly, Ovo and Tj ChIP-seq show complementary binding to the promoters of the germline and somatic genes (**Figure S6A**). Thus, our findings suggest that Tj regulates the somatic piRNA pathway in follicle cells, similar to Ovo's role in regulating the germline counterpart in nurse cells (**Figure S6B**). While Ovo's role in regulating germline piRNA factors is conserved from flies to humans, the somatic piRNA pathway governed by Tj appears to be a *Drosophila*-specific adaptation. This likely arose as a defense against the emergence of *gypsy*-family TEs, which possess the unique ability to infect the germline. Concordant activity of both Ovo and Tj within the *Drosophila* ovary, and possibly aided by additional

germline and somatic tissue TFs, would be essential for efficient suppression of transposons mobilization in both germline and soma respectively as well as from soma to germline (**Figure S6B**). Moreover, our findings demonstrate that Tj directly regulates *flam*, the major somatic piRNA cluster in *Drosophila*, which is essential for TE repression in follicle cells. This distinction highlights the unique role of Tj in the somatic piRNA pathway, indicating that Tj is not only regulating key piRNA pathway factors but is also intricately involved in the transcriptional control of piRNA cluster expression. *Cis*-regulatory elements harbouring Tj motifs near promoters of *flam* and soma-enriched piRNA factors may have co-evolved, resulting in the coordinated expression of piRNA pathway components in ovarian somatic cells. Since Tj controls the expression of key piRNA pathway components and the *flam* piRNA cluster in somatic cells of *Drosophila*, its gene-regulatory circuit functions similar to the mammalian A-MYB (MYBL1) transcription factor, which has been shown to initiate the transcription of both pre-pachytene piRNA clusters and piRNA genes, including MIWI, MILI, and VASA, during mouse spermatogenesis[51]. Both Tj and A-MYB act as master regulators, orchestrating entire pathways critical for safeguarding the genome by repressing TEs.

Relying on a single TF to regulate complex gene networks, as demonstrated with well-known regulators like OCT4, Zelda, and Pax3, which govern extensive gene sets across development and differentiation, can be highly efficient for controlling cell fate, coordinating developmental processes, and with A-MYB and Tj ensuring genome integrity[52–54]. While our findings establish Tj as a key player of the somatic piRNA pathway, the involvement of additional transcriptional regulators and co-factors cannot be ruled out, suggesting a more intricate and fine-tuned regulatory network. Other factors, such as Ci, Mirror or Stat92 – all DNA-binding proteins expressed in somatic follicle cells identified in our TRAP-based screening – could contribute to the regulation of *flam*, as is often described in gene regulatory networks where multiple transcription factors work together to ensure robustness and precise control across development[55,56]. One hypothesis could be that Tj acts as a pioneering transcription factor, akin to others like Zelda in *Drosophila* or FOXA1 in mammals[57,58], whose primary role is to initiate chromatin remodelling by binding to condensed, closed chromatin and making it accessible for

transcription. In this context, Tj would be indispensable for opening chromatin at key piRNA pathway loci, such as *flam*, allowing subsequent transcriptional activation, and once the chromatin is accessible, additional TFs could be recruited to the site to stabilize the open chromatin state, fine-tune the transcriptional output, and further specify gene expression. However, this hypothesis needs further experimental testing.

In addition to its role in regulating the somatic piRNA pathway, Tj appears to govern a broader network of genes that are critical for somatic cell development and function. Our ChIP-seq analysis revealed that Tj binds to approximately 8,000 genomic sites, many of which are located near the promoters of soma-expressed genes. These binding sites are enriched at somatic gene promoters while being depleted at germline-specific genes, further supporting Tj's role as a master regulator of a somatic expression program. Tj shares significant sequence similarity with its mammalian counterparts, c-Maf and MafB, which are known to regulate tissue-specific gene expression and cell differentiation by binding to target gene regulatory regions and interacting with other TFs. Like other members of the Maf family, Tj contains a leucine-zipper domain, a basic DNA-binding domain, and a Maf-specific extended homology domain in its C-terminal region. The leucine-zipper domain, known for facilitating the dimerization of Maf proteins, either with other Maf proteins or different bZIP transcription factors, suggests that Tj may also function as a dimer[33]. Furthermore, the presence of a palindromic motif within Tj binding sites, similar to motifs recognized by other dimeric transcription factors such as Estrogen Receptor (ER) and Heat Shock Factor 1 (HSF1), reinforces the idea that Tj may operate as a dimer[59,60]. Tj dimerization is also supported by AlphaFold3[61], similar to its homologs, which might allow it to bind symmetrically to DNA, thereby adding another layer of regulatory control that warrants further investigation. Future research should delve deeper into Tj's interaction with other TFs and chromatin modifiers to uncover the full extent of its influence on somatic gene networks.

Materials and Methods

Cell culture and growth media

Wild-type OSCs[34] were purchased from the *Drosophila* Genomics Resource Center (DGRC, Stock # 288) and cultured at 26 °C in Shields and Sang M3 Insect Medium (Sigma-Aldrich, S3652-6X1L) supplemented with 10% fetal bovine serum (Sigma-Aldrich, F9665-500ML), 10% fly extract (DGRC, Stock # 1645670), 0.6 mg/mL glutathione (Sigma-Aldrich, G6013-25G), and 10 mU/mL human insulin (Sigma-Aldrich, I9278-5ML) + 1% Penicillin-Streptomycin (Gibco, 15070063).

siRNA knockdowns in cell culture

Sense and antisense 21 nt siRNA sequences for target genes were designed using the DSIR tool (http://biodev.cea.fr/DSIR/DSIR.html) (**Table S3**) and ordered from IDT. The RNA oligonucleotides were resuspended to 400 µM final concentration in RNase-free water. Equal volumes of the resuspended sense and antisense siRNA were mixed then added to an equal amount of 2x annealing buffer (60 mM potassium acetate; 200 mM HEPES, pH 7.5). The mix was boiled for 5 min at 75 °C, then ramped down to 25 °C (-0.1 °C/s) to anneal the siRNA sequences into the siRNA duplex (100 µM final concentration). 2 µL of the final 100 µM siRNA duplex was mixed with 100 µL of Nucleofection solution V (Lonza, VVCA-1003) and transfected into 4 million cells (48 h knockdown) or 5 µL of 100 µM siRNA duplex was mixed with 100 µL of Nucleofection solution V and transfected into 10 million cells (96 h knockdown) using a NucleofectorTM 2b Device and program T-029 (Lonza). Cells were plated into 6-well plates (5 million cells) or 10 cm plates (10 million cells). RNA was either harvested at 48 h or nucleofection was repeated after 48 h and RNA was harvested at 96 h. RNA was either harvested at 48 h or nucleofection was repeated after 48 h and RNA was harvested at 96 h.

RNA isolation and RT-qPCR

OSCs were homogenized by addition of Buffer RLT directly on cell culture vessel surface. RNA was isolated using RNeasy Mini Kit (QIAGEN, 74106) with RNase-Free DNase Set (QIAGEN, 79254) for DNase digestion during RNA purification as per manufacturer's

instruction. Reverse transcription was performed with 100 ng to 1 µg RNA using SuperScript™ IV Reverse Transcriptase (Thermo Fisher Scientific, 18090010). RT-qPCR was performed on 1:10 diluted cDNA using Fast SYBR™ Green Master Mix (Thermo Fisher Scientific, 4385610) with Bio-Rad C1000 Thermal Cycler. Primers were designed against exon-exon junctions for genes with introns and *rpL32* was used as an internal standard (**Table S3**). Relative expression was analysed using the delta-delta Ct method[62]. When total RNAs were extracted from fresh ovaries we used TRIzol™ (Invitrogen) following the manufacturer's instructions. After DNase treatment, cDNA was synthesized using random priming of 1µg total RNA and SuperScript IV Reverse Transcriptase (Invitrogen). Quantitative PCR was performed with the LightCycler480 SYBR Green I Master mix system (Roche) and the data were analyzed using the LightCycler software (Roche). Each experiment was performed 6 times with technical duplicates. RNA levels were calculated with the relative quantification method with *rp49* housekeeping gene as the reference (for primer sequences, see **Table S3**). Statistical differences were evaluated using Kruskal-Wallis non-parametric test, with corrections for multiple comparisons (Holm–Bonferroni adjustment method).

RNA-seq

For mRNA-seq, poly(A) selection was performed using Poly(A) mRNA Magnetic Isolation Module (NEB, E7490) with 1 ug of total RNA as input. Libraries were prepared using The NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina (NEB, E7420L). Library quality was assessed using Agilent 2100 bioanalyzer (High Sensitivity DNA Kit). Libraries were quantified with KAPA Library Quantification Kit for Illumina (Kapa Biosystems, cat# KK4873) submitted to sequencing on a NovaSeq X System with 2 × 50 bp reads.

ATAC-seq

ATAC-seq was performed on 50,000 OSCs using a protocol described previously[63]. In brief, cells were lysed in ATAC-Resuspension Buffer (RSB) containing 0.1% NP40, 0.1% Tween20, and 0.01% Digitonin and washed out with cold ATAC-RSB containing 0.1% Tween-20. The transposition reaction was performed with 25 µl 2x TD buffer from and 2.5 µl transposase (100nM final) from Illumina Tagment DNA Enzyme and Buffer Small

Kit (Illumina, 20034197), 16.5 µl PBS, 0.5 µl 1% digitonin, 0.5 µl 10% Tween-20, 5 µl H2O at 37 °C for 30 min in a thermomixer with 1,000 RPM mixing. Zymo DNA Clean and Concentrator-5 Kit (Zymo, D4014) was used for clean-up and transposed fragments were amplified for 5 cycles [72 °C for 5 min, 98 °C for 30 s (98 °C for 10 s, 63 °C for 30 s, 72 °C for 1 min) × 5] using NEBNext 2xMasterMix (NEB, M0541S) with Ad1_noMX and v2_Ad2.* indexing primers followed by qPCR amplification to determine additional cycle numbers. Amplified DNA was purified using Zymo DNA Clean and Concentrator-5 Kit (Zymo, D4014) and eluted in 20 µl H_2O. AMPure XP beads (Beckman Coulter, A63881) were used for double-sided bead purification. 100-600 bp fragments were selected using Blue Pippin (Sage Science, NC1025035). The library fragment size distribution was assessed using Agilent 2100 bioanalyzer on a High Sensitivity DNA Kit. Libraries were quantified with KAPA Library Quantification Kit for Illumina (Kapa Biosystems, KK4873) and submitted to sequencing on Illumina MiSeq sequencing platform with 2 × 50 bp reads.

ATAC-seq data analysis

The read quality was assessed with FastQC (v0.11.8). The paired-end reads were trimmed of adapter sequences TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG and GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG using Cutadapt tool (v1.18; default parameters)[64]. Burrows-Wheeler Aligner (BWA) (v0.7.17, bwa mem -M -t 4)[65] was used to align the trimmed paired reads to 2014 (BDGP Release 6 + ISO1 MT/dm6) assembly of the *D. melanogaster* genome. Duplicates were marked using Picard tool (v2.9.0) (MarkDuplicates, validation stringency=lenient). SAMtools (v1.9) was used for indexing and filtering[65]. The quality metrics for the aligned ATAC-seq reads were assessed using ataqv (v1.0.0) (https://github.com/ParkerLab/ataqv)[66]. The ATAC-seq peaks were called with MACS2 (v2.1.1.20160309) using --nomodel --shift -37 --extsize 73 parameters and FDR cut-off $q \leq 0.05$[67]. RPKM normalized bigWig files were generated using deepTools (v3.5.1, bamCoverage)[68]. Genome browser visualizations were done using the UCSC Genome Browser[69].

RNA-seq data analysis

The reads were trimmed of adapter sequences using Cutadapt tool (v1.18; -m 1 specified to not have reads trimmed to zero)[64]. The trimmed reads were aligned to the genome assemblies using the RNA-seq aligner STAR (v2.7.3a)[70]. Gene counts were calculated with the featureCounts tool (Subread package v1.5.3; -s 2 for stranded libraries prepared by dUTP method)[71]. *D. melanogaster* gene annotations were taken from Drosophila_melanogaster.BDGP6.28.100.gtf in the Ensembl genome database[72] and annotations of orthologous *D. yakuba* genes were derived from *D. melanogaster* proteins mapped onto *D. yakuba* genome by chained tBLASTn (blastDm2FB in UCSC Table Browser). The read depth-normalized ovary RNA-seq counts were then normalized to species gene lengths. Consensus *Drosophila* transposon sequences were taken from (https://github.com/bergmanlab/transposons). Samtools (v1.9) was used to merge bam files from replicates[65]. Differential gene expression analysis was performed using DESeq2 (v1.30.1)[73]. The RPKM normalized bigWig files were generated for each strand using bamCoverage tool (--filterRNAstrand specified for dUTP stranded libraries) in deepTools (v3.5.1)[68]. Raw data for publicly available RNA-seq reads were downloaded from sources listed in **(Table S4)** and processed as above. Genome browser visualizations were done using the UCSC Genome Browser[69].

ChIP-seq data analysis

The reads were trimmed of adapter sequences using Cutadapt tool (v1.18)[64]. The trimmed reads were aligned to the genome assemblies using Burrows-Wheeler Aligner (BWA) (v0.7.17, bwa aln)[74]. SAMtools (v1.9) was used for sorting, merging, and indexing[65]. ChIP-seq peaks were called using MACS3 (v3.0.0a6) callpeak command (-q 0.01)[67] with data from ChIP input (the whole cell lysate) used as the control (-c). The reproducibility of ChIP-seq peaks between replicates was evaluated using the irreproducible discovery rate (IDR) tool (v2.0.2)[75]. The RPKM normalized bigWig files were generated using bamCoverage tool in deepTools (v3.5.1)[68] with the parameter (--extendReads 120) specified. The input-normalized bigWig files were generated using the bamCompare tool in the deepTools (v3.5.1)[68]. ChIP-seq heatmaps and profiles were generated using the plotHeatmap and plotProfile tools in the deepTools (v3.5.1)[68]. Raw data for publicly available ChIP-seq datasets were downloaded from sources listed in

(Table S4) and processed as describe above. Genome browser visualizations were done using the UCSC Genome Browser[69].

Motif discovery and ATAC-seq peak conservation

Motif scanning was performed using FIMO tool in MEME Suite (v5.4.1) (https://meme-suite.org/meme/index.html)[76]. *De novo* ChIP-seq motifs were generated using MEME-ChIP tool in MEME Suite (v5.4.1)[77]. List of *Drosophila* motifs were downloaded from FlyFactorSurvey (http://pgfe.umassmed.edu/TFDBS/) database[78]. fastaFromBed (v2.26.0) was used for conversion of bed file coordinates into fasta format using *D. melanogaster* BDGP6.28.dna.toplevel genome. Data was visualized with UCSC genome browser[69]. Orthologous ATAC-seq peak regions between the species genomes were derived using the UCSC LiftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver)[79]. *D. melanogaster* ATAC-seq peaks were classified as conserved if the peak orthologous LiftOver regions in the *D. yakuba* genome overlapped with the *D. yakuba* ATAC-seq peaks.

