

Manuscript EMBOR-2013-37898

## Transcriptional properties and splicing of the flamenco piRNA cluster

Coline Goriaux, Sophie Desset, Yoan Renaud, Chantal Vaury and Emilie Brasset

*Corresponding author: Emilie Brasset, GReD, CNRS, UMR 6293, INSERM U1103*

---

### Review timeline:

Submission date:	19 August 2013
Editorial Decision:	21 August 2013
Correspondence:	26 August 2013
Correspondence:	27 August 2013
Editorial Decision:	24 September 2013
Revision received:	20 December 2013
Editorial Decision:	21 January 2014
Revision received:	29 January 2014
Accepted:	31 January 2014

---

### Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

Editor: Esther Schnapp

1st Editorial Decision

21 August 2013

Thank you for the submission of your manuscript to EMBO reports. I have now had the opportunity to read and to discuss it with my colleagues here. I regret to say that we all agree that the manuscript is not suitable for our journal.

We note that your study reports that the transcription factor Cubitus interruptus (Ci) promotes expression of the flamenco (flam) piRNA cluster in *Drosophila*. You identify the TSS of flam and show that it contains the consensus initiator element (Inr) sequence that is required for the expression of a luciferase reporter driven by the TSS surrounding sequence. You show that Pol II transcribes endogenous flam, and you identify binding sites for several transcription factors around the flam TSS. Ci binds to the flam regulatory sequence and Ci depletion reduces flam transcription. You finally show that flam is alternatively spliced, but that all identified transcripts contain the first exon.

We appreciate that your findings indicate that Ci regulates flam transcription and that flam transcripts are alternatively spliced. We think that this will certainly be of interest to researchers in the field. However, we receive many more manuscripts than we can publish, and thus have the difficult task of selecting those we think most likely to compete successfully with other submissions

during the in-depth review process. In this case, we note that other transcription factors have been shown before to be required for the transcription of piRNA precursors in mice and *Drosophila*, as you mention. While we recognize that you show that Ci also promotes piRNA transcription, we think that this does not represent a sufficient conceptual advance over previous findings for consideration of the manuscript for publication here. We have therefore decided not to proceed with in-depth review.

In the interest of your manuscript and your time, I am providing you with an editorial decision on your manuscript that will allow you to submit it elsewhere without further delay. I am sorry to disappoint you on this occasion and I thank you once more for your interest in our journal.

---

Correspondence – author appeal

26 August 2013

We understand and accept your point of view.

As you point out, other transcription factors have been shown before to be required for the transcription of piRNA precursors in mice. However as far as I know no transcription factors that could regulate directly the transcription of piRNA cluster have been described in fly. Furthermore our data represent the first demonstration that the presume long piRNA precursors are alternatively spliced and are the precursor of mature piRNAs in fly illustrating then the dynamic nature of piRNA clusters resulting in distinct pools of piRNA precursors.

I know that I am not recognized in the field, it is effectively my first paper in corresponding author but I think our work is of interest to researchers in the field.

I know that you received many more manuscripts than you can publish but I was wondering if we could have the opportunity that you reconsider our work for in-depth review.

---

Correspondence – editor

27 August 2013

Thank you for your email. I have discussed your manuscript with a member of our Advisory Editorial Board now, and we have decided to send it for peer-review.

---

2nd Editorial Decision

24 September 2013

Thank you for your patience while your manuscript was under review. We have now received the full set of referee reports that is pasted below.

As you will see, the referees acknowledge that the findings are potentially interesting. However, they also all point out that the physiological relevance of Ci-controlled flamenco expression and piRNA production remains unclear. This is a substantial concern raised by all referees, which would need to be successfully addressed for consideration of the manuscript for publication here. It would need to be examined whether loss of Ci phenocopies loss of flamenco, whether piRNA production is affected and flamenco-controlled transposons are activated in the absence of Ci, and whether Ci promotes the expression of other piRNA precursors. Referees 1 and 2 further remark that the specificity of the Ci antibody needs to be demonstrated, that the abundance of splicing of flamenco transcripts should be investigated, and that it should be examined whether other piRNA cluster promoters contain similar, conserved regulatory elements. While referee 1 indicates in her/his cross-comments that it would not be essential to determine whether all flamenco transcripts are produced by RNA Pol II, s/he remarks that the control primers used and their location within RP49 need to be described.

Given these constructive comments, we would like to invite you to revise your manuscript with the understanding that the referee concerns (as mentioned above and in their reports) must be fully addressed and their suggestions taken on board. We can only consider a revised manuscript if it provides convincing evidence for a physiologically relevant role for Ci in piRNA production and transposon repression/germline function. I understand that addressing this essential point will require a lot of work, and please let me know if you think that it cannot be performed within three months, so that we can potentially extend the time for revision. If you decide to rather publish your manuscript more rapidly elsewhere, I would welcome a message to that effect.

