THE EMBO JOURNAL

A small targeting domain in Ty1 integrase is sufficient to direct retrotransposon integration upstream of tRNA genes

Amna Asif-Laidin, Christine Conesa, Amandine Bonnet, Camille Grison, Indranil Adhya, Rachid Menouni, Hélène Fayol, Noé Palmic, joel Acker, and Pascale LESAGE
DOI: N/A

Corresponding author(s): Pascale LESAGE (pascale.lesage@inserm.fr) , joel Acker (joel.acker@cea.fr)

Review Timeline:

Submission Date:	6th Jan 20
Editorial Decision:	9th Jan 20
Appeal Received:	13th Jan 20
Editorial Decision:	12th Feb 20
Revision Received:	13th May 20
Editorial Decision:	29th May 20
Revision Received:	9th Jun 20
Accepted:	18th Jun 20

Editor: Stefanie Boehm

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. Depending on transfer agreements, referee reports obtained elsewhere may or may not be included in this compilation. Referee reports are anonymous unless the Referee chooses to sign their reports.)

1st Editorial Decision 9th Jan 2020

Dr. Pascale LESAGE
INSERM U944, CNRS UMR 7212, Genomes & Cell Biology of Disease Unit, Institut de Recherche Saint-Louis, Université de Paris, Hôpital Saint-Louis
1 avenue Claude Vellefaux
Paris, Paris 75010
France

8th Jan 2020

Re: EMBOJ-2019-104337
The nuclear localization signal of Ty1 integrase targets retrotransposon integration to tRNA genes

Dear Dr. Lesage,

Thank you for submitting your manuscript (EMBOJ-2019-104337) to The EMBO Journal. I have now read your study carefully and discussed the work with the other members of the editorial team, including Senior Editor Hartmut Vodermaier and Chief Editor Bernd Pulverer. I regret to inform you that we have unfortunately decided not to pursue publication of this manuscript in The EMBO Journal.

We appreciate that you further characterize the interaction of Ty1 integrase (IN) and the RNAP III subunit AC40, and identify residues critical for the interaction within the bNLS sequence. Furthermore, this sequence is needed for the recruitment of IN to Pol III-transcribed genes and can also function as an independent module to target integration to Pol III-transcribed sites. We recognize that this study extends your previous work on Ty1 and AC40 and will likely raise interest in the immediate field. However, we also find that the molecular details of how this interaction affects integration at Pol III and/or Pol I-transcribed genes remains to be defined, as well as addressing a potential functional role. Thus, also in light of your previous study reporting the role of the IN1- AC40 interaction for integration per se, we have concluded that the mechanistic insight and conceptual advance provided for a broader audience is not sufficient to warrant further consideration for publication at The EMBO Journal.

That being said, given the interest in your study, I have taken the opportunity to discuss the manuscript with my colleague Andrea Leibfried, Executive Editor of our partner journal Life Science Alliance (http://www.life-science-alliance.org/), that we launched in partnership with Rockefeller University Press and Cold Spring Harbor Laboratory Press. I am pleased to confirm that Andrea would be interested in considering this work in its present form for in-depth external review at Life Science Alliance. Should you be interested in this option, please simply follow the link below for transfer; no reformatting is required.

Thank you for giving us the opportunity to consider your manuscript. I am sorry that we cannot be more positive on this occasion and I hope you will be interested in the transfer option.

Kind regards,

Stefanie Boehm

Stefanie Boehm, PhD
Editor
The EMBO Journal

*** As a service to authors, The EMBO Journal offers the possibility to directly transfer declined manuscripts to another EMBO Press title (EMBO Reports, EMBO Molecular Medicine, Molecular Systems Biology) or to the open access journal Life Science Alliance launched in partnership between EMBO Press, Rockefeller University Press and Cold Spring Harbor Laboratory Press. The full manuscript (including reviewer comments, where applicable and if chosen) will be automatically forwarded to the receiving journal, to allow for fast handling and a prompt decision on your manuscript. For more details of this service, and to transfer your manuscript to another EMBO title please follow this link:
Link Not Available

Dear Editors,

I appreciate the time you have devoted to analyze our work. However, we would like to challenge your view that it will primarily be of interest to these investigators in this immediate field.

1) Ty retrotransposons are at the forefront of research that has elucidated many fundamental mechanisms of retrotransposition. This is especially the case for the tethering of integrase by cellular factors, which underlies retroelement integration targeting. This tethering model also explains the HIV and MLV retrovirus integration site preferences, a point of crucial importance in human health.

2) Current literature on transposable elements is now exploding, due to their multiple impacts on genome biology, diseases and cancers. Historical models such as Ty1 are invaluable to decipher some of the molecular mechanisms at the heart of their effects on host cells, as nicely illustrated by our published and current work with Ty1.

3) The submitted manuscript puts forwards significant new data:

- For the first time, IN1 ChIP-seq is performed, allowing describing the genome-wide occupancy of IN1, including at Pol I genes.

- We identify a functional role for the IN1-AC40 interaction by showing that it is the determinant for IN1 interaction with Pol III and Pol I complexes, and IN1 recruitment at Pol III and Pol I-transcribed genes.