Fly stocks

All *D. melanogaster* stocks were grown on standard medium at 20 °C. RNAi lines expressing dsRNAs[80] or shRNAs[81] against *w* (33644), *tj* (25987, 34595), *Stat92E* (35600), *tango* (53351), *chd64* (62377), *mamo* (60111)*, schnurri (82982), piwi* (33724) were obtained from the Bloomington *Drosophila* stock center (BDSC). The *tj*-Gal4 driver and *tj*-Gal80ts-Gal4 driver were from iGReD collection, hsFlp; FRT40A-nls-RFP was a kind gift of V. Mirouse lab at iGReD and tj^{eo2} FRT40A/CyO, bw line, from Dorothea Godt lab, University of Toronto. The *tj*-Gal4 driver strain was crossed with the transgenic 14xUAS *flam*[-515; +482] *tomato* flies and with all RNAi strains. *tj*-Gal80ts-Gal4 line was crossed with RNAi lines except RNAi lines against *piwi*; the Gal80ts and Gal4 are driving by *tj*; Gal80ts is inactivated at 30 °C.

Clones and Constructs

All plasmids were constructed using NEB Gibson Assembly Cloning Kit (ref E5510) in accordance with the manufacturer's instructions and the recommendations of the

NEBuilder assembly tool (https://nebuilder.neb.com/). Inserts (listed in the **Table S3**) were amplified by PCR using Phusion taq polymerase (Thermofisher) using primers summarized in **Table S3**, and cloned in the W+attB plasmid (addgene #30326). Ten vectors have been then elaborated in W+attB backbone: nine of them possessed differents *flam* framents size upstream of tdTomato reporter sequence directly followed by polyA signal sequence from *K10* gene:
flam[-515; +30]TdTomato-K10 polyA, *flam[-515; +432]TdTomato-K10 polyA*, *flam[-515; +718]TdTomato-K10 polyA*, *flam[-515; +1200]TdTomato-K10 polyA*, *flam[-515; +2086]TdTomato-K10 polyA*, *flam[-515; +6334]TdTomato-K10 polyA*, *flam[-1626; +718]TdTomato-K10 polyA*, *flam[-1626; +2086]TdTomato-K10 polyA*, *flam[+770; +2086]TdTomato-K10 polyA*. The last construct contains the fourteen UAS sequence upstream of the *flam[-515; +432]TdTomato-K10 polyA fragment and is named 14xUAS flam[-515; +432]TdTomato-K10 polyA*.

Transgenic animal production

Plasmid constructions were purified using Plasmid Plus Maxi Kit (QIAGEN). Purified DNA was used for PhiC31 integrase-mediated transgenesis, which was carried out by BestGene (http://www.thebestgene.com/). Transgenes were integrated into the genomic cytological position 53B2.

Generation of *flam* knockout line by CRISPR/Cas9

To remove the entire promoter region of *flamenco* we decided to delete the specific locus ChrX:21631436-21632105 (-455; +215 around TSS). To do so, we used two flanking sgRNAs. The two gRNAs were designed using CRISPR optimal target finder[82]. The individual guide sequences are listed in the **Table S3**. Bolded sequences correspond to each sgRNA. Those primers were used on a pCFD5 template. The resulting PCR product was assembled with the linearised pCFD5 backbone in a single Gibson Assembly reaction. Then the plasmid was injected in Vasa-Cas9 embryos (BestGene, BL#51324). To establish KO lines, molecular characterisation of target loci was performed as described[83]. Briefly, genomic DNA was extracted from individual larvae using QuickExtract™ DNA Extraction Solution (Cambio) according to the supplier's instructions.

Mutant flies were selected by PCR reactions. Exact deletion of *flam* was precisely identified by PCR and sanger sequencing once homozygote mutant flies lines were obtained (primers listed in **Table S3**).

Mitotic clones of *tj* mutant analysis

The drosophila strain hsFlp; FRT40A - nls-RFP; has been crossed with tj^{eo2} FRT40A/CyO, bw strain and maintained at 25 °C on a rich medium. Three heat shocks at 37 °C for 1 h were applied to the F1 progeny, one day before eclosion (pupal stage), on the day of eclosion and one day after. The following day, ovaries were dissected and fixed for either FISH or IF experiments.

FISH and immunostaining

Five to ten ovary pairs were dissected into ice-cold PBS and fixed in formaldehyde solution (4% formaldehyde, 0.3% TritonX-100 in PBS) for 20 min at RT with agitation. The fixed ovaries were then washed three times for 10 min in 0.3% Triton X-100/PBS and permeabilized overnight at 4 °C in 70% ethanol. For probe hybridization, permeabilized ovaries were first rehydrated for 5 min in RNA FISH wash buffer (10% (v/w) formamide in 2x SSC). The ovaries were then resuspended in 50 µl hybridization buffer (10% (v/w) dextran sulphate and 10% (v/w) formamide in 2x SSC). Subsequently, 1.5µL of a homemade *flam* RNA probe labelled with Atto 488 (Sigma, ref 41698) or Atto 550 (Lumiprobe, ref 92835) was added with or without primary antibody. The samples were incubated at 37 °C with agitation. The ovaries were then rinsed twice for 15 min at RT in RNA FISH wash buffer. For FISH-IF, incubation with a secondary antibody conjugated to either Alexa-488, Cy3, or Cy5, was performed in 2x SCC for 1 h 30 at room temperature. After two washes in 2x SSC, DNA was stained with DAPI (1/500). The tissues were then mounted between a slide and coverslip using VectaShield mounting medium (Vector laboratories).

Immunofluorescence

Ovaries from 3- to 5-day-old flies were dissected in Schneider's Drosophila Medium, fixed in 4% formaldehyde/PBS for 20 min, rinsed three times with PBT (× 1 PBS, 0.1% Triton,

1% BSA), and incubated in PBT for at least 1 h 30. They were then incubated with primary antibodies overnight. After 3 washes in PBT, ovaries were incubated with the corresponding secondary antibodies (1:1,000), conjugated to Alexa-488, Cy3, or Cy5, respectively, for 90 min. After two washes, DNA was stained with DAPI. The tissues were then mounted between a slide and coverslip using VectaShield mounting medium (Vector laboratories). All antibodies used for immunostaining (IF) along with probe sequences for in situ hybridization (FISH) are listed in **Table S5**. FISH and IF experiments were imaged using Zeiss LSM 800 or LSM 980 Airyscan confocal with a 20x and 40x objectives and analyzed using OMERO figure software.

Small RNA sequencing and bioinformatics analysis of piRNAs

Total RNA was extracted from 80–100 pairs of ovaries from 3- to 5-day-old flies (for the analysis of piRNA production in somatic follicle cells) using TRIzol Reagent (Ambion). After 2S rRNA depletion, deep sequencing of 18–30-nt small RNAs was performed by Fasteris S.A. (Geneva/CH). Illumina small RNA-Seq reads were analysed with the small RNA pipeline sRNAPipe[84] to map them to various genomic sequence categories of the *D. melanogaster* genome (release 6.03). All libraries were normalized per million of either genome-mapped reads or unique genome-mapping piRNAs (0 mismatche). All the analyses were performed using genome-unique piRNAs mapped to piRNA clusters, as defined by Brennecke and colleagues [6] (no mismatch allowed).

The size profile for *flam* piRNA cluster was obtained by extracting the reads of 23–29 nt sequences uniquely mapped to this locus. Subsequent quantification of reads mapping to 1-kb tiles was done using bedtools, while relative quantification and plotting were done in R. Briefly, small RNA-seq tile signal was normalized to the estimated mappability scores for each 1-kb window. A pseudo-count of 1 was then added to each tile value and log2(fold-change) values relative to control samples was performed.

TRAP experiment and RNA sequencing analysis

Experiment was described by Vachias and colleagues [85] and data were available on GEO (GSE230452). Quality control of sequencing was evaluated using FastQC software. High quality reads were mapped to the drosophila melanogaster dm6 reference genome using

bowtie2 with default parameters[86]. Reads per gene were counted using HTSeqcount[87]. Normalization and differential gene expression analysis were performed using DESeq2[73]. Only genes with an adjusted p-value <0.05 were considered as differentially expressed between the two conditions. Gene expression data from Tj positive cells were compared to Nanos dataset by applying a fold change ≥ 2, $p < 0.05$ in order to generate list of enriched somatic genes. We computed and compared GO biological processes and molecular fonctions from this list using an R package cluster profiler from BioConductor[88].

Data and code availability

RNA-seq and ATAC-seq data generated in this study have been deposited to GEO database under series **GSE275646**.

Author Contributions

EB conception of the project, design of the experiments, supervision, writing the manuscript and funding acquisition. AM data analysis, Tj clonal experiment, functional assays with transcription factor knockdown, figures, and contribution to the manuscript's writing, NM performed all the cloning, generated the *Drosophila* transgenic lines, and conducted both the Tomato immunostaining and the corresponding small RNA sequencing. NG smFISH probe labelling. SM Tj clonal experiment and analysis, contribution to the manuscript's writing. VM and BB TRAP experiments and bioinformatics analyses. BB design of the CRISPR constructs, generation of the corresponding Drosophila mutant lines. AA experiments, computational data analysis, figures, and contribution to the manuscript's writing. JRS experiments, data analysis, figures. GJH supervision and funding acquisition. BCN conception of the project, design of the experiments, supervision, writing the manuscript and funding acquisition. NCL designed the Tj peptide, funded the antibody production, and supervised AJR in conducting affinity purification and specificity validation of the anti-Tj antibody[89].

Acknowledgments

We thank Brasset and Hannon group members for fruitful discussions, and I. Vaillant for comments on the manuscript and S. Jensen for participating in the bioinformatics training of students. We thank the Bloomington *Drosophila* Stock Center for providing fly lines. We thank Dorothea Godt for providing Tj mutant *Drosophila* stocks. We thank Y. Renaud and P. Pouchin for helpful advices and development of the iGReD bioinfomatics platform. We also thank the CLIC facility (Clermont Imagerie Confocale). We thank the Scientific Computing core at the CRUK Cambridge Institute for HPC resources, Research Instrumentation and Cell Services for training, liquid nitrogen storage and mycoplasma testing, and the Genomics core for sequencing services. This work was supported by grants from the Agence Nationale pour la Recherche (CHApiTRE ANR-20-CE12-0005, EB, BiopiC ANR-21-CE12-0022, EB). AM was supported by the Fondation pour la Recherche Médicale (ECO202006011583 and FDT202304016393, AM). This research was financed by the French government IDEX-ISITE initiative 16-IDEX-0001 (CAP 20-25). NCL is funded by NIH/NIGMS grant R01GM135215. GJH is a Royal Society Wolfson

Research Professor (RSRP\R\200001). This research was funded in whole, or in part, by Cancer Research UK (G101107) and the Wellcome Trust (110161/Z/15/Z and 226627/Z/22/Z).

Declaration of interests

The authors declare no competing interests.

References

(1) Bourque, G.; Burns, K. H.; Gehring, M.; Gorbunova, V.; Seluanov, A.; Hammell, M.; Imbeault, M.; Izsvák, Z.; Levin, H. L.; Macfarlan, T. S.; Mager, D. L.; Feschotte, C. Ten Things You Should Know about Transposable Elements. *Genome Biol.* **2018**, *19* (1), 199. https://doi.org/10.1186/s13059-018-1577-z.

(2) Burns, K. H.; Boeke, J. D. Human Transposon Tectonics. *Cell* **2012**, *149* (4), 740–752. https://doi.org/10.1016/j.cell.2012.04.019.

(3) Kazazian, H. H. Mobile Elements: Drivers of Genome Evolution. *Science* **2004**, *303* (5664), 1626–1632. https://doi.org/10.1126/science.1089670.

(4) Chuong, E. B.; Elde, N. C.; Feschotte, C. Regulatory Activities of Transposable Elements: From Conflicts to Benefits. *Nat. Rev. Genet.* **2017**, *18* (2), 71–86. https://doi.org/10.1038/nrg.2016.139.

(5) Cosby, R. L.; Chang, N.-C.; Feschotte, C. Host-Transposon Interactions: Conflict, Cooperation, and Cooption. *Genes Dev.* **2019**, *33* (17–18), 1098–1116. https://doi.org/10.1101/gad.327312.119.

(6) Brennecke, J.; Aravin, A. A.; Stark, A.; Dus, M.; Kellis, M.; Sachidanandam, R.; Hannon, G. J. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. *Cell* **2007**, *128* (6), 1089–1103. https://doi.org/10.1016/j.cell.2007.01.043.

(7) Malone, C. D.; Brennecke, J.; Dus, M.; Stark, A.; McCombie, W. R.; Sachidanandam, R.; Hannon, G. J. Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. *Cell* **2009**, *137* (3), 522–535. https://doi.org/10.1016/j.cell.2009.03.040.

(8) Théron, E.; Dennis, C.; Brasset, E.; Vaury, C. Distinct Features of the piRNA Pathway in Somatic and Germ Cells: From piRNA Cluster Transcription to piRNA Processing and Amplification. *Mob. DNA* **2014**, *5* (1), 1–11. https://doi.org/10.1186/s13100-014-0028-y.

(9) Malone, C. D.; Hannon, G. J. Small RNAs as Guardians of the Genome. *Cell* **2009**, *136* (4), 656–668. https://doi.org/10.1016/j.cell.2009.01.045.

(10) Sienski, G.; Batki, J.; Senti, K.-A.; Dönertas, D.; Tirian, L.; Meixner, K.; Brennecke, J. Silencio/CG9754 Connects the Piwi–piRNA Complex to the Cellular Heterochromatin Machinery. *Genes Dev.* **2015**, *29* (21), 2258–2271. https://doi.org/10.1101/gad.271908.115.

(11) Yu, Y.; Gu, J.; Jin, Y.; Luo, Y.; Preall, J. B.; Ma, J.; Czech, B.; Hannon, G. J. Panoramix Enforces piRNA-Dependent Cotranscriptional Silencing. *Science* **2015**, *350* (6258), 339–342. https://doi.org/10.1126/science.aab0700.

(12) Fabry, M. H.; Ciabrelli, F.; Munafò, M.; Eastwood, E. L.; Kneuss, E.; Falciatori, I.; Falconio, F. A.; Hannon, G. J.; Czech, B. piRNA-Guided Co-Transcriptional Silencing Coopts Nuclear Export Factors. *eLife* **2019**, *8*, e47999. https://doi.org/10.7554/eLife.47999.

(13) Batki, J.; Schnabl, J.; Wang, J.; Handler, D.; Andreev, V. I.; Stieger, C. E.; Novatchkova, M.; Lampersberger, L.; Kauneckaite, K.; Xie, W.; Mechtler, K.; Patel, D. J.; Brennecke, J. The Nascent RNA Binding Complex SFiNX Licenses piRNA-Guided Heterochromatin Formation. *Nat. Struct. Mol. Biol.* **2019**, *26* (8), 720–731. https://doi.org/10.1038/s41594-019-0270-6.

(14) Murano, K.; Iwasaki, Y. W.; Ishizu, H.; Mashiko, A.; Shibuya, A.; Kondo, S.; Adachi, S.; Suzuki, S.; Saito, K.; Natsume, T.; Siomi, M. C.; Siomi, H. Nuclear RNA Export Factor Variant Initiates piRNA-Guided Co-Transcriptional Silencing. *EMBO J.* **2019**, *38* (17), e102870. https://doi.org/10.15252/embj.2019102870.

(15) Schnabl, J.; Wang, J.; Hohmann, U.; Gehre, M.; Batki, J.; Andreev, V. I.; Purkhauser, K.; Fasching, N.; Duchek, P.; Novatchkova, M.; Mechtler, K.; Plaschka, C.; Patel, D. J.; Brennecke, J. Molecular Principles of Piwi-Mediated Cotranscriptional Silencing through the Dimeric SFiNX Complex. *Genes Dev.* **2021**, *35* (5–6), 392–409. https://doi.org/10.1101/gad.347989.120.