Acceptance of the manuscript will depend on a positive outcome of a second round of review. It is EMBO reports policy to allow a single round of revision only and acceptance or rejection of the manuscript will therefore depend on the completeness of your responses included in the next, final version of the manuscript.

Revised manuscripts should be submitted within three months of a request for revision; they will otherwise be treated as new submissions. Also, the revised manuscript may not exceed 30,000 characters (including spaces and references) and 5 main plus 5 supplementary figures, which should directly relate to their corresponding main figure.

Regarding data quantification, can you please specify the number "n" for how many experiments were performed, the bars and error bars (e.g. SEM, SD) and the test used to calculate p-values in the respective figure legends? This information is currently incomplete and needs to be provided in the figure legends.

We recently decided to offer the authors the possibility to submit "source data" with their revised manuscript that will be published in a separate source data file online along with the accepted manuscript. If you would like to use this opportunity, please submit the source data (for example entire gels or blots, data points of graphs, additional images, etc.) of your key experiments together with the revised manuscript. Please include size markers for scans of entire gels, label the scans with figure and panel number, and send one file per figure.

As part of the EMBO publication's Transparent Editorial Process, EMBO reports publishes online a Review Process File to accompany accepted manuscripts. This File will be published in conjunction with your paper and will include the referee reports, your point-by-point response and all pertinent correspondence relating to the manuscript.

You are able to opt out of this by letting the editorial office know ([emboreports@embo.org](mailto:emboreports@embo.org)). If you do opt out, the Review Process File link will point to the following statement: "No Review Process File is available with this article, as the authors have chosen not to make the review process public in this case."

I look forward to seeing a revised version of your manuscript when it is ready. Please let me know if you have any further questions or comments regarding the revision.

#### REFeree REPORTS:

Referee #1:

In the manuscript titled 'the flamenco piRNA cluster is alternatively spliced before being processed into piRNAs' Goriaux and colleagues characterize the site and partial genetic requirements of transcription initiation of the flamenco piRNA cluster. In addition they show that transcripts initiated from this site are often spliced prior to processing into small RNAs for the piRNA pathway. Although certainly of interest to those in the piRNA and transposon regulation field, I do not feel that the current state of this manuscript supports publication in EMBO reports. However, if the informatics analysis was extended to describe a broader representation of piRNA cluster transcription initiation and splicing, and other major issues were addressed, a revised manuscript could be considered for publication.

## Major Issues:

Figures 1 and 2 identify flamenco transcripts using mRNA cap-based 5' RACE cloning, and then verify the role of RNA pol II using the specific inhibitor, alpha-amanitin. One issue here is that the only experiments performed prove the presence of RNA pol II and its transcripts. It would be helpful to use a RNA pol I, RNA pol III or general RNA polymerase inhibitor to determine if all flamenco transcripts are products of RNA pol II, or only a subset. Also, your alpha-amanitin experiments were normalized to RP49. Where are these primers in relation to the RP49 transcription start site, and since you are assessing transcription across several positions on the flamenco, does the RP49 signal change depending on primer position during treatment?

There are many issues with the CI portion of this manuscript. The staining of CI is clearly cytoplasmic yet is being reported to act as a transcription factor, which would presumably act in the nucleus. Mutant clones of CI are generated so it would be helpful to do a staining on these to show that CI is actually lost. Additionally, the somatic phenotype of a flamenco mutation is strongly affects somatic tissues of the ovary. Can somatic clones of flamenco mutant alleles be used to assess whether CI mutant clones phenocopy flamenco mutant clones? It is reported that the P-element insertion in the Flam BG line aligns to disrupt the CI binding site. This would also support loss of CI to phenocopy loss of flamenco in the ovarian soma. The Flam KG allele also disrupts flamenco, and is often used as a trans-heterozygote with Flam BG. Does this also disrupt the CI binding region and if not, what is the explanation for how that allele disrupts flamenco transcription? Finally, if CI is upstream of flamenco transcription, then are you able to show that any transposons controlled by flamenco are derepressed when CI function is disrupted?

The informatics analysis should include other clusters in *Drosophila*. It seems that there is ample small RNA and RNA-seq analysis to easily extend the analysis of splicing and exon-junction-spanning piRNAs to other clusters expressed in ovary and OSS cells (which includes cluster 2). At least determining how many other potential cluster promoters contain similar transcription factor and promoter sequences should be reasonable and would broaden the interest of this work.

