EMBO
reports

Manuscript EMBO-2015-41866

Regulating retrotransposons via alternative transcription start sites

Jenna Persson, Babett Steglich, Agata Smialowska, Mette Boyd, Jette Bornholdt, Robin Andersson, Catherine Schurra, Benoit Arcangioli, Albin Sandelin, Olaf Nielsen and Karl Ekwall

Corresponding author: Karl Ekwall, Karolinska Institutet

Review timeline:

Original submission:	12 February 2015
Editorial Decision:	16 March 2015
Transfer date:	03 December 2015
Editorial Decision:	04 January 2016
Revision received:	06 January 2016
Accepted:	13 January 2016

Editor: Esther Schnapp

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.)

1st Editorial Decision 16 March 2015

Thank you for submitting your manuscript for consideration by The EMBO Journal. It has now been seen by three referees whose comments are shown below. As you will see, while the referees all express interest in the work and topic in principle, they do not offer strong support for publication in The EMBO Journal.

I will not repeat all their individual points of criticism here, but it becomes clear that while referee #1 is more positive, referee #2 and #3 raise concerns about the nature of the switch in RNA formation and about the specific mechanism of action for Fft2/3. Consequently, they are both hesitant to support publication of your study in The EMBO Journal.

In light of the slightly diverse concerns raised we also conducted cross-referee commenting, and based on that ref #1 and ref #3 both expressed agreement with the points raised by ref#2. In addition, ref#3 added the following paragraph in response to the comments from ref#2 on a putative manuscript focusing on the Fft2/3 and leaving out the origin of the truncated RNA::

'I note that this referee (probably an RNA expert) considers the conclusions relating to Fft proteins to be of great interest. With stronger expertise on the chromatin side, I'd disagree and think the main novelty stems from a stress induced change in retroviral transcription rather than fundamental new insight into these remodelling proteins.'

Which in turn lead to the following comment from ref #2:

'The point by Ref #3 about evolutionary interest is not a strong feature of this current paper'

Given the overall rather negative opinions expressed in the referees' concerns on both the conclusiveness of the model proposed and the more general scope of the study, I am therefore afraid we are unable to offer further steps towards publication in The EMBO Journal.

****
REFEREE REPORTS

Referee #1:

In this interesting article, the authors demonstrate that Tf2 transcription initiation is regulated by the factors Fft2 and Fft3, which control nucleosome position and clustering. When these factors are lost, there is a change in nucleosome position, clustering is lost, and transcription now initiates at an alternative start site, further 5', resulting in a full-length and productive transcript. Overall, the experiments are convincing and are clearly presented. Minor comments are below.

1. Ty1 elements of S. cerevisiae are also controlled by alternative transcription initiation. See

A trans-dominant form of Gag restricts Ty1 retrotransposition and mediates copy number control. Saha A, Mitchell JA, Nishida Y, Hildreth JE, Ariberre JA, Gilbert WV, Garfinkel DJ. J Virol. 2015 Jan 21. pii: JVI.03060-14. [Epub ahead of print]

and earlier studies form this lab. Although a different type of change, this work should be cited as a relevant comparison.

2. page 2 - This sentence: "Host cells have developed numerous silencing mechanisms to repress these elements, including DNA methylation, RNA interference (RNAi), repressive histone modifications and ATP-dependent chromatin remodeling (Thayer et al, 1993; Law & Jacobsen, 2010; Rafati et al, 2011)" should have better literature citrations. Two of the three appear to be focused on methylation based on the titles. There must be more appropriate reviews.

3. page 3 - This sentence: "Elevation of Tf2 open reading frame (ORF) RNA was more modest: 4-6-fold in the single mutants and 11-fold in the double mutant," was not clear. Please explain what is meant by ORF RNA.

Referee #2:

Review of Regulating retrotransposons via alternative transcription start sites by Persson et al.

In this manuscript the authors find the transcription of the retrotransposon Tf2 is controlled by the nucleosome remodeling factors fft2 and fft3. They discover that in wild type cells the 5' end of the dominant transcription is significantly downstream of the site in the LTR where transcription must initiate to produce the unique mRNA sequence necessary for self-primed reverse transcription. They conclude correctly that this dominant mRNA of Tf2 is incapable of initiating reverse transcription. They find that fft2 and fft3 bind the LTR sequence and recruit nucleosomes that inhibit transcription initiation at the LTR site. Importantly, they find the same LTR initiated transcript can be induced in wild type cells that are subjected to stress conditions. From these results they propose fft2 and fft3 regulate Tf2 transposition by allowing the self-priming mRNA to be produced when cells experience stress. They suggest that under conditions of stress fft2 and fft3 levels are down regulated allowing lower nucleosome occupancy in the LTR and this allows RNA pol II to initiate transcription from the LTR.