(16) Eastwood, E. L.; Jara, K. A.; Bornelöv, S.; Munafò, M.; Frantzis, V.; Kneuss, E.; Barbar, E. J.; Czech, B.; Hannon, G. J. Dimerisation of the PICTS Complex via LC8/Cut-up Drives Co-Transcriptional Transposon Silencing in Drosophila. *eLife* **2021**, *10*, e65557. https://doi.org/10.7554/eLife.65557.

(17) 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* (6), 1364–1379. https://doi.org/10.1016/j.cell.2014.04.031.
(18) Gebert, D.; Neubert, L. K.; Lloyd, C.; Gui, J.; Lehmann, R.; Teixeira, F. K. Large Drosophila Germline piRNA Clusters Are Evolutionarily Labile and Dispensable for Transposon Regulation. *Mol. Cell* **2021**, *81* (19), 3965-3978.e5. https://doi.org/10.1016/j.molcel.2021.07.011.
(19) Akkouche, A.; Kneuss, E.; Bornelöv, S.; Renaud, Y.; Eastwood, E. L.; Lopik, J. van; Gueguen, N.; Jiang, M.; Creixell, P.; Maupetit-Mehouas, S.; Nicholson, B. C.; Brasset, E.; Hannon, G. J. A Dual Histone Code Specifies the Binding of Heterochromatin Protein Rhino to a Subset of piRNA Source Loci. bioRxiv January 11, 2024, p 2024.01.11.575256. https://doi.org/10.1101/2024.01.11.575256.
(20) Brasset, E.; Taddei, A. R.; Arnaud, F.; Faye, B.; Fausto, A. M.; Mazzini, M.; Giorgi, F.; Vaury, C. Viral Particles of the Endogenous Retrovirus ZAM from Drosophila Melanogaster Use a Pre-Existing Endosome/Exosome Pathway for Transfer to the Oocyte. *Retrovirology* **2006**, *3*, 1–9. https://doi.org/10.1186/1742-4690-3-25.
(21) Barckmann, B.; El-Barouk, M.; Pélisson, A.; Mugat, B.; Li, B.; Franckhauser, C.; Lavier, A. S. F.; Mirouze, M.; Fablet, M.; Chambeyron, S. The Somatic piRNA Pathway Controls Germline Transposition over Generations. *Nucleic Acids Res.* **2018**, *46* (18), 9524–9536. https://doi.org/10.1093/nar/gky761.
(22) Duc, C.; Yoth, M.; Mouniée, N.; Jensen, S.; Vaury, C.; Brasset, E. Trapping a Somatic Endogenous Retrovirus into a Germline piRNA Clus- Ter Immunizes the Germline against Further Invasion Keywords. **2019**.
(23) Hirakata, S.; Ishizu, H.; Fujita, A.; Tomoe, Y.; Siomi, M. C. Requirements for Multivalent Yb Body Assembly in Transposon Silencing in Drosophila. *EMBO Rep.* **2019**, *20* (7), e47708. https://doi.org/10.15252/embr.201947708.
(24) Ishizu, H.; Kinoshita, T.; Hirakata, S.; Komatsuzaki, C.; Siomi, M. C. Distinct and Collaborative Functions of Yb and Armitage in Transposon-Targeting piRNA Biogenesis. *Cell Rep.* **2019**, *27* (6), 1822-1835.e8. https://doi.org/10.1016/j.celrep.2019.04.029.
(25) Zhao, K.; Cheng, S.; Miao, N.; Xu, P.; Lu, X.; Zhang, Y.; Wang, M.; Ouyang, X.; Yuan, X.; Liu, W.; Lu, X.; Zhou, P.; Gu, J.; Zhang, Y.; Qiu, D.; Jin, Z.; Su, C.; Peng, C.; Wang, J.-H.; Dong, M.-Q.; Wan, Y.; Ma, J.; Cheng, H.; Huang, Y.; Yu, Y. A Pandas Complex Adapted for piRNA-Guided Transcriptional Silencing and Heterochromatin Formation. *Nat. Cell Biol.* **2019**, *21* (10), 1261–1272. https://doi.org/10.1038/s41556-019-0396-0.
(26) Iwasaki, Y. W.; Sriswasdi, S.; Kinugasa, Y.; Adachi, J.; Horikoshi, Y.; Shibuya, A.; Iwasaki, W.; Tashiro, S.; Tomonaga, T.; Siomi, H. Piwi-piRNA Complexes Induce Stepwise Changes in Nuclear Architecture at Target Loci. *EMBO J.* **2021**, *40* (18), e108345. https://doi.org/10.15252/embj.2021108345.
(27) Andreev, V. I.; Yu, C.; Wang, J.; Schnabl, J.; Tirian, L.; Gehre, M.; Handler, D.; Duchek, P.; Novatchkova, M.; Baumgartner, L.; Meixner, K.; Sienski, G.; Patel, D. J.; Brennecke, J. Panoramix SUMOylation on Chromatin Connects the piRNA Pathway to the Cellular Heterochromatin Machinery. *Nat. Struct. Mol. Biol.* **2022**, *29* (2), 130–142. https://doi.org/10.1038/s41594-022-00721-x.
(28) Duc, C.; Yoth, M.; Jensen, S.; Mouniée, N.; Bergman, C. M.; Vaury, C.; Brasset, E. Trapping a Somatic Endogenous Retrovirus into a Germline piRNA Cluster Immunizes the Germline against Further Invasion. *Genome Biol.* **2019**, *20*, 127. https://doi.org/10.1186/s13059-019-1736-x.
(29) van Lopik, J.; Alizada, A.; Trapotsi, M.-A.; Hannon, G. J.; Bornelöv, S.; Czech Nicholson, B. Unistrand piRNA Clusters Are an Evolutionarily Conserved Mechanism to Suppress Endogenous Retroviruses across the Drosophila Genus. *Nat. Commun.* **2023**, *14*, 7337. https://doi.org/10.1038/s41467-023-42787-1.
(30) Czech, B.; Munafò, M.; Ciabrelli, F.; Eastwood, E. L.; Fabry, M. H.; Kneuss, E.; Hannon, G. J. piRNA-Guided Genome Defense: From Biogenesis to Silencing. *Annu. Rev. Genet.* **2018**, *52*, 131–157. https://doi.org/10.1146/annurev-genet-120417-031441.
(31) Mével-Ninio, M.; Pelisson, A.; Kinder, J.; Campos, A. R.; Bucheton, A. The Flamenco Locus

Controls the Gypsy and ZAM Retroviruses and Is Required for Drosophila Oogenesis. *Genetics* **2007**, *175* (4), 1615–1624. https://doi.org/10.1534/genetics.106.068106.
(32) Alizada, A.; Hannon, G. J.; Nicholson, B. C. Ovo Is a Master Regulator of the piRNA Pathway in Animal Ovarian Germ Cells. bioRxiv April 23, 2024. https://doi.org/10.1101/2024.04.23.590802.
(33) Li, M. A.; Alls, J. D.; Avancini, R. M.; Koo, K.; Godt, D. The Large Maf Factor Traffic Jam Controls Gonad Morphogenesis in Drosophila. *Nat. Cell Biol.* **2003**, *5* (11), 994–1000. https://doi.org/10.1038/ncb1058.
(34) Saito, K.; Inagaki, S.; Mituyama, T.; Kawamura, Y.; Ono, Y.; Sakota, E.; Kotani, H.; Asai, K.; Siomi, H.; Siomi, M. C. A Regulatory Circuit for Piwi by the Large Maf Gene Traffic Jam in Drosophila. *Nature* **2009**, *461* (7268), 1296–1299. https://doi.org/10.1038/nature08501.
(35) Niki, Y.; Yamaguchi, T.; Mahowald, A. P. Establishment of Stable Cell Lines of Drosophila Germ-Line Stem Cells. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (44), 16325–16330. https://doi.org/10.1073/pnas.0607435103.
(36) Handler, D.; Meixner, K.; Pizka, M.; Lauss, K.; Schmied, C.; Gruber, F. S.; Brennecke, J. The Genetic Makeup of the Drosophila piRNA Pathway. *Mol. Cell* **2013**, *50* (5), 762–777. https://doi.org/10.1016/j.molcel.2013.04.031.
(37) modENCODE Consortium; Roy, S.; Ernst, J.; Kharchenko, P. V.; Kheradpour, P.; Negre, N.; Eaton, M. L.; Landolin, J. M.; Bristow, C. A.; Ma, L.; Lin, M. F.; Washietl, S.; Arshinoff, B. I.; Ay, F.; Meyer, P. E.; Robine, N.; Washington, N. L.; Di Stefano, L.; Berezikov, E.; Brown, C. D.; Candeias, R.; Carlson, J. W.; Carr, A.; Jungreis, I.; Marbach, D.; Sealfon, R.; Tolstorukov, M. Y.; Will, S.; Alekseyenko, A. A.; Artieri, C.; Booth, B. W.; Brooks, A. N.; Dai, Q.; Davis, C. A.; Duff, M. O.; Feng, X.; Gorchakov, A. A.; Gu, T.; Henikoff, J. G.; Kapranov, P.; Li, R.; MacAlpine, H. K.; Malone, J.; Minoda, A.; Nordman, J.; Okamura, K.; Perry, M.; Powell, S. K.; Riddle, N. C.; Sakai, A.; Samsonova, A.; Sandler, J. E.; Schwartz, Y. B.; Sher, N.; Spokony, R.; Sturgill, D.; van Baren, M.; Wan, K. H.; Yang, L.; Yu, C.; Feingold, E.; Good, P.; Guyer, M.; Lowdon, R.; Ahmad, K.; Andrews, J.; Berger, B.; Brenner, S. E.; Brent, M. R.; Cherbas, L.; Elgin, S. C. R.; Gingeras, T. R.; Grossman, R.; Hoskins, R. A.; Kaufman, T. C.; Kent, W.; Kuroda, M. I.; Orr-Weaver, T.; Perrimon, N.; Pirrotta, V.; Posakony, J. W.; Ren, B.; Russell, S.; Cherbas, P.; Graveley, B. R.; Lewis, S.; Micklem, G.; Oliver, B.; Park, P. J.; Celniker, S. E.; Henikoff, S.; Karpen, G. H.; Lai, E. C.; MacAlpine, D. M.; Stein, L. D.; White, K. P.; Kellis, M. Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. *Science* **2010**, *330* (6012), 1787–1797. https://doi.org/10.1126/science.1198374.
(38) Panchal, T.; Chen, X.; Alchits, E.; Oh, Y.; Poon, J.; Kouptsova, J.; Laski, F. A.; Godt, D. Specification and Spatial Arrangement of Cells in the Germline Stem Cell Niche of the Drosophila Ovary Depend on the Maf Transcription Factor Traffic Jam. *PLoS Genet.* **2017**, *13* (5), e1006790. https://doi.org/10.1371/journal.pgen.1006790.
(39) Li, M.; Hu, X.; Zhang, S.; Ho, M. S.; Wu, G.; Zhang, L. Traffic Jam Regulates the Function of the Ovarian Germline Stem Cell Progeny Differentiation Niche during Pre-Adult Stage in Drosophila. *Sci. Rep.* **2019**, *9*, 10124. https://doi.org/10.1038/s41598-019-45317-6.
(40) Arnold, C. D.; Gerlach, D.; Stelzer, C.; Boryń, Ł. M.; Rath, M.; Stark, A. Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-Seq. *Science* **2013**, *339* (6123), 1074–1077. https://doi.org/10.1126/science.1232542.
(41) Arnold, C. D.; Gerlach, D.; Spies, D.; Matts, J. A.; Sytnikova, Y. A.; Pagani, M.; Lau, N. C.; Stark, A. Quantitative Genome-Wide Enhancer Activity Maps for Five Drosophila Species Show Functional Enhancer Conservation and Turnover during Cis-Regulatory Evolution. *Nat. Genet.* **2014**, *46* (7), 685–692. https://doi.org/10.1038/ng.3009.
(42) Zabidi, M. A.; Arnold, C. D.; Schernhuber, K.; Pagani, M.; Rath, M.; Frank, O.; Stark, A. Enhancer-Core-Promoter Specificity Separates Developmental and Housekeeping Gene Regulation. *Nature* **2015**, *518* (7540), 556–559. https://doi.org/10.1038/nature13994.
(43) Goriaux, C.; Renaud, Y.; Brasset, E.; Desset, S.; Vaury, C. Transcriptional Properties and Splicing of the Flamenco piRNA Cluster. *EMBO Rep.* **2014**, *15* (4), 411–418. https://doi.org/10.1002/embr.201337898.