## Minor Issues:

The title is a slight overstatement and is misleading given the content of the manuscripts. Here it is shown that some, but not all flamenco transcripts are spliced prior to processing into piRNAs. The current title implies that splicing is an essential step in processing of a transcript. Also, it also only references the final figure and fails to even hint at the conclusions drawn from the first four figures.

The second introductory paragraph references transposons being in the 'minus orientation' of transposons and then plus strand of flamenco transcripts. Since orientation is somewhat relative and this sentence reads slightly awkward, a simplification to something like 'the transposon fragments in flamenco are similarly oriented to produce antisense transcripts capable of silencing active transposon transcripts'. This was just a thought when reading through.

Later in this paragraph it states there is little known about the upstream steps in piRNA cluster transcription. I agree that the molecular details are still being worked out, but there are now several papers out of the Brennecke, Theurkauf, Aravin and Hannon labs that address this issue directly, as well as through screening which could identify chromatin and transcriptional components relating to this pathway. Please expand the introduction of what is know and include appropriate citations.

Figure 3A shows that the region between -515 and -356 is predominantly responsible for promoting transcription from the flamenco *inr*/DPE. It would be useful to control for the size of this construct driving luciferase by using a known non-promoting sequence to fill that space (as a negative control).

Malone et al 2009 Cell showed mapping of piRNAs upstream of DIP1 in a 'flamenco-like' cluster of *Drosophila erecta*. Can you place the transcription elements found in your analysis of *erecta* to the onset of piRNA production, and compare this to that found in *Drosophila melanogaster*?

## Referee #2:

In *Drosophila* ovarian somatic cells, most of the piRNAs are produced from a long single-stranded non-coding RNA that is transcribed from the piRNA cluster flamenco (flam). In this study, the authors demonstrated that flam transcription is initiated from an RNA Pol II promoter SF1 containing Inr and DPE, and requires the transcriptional factor Cubitus interruptus (Ci). Inr and Ci elements are highly conserved across various *Drosophila* species. The authors also showed that the flam primary transcripts undergo alternative splicing to generate diverse RNA precursors.

I found this study interesting and the data are mostly clear and convincing. Once the reviewer's concerns are addressed properly, I would recommend this manuscript be considered for publication in EMBO Reports.

- 1) The authors must examine how phenotypes of the Ci mutant are similar to those of flam mutant.
- 2) The authors must examine if the production of genic piRNAs is affected by loss of Ci function.
- 3) Can protein-coding RNAs and non-coding RNAs other than flam transcripts driven by Ci produce piRNAs in vivo?
- 4) (21,502,918..21,203,349) (page 8) should be (21,502,918..21,503,349).
- 5) Show the specificity of anti-Ci antibody (Fig. 4A).
- 6) The data in Fig. 4B shows that Ci binds flam503 as well as flam502. Why does Ci bind also to flam503?
- 7) The detailed information of the exons and the introns are required in Fig. 5A.
- 8) What is the ratio of individual alternative transcripts to the unspliced transcript? How abundant is the unspliced transcript in cells?

Referee #3:

Goriaux et al report a detailed characterization of the structure and regulation of the transcriptional unit that generates piRNA precursors in the flamenco cluster. The authors identify RNA pol II-driven transcription start sites, functional promoter elements (including various lines of evidence for the involvement of the Cubitus interruptus (Ci) transcription factor in transcriptional activation of the locus) and a variety of transcript isoforms generated by alternative splicing that diversify the piRNA precursors.

Understanding the generation and regulation of piRNAs is a topic of great relevance to explain the important functions of these RNAs in silencing foreign DNA elements and maintaining genomic integrity in the germline. The results of Goriaux et al undoubtedly contribute towards this goal. The less satisfying aspect of the manuscript is the rather descriptive nature of the results reported. It should be an editorial decision whether the information provided in the manuscript covers sufficient ground as to be published in EMBO Reports.

Specific points:

1. First, it is not entirely surprising that flamenco piRNA precursors are capped, pol II transcripts generated from a distal promoter (Sienski et al. (2012) Cell, 151, 964-980. Brennecke et al. (2007) Cell, 128, 1089-1103. Robert et al. (2001) Genetics, 158, 701-713.).
2. It is of interest to implicate Ci in flamenco transcription and document Ci expression in follicle cells, but the physiological role of this transcription factor remains to be firmly established: e.g. does the decrease in flamenco transcripts in the Ci mutant described in Figure 4C result in lower levels of the corresponding piRNAs and does this decrease lead to observable phenotypes on transposon mobilization / germline function? It would also be interesting to know the extent to which Ci contributes to the expression of other piRNA precursors.
3. While the generation of piRNAs from spliced transcripts containing multiple isoforms was reported before (e.g. Gan et al. (2011) RNA 17: 1191-1203), the possibility that alternative patterns of splicing can diversify the repertoire of piRNAs is of significant interest. However, the physiological role of this diversity, e.g. through the impact of regulation of their production, remain to be established.