- That Ty1 targeting sequence is sufficient to confer Ty1 integration site preferences to the Ty5 retrotransposon is both striking, novel and elegant.

Therefore, we would be grateful if you would reconsider your decision. We believe that external review could convince you not only of the intrinsic quality of the data, but also of its broad interest to other investigators in the field. In that respect, we would like to emphasize that since its deposit in BioRXiv on December 18th, 2019, the manuscript has received great attention with over 760 abstract views and 130 PDF downloads (even during the Christmas holidays!), revealing its broad interest to the community.

2nd Editorial Decision 12th Feb 2020

Dr. Pascale LESAGE
INSERM U944, CNRS UMR 7212, Genomes & Cell Biology of Disease Unit, Institut de Recherche Saint-Louis, Université de Paris, Hôpital Saint-Louis
1 avenue Claude Vellefaux
Paris, Paris 75010
France

12th Feb 2020

Re: EMBOJ-2019-104337R-Q
The nuclear localization signal of Ty1 integrase targets retrotransposon integration to tRNA genes

Dear Dr. Lesage,

Thank you again for submitting your manuscript on Ty1 integrase for consideration by The EMBO Journal, as well as sending your comments regarding our previous decision. As mentioned, after discussing these points with an external advisor of the journal, we sent the study to three experts in the field. We have now received the reports from these experts, which are included below for your information.

As you will see, the reviewers express an interest in the findings and appreciate an additional insight into Ty1 integration. However, they also raise several major concerns that would need to be addressed in a revised version of the manuscript. Specifically, it will be important to further clarify and discuss if and to which extent (preferential) integration is occurring at Pol I-transcribed genes, also taking into account previous studies and sequencing data, as well as including further possible explanations in addition to the proposed instability of Ty1 insertion at Pol I-transcribed genes. Furthermore in this context, referee #3's point regarding integration at Pol II genes should also be addressed (ref#3- major concern). In addition, you should reconsider the nomenclature of NLS/bNLS as referee #1 suggests, as well as experimentally addressing the issues s/he mentions regarding Fig. 3A (point 9). If you are able to fully resolve these key issues, as well as adequately responding to all of the other points raised by the referees, then we would be happy to consider the study further for publication. Therefore I would now like to invite you to prepare and submit a revised manuscript. Please note that it is EMBO Journal's policy to allow only a single round of major revision and that it is therefore important to clarify all key concerns raised at this stage.

Please feel free to contact me should you have any further questions regarding the revision. Thank you for the opportunity to consider your work for publication. I look forward to receiving your revised manuscript.

Kind regards,

Stefanie Boehm

Stefanie Boehm
Editor
The EMBO Journal

Referee #1:

Review of "The nuclear localization signal of Ty1 integrase targets retrotransposon integration to tRNA genes" by Asif-Laidin et al

This manuscript extends our understanding of Ty1 integration site selection from work previously published by the Lesage lab in Science 2015 where they show interaction between Ty1 integrase and the RNA pol III subunit AC40 is important for integration upstream of tRNA genes and when this interaction is impaired, Ty1 inserts into subtelomeric sites. In this new manuscript by the Lesage lab they map the interaction with AC40 to five amino acids at the C-terminus of integrase that are each required in two-hybrid and co-IP experiments for the interaction. Surprisingly, the interacting residues exist in between the two parts of a bipartite NLS. This is surprising in part because the authors show the NLS functions independently of the five residues needed to bind AC40. The five residues in integrase, which I will call the targeting domain (TD) were further characterized using results of ChIP-seq data mapping integrase binding genome-wide. This powerful set of data showed ectopically expressed integrase uniquely binds the start sites of tRNA and other pol III transcribed genes, a close match to the integration pattern. Further, this binding pattern depends on the TD residues. What is particularly interesting is integrase also occupied the genes of the rDNA repeats which are known to be transcribed by pol I. This finding addresses a controversy remaining from the 2015 paper. Which is their statement that Ty1 does not integrate into genes transcribed by pol I even though AC40 is associated with pol I. The ChIP-seq mapping of integrase at the rDNA suggests that Ty1 may indeed integrate there. In an elegant demonstration of specificity, the authors use high throughput mapping of insertion sites to show that the TD of Ty1 is sufficient to redirect integration of Ty5 from heterochromatin to instead integrate at pol III transcribed genes. This is perhaps the strongest example yet describing a targeting mechanism because it demonstrates how a small sequence of amino acids is fully sufficient to direct integration to select sites of the genome.

Overall the quality of the data is very high, and the writing is clear and organized. I have two suggestions/concerns that would improve the paper. The authors have the data to resolve the controversy regarding integration in rDNA repeats. Having found that integrase readily binds the rDNA the authors should use their high throughput integration data to measure what fraction of integration occurs there. The authors make conflicting statements about integration in rDNA repeats in the manuscript that I point out below. My second concern is that the authors overemphasize the connection between the TD and the NLS. Having shown that they function independently the go on to use the label bNLS to describe the targeting domain. This will mislead the readers to think there is a connection. Perhaps there is, but its not shown in this work. As a result, I would rename the domain TD instead of bNLS in the title and throughout. Also, the proposed connection in Fig. 6 between NLS function and AC40 binding is too premature to include in a figure.