(44) Homolka, D.; Pandey, R. R.; Goriaux, C.; Brasset, E.; Vaury, C.; Sachidanandam, R.; Fauvarque, M.-O.; Pillai, R. S. PIWI Slicing and RNA Elements in Precursors Instruct Directional Primary piRNA Biogenesis. *Cell Rep.* **2015**, *12* (3), 418–428. https://doi.org/10.1016/j.celrep.2015.06.030.
(45) Dennis, C.; Zanni, V.; Brasset, E.; Eymery, A.; Zhang, L.; Mteirek, R.; Jensen, S.; Rong, Y. S.; Vaury, C. "Dot COM", a Nuclear Transit Center for the Primary piRNA Pathway in Drosophila. *PLoS ONE* **2013**, *8* (9), e72752. https://doi.org/10.1371/journal.pone.0072752.
(46) Dennis, C.; Brasset, E.; Sarkar, A.; Vaury, C. Export of piRNA Precursors by EJC Triggers Assembly of Cytoplasmic Yb-Body in Drosophila. *Nat. Commun.* **2016**, *7*, 13739. https://doi.org/10.1038/ncomms13739.
(47) Heiman, M.; Schaefer, A.; Gong, S.; Peterson, J. D.; Day, M.; Ramsey, K. E.; Suárez-Fariñas, M.; Schwarz, C.; Stephan, D. A.; Surmeier, D. J.; Greengard, P.; Heintz, N. A Translational Profiling Approach for the Molecular Characterization of CNS Cell Types. *Cell* **2008**, *135* (4), 738–748. https://doi.org/10.1016/j.cell.2008.10.028.
(48) Sokolova, O. A.; Ilyin, A. A.; Poltavets, A. S.; Nenasheva, V. V.; Mikhaleva, E. A.; Shevelyov, Y. Y.; Klenov, M. S. Yb Body Assembly on the Flamenco piRNA Precursor Transcripts Reduces Genic piRNA Production. *Mol. Biol. Cell* **2019**, *30* (12), 1544–1554. https://doi.org/10.1091/mbc.E17-10-0591.
(49) McGuire, S. E.; Mao, Z.; Davis, R. L. Spatiotemporal Gene Expression Targeting with the TARGET and Gene-Switch Systems in Drosophila. *Sci. STKE Signal Transduct. Knowl. Environ.* **2004**, *2004* (220), pl6. https://doi.org/10.1126/stke.2202004pl6.
(50) Darricarrère, N.; Liu, N.; Watanabe, T.; Lin, H. Function of Piwi, a Nuclear Piwi/Argonaute Protein, Is Independent of Its Slicer Activity. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (4), 1297–1302. https://doi.org/10.1073/pnas.1213283110.
(51) Li, X. Z.; Roy, C. K.; Dong, X.; Bolcun-Filas, E.; Wang, J.; Han, B. W.; Xu, J.; Moore, M. J.; Schimenti, J. C.; Weng, Z.; Zamore, P. D. An Ancient Transcription Factor Initiates the Burst of piRNA Production During Early Meiosis in Mouse Testes. *Mol. Cell* **2013**, *50* (1), 67–81. https://doi.org/10.1016/j.molcel.2013.02.016.
(52) Babaie, Y.; Herwig, R.; Greber, B.; Brink, T. C.; Wruck, W.; Groth, D.; Lehrach, H.; Burdon, T.; Adjaye, J. Analysis of Oct4-Dependent Transcriptional Networks Regulating Self-Renewal and Pluripotency in Human Embryonic Stem Cells. *Stem Cells Dayt. Ohio* **2007**, *25* (2), 500–510. https://doi.org/10.1634/stemcells.2006-0426.
(53) Nien, C.-Y.; Liang, H.-L.; Butcher, S.; Sun, Y.; Fu, S.; Gocha, T.; Kirov, N.; Manak, J. R.; Rushlow, C. Temporal Coordination of Gene Networks by Zelda in the Early Drosophila Embryo. *PLoS Genet.* **2011**, *7* (10), e1002339. https://doi.org/10.1371/journal.pgen.1002339.
(54) Esteves de Lima, J.; Relaix, F. Master Regulators of Skeletal Muscle Lineage Development and Pluripotent Stem Cells Differentiation. *Cell Regen. Lond. Engl.* **2021**, *10* (1), 31. https://doi.org/10.1186/s13619-021-00093-5.
(55) Yang, Q.; Li, R.; Lyu, Q.; Hou, L.; Liu, Z.; Sun, Q.; Liu, M.; Shi, H.; Xu, B.; Yin, M.; Yan, Z.; Huang, Y.; Liu, M.; Li, Y.; Wu, L. Single-Cell CAS-Seq Reveals a Class of Short PIWI-Interacting RNAs in Human Oocytes. *Nat. Commun.* **2019**, *10* (1), 3389. https://doi.org/10.1038/s41467-019-11312-8.
(56) Kim, D.; Tran, A.; Kim, H. J.; Lin, Y.; Yang, J. Y. H.; Yang, P. Gene Regulatory Network Reconstruction: Harnessing the Power of Single-Cell Multi-Omic Data. *NPJ Syst. Biol. Appl.* **2023**, *9* (1), 51. https://doi.org/10.1038/s41540-023-00312-6.
(57) Cirillo, L. A.; McPherson, C. E.; Bossard, P.; Stevens, K.; Cherian, S.; Shim, E. Y.; Clark, K. L.; Burley, S. K.; Zaret, K. S. Binding of the Winged-Helix Transcription Factor HNF3 to a Linker Histone Site on the Nucleosome. *EMBO J.* **1998**, *17* (1), 244–254. https://doi.org/10.1093/emboj/17.1.244.
(58) Cirillo, L. A.; Lin, F. R.; Cuesta, I.; Friedman, D.; Jarnik, M.; Zaret, K. S. Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **2002**, *9* (2), 279–289. https://doi.org/10.1016/s1097-2765(02)00459-8.
(59) Patel, A. K. M.; Vilela, P.; Shaik, T. B.; McEwen, A. G.; Hazemann, I.; Brillet, K.; Ennifar, E.; Hamiche, A.; Markov, G. V.; Laudet, V.; Moras, D.; Klaholz, B. P.; Billas, I. M. L. Asymmetric

Dimerization in a Transcription Factor Superfamily Is Promoted by Allosteric Interactions with DNA. *Nucleic Acids Res.* **2023**, *51* (16), 8864–8879. https://doi.org/10.1093/nar/gkad632.
(60) Akerfelt, M.; Morimoto, R. I.; Sistonen, L. Heat Shock Factors: Integrators of Cell Stress, Development and Lifespan. *Nat. Rev. Mol. Cell Biol.* **2010**, *11* (8), 545–555. https://doi.org/10.1038/nrm2938.
(61) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C.-C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Žemgulytė, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Žídek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature* **2024**, *630* (8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w.
(62) Livak, K. J.; Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods San Diego Calif* **2001**, *25* (4), 402–408. https://doi.org/10.1006/meth.2001.1262.
(63) Corces, M. R.; Trevino, A. E.; Hamilton, E. G.; Greenside, P. G.; Sinnott-Armstrong, N. A.; Vesuna, S.; Satpathy, A. T.; Rubin, A. J.; Montine, K. S.; Wu, B.; Kathiria, A.; Cho, S. W.; Mumbach, M. R.; Carter, A. C.; Kasowski, M.; Orloff, L. A.; Risca, V. I.; Kundaje, A.; Khavari, P. A.; Montine, T. J.; Greenleaf, W. J.; Chang, H. Y. An Improved ATAC-Seq Protocol Reduces Background and Enables Interrogation of Frozen Tissues. *Nat. Methods* **2017**, *14* (10), 959–962. https://doi.org/10.1038/nmeth.4396.
(64) Martin, M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet.journal* **2011**, *17* (1), 10–12. https://doi.org/10.14806/ej.17.1.200.
(65) Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map Format and SAMtools. *Bioinforma. Oxf. Engl.* **2009**, *25* (16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
(66) Orchard, P.; Kyono, Y.; Hensley, J.; Kitzman, J. O.; Parker, S. C. J. Quantification, Dynamic Visualization, and Validation of Bias in ATAC-Seq Data with Ataqv. *Cell Syst.* **2020**, *10* (3), 298-306.e4. https://doi.org/10.1016/j.cels.2020.02.009.
(67) Zhang, Y.; Liu, T.; Meyer, C. A.; Eeckhoute, J.; Johnson, D. S.; Bernstein, B. E.; Nusbaum, C.; Myers, R. M.; Brown, M.; Li, W.; Liu, X. S. Model-Based Analysis of ChIP-Seq (MACS). *Genome Biol.* **2008**, *9* (9), R137. https://doi.org/10.1186/gb-2008-9-9-r137.
(68) Ramírez, F.; Dündar, F.; Diehl, S.; Grüning, B. A.; Manke, T. deepTools: A Flexible Platform for Exploring Deep-Sequencing Data. *Nucleic Acids Res.* **2014**, *42* (Web Server issue), W187-191. https://doi.org/10.1093/nar/gku365.
(69) Kent, W. J.; Sugnet, C. W.; Furey, T. S.; Roskin, K. M.; Pringle, T. H.; Zahler, A. M.; Haussler, D. The Human Genome Browser at UCSC. *Genome Res.* **2002**, *12* (6), 996–1006. https://doi.org/10.1101/gr.229102.
(70) Dobin, A.; Davis, C. A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T. R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinforma. Oxf. Engl.* **2013**, *29* (1), 15–21. https://doi.org/10.1093/bioinformatics/bts635.
(71) Liao, Y.; Smyth, G. K.; Shi, W. featureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features. *Bioinformatics* **2014**, *30* (7), 923–930. https://doi.org/10.1093/bioinformatics/btt656.
(72) Martin, F. J.; Amode, M. R.; Aneja, A.; Austine-Orimoloye, O.; Azov, A. G.; Barnes, I.; Becker, A.; Bennett, R.; Berry, A.; Bhai, J.; Bhurji, S. K.; Bignell, A.; Boddu, S.; Branco Lins, P. R.; Brooks, L.; Ramaraju, S. B.; Charkhchi, M.; Cockburn, A.; Da Rin Fiorretto, L.; Davidson, C.; Dodiya, K.; Donaldson, S.; El Houdaigui, B.; El Naboulsi, T.; Fatima, R.; Giron, C. G.; Genez, T.; Ghattaoraya, G. S.; Martinez, J. G.; Guijarro, C.; Hardy, M.; Hollis, Z.; Hourlier, T.; Hunt, T.; Kay, M.; Kaykala, V.; Le, T.;

Lemos, D.; Marques-Coelho, D.; Marugán, J. C.; Merino, G. A.; Mirabueno, L. P.; Mushtaq, A.; Hossain, S. N.; Ogeh, D. N.; Sakthivel, M. P.; Parker, A.; Perry, M.; Piližota, I.; Prosovetskaia, I.; Pérez-Silva, J. G.; Salam, A. I. A.; Saraiva-Agostinho, N.; Schuilenburg, H.; Sheppard, D.; Sinha, S.; Sipos, B.; Stark, W.; Steed, E.; Sukumaran, R.; Sumathipala, D.; Suner, M.-M.; Surapaneni, L.; Sutinen, K.; Szpak, M.; Tricomi, F. F.; Urbina-Gómez, D.; Veidenberg, A.; Walsh, T. A.; Walts, B.; Wass, E.; Willhoft, N.; Allen, J.; Alvarez-Jarreta, J.; Chakiachvili, M.; Flint, B.; Giorgetti, S.; Haggerty, L.; Ilsley, G. R.; Loveland, J. E.; Moore, B.; Mudge, J. M.; Tate, J.; Thybert, D.; Trevanion, S. J.; Winterbottom, A.; Frankish, A.; Hunt, S. E.; Ruffier, M.; Cunningham, F.; Dyer, S.; Finn, R. D.; Howe, K. L.; Harrison, P. W.; Yates, A. D.; Flicek, P. Ensembl 2023. *Nucleic Acids Res.* **2023**, *51* (D1), D933–D941. https://doi.org/10.1093/nar/gkac958.

(73) Love, M. I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15* (12), 550. https://doi.org/10.1186/s13059-014-0550-8.

(74) Li, H.; Durbin, R. Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinforma. Oxf. Engl.* **2009**, *25* (14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

(75) Li, Q.; Brown, J. B.; Huang, H.; Bickel, P. J. Measuring Reproducibility of High-Throughput Experiments. *Ann. Appl. Stat.* **2011**, *5* (3). https://doi.org/10.1214/11-AOAS466.

(76) Grant, C. E.; Bailey, T. L.; Noble, W. S. FIMO: Scanning for Occurrences of a given Motif. *Bioinforma. Oxf. Engl.* **2011**, *27* (7), 1017–1018. https://doi.org/10.1093/bioinformatics/btr064.

(77) Ma, W.; Noble, W. S.; Bailey, T. L. Motif-Based Analysis of Large Nucleotide Data Sets Using MEME-ChIP. *Nat. Protoc.* **2014**, *9* (6), 1428–1450. https://doi.org/10.1038/nprot.2014.083.

(78) Zhu, L. J.; Christensen, R. G.; Kazemian, M.; Hull, C. J.; Enuameh, M. S.; Basciotta, M. D.; Brasefield, J. A.; Zhu, C.; Asriyan, Y.; Lapointe, D. S.; Sinha, S.; Wolfe, S. A.; Brodsky, M. H. FlyFactorSurvey: A Database of Drosophila Transcription Factor Binding Specificities Determined Using the Bacterial One-Hybrid System. *Nucleic Acids Res.* **2011**, *39* (Database issue), D111-117. https://doi.org/10.1093/nar/gkq858.

(79) Hinrichs, A. S.; Karolchik, D.; Baertsch, R.; Barber, G. P.; Bejerano, G.; Clawson, H.; Diekhans, M.; Furey, T. S.; Harte, R. A.; Hsu, F.; Hillman-Jackson, J.; Kuhn, R. M.; Pedersen, J. S.; Pohl, A.; Raney, B. J.; Rosenbloom, K. R.; Siepel, A.; Smith, K. E.; Sugnet, C. W.; Sultan-Qurraie, A.; Thomas, D. J.; Trumbower, H.; Weber, R. J.; Weirauch, M.; Zweig, A. S.; Haussler, D.; Kent, W. J. The UCSC Genome Browser Database: Update 2006. *Nucleic Acids Res.* **2006**, *34* (Database issue), D590–D598. https://doi.org/10.1093/nar/gkj144.

(80) Perkins, L. A.; Holderbaum, L.; Tao, R.; Hu, Y.; Sopko, R.; McCall, K.; Yang-Zhou, D.; Flockhart, I.; Binari, R.; Shim, H.-S.; Miller, A.; Housden, A.; Foos, M.; Randkelv, S.; Kelley, C.; Namgyal, P.; Villalta, C.; Liu, L.-P.; Jiang, X.; Huan-Huan, Q.; Wang, X.; Fujiyama, A.; Toyoda, A.; Ayers, K.; Blum, A.; Czech, B.; Neumuller, R.; Yan, D.; Cavallaro, A.; Hibbard, K.; Hall, D.; Cooley, L.; Hannon, G. J.; Lehmann, R.; Parks, A.; Mohr, S. E.; Ueda, R.; Kondo, S.; Ni, J.-Q.; Perrimon, N. The Transgenic RNAi Project at Harvard Medical School: Resources and Validation. *Genetics* **2015**, *201* (3), 843–852. https://doi.org/10.1534/genetics.115.180208.

(81) Ni, J.-Q.; Zhou, R.; Czech, B.; Liu, L.-P.; Holderbaum, L.; Yang-Zhou, D.; Shim, H.-S.; Tao, R.; Handler, D.; Karpowicz, P.; Binari, R.; Booker, M.; Brennecke, J.; Perkins, L. A.; Hannon, G. J.; Perrimon, N. A Genome-Scale shRNA Resource for Transgenic RNAi in Drosophila. *Nat. Methods* **2011**, *8* (5), 405–407. https://doi.org/10.1038/nmeth.1592.

(82) Gratz, S. J.; Ukken, F. P.; Rubinstein, C. D.; Thiede, G.; Donohue, L. K.; Cummings, A. M.; O'Connor-Giles, K. M. Highly Specific and Efficient CRISPR/Cas9-Catalyzed Homology-Directed Repair in *Drosophila*. *Genetics* **2014**, *196* (4), 961–971. https://doi.org/10.1534/genetics.113.160713.

(83) Port, F.; Chen, H.-M.; Lee, T.; Bullock, S. L. Optimized CRISPR/Cas Tools for Efficient Germline and Somatic Genome Engineering in Drosophila. *Proc. Natl. Acad. Sci.* **2014**, *111* (29), E2967–E2976. https://doi.org/10.1073/pnas.1405500111.

(84) Pogorelcnik, R.; Vaury, C.; Pouchin, P.; Jensen, S.; Brasset, E. sRNAPipe: A Galaxy-Based Pipeline for Bioinformatic in-Depth Exploration of Small RNAseq Data. *Mob. DNA* **2018**, *9*, 25.

https://doi.org/10.1186/s13100-018-0130-7.
(85) Vachias, C.; Tourlonias, C.; Grelée, L.; Gueguen, N.; Renaud, Y.; Venugopal, P.; Richard, G.; Pouchin, P.; Brasset, É.; Mirouse, V. Gap Junction and Amino Acid Import in Somatic Cells Promote Germ Cell Growth. bioRxiv January 26, 2024, p 2023.05.15.540837. https://doi.org/10.1101/2023.05.15.540837.
(86) Langmead, B.; Salzberg, S. L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* **2012**, *9* (4), 357–359. https://doi.org/10.1038/nmeth.1923.
(87) Anders, S.; Pyl, P. T.; Huber, W. HTSeq--a Python Framework to Work with High-Throughput Sequencing Data. *Bioinforma. Oxf. Engl.* **2015**, *31* (2), 166–169. https://doi.org/10.1093/bioinformatics/btu638.
(88) Yu, G.; Wang, L.-G.; Han, Y.; He, Q.-Y. clusterProfiler: An R Package for Comparing Biological Themes among Gene Clusters. *Omics J. Integr. Biol.* **2012**, *16* (5), 284–287. https://doi.org/10.1089/omi.2011.0118.
(89) Rivera, A.; Lee, J.-H. R.; Gupta, S.; Yang, L.; Goel, R. K.; Zaia, J.; Lau, N. C. Traffic Jam Activates the Flamenco piRNA Cluster Locus and the Piwi Pathway to Ensure Transposon Silencing and Drosophila Fertility. bioRxiv August 16, 2024, p 2024.08.15.608167. https://doi.org/10.1101/2024.08.15.608167.