In the revised version we have carefully considered the comments made by the reviewers (see the point by point response below). We made several changes in the text and figures and included additional experiments as requested by referees. We hope that this improved manuscript is now acceptable for publication in *EMBO reports*.

#### Referee #1:

In the manuscript titled 'the flamenco piRNA cluster is alternatively spliced before being processed into piRNAs' Goriaux and colleagues characterize the site and partial genetic requirements of transcription initiation of the flamenco piRNA cluster. In addition they show that transcripts initiated from this site are often spliced prior to processing into small RNAs for the piRNA pathway. Although certainly of interest to those in the piRNA and transposon regulation field, I do not feel that the current state of this manuscript supports publication in *EMBO reports*. However, if the informatics analysis was extended to describe a broader representation of piRNA cluster transcription initiation and splicing, and other major issues were addressed, a revised manuscript could be considered for publication.

#### Major Issues:

##### Comment 1

Figures 1 and 2 identify flamenco transcripts using mRNA cap-based 5' RACE cloning, and then verify the role of RNA pol II using the specific inhibitor, alpha-amanitin.

One issue here is that the only experiments performed prove the presence of RNA pol II and its transcripts. It would be helpful to use a RNA pol I, RNA pol III or general RNA polymerase inhibitor to determine if all flamenco transcripts are products of RNA pol II, or only a subset.

We performed the experiment requested by the reviewer (supplementary Fig S1): we quantified the *flamenco* transcripts in the presence of :

- RNA pol III inhibitor to inhibit Pol III transcription: ML-60218 (CAS number 577784-91-9, Tebu-bio) (Yee *et al.*, 2012)
- low concentrations of actinomycin D (A1410, Sigma), which primarily inhibits RNA pol I (Bensaude, 2011)

We show that the level of *flamenco* transcript is not affected by both types of inhibitors. These results are included in a revised supplementary Fig S1 and mentioned in the text.

##### Comment 2

Also, your alpha-amanitin experiments were normalized to RP49. Where are these primers in relation to the RP49 transcription start site, and since you are assessing transcription across several positions on the flamenco, does the RP49 signal change depending on primer position during treatment?

The alpha-amanitin experiments were normalized against the 18S housekeeping gene (known to be transcribed by Pol I) and not against the *rp49* housekeeping gene (known to be transcribed by Pol II). However *rp49* were used as a positive control of the experiment. To clarify this point we added this sentence in the text « 18S ribosomal RNA known to be transcribed by Pol I was used as reference gene for normalization ».

The amount of *rp49* RNA is shown as a positive control of the experiment. As suggested by

the reviewer we added RT-qPCR of *rp49* transcript level using two primer pairs located at different positions on the gene: in the exon 3 and in the exon 1, 400bp downstream of the TSS. The experiment showed that *rp49* RNA levels are reduced more than two fold upon alpha amanitin treatment and it is not depending on the primer position. These results are included in the revised Figure 2A.

### Comment 3

There are many issues with the CI portion of this manuscript. The staining of CI is clearly cytoplasmic yet is being reported to act as a transcription factor, which would presumably act in the nucleus.

It was often reported that although Ci is a zinc-finger protein and a transcription factor, it is present at high levels in the cytoplasm and the staining appears to be excluded from the nucleus in the somatic follicular cells, as it is in embryonic and disk cells (Forbes AJ, *et al*, Development. 1996). None of the known anti-Ci antibodies shows nuclear staining in embryos or disks (Motzney and Holmgren, 1995).

### Comment 4

Mutant clones of CI are generated so it would be helpful to do a staining on these to show that CI is actually lost.

In the supplementary Fig S2, we replaced the immunostaining of Myc that showed Ci mutant clones by new immunostaining experiments using both an anti-Myc and an anti-Ci antibodies. We show that Ci expression is lost in the mutant clones (Homozygote Myc cells).

### Comment 5

Additionally, the somatic phenotype of a flamenco mutation is strongly affects somatic tissues of the ovary. Can somatic clones of flamenco mutant alleles be used to assess whether CI mutant clones phenocopy flamenco mutant clones?

We agree with reviewer 1 that this comparison would be informative. However it is not possible to perform.

Homozygote *flam* mutants are female sterile and display atrophic ovaries.

Since Ci is involved in controlling the transcription of Drosophila Hh target genes and thus represents the most downstream component known in this pathway (Méthot N, Basler K, development 2001), its loss of function (*ci*<sup>94</sup>) is lethal and mutants die before the end of the pupal stage. This phenotype does not allow comparison between *flam* and *ci* mutant ovaries.

