Scientific Report

TRANSPARENT PROCESS

EMBO reports

Transcriptional properties and splicing of the *flamenco* piRNA cluster

Coline Goriaux[1,2,3,†], Sophie Desset[1,2,3,†], Yoan Renaud[1,2,3], Chantal Vaury[1,2,3,**] & Emilie Brasset[1,2,3,*]

Abstract

In *Drosophila*, the piRNA cluster, *flamenco*, produces most of the piRNAs (PIWI-interacting RNAs) that silence transposable elements in the somatic follicle cells during oogenesis. These piRNAs are thought to be processed from a long single-stranded precursor transcript. Here, we demonstrate that *flamenco* transcription is initiated from an RNA polymerase II promoter containing an initiator motif (Inr) and downstream promoter element (DPE) and requires the transcription factor, Cubitus interruptus. We show that the *flamenco* precursor transcript undergoes differential alternative splicing to generate diverse RNA precursors that are processed to piRNAs. Our data reveal dynamic processing steps giving rise to piRNA cluster precursors.

Keywords *Drosophila*; *flamenco*; piRNA clusters; transcription; transposable elements

Subject Categories RNA Biology; Transcription

DOI 10.1002/embr.201337898 | Received 19 August 2013 | Revised 29 January 2014 | Accepted 31 January 2014 | Published online 21 February 2014

EMBO Reports (2014) 15, 411–418

Introduction

Small non-coding RNAs can induce gene silencing through specific base pairing with target molecules. A subclass of small non-coding RNAs (23–29 nt) that interact specifically with the PIWI clade of Argonaute proteins, the PIWI-interacting RNAs (piRNAs), ensures genomic stability by repressing the expression and transposition of transposable elements (TE) in reproductive tissues including *Drosophila* germline and surrounding somatic follicle cells. Most piRNAs are derived from presumed long, single-stranded precursor transcripts encoded by genomic loci known as piRNA clusters [1].

In somatic cells, a major piRNA cluster, *flamenco* (*flam*), controls several TEs such as *gypsy*, *Idefix* and *ZAM* [1–5]. This 180-kb locus is located at the boundary between euchromatin and pericentromeric heterochromatin on the *Drosophila* X-chromosome, proximal to the *DIP1* gene. It harbours many defective transposons similarly oriented to produce antisense transcripts capable of silencing active transposon mRNAs. We recently reported that the *flam* RNA precursor is transported from the genomic site where it is produced to a perinuclear structure called Dot COM juxtaposed with cytoplasmic Yb bodies where primary piRNA biogenesis occurs [6,7]. Promoters and transcription factors involved in piRNA cluster transcription are starting to be identified. In *Drosophila melanogaster*, Rhino and Cutoff are required for transcription/processing of germinal bidirectional piRNA clusters. In mice, the transcription factor MYB-related protein A has been reported to drive transcription of specific piRNA clusters [8–10]. To provide further understanding of piRNA cluster transcription, we undertook a comprehensive characterization of *flam* expression. We identified its transcription start site (TSS) and a transcription factor critical for its transcription in follicle cells. Our results also demonstrated that the *flam* transcript is alternatively spliced to generate multiple and distinct precursors.

Results and Discussion

An Inr-DPE Pol II promoter promotes *flam* piRNA cluster transcription

To identify the *flam* TSS, we performed 5′RACE experiments on four independent RNA extracts from *Drosophila* ovaries and ovarian somatic stem (OSS) cells (Supplementary Table S1). From the capped RNA fraction, a TSS located at position 21,502,918 (flybase version FB2011_08) was identified in all the independent amplifications from both ovary and OSS cell RNA extracts (Fig 1A). Several other TSSs (a total of 10) were occasionally amplified but were found in only one of the experiments performed. These data suggest that the *flam* transcripts are initiated from a major promoter located 1733 bp upstream of *DIP1*.

To gain a better understanding of the core promoter of *flam*, we examined the motifs located upstream and downstream of the TSS. Based on the consensus initiator element (Inr) sequence TCAGTY obtained by computational analysis of thousands of *Drosophila* core

1 Clermont Université, Université d'Auvergne, Clermont-Ferrand, France
2 Inserm, Unité 1103, Clermont-Ferrand, France
3 CNRS, UMR 6293, Clermont-Ferrand, France
*Corresponding author. Tel: +33 4 7317 8184; Fax: +33 4 7327 6132; E-mail: emilie.brasset@udamail.fr
**Corresponding author. Tel: +33 4 7317 8184; Fax: +33 4 7327 6132; E-mail: chantal.vaury@udamail.fr
†These authors contributed equally to this work.