This paper outlines an interesting hypothesis that the sites of transcription initiation of Tf2 are switched by fft2/3 in response to stress. This mechanism would be of great significance if true. However, their data falls short of demonstrating key elements of the model. While they provide strong evidence Tf2 transcription is initiated at the self-primer sequence when fft2/3 are mutated, or when cells are stressed, they do not demonstrate the dominant Tf2 RNA in wild type cells results from a downstream initiation event. This RNA could be the result of post-transcriptional processing. In fact, Ekwall argued in Durand-Dubief et al 2007, the truncated Tf2 RNA resulted from a cleavage event. In the current paper the authors argue they have discovered a stress-induced switch of TSSs. To demonstrate this interesting possibility they must show that the shorter Tf2 RNA has a cap structure at the 5' end. It is therefore a significant concern that in Fig. 2 their CAGE data designed to map 5' capped mRNA does not detect the shortened RNA. Also, their 5'RACE method appears not to have used the 5' phosphatase technique that degrades all RNA that lacks 5' caps. Absent CAGE or phosphatase treated RACE data the authors are not justified in concluding the downstream site is actually a transcription start site. The inconsistency in the CAGE data should be addressed and phosphatase treated RACE data should be added to support their model. Another question is whether the increases in Tf2 RNA under conditions of stress result from reduced fft2/3 activity. The authors could address this by using their qPCR amplicon method with cells subjected to heat shocked. In the absence of fft2/3 there should be no increase in the mRNA that initiates in the LTR.

If the authors cannot add data to support their conclusion that the truncated RNA is due to transcription initiation I can imagine a manuscript that describes the function of fft2/3 that would be of great interest. Such a manuscript would drop the idea that TSSs were switched and leave open the origin of the truncated RNA.

The following are additional concerns that should be addressed.

Abstract, and last paragraph of the intro, The claim that this is a new mechanism of retrotransposon regulation by transcription start site control is not true. The authors are apparently unaware of the work by David Garfinkel's lab on Ty1. Saha et al identified a downstream TSS that expresses a protein that inhibits Ty1 (Journal of Virology 2015, Jan 21). The downstream TSS is regulated by a subunit of the SATA complex. spt3.

© European Molecular Biology Organization

Page 2, the comment "retrotransposons play a crucial role in plasticity" is not true. While there is mounting evidence that L1 transposes during neurogenesis, there is no direct evidence that these insertions have a biological function.

Fig. 1A. Please distinguish in the figure between LTRs and antisense Tf2 sequences.

Fig. 1B. The figure legend does not describe the additional lane of molecular wt markers. I assume its just a diluted version of the markers.

Page 3, 11 lines from bottom, The logic of the qPCR should be described. Its not clear how the authors arrived at the conclusion the ftt mutants resulted in additional sequence at the 5' end. For example, why does amplicon 1 show much greater increases compared to amplicon 2 if both are normalized to RNA in wild type cells? The language "we confirm the size shift" is too strong for this indirect evidence.

Page 3, 9 lines from bottom, the comment RNA-sequencing shows that the full 5'UTR is not transcribed for any of the 13 copies of Tf2 in wild type (WT) cells (Rhind et al, 2011). This wording is misleading. While none of the Rhind transcripts were mapped to the exact start of the self-priming sequence 2 mapped close to it indicating the upstream TSS is active. Indicate in the text that the Tf2 starts from Rhind are shown on Fig 2B.

Fig. 2. The authors should comment on the peak of CAGE counts at 580-600 and the peak at 370. How are these interpreted? Also, please state whether CAGE counts are strand-specific and represent starts just in the plus strand. And as indicated above, I am concerned that the truncated 5' RNA was not detected by CAGE.

Figure 3C and B. The difference between "mean occupancy and "in vivo occupancy" should to be explained in the legend.

Page 4, line 16. Other than at Tf2, where else dose fft2 and fft3 bind. Are the transposons the primary site of binding? Or does fft2/3 bind to other genes. Does fft2/3 bind the coding genes shown in Fig.1A to be induced in the fft mutants? And finally, what are the other genes in Fig. 1A that are induced in the fft mutants, are they stress genes.

Page 5, 18 lines from bottom. Is the function of fft2 and fft3 to reduce nucleosome occupancy during stress? The reduction in the occupancy for the double mutant is very small. Can this account for the 50-fold increase in transcription caused by heat shock. Does the head shock for 60 min result in less ftt binding to the LTRs.

Page 6, top and Figure 6B. Why is the increase in amplicon 1 in the fft2,fft3 double mutant only showing a seven fold increase in this figure when in Fig. 1D it was 80 fold? Also, there is much more variation in replicas in Fig. 6B. The authors should demonstrate that the differences in the amplicon levels are statistical significant.

Figure 6B and C. The axes of both graphs need to be labeled.

Page 7, line 4. The text mentions hERV regulation, the subject of the reference listed however, (Zhou et al. 2013), is on Tf2 regulation. Please correct either the text or the reference.

Referee #3:

• what are the major claims and how significant are they?

The manuscript shows that under conditions of stress the transcription of Schizosaccharomyces pombe Tf2 elements initiates from an additional promoter that includes the primer for reverse transcription. This potentially represents an interesting adaptation that could enable an accelerated rate of genomic/evolutionary plasticity during stress. This potentially represents a new insight into the relationship between evolution and stress.

The study also shows that transcripts initiate from the alternate promoter in the LTR following deletion of two paralogs within the Fun30/SMARCAD1 family of remodelling enzymes. The aspect covering the involvement of these proteins is not quite so compelling.

• are the claims novel and convincing?

The manuscript does establish that there is a change to chromatin over the U3 region of the LTR in a Fft2D, Fft3D strain. There is also ChIP enrichment for both proteins just upstream of this region. However, this is not sufficient evidence to conclude that these enzymes directly act to reposition this nucleosome. There is also enrichment for both proteins at the 3' end of the ORF but no change to chromatin. It is likely that many factors are recruited to the 5' LTR region. The Fft proteins are presumably require for a step that contributes to the reduction in nucleosome occupancy. Could they for example be involved in regulating the distribution of a histone variant or modification that is itself required to regulate occupancy of this nucleosome? Note that Htz1 occupancy is affected in a fun30 delete in budding yeast. There are many possible explanations for the observations especially as the change in nucleosome occupancy is partial (about 2 fold change in occupancy) and the Fft proteins have a significant ATP-independent function in regulating LTR transcription. There are not obvious experiments that can be done to address this, so it is better to discuss this and moderate all sections of the manuscript that attribute a direct effect of Fft proteins in regulating nucleosome positioning.