We were not able to compare *flam* and *ci* mutant clones either. Indeed, mutant clones of *flam* are currently impossible to perform due to the pericentromeric position of the locus. It is localized to the 20A region on the X chromosome whereas the closest FRT site to the centromere (from the Kyoto FRT strains) is localized at position 19A. To date all our attempts to introduce an FRT sequence between *flam* and the centromere have been unsuccessful.

### Comment 6

It is reported that the P-element insertion in the Flam BG line aligns to disrupt the CI binding site.

This would also support loss of CI to phenocopy loss of flamenco in the ovarian soma. The Flam KG allele also disrupts flamenco, and is often used as a trans-heterozygote with Flam BG. Does this also disrupt the CI binding region and if not, what is the explanation for how that allele disrupts flamenco transcription?

In the *flam*<sup>BG</sup> mutant, the BG transgene is localized at position 21.502.538 and disrupts the Ci binding site. In *flam*<sup>KG</sup> mutant, the KG transgene does not disrupt the Ci binding site. It is localized at position 21.505.285 downstream of the TSS, at the beginning of intron 2. Nevertheless, both homozygote *flam*<sup>BG</sup> and *flam*<sup>KG</sup> mutant females exhibit atrophic ovaries,

*flam*<sup>BG</sup> females showing the most severe phenotype (Mevel-Ninio et al, 2007).

Through RT-qPCR experiments we have shown that in both mutants, *flam* transcription is strongly affected. If the reason why *flam*<sup>BG</sup> transcription is affected can be explained by the disruption of the Ci binding site, the reason why *flam* transcription is also affected in the *flam*<sup>KG</sup> 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 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 our analysis contain this spliced border, it can be anticipated that this donor site plays a crucial role in generating the pool of alternative spliced RNAs. *flam* mutation due to KG insertion might then lead to unstable *flam* transcripts and thus, as for the BG insertion, to a phenotype of atrophic ovaries.

Further work will be necessary to better characterize the link between the structure and the phenotype of this *flam* allele.

Comment 7:

Finally, if Ci is upstream of flamenco transcription, then are you able to show that any transposons controlled by flamenco are derepressed when Ci function is disrupted?

To address this point, we transfected OSS cells with siRNAs against *Ci* to knock down its expression and analyze the impact on *flam* and TE transcripts, and on small RNA populations. siRNA-mediated knockdown of *Ci* led to a decrease of *flam* transcripts two days post-transfection (Figure 4D). In contrast, the production of piRNAs and the TE mRNA levels were not significantly affected. However, a clear decrease in the silencing exerted on TEs known to be piRNA targets (*Tabor*, *Pifo*, *Blood*, *Gypsy*, *Idefix*, *roo*, *mdg1*) was observed four days post-transfection (supplementary Figure S3). These data suggest that a delay exists between *flam* knock down and the deregulation of TE expression. Although it is too early to fully interpret these results, the abundance and a possible stability of piRNAs produced by *flam* might explain this delay. These results are included in the revised Figure 4, in the supplementary Figure S3 and discussed in the manuscript.

Comment 8:

The informatics analysis should include other clusters in Drosophila. It seems that there is ample small RNA and RNA-seq analysis to easily extend the analysis of splicing and exon-junction-spanning piRNAs to other clusters expressed in ovary and OSS cells (which includes cluster 2).

We have now extended the bioinformatic analyses to other piRNA clusters expressed in ovary and OSS cells as suggested by the reviewer. We used available RNA seq data set (Sienski et al. 2012). Our analysis includes 21 major piRNA clusters identified in Brennecke et al. 2007 that are the source of up to 70% of the wild-type ovarian piRNA pool. From these available data, reads containing introns were identified in 7 clusters. We did not detect introns in reads reported for the somatic piRNA cluster 2. We have introduced a new supplementary Figure S5 to present these data and we discuss these results in the text.

Comment 9:

At least determining how many other potential cluster promoters contain similar transcription factor and promoter sequences should be reasonable and would broaden the interest of this work.

This information is currently difficult to extract because nothing is known on the 5' and 3' ends of piRNA clusters. Indeed, 5' and 3' RACE PCR aiming at determining the exact positioning of their start and end have not been performed yet (or not reported). The current definition of piRNA cluster is only based on computational analysis of mature piRNA sequences.

Nevertheless, as suggested by the reviewer we carefully analyzed the piRNA distribution of cluster 2 (using a personal OSS cell small RNA data set). We found an Inr consensus sequence (21.390.615), 108 bp upstream of the first piRNA mapped to this cluster and a Ci



binding site 2.846 bp upstream of the *Inr* consensus sequence. These data are reported in supplementary Figure S4

#### Minor Issues:

The title is a slight overstatement and is misleading given the content of the manuscripts. Here it is shown that some, but not all flamenco transcripts are spliced prior to processing into piRNAs. The current title implies that splicing is an essential step in processing of a transcript. Also, it also only references the final figure and fails to even hint at the conclusions drawn from the first four figures.