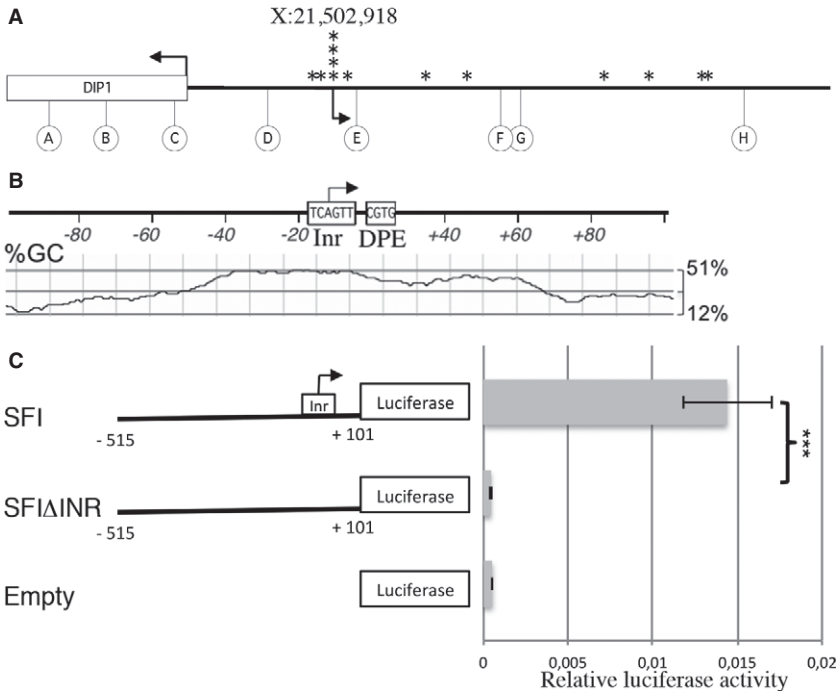

Figure 1. The *flam* locus is transcribed from an Inr-downstream promoter element (DPE) promoter.

A 5′RACE experiment on total RNA from ovary or ovarian somatic stem (OSS) cells. The arrowheads indicate the transcription start site (TSS) of the *DIP1* gene (X:21,501,185), the major TSS that initiates *flam* transcription (X:21,502,918), and the stars indicate the additional TSSs. Circles indicate primers used for 5′ RACE.

B *In silico* sequence analysis of the major *flam* TSS. The Inr and DPE motif are depicted, and the GC content is represented below (window size: 30 nucleotides).

C Schematic representation of the SF reporter constructs (left panel). Numbers indicate the fragment length from the TSS. SFΔInr carries a deletion of the predicted Inr sequence from −2 to +4. OSS cells were co-transfected with the firefly luciferase reporters carrying different *flam* promoter fragments and with an actin Renilla luciferase reporter construct (Supplementary Methods). Firefly luciferase (Fluc) activity was normalized to that of the Renilla luciferase (Rluc). Data are presented as means ($n = 4$). Error bars represent ± s.d.; Students *t*-test: ***$P < 0.001$.

promoters [11,12], we found that only the major TSS contains a consensus Inr sequence TCAGTT. In this Inr element, the A nucleotide corresponds to the +1 position of the core promoter (Fig 1B). Further analysis did not reveal a consensus TATA box, where the upstream T is usually located at −31 or −30 nt relative to the A +1 (or G +1) position in the Inr. However, a CGTG tetramer was characterized at +23 to +26 bp of the major TSS as a downstream promoter element (DPE), which is typically over-represented in many *Drosophila* TATA-less promoters. Like many *Drosophila* and mammalian promoters [13,14], a wide area in the vicinity of the major *flam* TSS (from −50 to +70 bp) displays a significant increase in GC content, which is known as a "GC hill." Aside from this major TSS, no other TSSs identified in this experiment displayed such promoter characteristics. Overall, these data designate the TSS located at 21,502,918 as the main promoter of the *flam* piRNA cluster.

To assess the potential of the *flam* Inr core promoter to drive transcription, the promoter region (SFI) including 515 bp upstream of the TSS and 101 bp of the transcribed sequence was cloned upstream of the luciferase reporter gene at the ATG start codon of the coding region. Transcriptional activities were measured in transient transfection experiments in OSS cells. Our results indicate that this *flam* fragment is sufficient to promote high-level expression of the *luciferase* reporter gene since an almost 30-fold enhancement of transcription of the firefly *luciferase* gene was observed compared to the empty plasmid (Fig 1C). Then, we generated a new reporter, SFIΔInr, that lacks the Inr sequence. This deleted reporter resulted in a significant decrease in luciferase expression compared to the transcriptional enhancement exhibited by the wild-type SFI. These results confirm the importance of the Inr sequence for promoting transcription of the *flam* locus.

The presence of an Inr core promoter and a cap structure indicates that RNA polymerase II (Pol II) could be responsible for *flam* transcription. In order to test this hypothesis, we treated OSS cells with alpha-amanitin, an inhibitor of initiation and elongation of Pol II. Transcription efficiency of the *flam* locus was determined by RT-qPCR using primer pairs spanning three different regions of *flam*. *18S* ribosomal RNA known to be transcribed by RNA polymerase I (Pol I) was used as a reference gene for normalization.

We found up to tenfold decreases in *flam*-derived long RNAs in cells cultured in the presence of alpha-amanitin, indicating that *flam* transcription is indeed Pol II dependent (Fig 2A). The amount of *rp49* transcripts (known to be transcribed by Pol II) is shown as a positive control. Moreover, using Pol I or Pol III inhibitors [15,16], we confirmed that *flam* transcripts are indeed products of Pol II (Supplementary Fig S1).