© European Molecular Biology Organization

As a minor point there is some discussion of the possibility that the copy number of Fun30 family members may have increased to compensate for a loss of ISWI proteins. This appears very speculative as it is more likely that duplication of the more closely related Chd1 paralogs substitutes for the loss of an ISWI member.

The plots showing propensity for nucleosome formation in figure 3C seem unnecessary as they add little to the manuscript.

• are the claims appropriately discussed in the context of earlier literature?
yes
• is the study of interest to more than a specialised audience?

I think the major interest is from an evolutionary perspective.

• does the paper stand out in some way from the others in its field?

• are there other experiments that would strengthen the paper?

Nothing reasonably feasible.

Manuscript Transfer - authors' response 03 December 2015

Response to reviewers

We would like to thank all three reviewers for constructive criticisms which we have addressed in a revised version of this manuscript for consideration in EMBO Reports. Below we provide a point-by-point response to the issues raised. We would like to point out the addition of a Tf2 in vivo retro-mobility assay (new Figure 8), which provides further support for the proposed model.

Referee #1:

In this interesting article, the authors demonstrate that Tf2 transcription initiation is regulated by the factors Fft2 and Fft3, which control nucleosome position and clustering. When these factors are lost, there is a change in nucleosome position, clustering is lost, and transcription now initiates at an alternative start site, further 5', resulting in a full-length and productive transcript. Overall, the experiments are convincing and are clearly presented. Minor comments are below.

1. Ty1 elements of S. cerevisiae are also controlled by alternative transcription initiation. See A trans-dominant form of Gag restricts Ty1 retrotransposition and mediates copy number control. Saha A, Mitchell JA, Nishida Y, Hildreth JE, Ariberre JA, Gilbert WV, Garfinkel DJ. J Virol. 2015 Jan 21. pii: JVI.03060-14. [Epub ahead of print] and earlier studies form this lab. Although a different type of change, this work should be cited as a relevant comparison.

Thanks a lot for pointing this out - the reference is inserted and discussed on page 3.

2. page 2 - This sentence: "Host cells have developed numerous silencing mechanisms to repress these elements, including DNA methylation, RNA interference (RNAi), repressive histone modifications and ATP-dependent chromatin remodeling (Thayer et al, 1993; Law & Jacobsen, 2010; Rafati et al, 2011)" should have better literature citrations. Two of the three appear to be focused on methylation based on the titles. There must be more appropriate reviews.

We have added two additional references to review articles on retrotransposons in yeast and plants in this section.

3. page 3 - This sentence: "Elevation of Tf2 open reading frame (ORF) RNA was more modest: 4-6-fold in the single mutants and 11-fold in the double mutant," was not clear. Please explain what is meant by ORF RNA.

Page 4: Corrected to 'protein coding RNA'

© European Molecular Biology Organization

Referee #2:

Review of Regulating retrotransposons via alternative transcription start sites by Persson et al.

In this manuscript the authors find the transcription of the retrotransposon Tf2 is controlled by the nucleosome remodeling factors fft2 and fft3. They discover that in wild type cells the 5' end of the dominant transcription is significantly downstream of the site in the LTR where transcription must initiate to produce the unique mRNA sequence necessary for self-primed reverse transcription. They conclude correctly that this dominant mRNA of Tf2 is incapable of initiating reverse transcription. They find that fft2 and fft3 bind the LTR sequence and recruit nucleosomes that inhibit transcription initiation at the LTR site. Importantly, they find the same LTR initiated transcript can be induced in wild type cells that are subjected to stress conditions. From these results they propose fft2 and fft3 regulate Tf2 transposition by allowing the self-priming mRNA to be produced when cells experience stress. They suggest that under conditions of stress fft2 and fft3 levels are down regulated allowing lower nucleosome occupancy in the LTR and this allows RNA pol II to initiate transcription from the LTR.

This paper outlines an interesting hypothesis that the sites of transcription initiation of Tf2 are switched by fft2/3 in response to stress. This mechanism would be of great significance if true. However, their data falls short of demonstrating key elements of the model. While they provide strong evidence Tf2 transcription is initiated at the self-primer sequence when fft2/3 are mutated, or when cells are stressed, they do not demonstrate the dominant Tf2 RNA in wild type cells results from a downstream initiation event. This RNA could be the result of post-transcriptional processing. In fact, Ekwall argued in Durand-Dubief et al 2007, the truncated Tf2 RNA resulted from a cleavage event. In the current paper the authors argue they have discovered a stress-induced switch of TSSs. To demonstrate this interesting possibility they must show that the shorter Tf2 RNA has a cap structure at the 5' end. It is therefore a significant concern that in Fig. 2 their CAGE data designed to map 5' capped mRNA does not detect the shortened RNA. Also, their 5'RACE method appears not to have used the 5' phosphatase technique that degrades all RNA that lacks 5' caps. Absent CAGE or phosphatase treated RACE data the authors are not justified in concluding the downstream site is actually a transcription start site. The inconsistency in the CAGE data should be addressed and phosphatase treated RACE data should be added to support their model.