Comments and concerns

1. The title is inconsistent with the results. You show the NLS and the targeting domain function independently. The only connection is the proximity of the NLS and the AC40 binding residues.
2. The term bNLS is used for the 57 amino acid domain with the interaction with AC40. This region is more than just the bipartite nuclear localization signal. I recommend another name is used such as TD for targeting domain. This reflects the point #1, that the NLS and the TD operate independently. This name leads to problems as in line 350 which suggests the NLS is responsible for targeting. And on line 435 which says the NLS is involved in integration targeting. This is

misleading.
3. Line 46, Huang et al is not the correct reference.
4. Line 60, It would be more accurate to say upstream of transfer RNA genes.
5. Line 99, typo, domain
6. Line 132 and Fig 1B, To conclude the alanine mutations disrupt interaction with AC40 one would need to show by western that all the mutants are expressed comparable to wt IN1
7. Line 140, please describe the reason for the EPEA residues added to IN1.
8. Line 142, typo, conservation
9. Line 187 and Fig. 3A, this figure shows the CD mutations disrupts the interaction with one pol III subunit C160. But Cheung et al previously showed other subunits (Rpc31, Rpc34, and Rpc53) interact directly with integrase. Its therefore important to probe for some of these subunits in the pulldown to determine whether their interaction depends on AC40.
10. Line 209, this statement is inconsistent with Fig. 3H showing HA-IN1 binds to the rDNA repeats.
11. Line 230, It would be helpful to show or indicate results of pol III ChIP-seq in Fig. 3E. This would show how closely IN1 matches binding of pol III. This is important because Ty1 integrates upstream of the tRNA genes.
12. Line242 and Fig. 3H, here too it is important to know how well this distribution of IN1 matches the ChIPseq pattern of pol III on the rDNA repeats. This would greatly support your statement on line 249, that AC40-IN1 interaction is necessary for interacting with pol III.
13. Line 255, the names of the supplementary tables in the text don't match the names of the files. Legends would be very helpful to explain the content of the excel files. Specifically, its important to say whether the IN1 mutants increased occupancy at the subtelomere repeats and at SEO1 where integration increases with the IN1 muts.
14. Line 274, it would be helpful to say what the Z-score is and what it measures. To gage the power of this statistic please provide the total number of insertions in each library.
15. Line 275, Please comment on the integration in the rDNA repeats. Your Z score suggest that there is no integration in rDNA above random. This is in contrast to the binding of IN1 to rDNA and to previous papers that detected integration of Ty1 in rDNA. This Z score is particularly confusing because in line 359 you say the muts in IN1 caused a decrease in integration in Pol I loci.
16. Fig. 4D, can you explain the high integration of WT Ty1 at the ch II-L sub-telomere region. Is there a pol III gene there?
17. Line 345, The difference in banding patterns is difficult to appreciate. It would be good to show a plot of this as in Fig. 4C.
18. Line 359, there is no information in the paper saying the IN1 muts reduce integration at the Pol I loci (rDNA?).
19. Line 375, the integration in rDNA was not so rare in Bryk et al. About 7 out of 75. But the point is you have to data to say what amount of integration occurs in the rDNAs and describe what happens in the mutants.
20. Line 388, this statement is too strong. Retargeting has been seen for HIV-1 when the targeting domain (IBD) was fused to other chromatin proteins causing integration to occur at novel loci (Ferris et al, 2009 PNAS, PMID: 20133638), and for Tf1 when a chromodomain was added to the C-terminus of IN causing integration to occur at heterochromatin (Gao et al, 2008 Genome Research, PMID: 18256242)
21. Line 442, These references should be explained in context of your model.

Referee #2:

Lesage and colleagues report here an extensive study of the interaction between the Ty1

integrase protein, In1, and the RNA Polymerase III subunit, AC40, that dictates preferential integration of Ty1 near tRNA genes. Specifically, they demonstrate the critical role in Ty1 targeting of several amino acids in the C-terminal region of IN1 that reside between the bipartite moieties of the IN1 nuclear localization signal.
The work is thorough and rigorous. My initial impression on reading the abstract was that the work might, however, be of limited general interest. Given that the authors and Cheung et al. had demonstrated convincingly that interactions between Pol III and IN1 are critical for targeting Ty1 to Pol III transcribed genes, it seemed that narrowing this interaction to a specific region of IN1 might be more appropriate for a specialty journal. However, the authors report several other interesting and novel observations that raise the general interest of the work, in my opinion. They perform ChIP-seq of mutant and wild type versions of IN1 and compare this to the integration profiles of Ty1 expressing the same mutant (and wild type) versions of IN1, and also show that Ty1 integrates at Pol I transcribed genes and that this also depends on the same interactions between AC40 (which is also a Pol I subunit) and IN1 as for Pol III targeted integration. I would accordingly recommend that the authors revise their abstract to do better justice to the findings they report. (Even the keywords do not give any indication that genome-wide experiments have been done to link IN1 association to integration sites.)
One important point is that the sequence deposition numbers appear to be incorrect. The ChIP-seq deposition at ArrayExpress cited contains data from work on interaction between Mediator and TFIIB, while a search for the SRA number returned no found items. The authors need to provide correct information for this, and should state the number of replicate experiments performed. Although standards are beginning to dictate that replicates be performed for all ChIP-seq experiments, I would not demand that in this instance (if it has not been done), as the large amount of corroborative data supports the conclusions derived from ChIP-seq. Similarly, although technically it is best to swap bait and prey in 2-hybrid experiments, I do not think that is necessary here given the amount of follow-up, independent experiments that support results of 2-hybrid experiments.