Then, we performed ChIP-qPCR experiments using an antibody against the initiating form of Pol II. We found that Pol II was more strongly recruited immediately downstream of the *flam* TSS than elsewhere within the gene body (Fig 2B). Thus, Pol II is the polymerase involved in *flam* piRNA cluster transcription. These results extend findings obtained in mouse testes, in which piRNA precursor

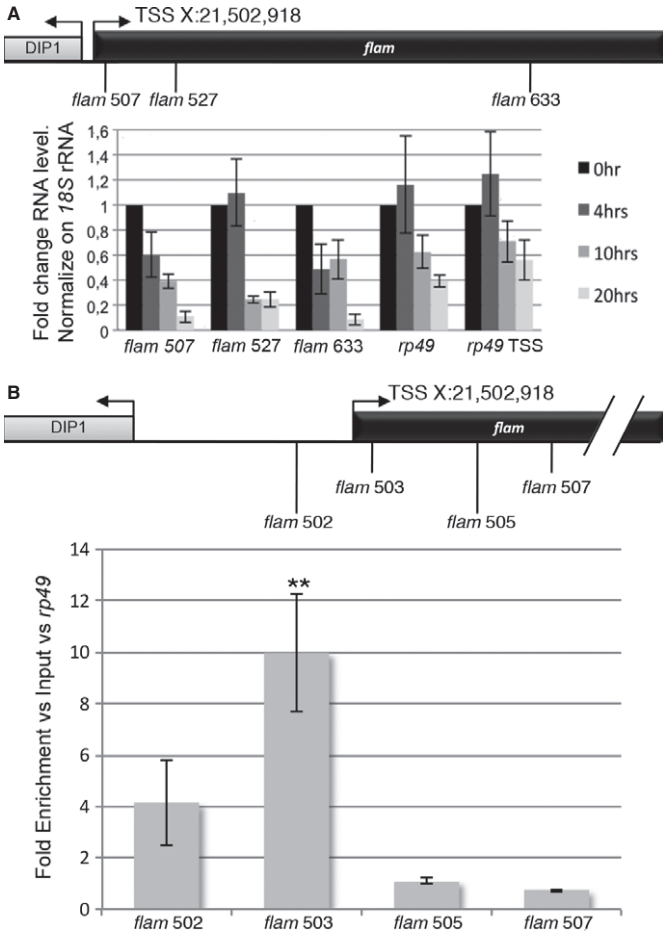

Figure 2. The *flam* locus is transcribed by RNA polymerase II.

A RT-qPCR on total RNAs from ovarian somatic stem (OSS) cells treated with α-amanitin from 0 to 20 h. The positions of primers along the *flam* sequence are indicated above. The amount of *rp49* transcript is shown as a positive control. The first primer set is located in exon 3, and the second primer set in exon 1 of the *rp49 gene*. qPCR data were normalized to *18S* rRNA. Data are presented as means (n = 4). Error bars represent ± s.d. Expression in non-treated OSS cells was set to one.

B ChIP-qPCR analysis was carried out on the *flam* promoter with a specific antibody against phospho-S5 RNA polymerase II. The positions of primers used for PCR amplification are indicated above. Enrichments were calculated versus Input and *rp49*. Data are presented as means (n = 5). Error bars represent ± s.e.m.; Students t-test: $^{**}P < 0.01$.

transcripts have been described to be canonical Pol II transcripts bearing 5′caps and 3′ poly(A) [10].

The transcription factor, Cubitus interruptus, is required to activate transcription of the *flam* locus

To identify cis-regulatory sequences, we constructed serially deleted promoter-*luciferase* reporter plasmids containing various lengths of the *flam* promoter region from either −1,624 bp (SF), −515 bp (SFI) or −356 bp (SFII) upstream to +101 bp downstream of the TSS. When the SF construct was used for transfection, efficient reporter activity was detected (Fig 3A). Deletion of the region from −1,624 to −515 (SFI) did not result in any significant change in promoter activity. On the contrary, further deletion to −356 (SFII) caused an eightfold decrease in promoter activity compared to the SFI construct. Finally, a NC construct corresponding to SFI in which the *flam* fragment comprised between −515 and −356 has been replaced by a 159-bp fragment of a non-promoting sequence, confirmed that the region located downstream of position X: 21,502,403 (−515 bp) and upstream of position X: 21,502,562 (−356 bp) contains critical *cis*-elements required for the transcriptional activation of the locus.

Within the −515; −356 region, nine potential transcription factor-binding sites were identified using genomatix MatInspector (Fig 3B). Based on the modENCODE dataset, four of them are expressed in OSS cells: Broad (Br), Big-brother (BgB), Doublesex (Dsx) and Cubitus interruptus (Ci). To specifically analyse the involvement of these factors in *flam* transcription, we performed successive deletions of each of their predicted binding sites (Fig 3C). The expression of each construct significantly decreased when compared with the SFI control but the most severe reduction (tenfold) was observed with SFI deleted for the Ci binding site, which was similar to the levels seen with the SFII construct. This suggests that the Ci binding site is necessary for the activation of *flam* transcription.

Several lines of evidence further implicated Ci in regulating *flam* transcription. First, Ci is expressed in follicle cells from the germarium to stage 6 egg chambers (Fig 4A) (Supplementary Fig S2) [17]. Second, based on ChIP assays, we found that Ci is 10- to 12-fold more recruited around the TSS and its predicted binding site than elsewhere in the locus (Fig 4B). Third, mutant clones generated by mitotic recombination using flies [*y-hs-flp*; *FRT42D P*[*Ci*+] /*FRT42D hs-MYC 45*; *Ci94*/*Ci94*] indicated that the *flam* transcript level decreases in *Ci* mutants in a manner similar to the decrease observed for *ptc* transcripts, a gene known to be activated by Ci, but not producer of piRNAs [18] (Supplementary Fig S2). Fourth, siRNA-mediated knockdown of *Ci* in OSS cells led to a decrease in *flam* transcripts two days post-transfection (Fig 4C). In contrast, the production of piRNAs and the TE mRNA levels were not significantly affected (Supplementary Fig S3). However, an upregulation of TE expression was observed 4 days post-infection (Fig 4D). A delay is observed between disruption of *flam* transcription and TE deregulation possibly due to stability and abundance of *flam* piRNAs.