A new title was proposed

Transcriptional properties of the *flamenco* piRNA cluster

The second introductory paragraph references transposons being in the 'minus orientation' of transposons and then plus strand of flamenco transcripts. Since orientation is somewhat relative and this sentence reads slightly awkward, a simplification to something like 'the transposon fragments in flamenco are similarly oriented to produce antisense transcripts capable of silencing active transposon transcripts'. This was just a thought when reading through.

We rephrased this sentence according to the referee's advice.

Later in this paragraph it states there is little known about the upstream steps in piRNA cluster transcription. I agree that the molecular details are still being worked out, but there are now several papers out of the Brennecke, Theurkauf, Aravin and Hannon labs that address this issue directly, as well as through screening which could identify chromatin and transcriptional components relating to this pathway. Please expand the introduction of what is known and include appropriate citations.

We have changed the text and introduced appropriate citations according to the reviewer's suggestions.

The text has been changed as follows:

"Promoters and transcription factors involved in piRNA cluster transcription are starting to be identified. In *D. melanogaster*, Rhino and Cutoff are required for transcription/processing of germinal bidirectional piRNA cluster. In mice the transcription factor MYB-related protein A has been shown 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"

Figure 3A shows that the region between -515 and -356 is predominantly responsible for promoting transcription from the flamenco *Inr*/DPE. It would be useful to control for the size of this construct driving luciferase by using a known non-promoting sequence to fill that space (as a negative control).

We performed this experiment.

A NC (Negative Control) construct corresponding to SFI in which the *flam* fragment comprised between -515 and -356 has been replaced by a 159 bp fragment taken in a known non-promoting sequence within the *gfp* gene, confirmed that the region located downstream of position X: 21,502,403 (-515bp) and upstream of position X: 21,502,562 (-356bp) contains critical *cis*-elements required for the transcriptional activation of the locus.

These results are included in a revised Figure 3A and mentioned in the text.

Malone et al 2009 Cell showed mapping of piRNAs upstream of DIP1 in a 'flamenco-like' cluster of *Drosophila erecta*. Can you place the transcription elements found in your analysis of *erecta* to the onset of piRNA production, and compare this to that found in

*Drosophila melanogaster*?

We performed an analysis of the piRNA distribution along the *flamenco* locus in *Drosophila melanogaster* and *Drosophila erecta*. We interrogated publicly available small-RNA-seq libraries (Malone et al, 2009). Uniquely mapping *D. erecta* piRNAs have been plotted over the putative *flamenco* cluster. We found that sequence motifs similar to the Inr, DPE and Ci binding sites found in the *flam* cluster of *D. melanogaster* were also detected in *D. erecta* 1 kb upstream of the first piRNAs.

These results are included in the supplementary Figure S4 and discussed in the text.

Referee #2:

In *Drosophila* ovarian somatic cells, most of the piRNAs are produced from a long single-stranded non-coding RNA that is transcribed from the piRNA cluster *flamenco* (*flam*). In this study, the authors demonstrated that *flam* transcription is initiated from an RNA Pol II promoter SF1 containing Inr and DPE, and requires the transcriptional factor *Cubitus interruptus* (Ci). Inr and Ci elements are highly conserved across various *Drosophila* species. The authors also showed that the *flam* primary transcripts undergo alternative splicing to generate diverse RNA precursors.

I found this study interesting and the data are mostly clear and convincing. Once the reviewer's concerns are addressed properly, I would recommend this manuscript be considered for publication in EMBO Reports.

- 1) The authors must examine how phenotypes of the Ci mutant are similar to those of *flam* mutant.

Please see comment 6 in response to referee 1.

- 2) The authors must examine if the production of genic piRNAs is affected by loss of Ci function.

Please see comment 7 addressed to referee 1

- 3) Can protein-coding RNAs and non-coding RNAs other than *flam* transcripts driven by Ci produce piRNAs in vivo?

We analyzed piRNAs distributed along the *patched* (*ptc*) protein-coding mRNA. Indeed, *ptc* expression is activated by Ci in several tissues including *Drosophila* ovaries.

We found that the *ptc* mRNAs are not processed in piRNAs. These results confirm that transcripts driven by Ci do not systematically generate piRNAs.

These results are included in the revised supplementary Fig S2 and mentioned in the text.

- 4) (21,502,918..21,203,349) (page 8) should be (21,502,918..21,503,349).

We thank reviewer for pointing these errors, this statement was corrected.