We have performed 5'-cap sensitive RNA ligase mediated RACE (RLM-RACE) to validate the capped RNA produced in wild type cells (Table 1). The RLM-CAGE reads coincide with CAGE mapped TSS in region 600 bp downstream of the 5' end of the LTR (compare Table 1; Figure 2)

Another question is whether the increases in Tf2 RNA under conditions of stress result from reduced fft2/3 activity. The authors could address this by using their qPCR amplicon method with cells subjected to heat shocked. In the absence of fft2/3 there should be no increase in the mRNA that initiates in the LTR.

Thanks for suggesting this key experiment. We have studied Tf2 RNA levels in fft2 and fft3 mutants and wt controls during heat shock conditions (30 min and 60 min). We detect no significant increase of full length Tf2 mRNA in mutant backgrounds (Figure 5C). This indicates that Fft2 and Fft3 are epistatic to stress signaling supporting the proposed model.

If the authors cannot add data to support their conclusion that the truncated RNA is due to transcription initiation I can imagine a manuscript that describes the function of fft2/3 that would be of great interest. Such a manuscript would drop the idea that TSSs were switched and leave open the origin of the truncated RNA.

The following are additional concerns that should be addressed.

Abstract, and last paragraph of the intro, The claim that this is a new mechanism of retrotransposon regulation by transcription start site control is not true. The authors are apparently unaware of the work by David Garfinkel's lab on Ty1. Saha et al identified a downstream TSS that expresses a protein that inhibits Ty1 (Journal of Virology 2015, Jan 21). The downstream TSS is regulated by a subunit of the SATA complex. spt3.

This reference is inserted and discussed on page 3. 'It is known from studies of the Ty1 retrotransposon in budding yeast that a shorter transcript, giving rise to a truncated Gag protein is involved in copy number control at the level of virus-like particle formation (Saha et al, 2015).'

Page 2, the comment "retrotransposons play a crucial role in plasticity" is not true. While there is mounting evidence that L1 transposes during neurogenesis, there is no direct evidence that these insertions have a biological function.

We have softened the text to 'retrotransposons may be involved in plasticity'

Fig. 1A. Please distinguish in the figure between LTRs and antisense Tf2 sequences.

Corrected. The grey arrows represent antisense Tf2 sequences.

Fig. 1B. The figure legend does not describe the additional lane of molecular wt markers. I assume its just a diluted version of the markers.

Yes, two different exposures of the molecular size marker are shown. This is now mentioned in the legend. The lighter marker lane is from the same exposure as the rest of the blot, while the longer exposure has been included for clarity.

Page 3, 11 lines from bottom, The logic of the qPCR should be described. Its not clear how the authors arrived at the conclusion the ftt mutants resulted in additional sequence at the 5' end. For example, why does amplicon 1 show much greater increases compared to amplicon 2 if both are normalized to RNA in wild type cells? The language "we confirm the size shift" is too strong for this indirect evidence.

We have clarified this paragraph:
'Next, we performed reverse transcription and qPCR for different regions of the Tf2 mRNA (Durand-Dubief et al, 2007) (Fig 1C). In this assay the different regions of the RNA are represented in different ratios in mutant vs wt samples. In wt, RNA molecules that contain amplicon 2 are abundant, while amplicon 1 is rare. In fft mutants, however, amplicon 1 is upregulated, leading to a strong ratio increase. This assay thus validated the size shift and suggested that the additional sequence in the mutant samples is at the 5' end of the transcript (Fig 1C,D). '

Page 3, 9 lines from bottom, the comment RNA-sequencing shows that the full 5'UTR is not transcribed for any of the 13 copies of Tf2 in wild type (WT) cells (Rhind et al, 2011). This wording is misleading. While none of the Rhind transcripts were mapped to the exact start of the self-priming sequence 2 mapped close to it indicating the upstream TSS is active. Indicate in the text that the Tf2 starts from Rhind are shown on Fig 2B.

Changed to:
'Interestingly, RNA-sequencing shows that the full 5'UTR is not transcribed for any of the 13 copies of Tf2 in wild type (WT) cells (Rhind et al, 2011; shown in Fig 2B).'

Fig. 2. The authors should comment on the peak of CAGE counts at 580-600 and the peak at 370. How are these interpreted? Also, please state whether CAGE counts are strand-specific and represent starts just in the plus strand. And as indicated above, I am concerned that the truncated 5' RNA was not detected by CAGE.

See changes on page 4:
'To validate that the shorter RNA produced in wt cells are capped we used 5'-cap sensitive RNA ligase mediated RACE (RLM-RACE). A total of 21 wt RLM-RACE clones cDNA clones were sequenced and they verified the CAGE results with capped mRNA with most sequence reads starting in the region around 600 bp relative to the 5' end of the LTR coinciding with wt CAGE signals (Table 1; Figure 2).'

Figure 3C and B. The difference between "mean occupancy and "in vivo occupancy" should to be explained in the legend.

The panel Fig 3C has been removed since referee 3 thinks it is unnecessary and we agree. Also we have realized that the algorithms we used are not optimized to S. pombe nucleosome positions (see Lantermann et al Nat Struct Mol Biol. 2010 Feb;17(2):251-7.).