Finally, evidence that Ci is involved in *flam* transcription was also provided by an analysis of the *flam* mutation present in the BG lines [19]. In this line, a P-element insertion at the 5′ end of *flam* results in an absence of the precursor transcripts encoded by *flam* [1]. When examined in detail, we found that the P-insertion occurred at position X:21,502,538 (−380 bp from the TSS), a position that disrupts the Ci binding site. Considered together, these data strongly suggest a role for Ci in the activation of *flam* transcription.

In *Drosophila* somatic follicle cells, the major sources of piRNAs are the *flam* locus and the cluster 2. Thus, we examined the cluster 2 promoter and found an Inr consensus sequence (21,390,615) 108 bp upstream of the first piRNA, and a Ci binding site 2,846 bp upstream of the Inr (Supplementary Fig S4). Furthermore, Ci mutants led to a decrease in *cluster 2* expression (Fig 4C and Supplementary Fig S2). These data suggest that Ci might also contribute to the transcription of other piRNA clusters in these cells.

A comparative analysis of the *flam* promoter region performed across several *Drosophila* species, *D. sechellia, D. simulans, D. yakuba, D. erecta,* was then performed. These species diverged from a common ancestor approximately 10 million years ago

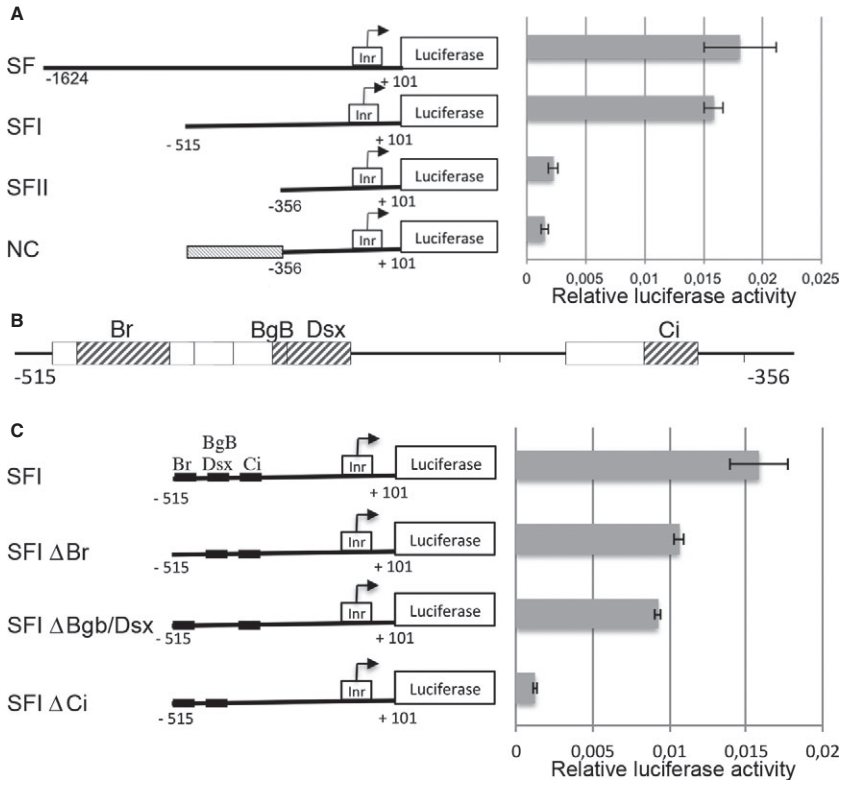

Figure 3. A functional Ci binding site is required for *flam* transcription.

A Schematic representations of the SF reporter constructs (left panel). The relative luciferase activity (Fluc/Rluc) was measured as described previously (right panel). Data are presented as means (n = 4). Error bars represent ± s.d. In the negative control construct (NC), a 159-bp fragment of a non-promoting sequence taken within the *gfp* gene was cloned upstream the −356 to fill the space between −515 and −356.

B Genomatix *in silico* analysis of the region from −515 to −356 upstream of the transcription start site (TSS). Boxes indicate transcription factor-binding sites. Grey boxes indicate transcription factors known to be expressed in ovarian somatic stem (OSS) cells (modEncode data).

C Schematic representation of firefly reporters carrying deletions of each predicted transcription factor-binding site. The Fluc/Rluc activity was measured as described previously (right panel). Data are presented as means (n = 4). Error bars represent ± s.d.