Figure legends

Figure 1: Traffic jam regulates the expression of somatic piRNA pathway genes in *Drosophila* ovarian somatic cells

(A) Heatmap showing germline-specific and soma+germline-expressed piRNA pathway genes based on ranking of gene expression fold-enrichments in the Vas-GFP+ germline cells compared to the Vas-GFP- somatic cells from the FACS-sorted transgenic Vas-GFP fly ovaries (*<0.01, DEseq2, RNA-seq n=3 replicates from distinct samples). The normalized RNA-seq counts (DEseq2) from OSCs, Vas-GFP+ germline cells, and Vas-GFP- somatic cells are also shown. The presence of Tj motifs within ATAC-seq peaks and Tj ChIP-seq peaks (ENCODE, whole-fly) within ±500 bp of gene TSS is depicted with dots.

(B) Venn diagram showing the Tj motif as the only motif shared between the ATAC-seq peaks of the promoter regions (±500 bp of TSS) of the somatic-enriched piRNA pathway genes.

(C) Tj ChIP-seq (Encode, female whole-fly) showing Tj binding events to the promoters of the somatic piRNA pathway genes. Grey bars indicate Tj ChIP-seq peaks with the central red lines showing peak summits.

(D) RT-qPCR plot showing siRNA knockdown of *tj* in OSCs resulting in downregulation of the somatic piRNA pathway genes by 48 h normalised to *rpl32* housekeeping control (n=3 replicates from distinct samples; error bars indicate standard deviation).

(E) Volcano plot showing downregulation of the somatic piRNA pathway genes by 48 h of *tj* siRNA knockdowns in OSCs using differential RNA-seq analysis (DEseq2) between *tj* and *renilla* siRNA knockdowns (n=3 replicates from distinct samples). Blue dots are showing the somatic-enriched piRNA pathway genes (e.g., *fs(1)Yb*, *nxf2*, *panx*, *soYb* and *armi*); green dots showing the soma-germ-shared piRNA factors (e.g., *piwi*) and red dots are showing the siRNA target *tj*.

(F) Genomic binding profiles of Tj ChIP-seq (Encode; whole-fly; n=3) at the promoters of germline- and soma-enriched genes using normalized Tj ChIP-seq signal via input subtraction (top) and $\log_2$ of chip rpm/input rpm (bottom).

(G) The top-scoring *de novo* motifs and their distribution within Tj ChIP-seq peaks

(MEME-ChIP).

(H) Confocal images of *tj*^{eo2} clonal egg chamber showing Armi (black) and Tj (green) expression via immunostaining. Clone is circled in white. Nuclei are labelled with DAPI (white). Scale bars: 10 µm.

Figure 2: Tj binds to conserved *cis*-regulatory elements in *flamenco* and prevents *gypsy* TE expression

(A) Conservation of ovary ATAC-seq peaks in the *flam* promoter region of *D. melanogaster* (*dm6*) and *D. yakuba* (*droYak2*) species (orthologous regions derived with UCSC LiftOver and chain files). OSC ATAC-seq, OSC STARR-seq peaks and Tj ChIP-seq (Encode, whole-fly) indicate conserved enhancer activity and Tj binding for the conserved ATAC-seq peak #7.

(B) Venn diagram showing motifs for 21 TFs shared between *D. melanogaster* ATAC-seq peak #7 and its orthologous ATAC-seq peak in *D. yakuba*. The dot plot showing expression of the 21 candidate transcription factors in the ovaries of *D. melanogaster* and *D. yakuba* species (read depth and gene length-normalized DEseq2 counts, RNA-seq n=4).

(C) The somatic and germline enrichment of the 21 candidate TFs based on fold-enrichments of gene expression in the Vas-GFP- somatic cells compared to the Vas-GFP+ germline cells from the FACS-sorted transgenic Vas-GFP fly ovaries (DEseq2, RNA-seq, n=3).

Figure 3: An enhancer of *flam* is located downstream of the TSS

(A) Schematic representation of the transgenic constructs.

(B) Confocal images of ovarioles of the corresponding *Drosophila* transgenic line showing Tomato (red) expression via immunostaining. Nuclei are stained with DAPI (white). Scale bars: 10 µm.

(C) Quantification of relative transgene expression levels in transgenic ovaries by qRT-PCR. *tomato* expression levels were normalized to the expression of the housekeeping gene *rp49*. Data are presented as the mean (n=6) with error bars representing standard

deviation. Statistical significance was determined using the Mann-Whitney test, with ** indicating a P-value < 0.005 and * a P-value < 0.05.

(D) Amount of piRNAs produced from the *tomato* reporter sequence, expressed in reads per million (RPM). piRNAs with 1U, 1G, 1C, and 1A are represented by black, dark grey, intermediate grey, and light grey, respectively.

Figure 4: The minimal *flam* promoter is crucial for its expression and piRNA biogenesis

(A) Schematic representation of the transgenic constructs and deletion in *flam* (*flam*△*prom* locus).

(B) Confocal images of ovarioles of *Drosophila* transgenic line showing Tomato (red) expression by immunostaining. Nuclei are labelled with DAPI (white). Scale bars: 10 µm.

(C) Quantification of relative transgene expression levels in transgenic ovaries by RT-qPCR. Tomato expression was normalized to the housekeeping gene *rp49*. Data are presented the mean (n=6), with error bars representing standard deviation. Statistical analysis was performed using the Mann-Whitney test, with ** indicating a P-value of 0.005.

(D) Boxplot showing changes in piRNA levels for the indicated clusters (split into 1-kb tiles) in *flam* mutant (*flam*△*prom*) relative to controls (*flam* WT). The boxplot displays the median (line) and the first and third quartiles (box).

(E) Uniquely mapped *flam* piRNAs are plotted displaying the piRNA profiles of heterozygous (top) and homozygous *flam* mutant (bottom) ovaries. Normalized per million of genome-mapped reads (0 mismatches).

(F) Image of a pair of *Drosophila* ovaries from *flam* heterozygous (top) or homozygous adult mutant (bottom).

Figure 5: Tj controls *gypsy-family TEs* silencing through regulation of *flam* expression

(A) Confocal images of tj^{eo2} clonal egg chamber showing *flam* RNA (magenta) by smFISH RNA and Tj (green) by immunostaining. The clone are circled in white. Nuclei are labelled with DAPI (white). Scale bars: 10 µm.

(B) Uniquely mapped piRNAs are plotted over the *42AB*, *20A* and *flam* piRNA clusters. The piRNA profiles of *tj*-Gal80ts-Gal4; *w-sh* (control) and *tj*-Gal80ts-Gal4; *tj-sh* adult ovaries are shown. Normalized per million of unique genome-mapping piRNAs (0 mismatches).

(C) Volcano plot showing upregulation of *gypsy*-family TEs by 96 h of *tj* siRNA knockdowns in OSCs using differential mRNA-seq analysis (DEseq2) between *tj* and *renilla* siRNA knockdowns (mRNA-seq, n=3 replicates from distinct samples). Grey circles are showing the genes; blue triangles are showing the transposons.

Figure 6: Tj regulates *flam* transcription independently of Piwi

(A) Morphology of a pair of ovaries from an adult *Drosophila* with genotypes *tj*-Gal4; *w-sh* (top) and *tj*-Gal4; *piwi-sh* (bottom).

(B) Confocal images of egg chambers with indicated genotype showing *flam* RNA (magenta) and Tj (green) by Immuno-FISH. A zoom on one egg chamber (white square) is presented in right. Scale bars: 10 µm.

(C) Quantification of relative transgene expression levels in transgenic ovaries by RT-qPCR. *tomato* expression levels were normalized to the expression of the housekeeping gene *rp49*. Data are presented as the mean (n=6) with error bars representing standard deviation.

Supplementary Figure legends

Figure S1: Tj motifs are enriched at soma-expressed genes and knockdown of *tj* in OSCs downregulates somatic piRNA pathway genes.

(A) Volcano plot showing downregulation of the somatic piRNA pathway genes by 96 h of *tj* siRNA knockdowns in OSCs using differential RNA-seq analysis (Deseq2) between *tj* and *renilla* siRNA knockdowns (n=3 replicates from distinct samples). Blue dots are showing the somatic-enriched piRNA pathway genes (e.g., *fs(1)Yb*, *nxf2*, *panx*, *soYb* and *armi*); green dots showing the soma-germ-shared piRNA factors (e.g., *piwi*) and red dots are showing the siRNA target *tj*.

(B) Fold-changes in *gypsy* TE expression upon the indicated siRNA knockdown in OSCs after 48 h are shown by RT-qPCR (*rpl32* housekeeping control, n=3, error bars indicate standard deviation).

(C) As in **(B)** but comparing fold-changes between 48 h and 96 h of *tj* knockdown

(D)-(F) Volcano plots showing upregulation of transposons by 48 h of *tj* siRNA in OSCs using differential mRNA-seq analysis (DEseq2) compared to *renilla* siRNA knockdowns (mRNA-seq, n=3 replicates from distinct samples). Grey circles are showing the genes; blue triangles are showing the transposons.

Figure S2: The region [+770; +2086] of *flam* is not sufficient to drive the transcription of the cluster.

(A) Schematic representation of the transgenic constructs.

(B) Confocal images of egg chambers with the indicated genotype showing Tomato (red) by immunostaining. The region [+770; +2086] of *flam* is not sufficient to drive the transcription of the cluster. Nuclei are stained with DAPI (white). Scale bars: 10 µm.

Figure S3: The loss of Traffic jam is sufficient to loss *flam* expression.

Confocal images of tj^{eo2} clonal egg chambers showing *flam RNA* (magenta) and Tj (green) by Immuno-FISH. tj^{eo2} clones are circled in white. Nuclei are labelled with DAPI (white). Scale bars: 10 µm.

Figure S4: TRAP experiments identify additional candidate transcripton factors regulating the somatic piRNA pathway

(A) Schematic representation of TRAP (Translating ribosome affinity purification) method performed on adult somatic cells carrying GFP tagged ribosomal protein RL10a, UAS-GFP::Rpl10AGFP; *tj*-Gal4 versus UAS-GFP::Rpl10AGFP; *nos*-Gal4.

(B)-(C). Gene ontology analyses of genes enriched in follicles cells. List of the significantly enriched GO Biological Process **(B)** and Molecular Function **(C)**.

Figure S5: Test of different candidates on *flam* expression

Confocal images of egg chambers with indicated genotype stained for the *flam* RNA using smRNA-FISH (magenta). The driver used is *tj*-Gal80ts-Gal4. Nuclei are labelled with DAPI (white). Scale bars: 10 µm.

Figure S6: Germline and somatic TF binding sites controlling the piRNA pathway are co-opted by *gypsy*-family TEs.

(A) Ovo and Tj ChIP-seq showing antagonistic binding at the germline and somatic genes.

(B) Model depiction of the germline and somatic transcriptional regulators of the piRNA pathway components in *Drosophila* ovaries and their co-option by transposons for mobilization and infection into the germline.

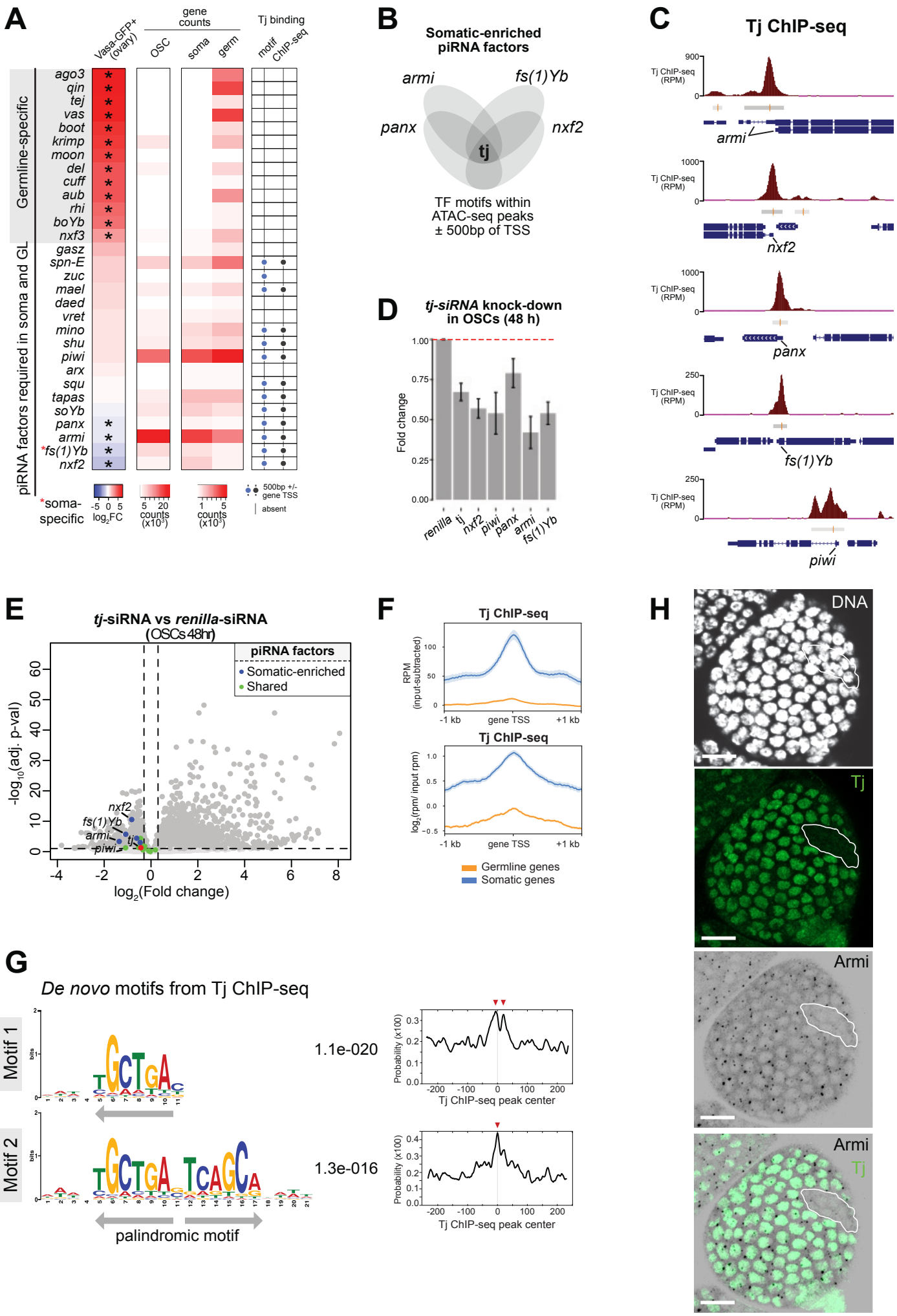

A
Vasa-GFP+ (ovary)
gene counts
OSC
soma
germ
Tj binding
motif
ChIP-seq
Germline-specific
piRNA factors required in soma and GL
ago3
qin
tej
vas
boot
krimp
moon
del
cuff
aub
rhi
boYb
nxf3
gasz
spn-E
zuc
mael
daed
vret
mino
shu
piwi
arx
squ
tapas
soYb
panx
armi
*fs(1)Yb
nxf2
*soma-specific
log2FC
counts (x10-3)
500bp +/- gene TSS
absent
B
Somatic-enriched piRNA factors
armi
fs(1)Yb
panx
nxf2
tj
TF motifs within ATAC-seq peaks ± 500bp of TSS
C
Tj ChIP-seq
Tj ChIP-seq (RPM)
armi
nxf2
panx
fs(1)Yb
piwi
D
tj-siRNA knock-down in OSCs (48 h)
Fold change
renilla
tj
nxf2
piwi
panx
armi
fs(1)Yb
E
tj-siRNA vs renilla-siRNA (OSCs 48hr)
piRNA factors
Somatic-enriched
Shared
-log10(adj. p-val)
log2(Fold change)
nxf2
fs(1)Yb
armi
tj
piwi
F
Tj ChIP-seq
RPM (input-subtracted)
-1 kb
gene TSS
+1 kb
log2(rpm/ input rpm)
Germline genes
Somatic genes
G
De novo motifs from Tj ChIP-seq
Motif 1
Motif 2
1.1e-020
1.3e-016
Probability (x100)
Tj ChIP-seq peak center
palindromic motif
H
DNA
Tj
Armi
Armi
Tj