- 5) Show the specificity of anti-Ci antibody (Fig. 4A).

We added the western blot showing the specificity of the anti-Ci antibody in supplementary Fig S2. The two isoforms of Ci protein are detected in OSS cells : the repressive form of 75 Kda and the active form of 140 Kda.

And see comment 4 addressed to referee 1.

- 6) The data in Fig. 4B shows that Ci binds flam503 as well as flam502. Why does Ci bind also to flam503?

The exact genomic positions of flam 502 and flam 503 are 21.502.584 and 21.502.944 respectively. The distance between flam 502 and flam 503 is around 360 bp. Chromatin extracted from ovaries were subjected to sonication. After this treatment, DNA fragments displayed an average 500 bp length. Even if flam 502 is enriched in Ci, it is not surprising to find some enrichment at flam 503.

- 7) The detailed information of the exons and the introns are required in Fig. 5A.

We added the requested data with all the detailed informations concerning exons as Table S4. The detailed informations on introns are present in Table S3.

- 8) What is the ratio of individual alternative transcripts to the unspliced transcript?

To answer this question, we added the ratio of individual alternative transcripts to total transcripts found by 5' rase experiments. They are presented on the left part of the figure 5A.

How abundant is the unspliced transcript in cells?

We have analysed in details the exon1-exon2 junction and found that 84% of the reads mapp to exon1-exon2 junction and only 16% to exon1-intron1 junction (the unspliced version of the transcript)

We have changed the y axis of the histogramme presented figure 5C and present the data in % of reads instead of read counts. In the revised version of the manuscript, we have added the following sentence : «Furthermore, we found that 84% and 16 % of reads mapp the first exon-exon and intron-exon junction respectively”.

Referee #3:

Goriaux et al report a detailed characterization of the structure and regulation of the transcriptional unit that generates piRNA precursors in the flamenco cluster. The authors identify RNA pol II-driven transcription start sites, functional promoter elements (including various lines of evidence for the involvement of the Cubitus interruptus (Ci) transcription factor in transcriptional activation of the locus) and a variety of transcript isoforms generated by alternative splicing that diversify the piRNA precursors.

Understanding the generation and regulation of piRNAs is a topic of great relevance to explain the important functions of these RNAs in silencing foreign DNA elements and maintaining genomic integrity in the germline. The results of Goriaux et al undoubtedly contribute towards this goal. The less satisfying aspect of the manuscript is the rather descriptive nature of the results reported. It should be an editorial decision whether the information provided in the manuscript covers sufficient ground as to be published in EMBO Reports.

Specific points:

1. First, it is not entirely surprising that flamenco piRNA precursors are capped, pol II transcripts generated from a distal promoter (Sienski et al. (2012) Cell, 151, 964-980. Brennecke et al. (2007) Cell, 128, 1089-1103. Robert et al. (2001) Genetics, 158, 701-713.).

We agree with this comment. Chromatin immunoprecipitation of RNA polymerase II suggests that piRNA precursor transcripts are transcribed by RNA polymerase II in mouse. However, to date, piRNA clusters from *Drosophila* have only been defined computationally using mature piRNA sequences. 5' and 3' RACE PCR aiming at determining the exact positioning of their start and end had not been performed and reported. The location of piRNA precursor transcription start sites (TSSs) and promoters or whether they are spliced was also unknown. Here, our data bring for the first time the exact location of a piRNA TSS and its enhancer motif in *Drosophila*. Interestingly, in the revised version of the manuscript, we also report that these characteristics of *flam* promoter are not specific to *flam* but are also shared with another piRNA cluster, Cluster 2, expressed in the same lineage, the somatic follicle cells of the ovaries.

2. It is of interest to implicate Ci in flamenco transcription and document Ci expression in follicle cells, but the physiological role of this transcription factor remains to be firmly established: e.g. does the decrease in flamenco transcripts in the Ci mutant described in Figure 4C result in lower levels of the corresponding piRNAs and does this decrease lead to observable phenotypes on transposon mobilization / germline function?

It would also be interesting to know the extent to which Ci contributes to the expression of other piRNA precursors.

Please see comment 7 addressed to referee 1

3. While the generation of piRNAs from spliced transcripts containing multiple isoforms was reported before (e.g. Gan et al. (2011) RNA 17: 1191-1203), the possibility that alternative patterns of splicing can diversify the repertoire of piRNAs is of significant interest. However, the physiological role of this diversity, e.g. through the impact of regulation of their production, remain to be established.

At the end of our manuscript, we indeed suggest that alternative splicing can diversify the repertoire of piRNAs. We agree with referee 3 that this remains to be demonstrated. We recently reported the high diversity and dynamic of TEs at the *flam* locus (Zanni et al. PNAS, 2013). It is tempting to think that these constant structural modifications impact *flam* splicing and thus the piRNA populations. However, we completely agree that this scenario remains to be established. The analysis of piRNAs produced from ovaries of different strains displaying variations in their *flam* genomic structure should help to address this issue.