[20,21]. We found that *flam* orthologs are located on the pericentromeric X-chromosome close to the *DIP1* gene in *D. simulans* and *D. erecta*, similar to *D. melanogaster*, whereas they are still assigned in a scaffold in *D. yakuba* and *D. sechellia* (Supplementary Table S2). A multiple alignment revealed two highly conserved regions located at positions (−14;+37) and (−398; −372) according to the *D. melanogaster flam* TSS. The first (−14;+37) corresponds to the Inr-DPE core promoter suggesting a high conservation of its function. The second (−398; −372) includes the Ci binding site (Supplementary Fig S4). Then, we plotted uniquely mapping piRNAs that could be assigned to the putative *D. erecta flam* locus [5]. We found that, like in *D. melanogaster*, the density of piRNAs is very low close to the *flam* presumptive promoter and it highly increases 1 kb downstream (Supplementary Fig S4). This analysis of the *flam* promoter sequence across several *Drosophila* species confirms that the Inr-DPE and the Ci binding site are necessary motifs for *flam* transcription.

The *flam* transcript is alternatively spliced and gives rise to multiple *flam* precursors

The *flam* piRNA cluster has been proposed to produce a long single-stranded precursor RNA that is processed into primary piRNAs in the cytoplasmic Yb bodies [6,22]. We sought to better characterize this proposed long precursor. Fragments amplified from the 5′RACE experiments described above to localize the TSS were systematically sequenced. This allowed the identification of an intron located between bases +432 and +2067 from the *flam* promoter. Then, RT-PCR experiments were performed using a 5′ primer taken either within the first or the second exon, and 3′ primers designed along the 180 kb of this cluster. Figure 5A shows structures of *flam* transcripts deduced from sequencing of RT-PCR products. Different patterns of intron splicing were detected. The intron sizes are extremely diverse and range from 0.7 kb to 158 kb. Interestingly, the first exon (exon 1: 21,502,918…21,503,349) was found to be constitutively spliced since it is always present within the processed RNAs. By contrast, downstream of this first common exon, the other exons differ indicating that they result from alternative splicing. Analysis of *flam* spliced transcripts revealed that the majority of the intron boundaries obey the GT-AG rule (Supplementary Tables S3 and S4).

To verify our findings, we interrogated publicly available RNA-seq libraries [23] and found that indeed very few reads corresponding to intron 1 have been reported compared to the number of reads mapping exon 1 or exon 2 (Fig 5B). We found that 84% and

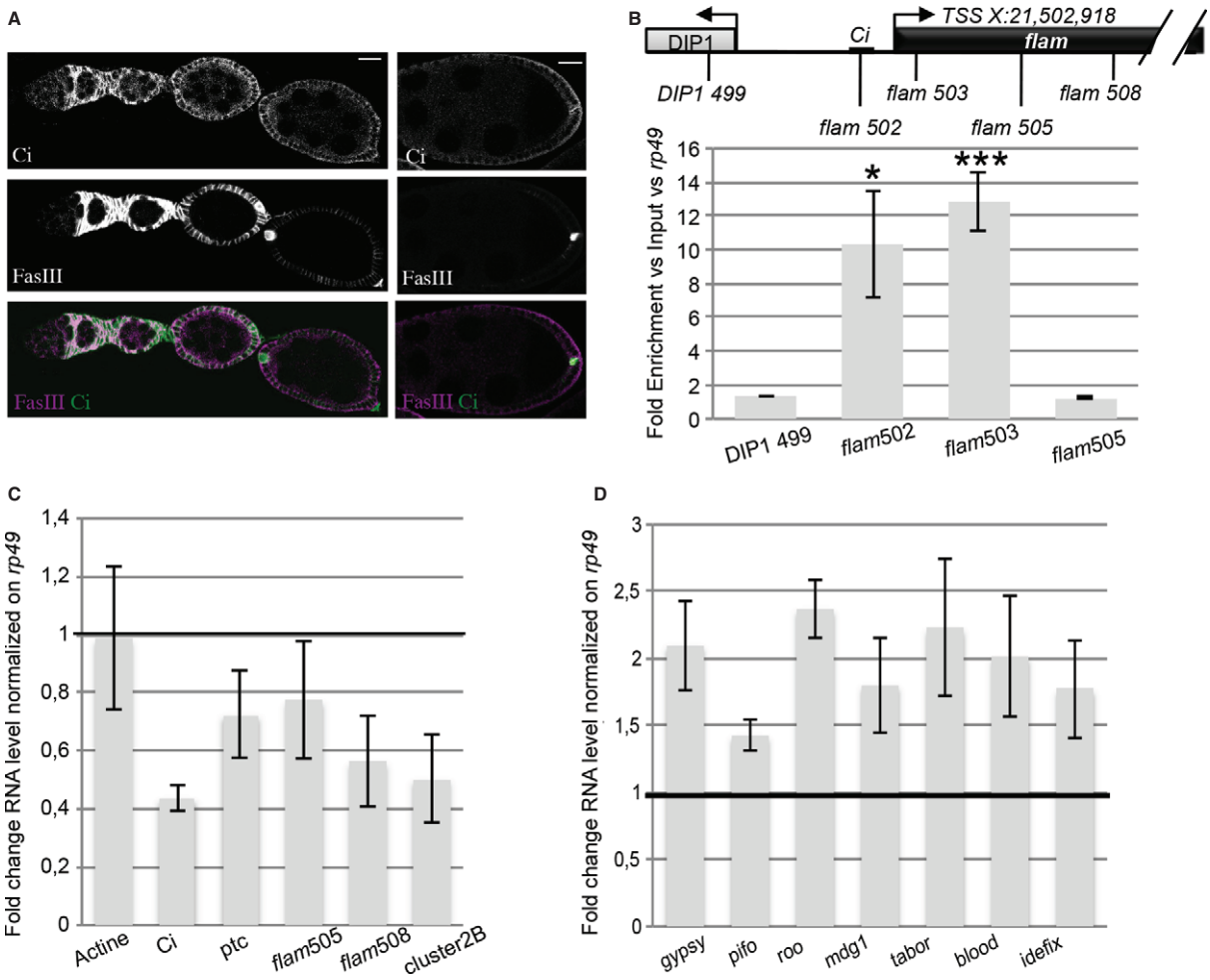

Figure 4. Ci activates *flam* transcription *in vivo*.