Figure 2

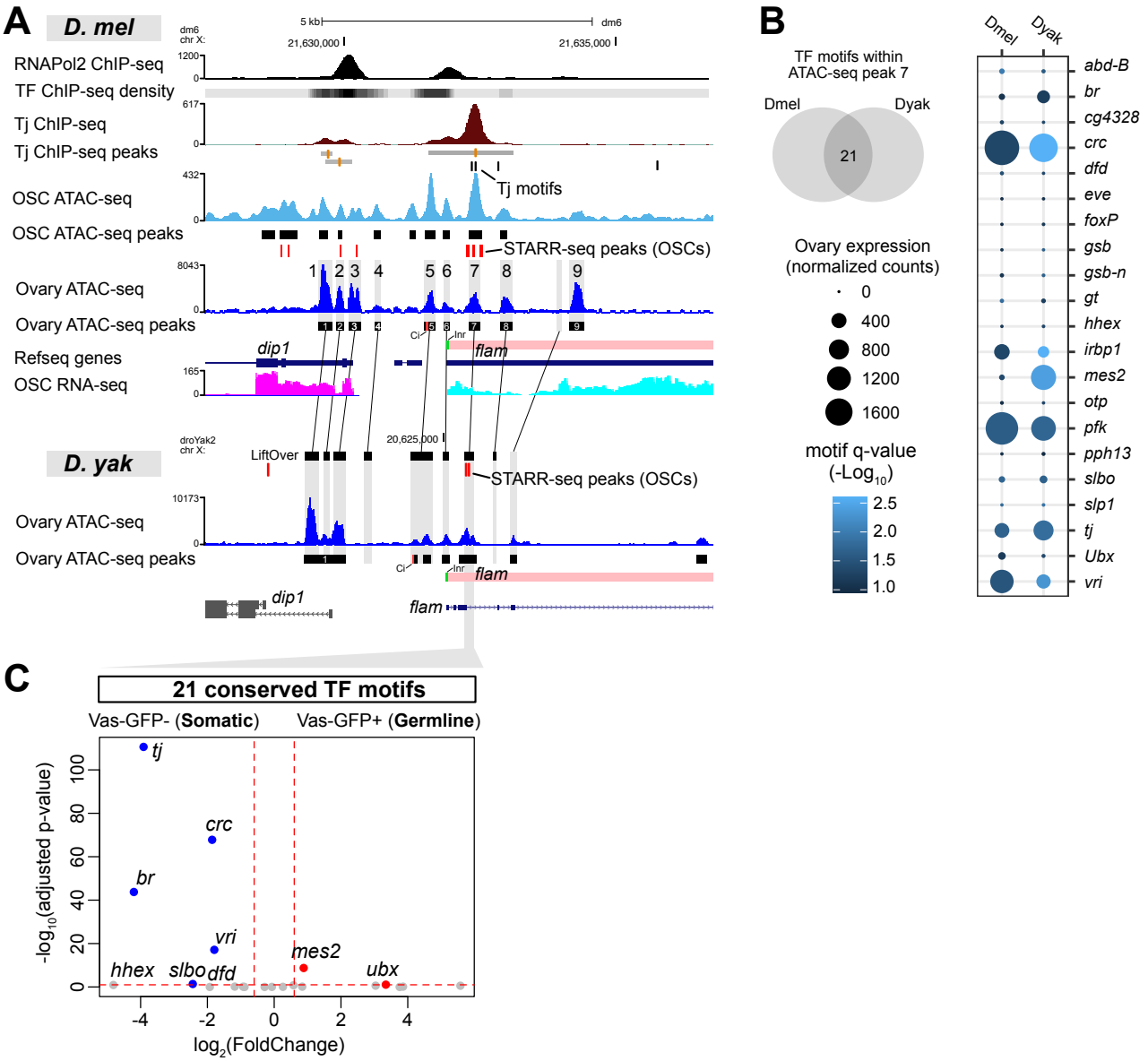

A
D. mel
RNAPol2 ChIP-seq
TF ChIP-seq density
Tj ChIP-seq
Tj ChIP-seq peaks
OSC ATAC-seq
OSC ATAC-seq peaks
Tj motifs
STARR-seq peaks (OSCs)
Ovary ATAC-seq
Ovary ATAC-seq peaks
Refseq genes
OSC RNA-seq
dip1
flam
D. yak
LiftOver
STARR-seq peaks (OSCs)
Ovary ATAC-seq
Ovary ATAC-seq peaks
B
TF motifs within ATAC-seq peak 7
Dmel
Dyak
21
Ovary expression (normalized counts)
0
400
800
1200
1600
motif q-value (-Log10)
2.5
2.0
1.5
1.0
abd-B
br
cg4328
crc
dfd
eve
foxP
gsb
gsb-n
gt
hhex
irbp1
mes2
otp
pfk
pph13
slbo
slp1
tj
Ubx
vri
C
21 conserved TF motifs
Vas-GFP- (Somatic)
Vas-GFP+ (Germline)
-log10(adjusted p-value)
log2(FoldChange)
tj
crc
br
vri
hhex
slbo
dfd
mes2
ubx

Figure 3

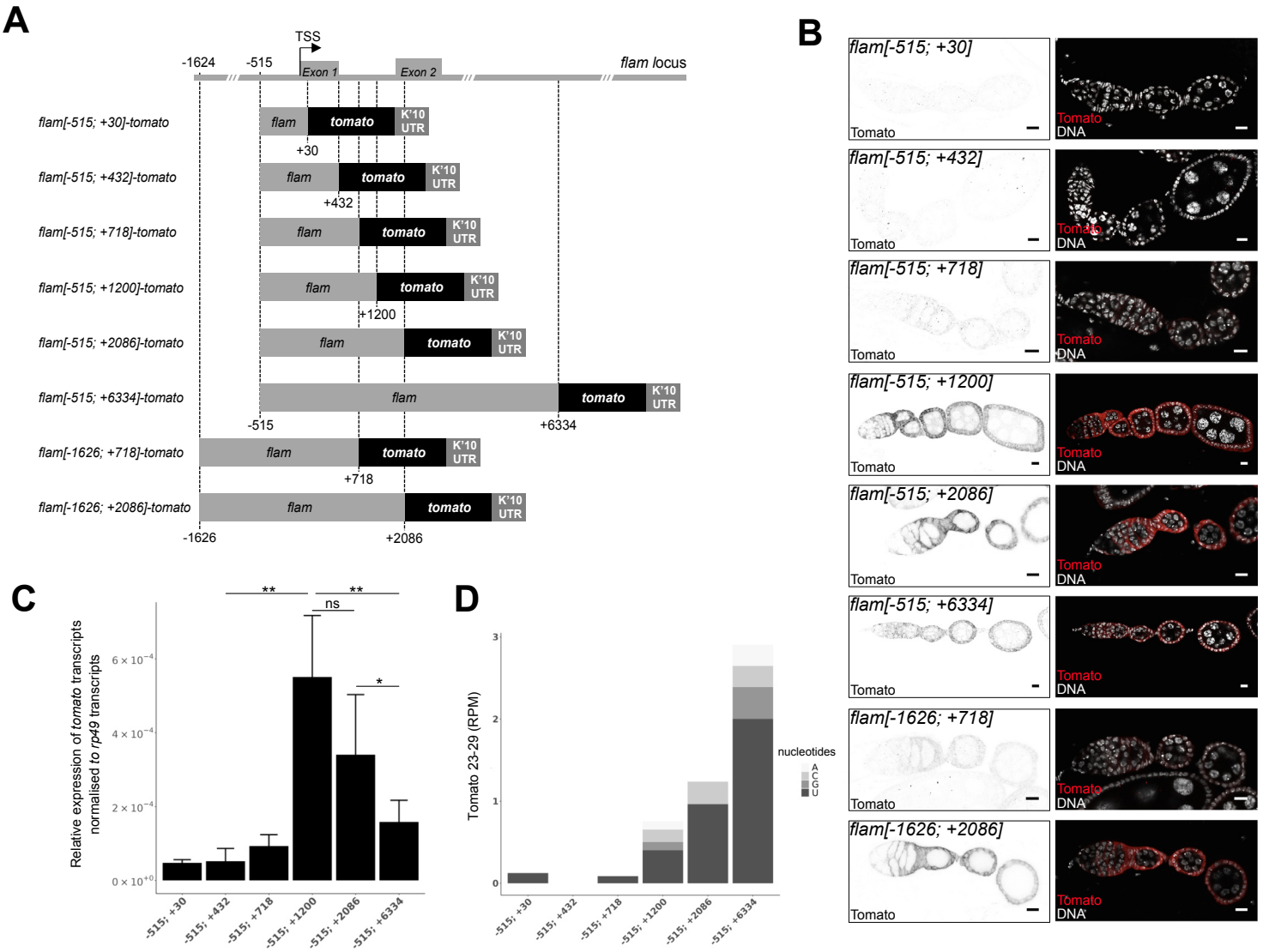

A
TSS
-1624
-515
Exon 1
Exon 2
flam locus
flam[-515; +30]-tomato
flam[-515; +432]-tomato
flam[-515; +718]-tomato
flam[-515; +1200]-tomato
flam[-515; +2086]-tomato
flam[-515; +6334]-tomato
flam[-1626; +718]-tomato
flam[-1626; +2086]-tomato
flam
tomato
K'10 UTR
+30
+432
+1200
+718
+2086
+6334
-1626
B
flam[-515; +30]
flam[-515; +432]
flam[-515; +718]
flam[-515; +1200]
flam[-515; +2086]
flam[-515; +6334]
flam[-1626; +718]
flam[-1626; +2086]
Tomato
DNA
C
Relative expression of tomato transcripts normalised to rp49 transcripts
ns
D
Tomato 23-29 (RPM)
nucleotides
A
C
G
U

Figure 4

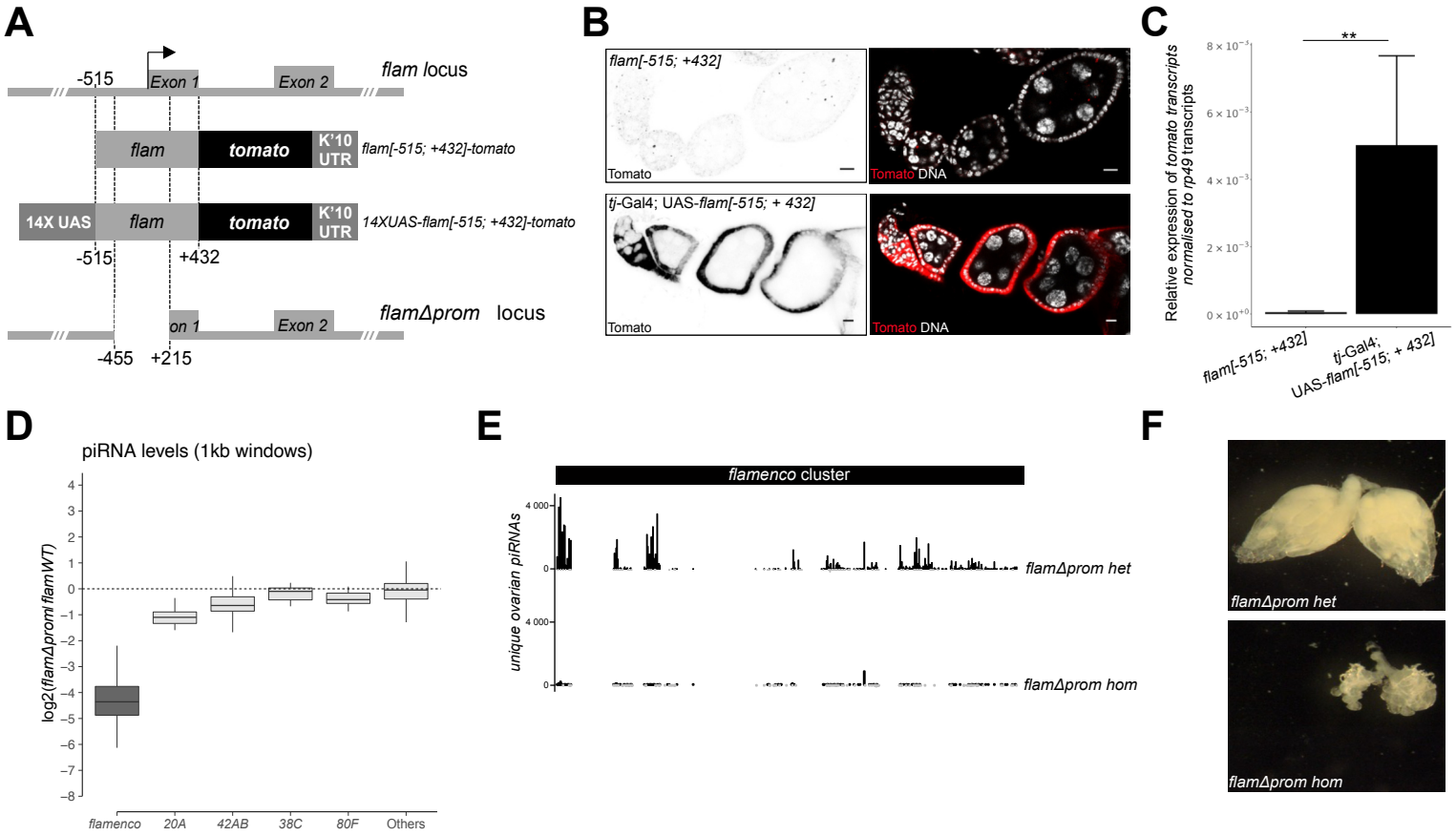

A
-515
Exon 1
Exon 2
flam locus
flam
tomato
K'10 UTR
flam[-515; +432]-tomato
14X UAS
flam
tomato
K'10 UTR
14XUAS-flam[-515; +432]-tomato
-515
+432
flamΔprom locus
on 1
Exon 2
-455
+215
B
flam[-515; +432]
Tomato
Tomato DNA
tj-Gal4; UAS-flam[-515; + 432]
Tomato
Tomato DNA
C
Relative expression of tomato transcripts normalised to rp49 transcripts
**
flam[-515; +432]
tj-Gal4; UAS-flam[-515; + 432]
D
piRNA levels (1kb windows)
log2(flamΔprom/flamWT)
flamenco
20A
42AB
38C
80F
Others
E
flamenco cluster
unique ovarian piRNAs
flamΔprom het
flamΔprom hom
F
flamΔprom het
flamΔprom hom