Other minor points:

Line 142, "conserved" should be "conservation"

Lines 175-6: "mutants that prevent the interaction with AC40"; this is too strong. In fact, at later points in the manuscript the authors recognize that expression level is likely to affect the outcome of the various experiments used to assess interactions (lines 230 ff). The authors should temporize conclusions regarding interactions at each relevant point (e.g. they should recognize that the expression levels in the 2-hybrid experiment are much higher than physiological).

The association of IN1relative to tDNAs appears possibly asymmetric (Fig. 3E). Is this correct? It could be that there is some association with the transcribing Pol III that gives a low level of "downstream" association, but that the predominant association occurs when Pol III is recruited but has not begun to transcribe. Some comment, perhaps?

Line 951, "biniding"

Lines 467-68 says SC-LEU-TRP for both growth and interaction; should be SC-LEU-TRP-HIS for interaction.

Line 562 ff: solubilized chromatin was recovered as the supernatant, correct? It reads as though the pellet might have been used.

Line 582: I believe this should read "Reads per Kilobase per Millions of Mapped reads".

In Figure 3C and D, it is unclear how the enrichment was calculated, and what region was used in the calculation-was it just the transcribed portion for Pol III-transcribed genes or a larger region? For Pol II genes, was it a specific upstream region, or the ORF? Was the data normalized for ORF length if ORFs were used? This should be explained in the Methods and/or the figure legend.

Referee #3:

Summary
In this work, Asif-Laidin et al. provide convincing evidence that the IN1 targeting determinant recognized by Pol III AC40 is located within the bipartite IN1 NLS. Importantly, the IN1 tethering determinant is necessary and sufficient to mediate preferential integration. The authors bring together other Pol III proteins that interact with IN1 as well as the recently reported NPC involvement in targeting into a coherent testable model for how Ty1 preferentially integrates. The default integrations at subtelomeric regions are also interesting as other features of chromatin organization may be revealed by further study. However, the author's should determine whether insertions at Pol II genes still occur when tethering to AC40 is defective. Lastly, the authors speculate that the AC40/Ty1 interaction could be incorporated into novel retroviral vectors that may minimize deleterious off-targeting in gene therapy applications. Although a long-shot as AC40sp cannot replace AC40sc for Ty1 targeting, it is worth testing.

Major concern
Line 300. It is important to know if an IN1 tethering mutant alters the Ty1 fraction and promoter bias among mutants at well studied target genes such as CAN1. This simple straightforward experiment based on analyzing Ty1-induced canavanine resistant mutants would directly address insertions affecting a target gene transcribed by RNA Pol II.

Minor concerns
Line 70. To continue to unify the transposon and retrovirus integration fields, please use intasome instead of PIC throughout.

Line 80. Please specify that Ty1 is a predominant retroelement in certain strains such as the S288c reference strain. Little is known about Ty1 copy number dynamics in other human associated strains or natural isolates.

Line 97. Typo, replace "to" with "in".

Line 145-147. Soften, use "strongly suggest". Although unlikely, E. coli may contain a bridging protein. The most convincing evidence is an in vitro interaction with purified components. Alternative is proteomic analysis of pull downs.

Line 167. Incorrect citation. A more informative reference is Curcio and Garfinkel Mol Cell Biol 1992.

Line 215. The authors claim that cDNA is not necessary for IN1 recruitment at Pol III-transcribed genes. However, they do not determine relative cDNA levels in their system at 30{degree sign}C. Soften interpretation as growth at 30d decreases but does not eliminate transposition, and the extent of temperature sensitivity depends on whether Ty1 is overexpressed or not (see Lawler et

al. JVirol 2002 and Garfinkel et al. Genetics 2005).

Line 224: conflicts with Figure 3 and Figure EV2 and legends: Pearson or Spearman correlation?

Figure 3C: label y-axis as fold enrichment. Unclear what the dashed line indicates.

Figure 3E legend: state that Pearson correlations are with WT.

First of all, we warmly thank the reviewers for their time and efforts to assess the work reported in our manuscript. We are most grateful for their appreciation that our study is important and of high quality. Their comments were very helpful in guiding us in the revision and improvement of the manuscript. Below is a point-by-point response to their comments. For clarity, all of our responses are highlighted in blue and changes are indicated in red in the manuscript.