A Ci and FasIII (a follicle cell marker) immunostaining of a *Drosophila melanogaster* ovariole. Left panel shows early stages and right panel a stage 8. Scale bar, 20 μm.

B ChIP-qPCR analysis was carried out on the *flam* promoter with a specific Ci antibody. The positions of primers used for PCR amplification are indicated above. Enrichments were calculated versus Input and *rp49*. Data are presented as means (n = 5). Error bars represent ± s.e.m., *P < 0.05, ***P < 0.001.

C RNA level in ovarian somatic stem (OSS) cells transfected with siRNAs against Ci as compared to OSS cells transfected with siRNAs against GFP. The positions of primers used for PCR amplification are indicated above. *ptc* is a gene known to be activated by Ci in follicle cells. Data are presented as means (n = 4). Error bars represent ± s.e.m.

D Fold change in RNA level of diverse somatic transposable elements in OSS cells transfected with siRNAs against Ci and compared to OSS cells transfected with siRNAs against GFP, 4 days post-transfection. Data are normalized to *rp49* expression and are presented as means (n = 4). Error bars represent ± s.e.m.

16% of reads mapped the first exon–exon and intron–exon junction, respectively (Fig 5C). Then, we extended this analysis to 21 major piRNA clusters expressed in ovaries and found that seven of them contain introns (Supplementary Fig S5). These data suggest that several piRNA clusters including *flam* are transcribed as a long primary multi-kilobase RNA transcript before being spliced.

To determine whether these spliced RNAs are processed into piRNAs, we sequenced small RNAs from OSS cells and searched for reads that align uniquely to the identified *flam* spliced junctions. Reads spanning exon junctions were identified. Furthermore, we found that piRNAs encompassing the exon 1/intron 1 junction are under-represented compared to piRNAs matching the splice junction (Fig 5C). These results further indicate that *flam* transcripts are processed into piRNAs after the precursor is spliced. Although the diversity of alternatively spliced transcripts of *flam* is likely underestimated, it can be predicted that the multiple splicing events contribute to create a high diversity of *flam* precursors.

In $flam^{KG}$ mutant, the *KG* transgene is localized at position 21,505,285 downstream of the TSS, at the beginning of intron 2. Nevertheless, homozygote $flam^{KG}$ mutant females exhibit atrophic ovaries like $flam^{BG}$ females [24]. This ovarian phenotype has been attributed to an absence of *flam* transcription. If the reason why $flam^{BG}$ transcription is affected can be explained by disruption of the Ci binding site, the reason why *flam* transcription is also affected in the $flam^{KG}$ mutant remains obscure. It can be proposed that either the correct transcription of *flam* or the stability of its transcripts is affected. We have shown that the *KG* transgene is