Figure 5

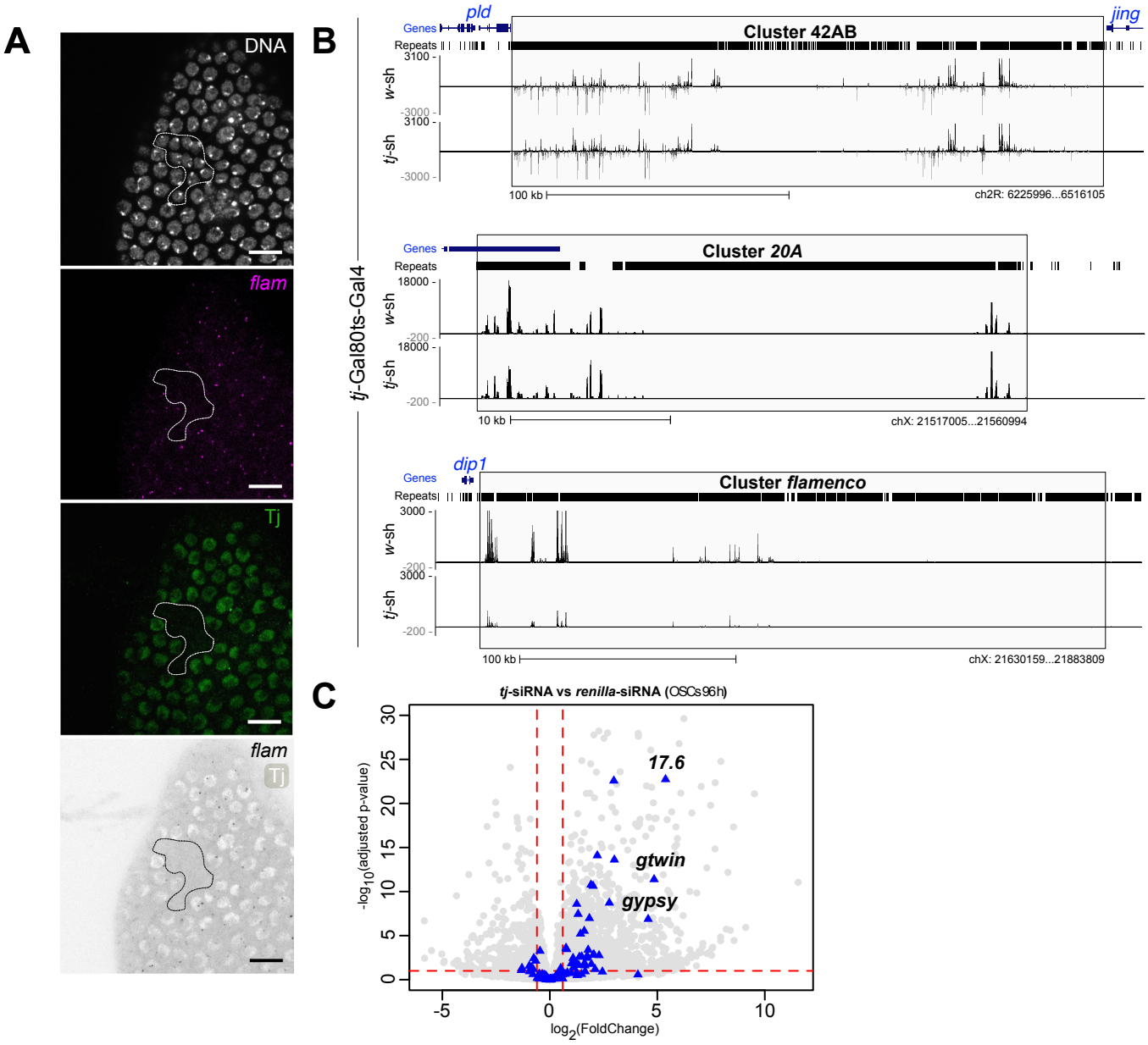

A
DNA
flam
Tj
flam
Tj
B
tj-Gal80ts-Gal4
Genes
Repeats
pld
jing
Cluster 42AB
w-sh
tj-sh
3100
-3000
100 kb
ch2R: 6225996...6516105
Genes
Repeats
Cluster 20A
18000
-200
10 kb
chX: 21517005...21560994
Genes
dip1
Repeats
Cluster flamenco
3000
-200
100 kb
chX: 21630159...21883809
C
tj-siRNA vs renilla-siRNA (OSCs 96h)
17.6
gtwin
gypsy
-log10(adjusted p-value)
log2(FoldChange)
0
5
10
15
20
25
30
-5
0
5
10

Figure 6

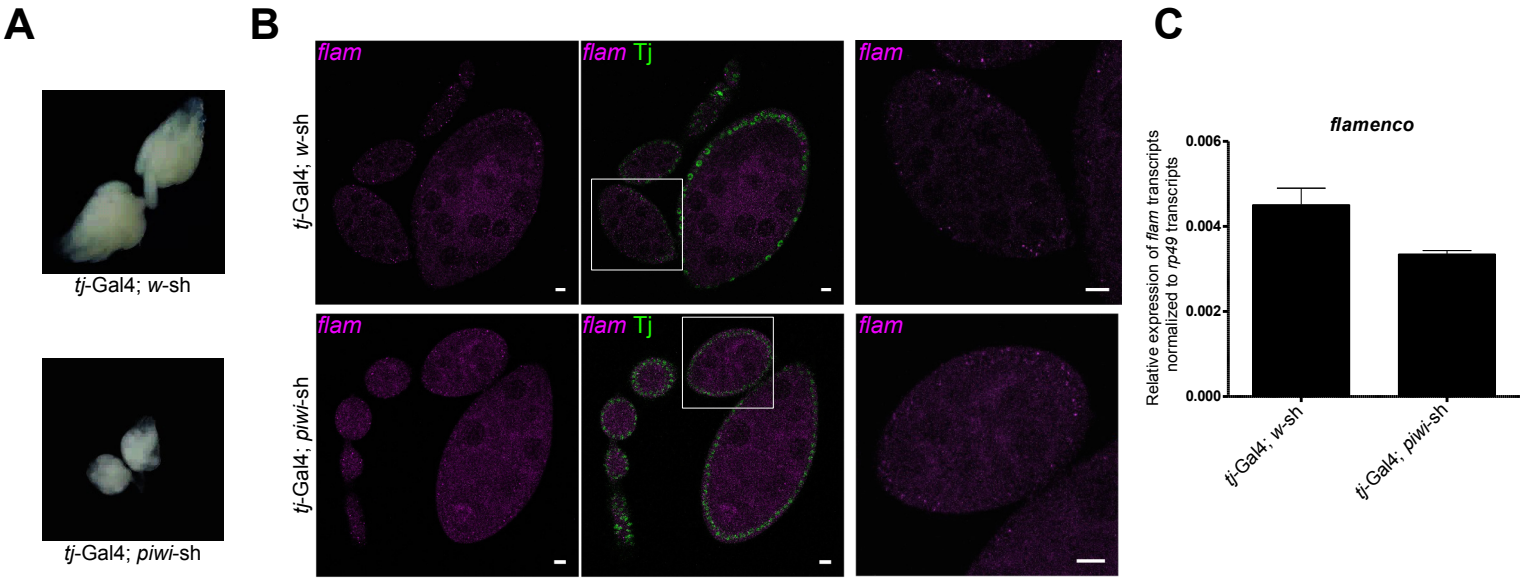

A
tj-Gal4; w-sh
tj-Gal4; piwi-sh
B
tj-Gal4; w-sh
tj-Gal4; piwi-sh
flam
flam Tj
flam
flam
flam Tj
flam
C
flamenco
Relative expression of flam transcripts normalized to rp49 transcripts
0.006
0.004
0.002
0.000
tj-Gal4; w-sh
tj-Gal4; piwi-sh

Figure S1

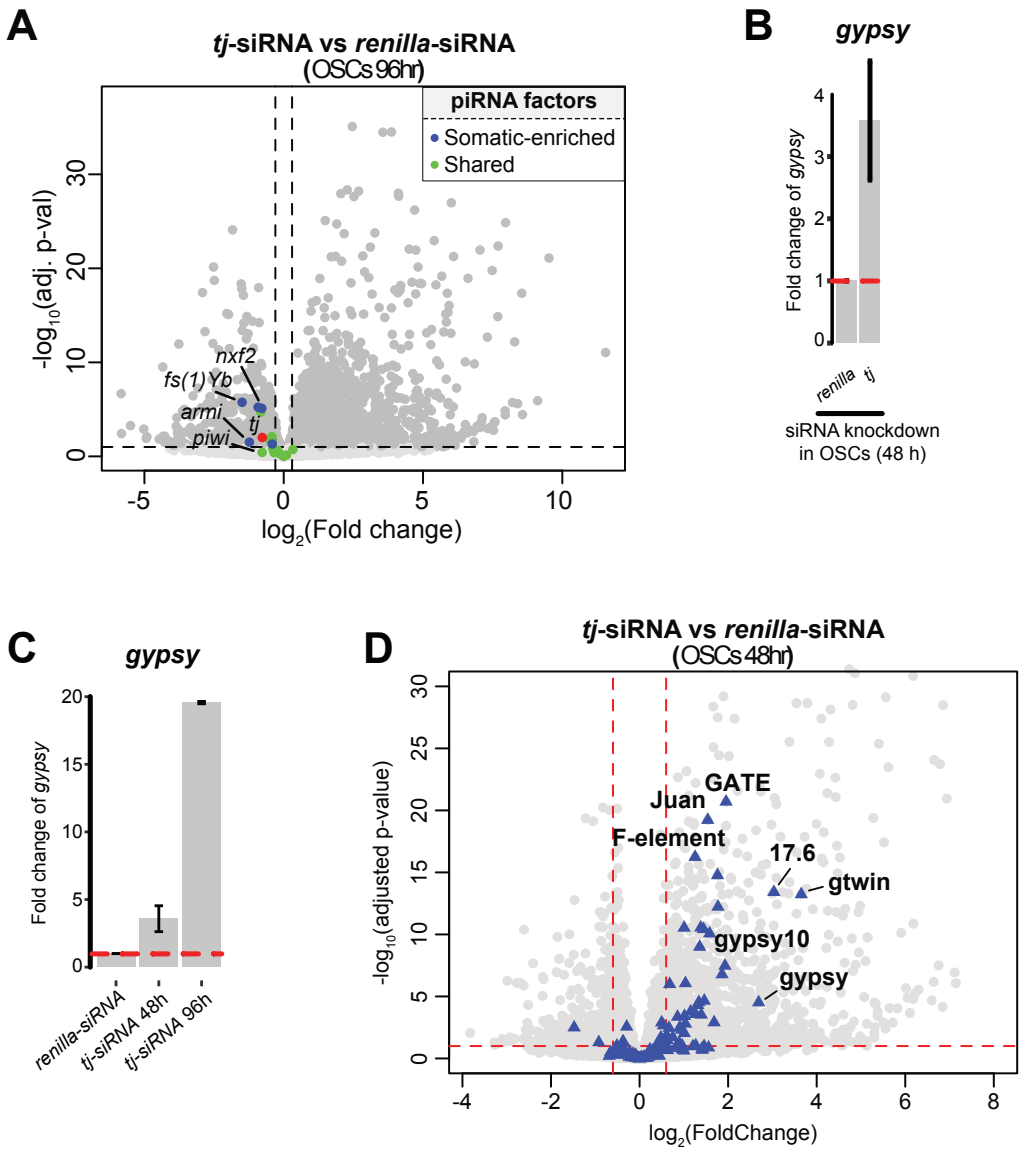

A
tj-siRNA vs renilla-siRNA
(OSCs 96hr)
piRNA factors
Somatic-enriched
Shared
nxf2
fs(1)Yb
armi
tj
piwi
-log10(adj. p-val)
log2(Fold change)
B
gypsy
Fold change of gypsy
renilla
tj
siRNA knockdown
in OSCs (48 h)
C
gypsy
Fold change of gypsy
renilla-siRNA
tj-siRNA 48h
tj-siRNA 96h
D
tj-siRNA vs renilla-siRNA
(OSCs 48hr)
GATE
Juan
F-element
17.6
gtwin
gypsy10
gypsy
-log10(adjusted p-value)
log2(FoldChange)

Figure S2

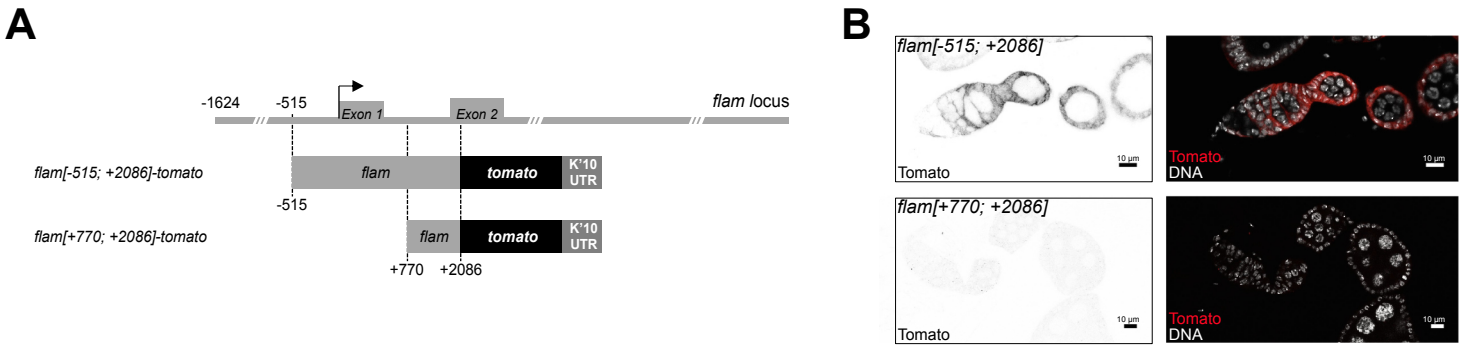

A
-1624
-515
Exon 1
Exon 2
flam locus
flam[-515; +2086]-tomato
flam
tomato
K'10 UTR
-515
flam[+770; +2086]-tomato
flam
tomato
K'10 UTR
+770
+2086
B
flam[-515; +2086]
Tomato
10 µm
Tomato
DNA
10 µm
flam[+770; +2086]
Tomato
10 µm
Tomato
DNA
10 µm