3rd Editorial Decision

21 January 2014

Thank you for your patience while your revised manuscript was re-reviewed. I am sorry for the delay in getting back to you, which is due to the recent Christmas break. We have now received the enclosed reports from all referees, and you will see that all of them support publication of the manuscript by EMBO reports. Referees 1 and 3 still have a few suggestions regarding the discussion and moving SF3C to the main manuscript file, which I would like you to address before we can proceed with the official acceptance of your manuscript.

In order to have a little more space for the additional discussions, you can move the method sections Plasmid construction, 5'-RACE and RT-PCR experiments, Cell culture and transfection and Chromatin immunoprecipitation assays to the supplementary information. It is also not a problem if the final manuscript has a little more than 30.000 characters.

Can you please define the error bars and "n" in the legend for Figure 3A, C, and add scale bars to Figure 4A and SF2B?

I also would like to suggest to change the title to:

Cubitus interruptus is required for the transcription of the flamenco piRNA cluster.

Or, alternatively:

Transcriptional properties and splicing of the flamenco piRNA cluster

Please let me know what you prefer.

I look forward to seeing a new revised version of your manuscript as soon as possible. Please let me know if you have any questions.

#### REFEREE REPORTS:

Referee #1:

The revised manuscript submitted by Goriaux and colleagues has dealt with all reviewer issues adequately. The added data analyzing additional clusters and across different species provides a broader understanding of piRNA cluster transcription and evolution in *Drosophila*, and makes the paper of more general interest. The manuscript text is a little short, and if the authors wish, it could be expanded with some additional explanation and discussion.

Although dealt with in the rebuttal, it would be beneficial to have some discussion of the Flamenco KG allele in light of this work, how it compares/contrasts with the BG allele, and mention (possibly in the discussion) the potential reasons for also having a phenotype.

I would recommend this manuscript for publication in EMBO reports with only limited revision of the text.

Referee #2:

I now suggest that the revised manuscript is considered for publication in EMBO Reports.

Referee #3:

While the manuscript remains mostly descriptive, I support publication in EMBO Reports because the results can be of significant interest for understanding the biosynthesis of piRNAs. The authors have discussed in the text some of the issues raised in my report and have carried out additional experiments which have strengthened the paper. My remaining comment concerns the functional impact of Ci knock down, which appears to show some delay regarding its effect on the expression on TEs known to be piRNA targets. I would bring the results now shown in Figure S3C to one of the main figures because they seem to be very relevant to document a physiological function of Ci on piRNA function, and I would not conclude that a "delay is required between disruption of flam transcription and TE regulation" but rather that a "delay is observed" and discuss possible reasons for this delay in the effects.

In the revised version we have carefully considered the comments made by you and the reviewers (see the point by point response below) :

- We moved SF3C to the main manuscript file (Figure 4D) and moved the Figure 4D to the supplementary Fig2.
- We added several points of discussion.
- As suggested we moved the method sections Plasmid construction, 5'-RACE and RT-PCR experiments, Cell culture and transfection and Chromatin immunoprecipitation assays to the supplementary information.
- We defined the error bars and « n » in the legend for Figure 3A,C. we added this sentence in the legend Figure 3A,C : « Data are presented as means (n=4). Error bars represent  $\pm$  S.D »
- We added scale bars to Figure 4A and SF2B and added in the legend « Scale bar, 20 $\mu$ m »
- As you have suggested we have changed the title. « Transcriptional properties and splicing of the flamenco piRNA cluster » is the one that we prefer.

Key words: piRNA clusters, transcription, transposable elements, somatic silencing, *Drosophila, flamenco*

We hope that this improved manuscript is now acceptable for publication in EMBO reports.

I am very pleased to accept your manuscript for publication in the next available issue of EMBO reports. Thank you for your contribution to our journal.

As part of the EMBO publication's Transparent Editorial Process, EMBO reports publishes online a Review Process File to accompany accepted manuscripts. As you are aware, this File will be published in conjunction with your paper and will include the referee reports, your point-by-point response and all pertinent correspondence relating to the manuscript.

If you do NOT want this File to be published, please inform the editorial office within 2 days, if you have not done so already, otherwise the File will be published by default [contact: emboreports@embo.org]. If you do opt out, the Review Process File link will point to the following statement: "No Review Process File is available with this article, as the authors have chosen not to make the review process public in this case."

Thank you again for your contribution to EMBO reports and congratulations on a successful publication. Please consider us again in the future for your most exciting work.