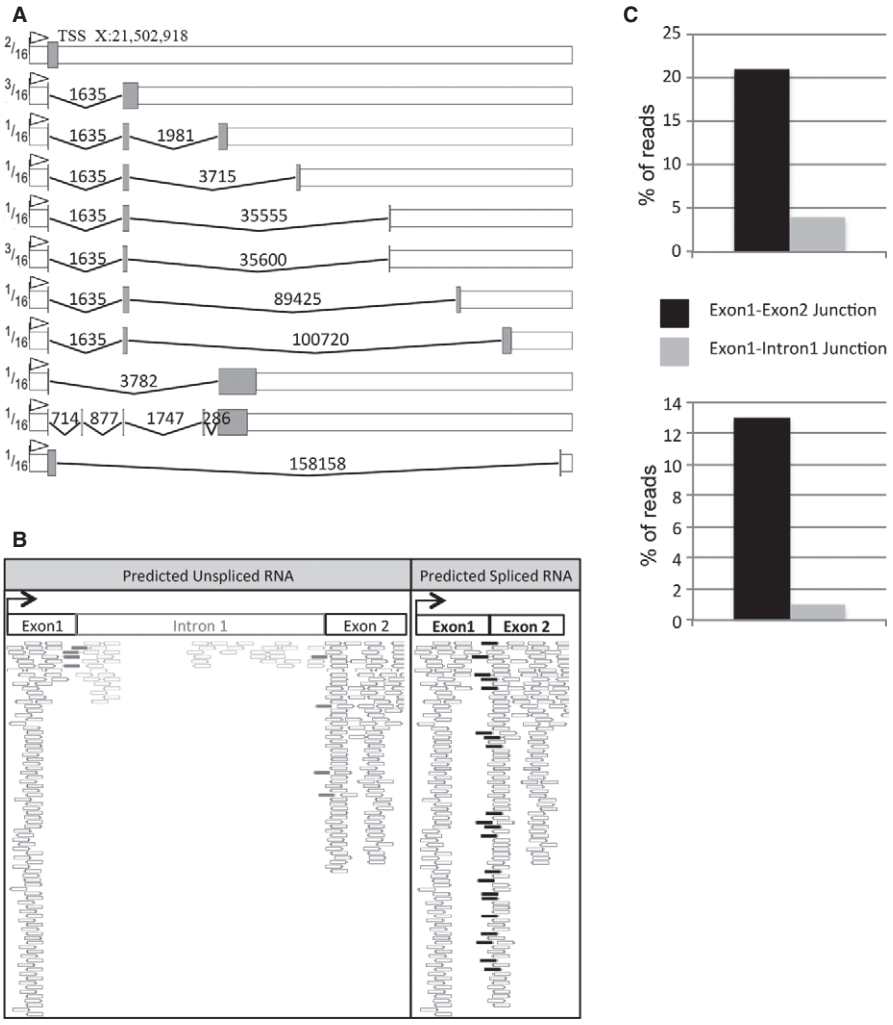

Figure 5. *flam* transcripts are alternatively spliced before piRNA processing.

A Representation of alternatively spliced RNAs identified by RT-PCR experiments. Grey boxes represent exons and peaked lines introns, numbers above introns indicate the length. The ratio of individual alternative transcripts to the total transcripts is indicated on the left part of the figure.

B Mapping of reads from Sienski *et al* (2012) sequencing data on exon–1–intron–1–exon–2 or on exon–1–exon–2 predicted *flam* transcripts. White reads mapped in exons, grey reads between exon 1 or 2 and intron 1, and black reads between the 2 exons.

C Percentage of reads (RNAs or piRNAs) corresponding either to the predicted non-spliced exon 1/intron 1 junction or to the predicted spliced exon 1/exon 2 junction in ovarian somatic stem (OSS) cells.

located at the border of the second intron. Disruption of this site might prevent its recognition as a donor site. Since almost all the spliced transcripts detected in WT *flam* alleles contain this spliced border, it might then be anticipated that this donor site plays a crucial role in generating the pool of alternative spliced RNAs. *flam* mutation due to *KG* insertion would then lead to unstable *flam* transcripts and thus, as for the *BG* insertion, to a phenotype of atrophic ovaries.

Overall, *flam* precursors display two characteristics: first, they display distinct structures resulting from alternative splicing, and second, they all share the first exon at their 5′ end. Future work is needed to elucidate the function of this common 5′ end. A likely hypothesis is that it helps to transfer RNA precursors from their site of transcription to Dot COM at the nuclear membrane facing the cytoplasmic Yb bodies, where they are processed to piRNAs. Recently, UAP56, a helicase of the exon junction complex (EJC), has been shown to play a role in the transport of germline precursor piRNA transcripts to the nuclear pore [25]. It remains to be clarified whether the recruitment of the EJC necessary for *flam* splicing also plays a role in the stabilization, surveillance and transport of the *flam* precursors.

Many TE families are known to originate from recent horizontal transfer between *Drosophila* species [26]. Recently, we have reported that many of these new TEs preferentially insert within heterochromatic regions such as the *flam* locus [27]. Thus, the dynamic nature of this piRNA cluster suggests that novel motifs for splicing are constantly gained or lost resulting in distinct pools of

flam precursors. Such stochastic splicing depending on structural modifications affecting piRNA loci might help genomes to rapidly react against new TE invasions.

Materials and Methods

Drosophila strains

ChIP experiments were performed on the W^{1118} line. Clonal analyses were performed on flies with the following genotype: *y-hs-flp; FRT42D P[Ci+]/FRT42D hs-Myc;; ci94/ci94*. Flies were heat-shocked three times in 12 h and then dissected 7 days later.

RNA extraction and RT-qPCR analysis

Total RNAs from 15 ovaries or OSS cells were extracted with Trizol. After DNase treatment, cDNA was synthesized from 1 μg RNA using random primers and SuperScript III Reverse Transcriptase. qPCR was performed to assay levels of *flam*. *18S* or *rp49* RNA was used for the normalization. Fold changes were calculated using the delta Ct method [28]. Primers are listed in Supplementary Table S5.

RNA analyses

Small RNAs from OSS cells were extracted by Trizol. Deep sequencing was performed by Fasteris S.A. (Geneva/CH) on an Illumina Hi-Seq 2000 (Fasteris). RNA-seq libraries were analysed with *bowtie* mappers and were visualized using http://genomeview.org/. Small RNA sequencing data were analysed with NucBase [29]. For researches of mRNA or piRNA across exon junctions, reads were mapped on the reconstituted junction using *bowtie* [30] for mRNA and NucBase for small RNA.

Supplementary information for this article is available online: http://embor.embopress.org

Acknowledgements

The OSS cell line was a gift of Yuzo Niki (Ibaraki University). Flies with the following genotypes *y-hs-FLP; FRT42D ci[+];;ci[94]* and *y; FRT42D hs-MYC 45/CyO;;ci[94]/y[+]* were kindly provided by Robert Holmgren. We are grateful to Françoise Pellissier and Agostinha De Sousa for technical assistance. CG received a graduate grant from the Ministere de lEnseignement Superieur et de la Recherche (MESR). This work was supported by grants from the Region Auvergne, European Union (FEDER), the Ligue régionale contre le cancer and the Association Nationale de la Recherche (ANR) (project plasTiSiPi). We thank all members of our group for helpful discussion.