Figure S3

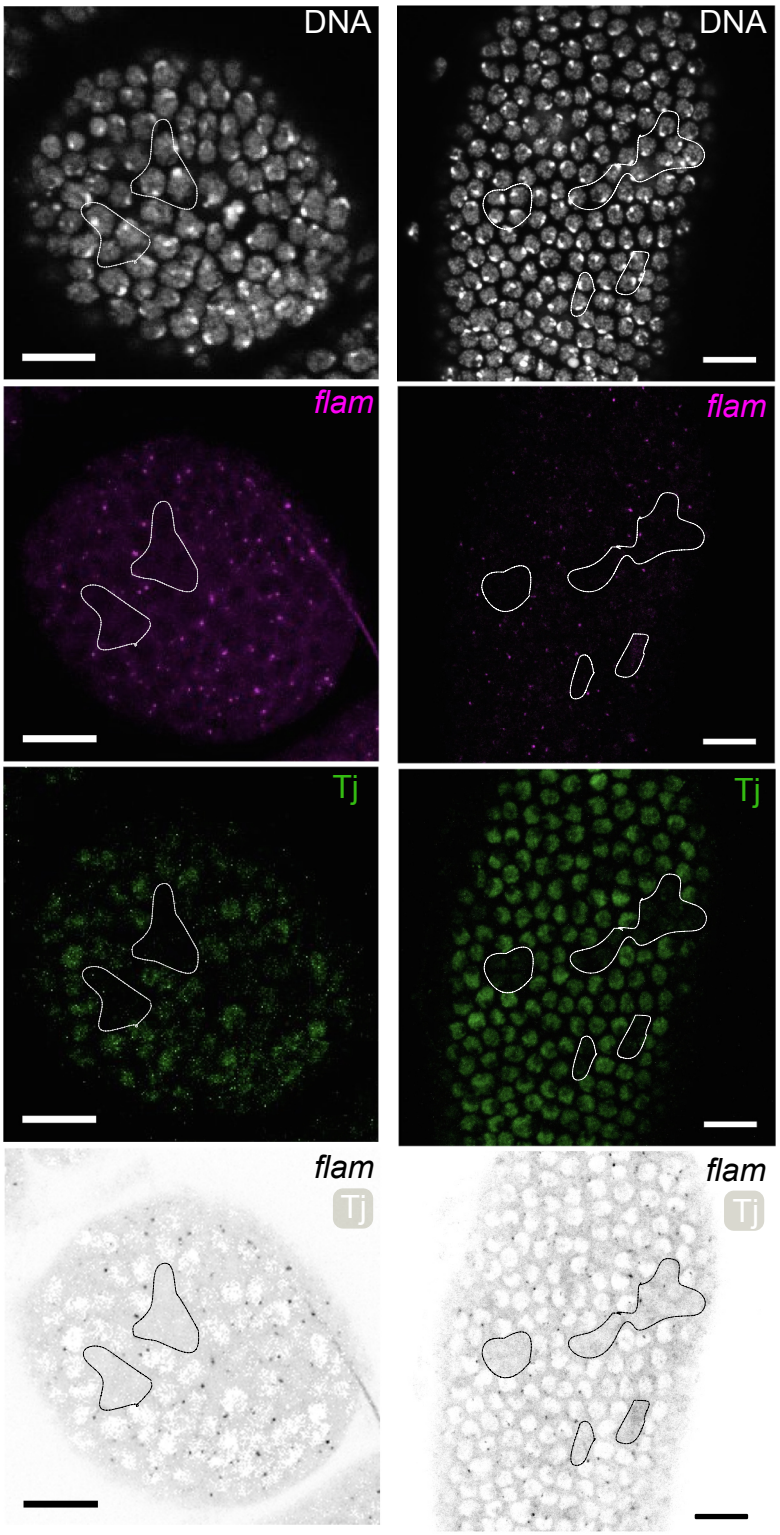


Figure S4

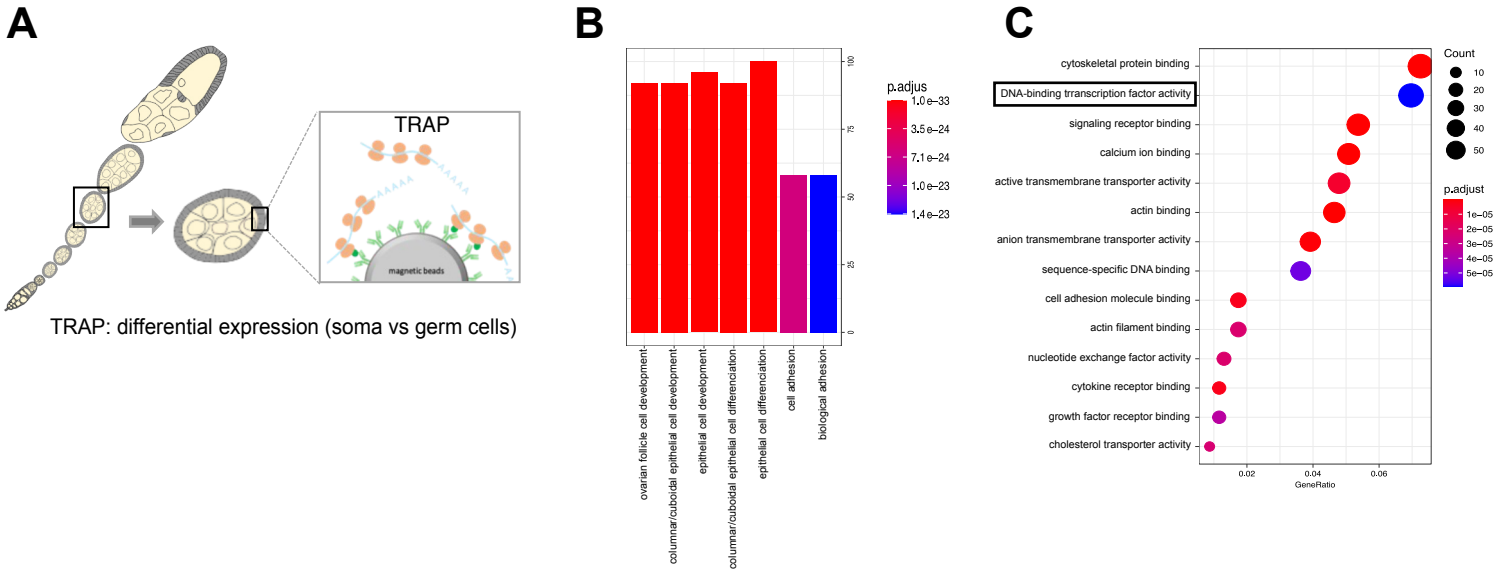

A
TRAP
magnetic beads
TRAP: differential expression (soma vs germ cells)
B
p.adjus
1.0e−33
3.5e−24
7.1e−24
1.0e−23
1.4e−23
ovarian follicle cell development
columnar/cuboidal epithelial cell development
epithelial cell development
columnar/cuboidal epithelial cell differenciation
epithelial cell differenciation
cell adhesion
biological adhesion
C
cytoskeletal protein binding
DNA-binding transcription factor activity
signaling receptor binding
calcium ion binding
active transmembrane transporter activity
actin binding
anion transmembrane transporter activity
sequence-specific DNA binding
cell adhesion molecule binding
actin filament binding
nucleotide exchange factor activity
cytokine receptor binding
growth factor receptor binding
cholesterol transporter activity
GeneRatio
Count
p.adjust

Figure S5

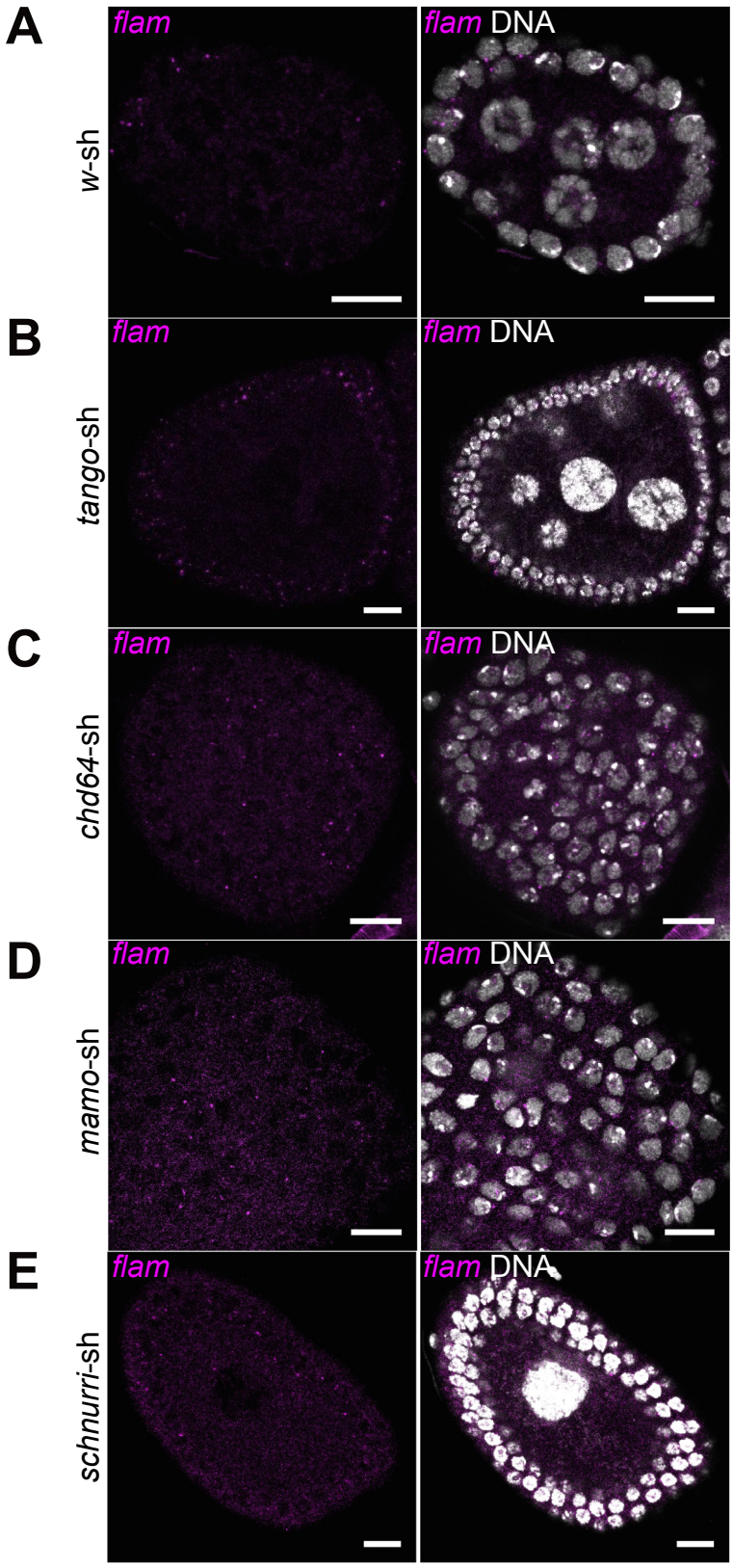
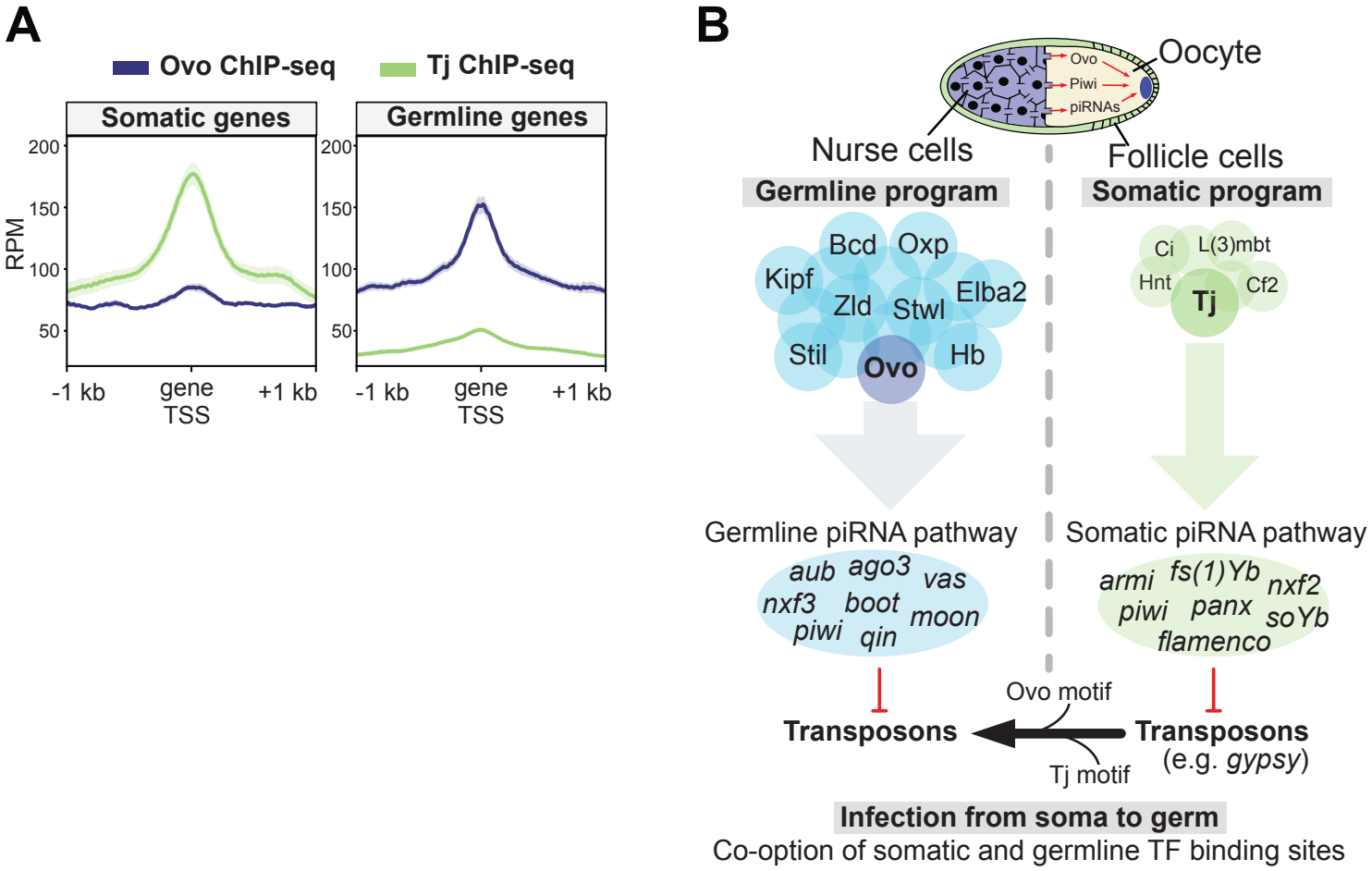

A
w-sh
flam
flam DNA
B
tango-sh
flam
flam DNA
C
chd64-sh
flam
flam DNA
D
mamo-sh
flam
flam DNA
E
schnurri-sh
flam
flam DNA

Figure S6


A
Ovo ChIP-seq
Tj ChIP-seq
Somatic genes
Germline genes
RPM
200
150
100
50
-1 kb
gene TSS
+1 kb
B
Oocyte
Ovo
Piwi
piRNAs
Nurse cells
Follicle cells
Germline program
Somatic program
Bcd
Oxp
Kipf
Zld
Stwl
Elba2
Stil
Ovo
Hb
Ci
L(3)mbt
Hnt
Tj
Cf2
Germline piRNA pathway
Somatic piRNA pathway
aub ago3 vas
nxf3 boot moon
piwi qin
armi fs(1)Yb nxf2
piwi panx soYb
flamenco
Ovo motif
Tj motif
Transposons
Transposons
(e.g. gypsy)
Infection from soma to germ
Co-option of somatic and germline TF binding sites