Page 4, line 16. Other than at Tf2, where else dose fft2 and fft3 bind. Are the transposons the primary site of binding? Or does fft2/3 bind to other genes. Does fft2/3 bind the coding genes shown in Fig.1A to be induced in the fft mutants? And finally, what are the other genes in Fig. 1A that are induced in the fft mutants, are they stress genes.

The genome-wide binding of Fft2 and Fft3 will be further explored in another manuscript (in preparation).

Page 5, 18 lines from bottom. Is the function of fft2 and fft3 to reduce nucleosome occupancy during stress? The reduction in the occupancy for the double mutant is very small. Can this account for the 50-fold increase in transcription caused by heat shock. Does the head shock for 60 min result in less ftt binding to the LTRs.

As mentioned on page 5: Fft2 and Fft3 are downregulated in heat stress conditions. We refer to published data regarding this finding (Chen et al, 2003). Exactly how Fft2 and Fft3 proteins are reduced by heat shock will be subject to a future study.

Page 6, top and Figure 6B. Why is the increase in amplicon 1 in the fft2,fft3 double mutant only showing a seven fold increase in this figure when in Fig. 1D it was 80 fold? Also, there is much more variation in replicas in Fig. 6B. The authors should demonstrate that the differences in the amplicon levels are statistical significant.

Yes, there is some variation from experiment to experiment and we provide error bars in each figure for the different amplicons to provide statistical significance. Please note that the data in Fig1 and Fig6 are not directly comparable since (for historical reasons) different control loci were used (SPAC1F8.07c and act1).

Figure 6B and C. The axes of both graphs need to be labeled.

Corrected

Page 7, line 4. The text mentions hERV regulation, the subject of the reference listed however, (Zhou et al. 2013), is on Tf2 regulation. Please correct either the text or the reference.

Corrected:
'Our model is consistent with observations that reduced histone production results in Tf2 upregulation (Zhou et al, 2013),…'

Referee #3:

• what are the major claims and how significant are they?

The manuscript shows that under conditions of stress the transcription of Schizosaccharomyces pombe Tf2 elements initiates from an additional promoter that includes the primer for reverse transcription. This potentially represents an interesting adaptation that could enable an accelerated rate of genomic/evolutionary plasticity during stress. This potentially represents a new insight into the relationship between evolution and stress.

The study also shows that transcripts initiate from the alternate promoter in the LTR following deletion of two paralogs within the Fun30/SMARCAD1 family of remodelling enzymes. The aspect covering the involvement of these proteins is not quite so compelling.

• are the claims novel and convincing?

The manuscript does establish that there is a change to chromatin over the U3 region of the LTR in a Fft2D, Fft3D strain. There is also ChIP enrichment for both proteins just upstream of this region. However, this is not sufficient evidence to conclude that these enzymes directly act to reposition this nucleosome. There is also enrichment for both proteins at the 3' end of the ORF but no change to chromatin. It is likely that many factors are recruited to the 5' LTR region. The Fft proteins are presumably require for a step that contributes to the reduction in nucleosome occupancy. Could they for example be involved in regulating the distribution of a histone variant or modification that is itself required to regulate occupancy of this nucleosome?

Yes, it is possible that the histone variant H2A.Z plays a role in nucleosomes at LTR elements. However we did not observe Tf2 derepression in swr1, msc1 mutants which affect H2A.Z levels nor in the pht1 (H2A.Z) mutant (Buchanan et al PloS genetics 2009). Therefore we prefer not to include this as a possibility in the text. The simplest model is a direct effect of Fun30 on occupancy of nucleosomes given the binding of Fun30 proteins to the LTR and the effect of catalytically dead mutants (Fig 6A).

Note that Htz1 occupancy is affected in a fun30 delete in budding yeast. There are many possible explanations for the observations especially as the change in nucleosome occupancy is partial (about 2 fold change in occupancy) and the Fft proteins have a significant ATP-independent function in regulating LTR transcription. There are not obvious experiments that can be done to address this, so it is better to discuss this and moderate all sections of the manuscript that attribute a direct effect of Fft proteins in regulating nucleosome positioning.

As a minor point there is some discussion of the possibility that the copy number of Fun30 family members may have increased to compensate for a loss of ISWI proteins. This appears very speculative as it is more likely that duplication of the more closely related Chd1 paralogs substitutes for the loss of an ISWI member.

This paragraph has been changed and ISWI proteins are not mentioned:
'From an evolutionary perspective, it is interesting to note that the radiation of the Fun30 remodeler subfamily in the fission yeasts occurred in parallel with other major evolutionary changes, including the elimination of most transposon families and a transition to non-transposon centromeres (Rhind et al, 2011). Given the involvement of the S. pombe Fun30 remodelers in both centromeric function and retrotransposon regulation, it is tempting to speculate that the diversification of this subfamily either allowed for or was favored by these changes.'

The plots showing propensity for nucleosome formation in figure 3C seem unnecessary as they add little to the manuscript.

We agree. The panel Fig 3C has been removed. Also we have realized that the algorithms we used are not optimized to S. pombe nucleosome positions (see Lantermann et al Nat Struct Mol Biol. 2010 Feb;17(2):251-7.).

• are the claims appropriately discussed in the context of earlier literature?
yes
• is the study of interest to more than a specialised audience?

I think the major interest is from an evolutionary perspective.

• does the paper stand out in some way from the others in its field?

• are there other experiments that would strengthen the paper?

Nothing reasonably feasible.

© European Molecular Biology Organization

2nd Editorial Decision 04 January 2016

Thank you for the transfer of your revised manuscript to EMBO reports. We have now received the comments from both referees, and I am happy to tell you that both support the publication of your revised study.