Author contributions

EB and CV conceived and designed the experiments. CG and SD performed most of the experiments. YR and CG analysed bioinformatic data. EB and CV wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* 128: 1089 – 1103
2. Prudhomme N, Gans M, Masson M, Terzian C, Bucheton A (1995) Flamenco, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*. *Genetics* 139: 697 – 711
3. Desset S, Meignin C, Dastugue B, Vaury C (2003) COM, a heterochromatic locus governing the control of independent endogenous retroviruses from *Drosophila melanogaster*. *Genetics* 164: 501 – 509
4. Sarot E, Payen-Groschêne G, Bucheton A, Pelisson A (2004) Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the *Drosophila melanogaster flamenco* gene. *Genetics* 166: 1313 – 1321
5. Malone CD, Brennecke J, Dus M, Stark A, Mccombie WR, Sachidanandam R, Hannon GJ (2009) Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell* 137: 522 – 535
6. Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, Siomi H, Siomi MC (2010) Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. *Genes Dev* 24: 2493 – 2498
7. Dennis C, Zanni V, Brasset E, Zhang L, Eymery A, Jensen S, Rong Y, Vaury C (2013) Dot COM, a nuclear transit center for the primary piRNA pathway in Drosophila. *PLoS ONE* 8: e72752
8. Klattenhoff C, Xi H, Li C, Lee S, Xu J, Khurana JS, Zhang F, Schultz N, Koppetsch BS, Nowosielska A *et al* (2009) The Drosophila HP1 homolog rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* 138: 1137 – 1149
9. Pane A, Jiang P, Zhao DY, Singh M, Schüpbach T (2011) The Cutoff protein regulates piRNA cluster expression and piRNA production in the Drosophila germline. *EMBO J* 30: 4601 – 4615
10. Li XZ, Roy CK, Dong X, Bolcun-Filas E, Wang J, Han BW, Xu J, Moore MJ, Schimenti JC, Weng Z *et al* (2013) An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Mol Cell* 50: 67 – 81
11. Lo K, Smale ST (1996) Generality of a functional initiator consensus sequence. *Gene* 182: 13 – 22
12. Ohler U, Liao G-C, Niemann H, Rubin GM (2002) Computational analysis of core promoters in the *Drosophila* genome. *Genome Biol* 3: research0087-0087.12
13. Arkhipova IR (1995) Promoter elements in *Drosophila melanogaster* revealed by sequence analysis. *Genetics* 139: 1359 – 1369.
14. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CAM, Taylor MS, Engström PG, Frith MC *et al* (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* 38: 626 – 635
15. Bensaude O (2011) Inhibiting eukaryotic transcription. Which compound to choose? How to evaluate its activity?. *Transcription* 2: 103 – 108.
16. Yee NS, Zhou W, Chun SG, Liang IC, Yee RK (2012) Targeting developmental regulators of zebrafish exocrine pancreas as a therapeutic approach in human pancreatic cancer. *Biol Open* 1: 295 – 307
17. Forbes AJ, Spradling AC, Ingham PW, Lin H (1996) The role of segment polarity genes during early oogenesis in Drosophila. *Development* 122: 3283 – 3294
18. Méthot N, Basler K (1999) Hedgehog controls limb development by regulating the activities of distinct transcriptional activator and repressor forms of Cubitus interruptus. *Cell* 96: 819 – 831

19. Bellen HJ (2004) The BDGP gene disruption project: single transposon insertions associated With 40% of Drosophila genes. *Genetics* 167: 761 – 781
20. Russo CA, Takezaki N, Nei M (1995) Molecular phylogeny and divergence times of drosophilid species. *Mol Biol Evol* 12: 391 – 404
21. Tamura K (2004) Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. *Mol Biol Evol* 21: 36 – 44
22. Qi H, Watanabe T, Ku H-Y, Liu N, Zhong M, Lin H (2011) The Yb body, a major site for PIWI-associated RNA biogenesis and a gateway for PIWI expression and transport to the nucleus in somatic cells. *J Biol Chem* 286: 3789 – 3797
23. Sienski G, Dönertas D, Brennecke J (2012) Transcriptional silencing of transposons by PIWI and maelstrom and its impact on chromatin state and gene expression. *Cell* 151: 964 – 980
24. Mével-Ninio M, Pelisson A, Kinder J, Campos AR, Bucheton A (2007) The flamenco locus controls the gypsy and ZAM retroviruses and is required for Drosophila oogenesis. *Genetics* 175: 1615 – 1624
25. Zhang F, Wang J, Xu J, Zhang Z, Koppetsch BS, Schultz N, Vreven T, Meignin C, Davis I, Zamore PD *et al* (2012) UAP56 Couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell* 151: 871 – 884
26. Bartolomé C, Bello X, Maside X (2009) Widespread evidence for horizontal transfer of transposable elements across Drosophila genomes. *Genome Biol* 10: R22
27. Zanni V, Eymery A, Coiffet M, Zytnicki M, Luyten I, Quesneville H, Vaury C, Jensen S (2013) Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. *Proc Natl Acad Sci* 110: 19842 – 19847.
28. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25: 402 – 408
29. Dufourt J, Pouchin P, Peyret P, Brasset E, Vaury C (2013) NucBase, an easy to use read mapper for small RNAs. *Mob DNA* 4: 1
30. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Meth* 9: 357 – 359