Referee #1:

Review of "The nuclear localization signal of Ty1 integrase targets retrotransposon integration to tRNA genes" by Asif-Laidin et al

This manuscript extends our understanding of Ty1 integration site selection from work previously published by the Lesage lab in Science 2015 where they show interaction between Ty1 integrase and the RNA pol III subunit AC40 is important for integration upstream of tRNA genes and when this interaction is impaired, Ty1 inserts into subtelomeric sites. In this new manuscript by the Lesage lab they map the interaction with AC40 to five amino acids at the C-terminus of integrase that are each required in two-hybrid and co-IP experiments for the interaction. Surprisingly, the interacting residues exist in between the two parts of a bipartite NLS. This is surprising in part because the authors show the NLS functions independently of the five residues needed to bind AC40. The five residues in integrase, which I will call the targeting domain (TD) were further characterized using results of ChIP-seq data mapping integrase binding genome-wide. This powerful set of data showed ectopically expressed integrase uniquely binds the start sites of tRNA and other pol III transcribed genes, a close match to the integration pattern. Further, this binding pattern depends on the TD residues. What is particularly interesting is integrase also occupied the genes of the rDNA repeats which are known to be transcribed by pol I. This finding addresses a controversy remaining from the 2015 paper. Which is their statement that Ty1 does not integrate into genes transcribed by pol I even though AC40 is associated with pol I. The ChIP-seq mapping of integrase at the rDNA suggests that Ty1 may indeed integrate there. In an elegant demonstration of specificity, the authors use high throughput mapping of insertion sites to show that the TD of Ty1 is sufficient to redirect integration of Ty5 from heterochromatin to instead integrate at pol III transcribed genes. This is perhaps the strongest example yet describing a targeting mechanism because it demonstrates how a small sequence of amino acids is fully sufficient to direct integration to select sites of the genome.

Overall the quality of the data is very high, and the writing is clear and organized. I have two suggestions/concerns that would improve the paper. The authors have the data to resolve the controversy regarding integration in rDNA repeats. Having found that integrase readily binds the rDNA the authors should use their high throughput integration data to measure what fraction of integration occurs there. The authors make conflicting statements about integration in rDNA repeats in the manuscript that I point out below. My second concern is that the authors overemphasize the connection between the TD and the NLS. Having shown that they function independently the go on to use the label bNLS to describe the targeting domain. This

will mislead the readers to think there is a connection. Perhaps there is, but its not shown in this work. As a result, I would rename the domain TD instead of bNLS in the title and throughout. Also, the proposed connection in Fig. 6 between NLS function and AC40 binding is too premature to include in a figure.

The main concerns of Reviewer # 1 are as follows:
- Ty1 integration into rDNA repeats,
- The distinction between the Ty1 targeting domain and the bipartite NLS,
- The relevance of Figure 6.
These concerns are addressed below in the point-by-point responses.

Comments and concerns

1. The title is inconsistent with the results. You show the NLS and the targeting domain function independently. The only connection is the proximity of the NLS and the AC40 binding residues.

To more accurately reflect the message of the article, we have changed the title to "**A small targeting domain in Ty1 integrase is sufficient to direct retrotransposon integration to tRNA genes**".

2. The term bNLS is used for the 57 amino acid domain with the interaction with AC40. This region is more than just the bipartite nuclear localization signal. I recommend another name is used such as TD for targeting domain. This reflects the point #1, that the NLS and the TD operate independently. This name leads to problems as in line 350 which suggests the NLS is responsible for targeting. And on line 435 which says the NLS is involved in integration targeting. This is misleading.

We thank the reviewer for pointing out this lack of clarity. We recognize that using the term bNLS to describe the region of interaction with AC40 could be misleading as we provide evidence that the two functions (Ty1 targeting and IN1 nuclear import) are dissociated despite the partial overlapping of their sequences. We have therefore reworded the title and the text and named the sequence of interaction with AC40 the Ty1 Targeting Domain of IN1 (TD1). Changes have been made in the manuscript (**13 changes**).

3. Line 46, Huang et al is not the correct reference.

Indeed, this was a mistake that we have corrected. The exact reference is Huang CR, Burns KH and Boeke JD, Annu Rev Genet 2012, 46:651-75.

4. Line 60, It would be more accurate to say upstream of transfer RNA genes.

Preferential integration has been described for retrotransposons upstream of *tRNA* genes in *S. cerevisiae* and upstream and downstream of tRNA genes in *D. discoideum* (described in Spaller *et al*. Mobile DNA, 2016). We have modified the text to provide this information and we have added the reference of Spaller *et al*. (**Lines 57-58 and 60).**

5. Line 99, typo, domain

Corrected (**Line 96**).

6. Line 132 and Fig 1B, To conclude the alanine mutations disrupt interaction with AC40 one would need to show by western that all the mutants are expressed comparable to wt IN1.

A western-blot is shown in **Figure EV1A** and indicates that all GAD-IN1 mutants are expressed at comparable levels to WT GAD-IN1. Reference to Figure EV1A was erroneously omitted. This has been corrected in the text (**Line 128**). References to **Figures EV1B-D** have been also included in the text (**Lines 132 and 137-138**).

7. Line 140, please describe the reason for the EPEA residues added to IN1.

EPEA (glutamic acid-proline-glutamic acid-alanine) is a four amino acid peptide tag also known as C-tag, which is recognized by a commercial antibody. To simplify, we have renamed the construct IN1-C-tag and explained what is C-tag in Materials and Methods (**Line 564**). Changes have been made in the text (**6 changes**) and in **Figure 1D**.