Referee 1 has several minor comments that I would like you to address before we can proceed with the official acceptance of your manuscript. Our supplementary figures are called expanded view (EV) figures now, can you therefore please change the names and cite them as EV1, 2, etc figures in the manuscript text? The tables should also be called table EV1 and EV2. Please add the legends for the EV figures to the end of the main manuscript file, and delete "supplemental material".

Regarding statistics, please specify the error bars and the test used to calculate the p value for Figures 4, 5, 6 and EV3. Please also specify "n" for Fig 7 and EV3. This information must be provided in the figure legends.
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.

I look forward to seeing a final version of your manuscript as soon as possible.

REFEREE REPORTS

Referee #1:

Review of "Regulating retrotransposons via alternative transcription start sites" by Persson et al.

In this manuscript the authors find the transcription of the retrotransposon Tf2 is controlled by the nucleosome remodeling factors fft2 and fft3. They discover that in wild type cells the 5' end of the dominant transcription is significantly downstream of the site in the LTR where transcription must initiate to produce the unique mRNA sequence necessary for self-primed reverse transcription. They conclude correctly that this dominant mRNA of Tf2 is incapable of initiating reverse transcription. They find that fft2 and fft3 bind the LTR sequence and recruit nucleosomes that inhibit transcription initiation at the LTR site. Importantly, they find the same LTR initiated transcript can be induced in wild type cells containing a marked Tf2 element and this leads to a substantial increase in transposition frequencies.

In my earlier review I argued that they had not confirmed that the shorter RNA was the result of a novel transcription start site. In the revised manuscript the authors added an experiment that mapped the start of 5' Capped mRNA and these coincided with the start sites mapped by their CAGE experiment. As a result the manuscript is greatly improved. In addition, the authors added another experiment that shows the stress-induced transcription of Tf2 requires fft2 and fft3. This revised draft presents a focused series of experiments showing a new mechanism for regulating LTR-retrotransposons. It will be of great interest to a wide readership. The following are minor considerations that should be addressed with modifications to the text.

1. Top of page 2, there is a typo with a 1 following the reference (Struhl & Segal, 2013).
2. Page 2, line 14. Of should be added to "restricted by the activity many" before many.
3. Page 2, line 24. The word in is repeated twice.
4. Page 2, 5 lines from bottom. This statement is incorrect. The PBS is not located in the LTR, its just downstream of the LTR. This is true in all LTR elements.
5. Page 3, 10 lines from bottom. It is not clear to me how Fig. S1 shows the ura4 reporter is derepressed in the fft double mutant. For this to be shown the strain with the Tf2-ura4 reporter and the fft mutations should grow better on -ura than without the fft deletions. This is not the case because the fft deletions cause slow growth. Unless Im missing something here the figure should be

removed.
6. Page 3, four lines from the bottom. The word evident is too strong. It is not obvious that the mobility has shifted. Use a word like suggests.
7. Page 3, last line. The word represented seems inappropriate here. Instead the word detected would be better.
8. Page 4, top. The statement "In wt, RNA molecules that contain amplicon 2 are abundant, while amplicon 1 is rare" can not be deduced by Fig. 1D because the RNA levels are all normalized. Please provide the absolute amounts of the amplicons (relative to another mRNA) to show there is more #2 than #1. The same point applies to 7 lines down. The absolute levels are needed to conclude the extended RNA is "virtually" absent.
9. Page 4, 12 lines from the bottom. The statement "this confirms that the short form of RNA that dominates in wt is produced from a downstream TSS" oversimplifies the result. There is not a single short RNA. The CAGE result and the 5' Capped RACE identify many shorter species. In fact the authors should say that the 5' end detected by 5' RACE in Fig. 1E does not correspond to the other shortened species in the CAGE experiment. It should be presented that there are multiple downstream TSS sites. Otherwise, the reads will be confused.
10. Figure 3B. The Y-axis labeled occupancy should include the word nucleosome.
11. Figure 4 lacks the A and B labels.
12. Page 8, bottom. The statement "Both human and mouse development exhibit close ties to transposable elements, with MuERV-L and HERV-H activation being shown to mark pluripotency or trigger and regulate embryonic development, respectively (Macfarlan et al, 2013; Peaston et al, 2004; Santoni et al, 2012)." Is too strong. There is no direct evidence that TEs or ERVs trigger or regulate development. This is a model and the authors should present it as such.
13. Page 9, line 3. Same point as #13. Fort et al presents a correlation not "proof" as the authors say.
14. Page 10, line 3. Same point as #s 13 and 14. The activities of ERVs and retrotransposons correlate with development. There is no direct evidence that they are "important" for mammalian development and plasticity. The authors should use the words surprising correlation or say the interesting possibility that ERVs may activate development.

Referee #2:

I am satisfied the authors have revised the manuscript to a standard where it is suitable for publication.

1st Revision - authors' response 06 January 2016

Changes in the revised manuscript (after manuscript transfer to EMBO reports)

Response to editor

Expanded View: Supplementary figures renamed to Figure EV1 to EV5
Tables renamed to Table EV1 to EV3
Legends for EV figures have been added to the end of the manuscript file

Reference style has been changed to the numbered EMBO reports style

A filled in author checklist is provided

Regarding statistics, please specify the error bars and the test used to calculate the p value for Figures 4, 5, 6 and EV3. Please also specify "n" for Fig 7 and EV3. This information must be provided in the figure legends.

Changes in Legends
Fig. 1D) Last sentence has been changed to: 'Error bars represent standard deviation of duplicate reverse transcriptions of biological triplicates.'

Fig. 3B) Inserted 'Error bars show standard deviation of nucleosome occupancy averaged over 13 tf2 elements.'

*Figs 4A) and 4B) Inserted 'A and B) Data shown as box plots, significance of difference between categories was assessed by Wilcoxon-Mann-Whitney test (** p<0.001).'*

Figs 5A) and 5C) Inserted at the end of each figure legend: 'Error bars represent standard deviation of reverse transcriptions of biological duplicates'

Fig 6A) Inserted: 'Error bars represent standard deviation of reverse transcriptions of biological duplicates'

Fig 6B) Inserted: 'Error bars represent standard deviation of nucleosome occupancy of biological duplicates.'

Fig 7B) Added: 'Tf2 clustering was counted in wt cells (n=663), fft2Δ (n=556), fft3Δ (n=493), fft2Δ fft3Δ (n=636). Significance of difference between strains was assessed by Chi-square test.'

Fig EV2) Changed to: 'Acetylation of H3K9 is elevated at Tf2 elements in fft2D, fft3D, and fft2D fft3D. Error bars represent the standard deviation of biological triplicate samples. Occupancy is relative to a control gene (SPAC1F8.07c) promoter and to WT.'

Fig EV3) Changed to: 'Transcripts with an upstream LTR are neither more nor less abundant than transcripts without an upstream LTR. Average signal intensity across genes of wildtype RNA hybridized to tiling array shown as box plot for each of the following categories: non-coding with LTR (n=69) / without LTR (n=1665); protein coding with LTR (n=246) / without LTR (n=4887); Tf2s (n=13).'

Fig EV5) The legend already has a sentence on error bars.

Response to Referee #1

Referee #1:

Review of "Regulating retrotransposons via alternative transcription start sites" by Persson et al.

In this manuscript the authors find the transcription of the retrotransposon Tf2 is controlled by the nucleosome remodeling factors fft2 and fft3. They discover that in wild type cells the 5' end of the dominant transcription is significantly downstream of the site in the LTR where transcription must initiate to produce the unique mRNA sequence necessary for self-primed reverse transcription. They conclude correctly that this dominant mRNA of Tf2 is incapable of initiating reverse transcription. They find that fft2 and fft3 bind the LTR sequence and recruit nucleosomes that inhibit transcription initiation at the LTR site. Importantly, they find the same LTR initiated transcript can be induced in wild type cells containing a marked Tf2 element and this leads to a substantial increase in transposition frequencies.

In my earlier review I argued that they had not confirmed that the shorter RNA was the result of a novel transcription start site. In the revised manuscript the authors added an experiment that mapped the start of 5' Capped mRNA and these coincided with the start sites mapped by their CAGE experiment. As a result the manuscript is greatly improved. In addition, the authors added another experiment that shows the stress-induced transcription of Tf2 requires fft2 and fft3. This revised draft presents a focused series of experiments showing a new mechanism for regulating LTR-retrotransposons. It will be of great interest to a wide readership. The following are minor considerations that should be addressed with modifications to the text.

1. Top of page 2, there is a typo with a 1 following the reference (Struhl & Segal, 2013).

Page 2: Corrected

2. Page 2, line 14. Of should be added to "restricted by the activity many" before many.

Page 2: Corrected

3. Page 2, line 24. The word in is repeated twice.

Page 2: Corrected

4. Page 2, 5 lines from bottom. This statement is incorrect. The PBS is not located in the LTR, its just downstream of the LTR. This is true in all LTR elements.

Page 2: Corrected and redrawn in Figure 9

5. Page 3, 10 lines from bottom. It is not clear to me how Fig. S1 shows the ura4 reporter is derepressed in the fft double mutant. For this to be shown the strain with the Tf2-ura4 reporter and the fft mutations should grow better on -ura than without the fft deletions. This is not the case because the fft deletions cause slow growth. Unless Im missing something here the figure should be removed.

The fft double mutant is growing slowly making the direct comparison on spotting assay difficult. However the derepression of Tf2-ura4+ in fft double mutant is seen as a completely inhibited growth on FOA plates while Ura+ colonies are still detected on –ura plates after 8 days. In contrast Wt cells with Tf2-ura4+ grow on FOA and –ura plates to similar extent.

To clarify this we have made a text change on Page 3 to 'A ura4+ reporter gene inserted 3' of Tf2-11 is also clearly derepressed in an fft2D fft3D double deletion mutant leading to reduced growth on counter-selective FOA plates (Fig EV1).'

6. Page 3, four lines from the bottom. The word evident is too strong. It is not obvious that the mobility has shifted. Use a word like suggests.

Changed to: 'A size shift was also suggested by the blot, with the Tf2 mRNA gel mobility being slightly decreased in the single and double mutants (Fig 1B).'

7. Page 3, last line. The word represented seems inappropriate here. Instead the word detected would be better.

Changed to: 'detected'

8. Page 4, top. The statement "In wt, RNA molecules that contain amplicon 2 are abundant, while amplicon 1 is rare" can not be deduced by Fig. 1D because the RNA levels are all normalized. Please provide the absolute amounts of the amplicons (relative to another mRNA) to show there is more #2 than #1. The same point applies to 7 lines down. The absolute levels are needed to conclude the extended RNA is "virtually" absent.