8. Line 142, typo, conservation

Corrected (**Line 143**).

9. Line 187 and Fig. 3A, this figure shows the CD (TD1) mutations disrupts the interaction with one pol III subunit C160. But Cheung et al previously showed other subunits (Rpc31, Rpc34, and Rpc53) interact directly with integrase. Its therefore important to probe for some of these subunits in the pulldown to determine whether their interaction depends on AC40.

We thank the reviewer for raising this interesting and important point. Pol III is a stable complex, and immunoprecipitation of a specific subunit results in the co-precipitation of the whole complex as described with HA-C160 (Oficjalska-Pham D *et al.* Mol cell, 2006) or Flag-C128 (Bhalla P *et al.* Gene, 2019). Thus, the IP of C160-HA shown in **Figure 3A** reveals an interaction between Pol III and IN1 and not between C160 and IN1.

In Cheung *et al.* (JBC, 2016), an interaction was detected *in vitro* between IN1 and the C31, C34 and C53 Pol III proteins purified in *E. coli* (i.e. Rpc31, Rpc34, and Rpc53). To ask whether these associations could be detected *in vivo* outside the Pol III complex, we have analyzed whether an interaction between IN1 and C34 or C53 could be detected in yeast cell extracts, using antibodies against C34 and C53 available in the lab. We have performed IN1 immunoprecipitation experiments in yeast cells expressing IN1-Strep (WT or K_{617}A mutant) and shown that both C53 and C34 could be co-immunoprecipitated with WT IN1 but not with IN1 TD1 mutant K_{617}A. Therefore, the association of C34 or C53 with IN1 is observed *in vivo* only when IN1 interacts with AC40. These novel data have now been included in the revised manuscript **(Lines 183-185, 190-196 and new Figures EV2A-B).**

10. Line 209, this statement is inconsistent with Fig. 3H showing HA-IN1 binds to the rDNA repeats.

We agree that the statement "We did not detect significant HA-IN1 binding at other chromosomal loci" is not at the right place in the text. The sentence has been deleted.

11. Line 230, It would be helpful to show or indicate results of pol III ChIP-seq in Fig. 3E. This would show how closely IN1 matches binding of pol III. This is important because Ty1 integrates upstream of the tRNA genes.

The reviewer is right, but the tRNA genes are too small to allow spatial resolution of the Pol III machinery localization by ChIP-PCR or ChIP-seq. Therefore, to show how well IN1 matches the Pol III binding, we have added a figure describing a quantitative ChIP PCR on *SCR1*, the longest Pol III-transcribed gene (522-bp). By analyzing ten real-time PCR amplicons distributed over the region, IN1 was detected all along the *SCR1* gene with maximum binding in the coding region but not upstream or downstream of the gene. A comparison of IN1 occupancy across *SCR1* with that of Pol III (published in Ghavi-Helm Y *et al.* Genes Dev. 2008) confirms the superposition of IN1 and Pol III binding.

These data have been added to the revised manuscript (**Lines 232-238 and new Fig 3G).**

12. Line 242 and Fig. 3H, here too it is important to know how well this distribution of IN1 matches the ChIP-seq pattern of Pol III on the rDNA repeats. This would greatly support your statement on line 249, that AC40-IN1 interaction is necessary for interacting with pol III.

All data concerning IN1 binding to *RDN1* have been put together in a new Figure (**Figure 4**). Figure 3H (**now Figure 4B**) shows the occupancy of HA-IN1 at the *RDN1* locus, which contains *RDN5* (transcribed by Pol III) and *RDN37* (transcribed by Pol I). WT HA-IN1 is distributed over both genes and its occupancy is decreased for the three IN1 TD mutants, particularly at the *RDN5* repeats. These data support our conclusion that the interaction with AC40 is important for IN1 binding to Pol I and Pol III-transcribed genes.

However, the Chip-seq data (**Figure 4B**) results in a poor resolution of the rDNA locus region. Thus, we have now included a quantitative ChIP-PCR of different amplicons distributed over *RDN37* (18S and 25S regions and promoter), *RDN5* (5S region) and the adjacent non-transcribed regions (NTS1 and NTS2). The data are presented in **Figure 4C** and show an enrichment of WT HA-IN1 at both Pol I and Pol III-transcribed loci, but not at NTS1 and NTS2, confirming that the presence of IN1 correlates with that of the two polymerases. These new data also indicate that the occupancy of WT HA-IN1 is lower at *RDN5* than at other Pol III loci (compare **Figures 3B and 4C**), which is consistent with the lower binding observed for Pol III on *RDN5* (Moqtaderi et al. MCB 2004). Since IN1 occupancy is similar at *RDN37* and *RDN5*, our results suggest that the recruitment of IN1 at the *RDN1* might be less efficient than at other preferred regions located outside of the *RDN1*.

To compare the recruitment of IN1 WT or mutants at Pol I-transcribed genes, we have also performed ChIP-PCR analysis. A decrease in the binding of $K_{617}A$, $S_{621}A$ and $L_{622}A$ IN1 mutants to 18S rDNA was detected supporting again our conclusion that the interaction with AC40 plays an important role in IN1 binding to *RDN1* repeats (**Figure 4D**).