We have softened the language here since we don't have absolute numbers. Changed to: 'In wt, RNA molecules that contain amplicon 2 are relatively abundant, while amplicon 1 is rare. In fft mutants, however, amplicon 1 is upregulated, leading to a strong ratio increase. This assay thus validated the size shift and suggested that the additional sequence in the mutant samples is at the 5' end of the transcript (Fig 1C,D). The full 5'UTR of fission yeast Tf1 and Tf2 elements is needed for the reverse transcription stage of the retrotransposon life cycle (Levin, 1995). Interestingly, RNA-sequencing shows that the full 5'UTR is not transcribed for any of the 13 copies of Tf2 in wild type (WT) cells (Rhind et al, 2011; shown in Fig 2B). Our results confirm that in WT cells, mRNA molecules with this extended 5'UTR are relatively rare from the population (Fig 1D).'

9. Page 4, 12 lines from the bottom. The statement "this confirms that the short form of RNA that dominates in wt is produced from a downstream TSS" oversimplifies the result. There is not a single short RNA. The CAGE result and the 5' Capped RACE identify many shorter species. In fact the

authors should say that the 5' end detected by 5' RACE in Fig. 1E does not correspond to the other shortened species in the CAGE experiment. It should be presented that there are multiple downstream TSS sites. Otherwise, the reads will be confused.

Changed to: 'The CAGE signals were highly consistent over three replicates, and since RLM-RACE uses a different chemistry than CAGE, this confirms that the shorter forms of RNA that dominate in wt are produced from a downstream TSS. There are multiple downstream TSS sites in a region around 600 bp from the 5' end of the LTR (Table EV1).'

10. Figure 3B. The Y-axis labeled occupancy should include the word nucleosome.

11. Figure 4 lacks the A and B labels.

12. Page 8, bottom. The statement "Both human and mouse development exhibit close ties to transposable elements, with MuERV-L and HERV-H activation being shown to mark pluripotency or trigger and regulate embryonic development, respectively (Macfarlan et al, 2013; Peaston et al, 2004; Santoni et al, 2012)." Is too strong. There is no direct evidence that TEs or ERVs trigger or regulate development. This is a model and the authors should present it as such.

Changed to: 'Both human and mouse development exhibit close ties to transposable elements, with MuERV-L and HERV-H activation being suggested to mark pluripotency or trigger and regulate embryonic development, respectively…'

13. Page 9, line 3. Same point as #13. Fort et al presents a correlation not "proof" as the authors say.

Changed to: 'These transcripts demonstrate cell-type specific regulation (Faulkner et al, 2009) and have, in some cases, been shown to correlate with pluripotency (Fort et al, 2014).'

14. Page 10, line 3. Same point as #s 13 and 14. The activities of ERVs and retrotransposons correlate with development. There is no direct evidence that they are "important" for mammalian development and plasticity. The authors should use the words surprising correlation or say the interesting possibility that ERVs may activate development.

Changed to: 'Strict TSS control, regulated by chromatin remodelers and nucleosome positioning, provides context for the interesting possibility that ERVs may activate mammalian development.'

2nd Editorial Decision 13 January 2016

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.

© European Molecular Biology Organization

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND

Corresponding Author Name: Karl Ekwall
Journal Submitted to: EMBO reports
Manuscript Number: EMBOR-2015-41866V1

Reporting Checklist For Life Sciences Articles (Rev. July 2015)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified
- Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.
- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple χ2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

Please ensure that the answers to the following questions are reported in the manuscript itself. We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

In the pink boxes below, provide the page number(s) of the manuscript draft or figure legend(s) where the information can be located. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable).

USEFUL LINKS FOR COMPLETING THIS FORM

http://www.antibodypedia.com
http://1degreebio.org
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repo
http://grants.nih.gov/grants/olaw/olaw.htm
http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm
http://ClinicalTrials.gov
http://www.consort-statement.org
http://www.consort-statement.org/checklists/view/32-consort/66-title
http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tun
http://datadryad.org
http://figshare.com
http://www.ncbi.nlm.nih.gov/gap
http://www.ebi.ac.uk/ega
http://biomodels.net/
http://biomodels.net/miriam/
http://jjj.biochem.sun.ac.za
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html
http://www.selectagents.gov/

B- Statistics and general methods

Please fill out these boxes (Do not worry if you cannot see all your text once you press return)

Question	Answer
1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	For both locus-specific experiments and genome-wide data, biological duplicate or (where feasible) triplicate experiments were analyzed and averaged. In FISH experiments, at least 450 cells were analyzed for each genotype.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	No animal studies were conducted
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	No samples were excluded
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	No
For animal studies, include a statement about randomization even if no randomization was used.	No animal studies were conducted
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	No
4.b. For animal studies, include a statement about blinding even if no blinding was done	No animal studies were conducted
5. For every figure, are statistical tests justified as appropriate?	Yes
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	To assess statistical significance of genome-wide data, the Wilcoxon-Mann-Whitney test was used, which does not assume normally distributed data and only requires independent observations. To compare cell counts in the FISH experiments, Pearson's Chi-squared test was used, which is applied to categorical data and requires independent observations of sufficient number.
Is there an estimate of variation within each group of data?	For genome-wide data, box-and-whiskers plots are used to show the spread of the data. In locus-specific experiments, variance is either shown through standard deviation when three replicates were used, or by showing values from two replicates as well as their mean.
Is the variance similar between the groups that are being statistically compared?	Yes