All these data have been described in the text (**Lines 255-264)**.

13. Line 255, the names of the supplementary tables in the text don't match the names of the files. Legends would be very helpful to explain the content of the excel files. Specifically, its important to say whether the IN1 mutants increased occupancy at the subtelomere repeats and at SEO1 where integration increases with the IN1 muts.

The reviewer is right and we apologize for the confusion caused by the incorrect annotation of the supplementary table files. This has been corrected and table titles have been provided (**Lines 1179-1184**).

We have not detected the binding of IN1 mutants to subtelomeric regions by ChIP-seq. This information was given in the discussion as well as our hypothesis that it could be due to the scattered dispersion of IN1 binding sites in subtelomeres (**lines 442-444)**. However, we agree that we should also indicate it in the result section and have therefore added a sentence (**Lines 265-267**).

14. Line 274, it would be helpful to say what the Z-score is and what it measures. To gage the power of this statistic please provide the total number of insertions in each library.

The description of what Z-score is and measures was indeed not explained. We have corrected the text and given the information in Material and Methods (**Lines 665-672**).

In **Figure 5B** (**referred in the text, Lines 292-293**), we compare Z-score values for four features (ORF, Pol I, up 1-kb Pol III and subtelomeres) for each insertion library. The total number of Ty1 insertions in each library is now provided in Materials and Methods (**Line 663**).

15. Line 275, Please comment on the integration in the rDNA repeats. Your Z score suggest that there is no integration in rDNA above random. This is in contrast to the binding of IN1 to rDNA and to previous papers that detected integration of Ty1 in rDNA. This Z score is particularly confusing because in line 359 you say the muts in IN1 caused a decrease in integration in Pol I loci.

We sincerely thank reviewer #1 for all her/his important questions regarding Ty1 integration in the genes transcribed by Pol I. We have re-analyzed our sequencing data and concluded that Ty1 integration is nearly random in Pol I-transcribed genes (see our complete response in points 18 and 19). These results actually consolidate the conclusions of the submitted article.

16. Fig. 4D, can you explain the high integration of WT Ty1 at the ch II-L sub-telomere region. Is there a pol III gene there?

The high level of WT Ty1-*HIS3* integration in this region is due to the presence of a *tRNA* gene (t*F(GAA)B* (chrII:36398..36488)). We have added an explanation for this high level in the legend to **Figure 5E** (**Lines 1076-1077**).

17. Line 345, The difference in banding patterns is difficult to appreciate. It would be good to

show a plot of this as in Fig. 4C.

Line 345 (now Line 388) referred to **Figure 6D**, which shows the result of a PCR assay to detect integration events of Ty1, Ty5 or $Ty5_{\Delta TD+bNLS}$ at specific loci. Figure 4C (**now Figure 5D**) was generated from the high-throughput sequencing data of Ty1-*HIS3* de novo insertion events. For Ty5 and $Ty5_{\Delta TD+bNLS}$, we did not perform high-throughput sequencing of integration events because the insertion frequency of $Ty5_{\Delta TD+bNLS}$*his3*AI was 1000-fold lower than that of Ty1*his3*AI, making difficult the recovery of a sufficient number of independent insertion events for sequencing. Therefore, we cannot provide the same type of plot.

To improve **Figure 6D** and better support our conclusion that the different banding patterns could reflect different affinity of Ty1 and $Ty5_{\Delta TD+bNLS}$ for nucleosomes, we have used the database of nucleosome position over the yeast genome (Broogard *et al.* Nature, 2012) to indicate the position of the nucleosomes at the different loci we have analyzed. Nucleosome positions are now indicated on the right of each panel.

18. Line 359, there is no information in the paper saying the IN1 muts reduce integration at the Pol I loci (rDNA?).

The reviewer is right. We have now addressed Ty1 integration at Pol I-transcribed genes and our conclusion is that Ty1 does not integrate preferentially in these genes, despite the IN1/Pol I interaction.

To analyze Ty1 integration at the *RDN1* repeats, we have used the sequencing reads mapping at multiple positions and detected Ty1-*HIS3 de novo* insertion events at both *RDN5* and *RDN37* (**Figure EV3A**). To determine if there is a Ty1 integration targeting preference for Pol I-transcribed genes, we have first estimated the number of *RDN1* copy transcribed within a cell based on two observations:
i) A yeast strain with only 20 *RDN1* copies is viable and displays no growth defect, suggesting that the level of Pol I transcription in WT may correspond to 20 actively transcribed copies (Ide et al Science 2010),
ii) Among the 150 -200 rDNA repeats present in the genome, about 30%–50%, i.e. approximately 60 copies are actively transcribed (Merz et al Genes and Dev, 2008).
Next, we have compared the percentage of integration in *RDN1* obtained in our libraries, with those of 1 $x10^5$ Ty1 random insertions generated *in silico* in artificial yeast genomes composed of 20 or 60 *RDN1* repeats.

To validate this approach, we have first analyzed Ty1 integration at *RDN5*. As shown in **Figure EV3A**, there is a clear enrichment of WT Ty1-*HIS3* insertions upstream of all Pol III-transcribed *RDN5* genes, which is less visible with the three TD1 mutants. When compared to the two control cases, our data (**Figure EV3B)** indicate a 7 to 20-fold enrichment of WT Ty1-*HIS3* insertions in a 1-kb window upstream of the *RDN5* locus, compared to the two situations where these insertions would be randomly distributed. In contrast, the percentage of insertions of Ty1 mutants decreased compared to WT but was still higher than random. We have also confirmed this effect by a qualitative PCR assay to detect Ty1-*HIS3* insertions upstream of the *RDN5* genes present in *RDN1* **(Figure 5C).** Therefore, with this approach, we reached the same conclusion that the Pol III-transcribed gene *RDN5* is a hotspot of Ty1 insertions, which depends on the interaction between IN1 TD1 and AC40.

We used the same approach to study integration at the *RDN37* locus transcribed by Pol I. Ty1 insertions are detected both by high-throughput sequencing (**Figure EV3A)** and qualitative PCR (**Figure EV3C**). However, the percentage of WT Ty1-*HIS3* insertion events at *RDN37* was close to the random situations, lower than at *RDN5* and not diminished by the $K_{617}A$, $S_{621}A$, or $L_{622}A$ mutations (**Figures 5C and EV2B**). These results indicate that Ty1 integration is not targeted to *RDN37* and that the recruitment of IN1 at Pol I-transcribed genes is not sufficient for subsequent Ty1 integration.

Ty1 integration at *RDN1* is now addressed in the manuscript in the result section (**Lines 300-323),** in the discussion (**Lines 428-437)** and shown in **Figures 5C, EV3A and EV3B**.

19. Line 375, the integration in rDNA was not so rare in Bryk et al. About 7 out of 75. But the point is you have to data to say what amount of integration occurs in the rDNAs and describe what happens in the mutants.

This is an important point. Please note that Bryk's article (Bryk *et al.* Genes and Dev 1997) describes five insertions of Ty1*his3*AI in the *RDN1*, of which three are unambiguously located upstream of *RDN5* (Pol III) and one at the end of *RDN37* (Pol I-transcribed). From their study it is impossible to estimate the frequency of Ty1 insertions in this specific region. Nevertheless, a useful observation is that in a *ubc2*Δ mutant, which derepresses specifically the expression of Ty1*his3*AI copies present in the *RDN1*, there is no global increase of Ty1 mRNA levels, suggesting that in laboratory strains there is little if any Ty1 copy in the *RDN1* repeats. Moreover, the authors did not detect mitotic instability of *de novo* Ty1 insertions in *RDN1* in WT cells, indicating that the low number of Ty1 insertions is probably not due to mitotic instability and actually support our data suggesting that Ty1 integration in Pol I-transcribed genes is disfavored. We have compared our data with Bryk's data and indicated that mitotic instability might not explain the low number of Ty1 insertions in *RDN37* **(Lines 430-437).**

20. Line 388, this statement is too strong. Retargeting has been seen for HIV-1 when the targeting domain (IBD) was fused to other chromatin proteins causing integration to occur at novel loci (Ferris et al, 2009 PNAS, PMID: 20133638), and for Tf1 when a chromodomain was added to the C-terminus of IN causing integration to occur at heterochromatin (Gao et al, 2008 Genome Research, PMID: 18256242)

The reviewer is right and we agree that retargeting has been already observed when the integrases of retroviruses or retrotransposons are fused to specific chromatin proteins or chromatin binding domains and we have added the information at the end of our discussion (**Lines 502-504**). Our point was that, to our knowledge, Ty1 is the only retrotransposon with two different targeting preferences: Pol III-transcribed genes and subtelomeres, the latter being favored when the interaction with AC40 is abolished. We have reworded the text to clarify this point (**lines 445-446**).

21. Line 442, These references should be explained in context of your model.

The references were used to describe the model presented in the former Figure 6.
This model describes a link between the nuclear import of Ty1 intasome and Ty1 integration at Pol III-transcribed genes. This is based on a hypothesis for which we provide no experimental data here. Thus, we agree with reviewer #1 that this model was too premature to be shown and have deleted it. In line with this, we have modified the discussion and deleted the references of Rothenbusch *et al.* (2012) and Passos *et al.* (2017).

Referee #2:

Lesage and colleagues report here an extensive study of the interaction between the Ty1 integrase protein, In1, and the RNA Polymerase III subunit, AC40, that dictates preferential integration of Ty1 near tRNA genes. Specifically, they demonstrate the critical role in Ty1 targeting of several amino acids in the C-terminal region of IN1 that reside between the bipartite moieties of the IN1 nuclear localization signal.
The work is thorough and rigorous. My initial impression on reading the abstract was that the work might, however, be of limited general interest. Given that the authors and Cheung et al. had demonstrated convincingly that interactions between Pol III and IN1 are critical for targeting Ty1 to Pol III transcribed genes, it seemed that narrowing this interaction to a specific region of IN1 might be more appropriate for a specialty journal. However, the authors report several other interesting and novel observations that raise the general interest of the work, in my opinion. They perform ChIP-seq of mutant and wild type versions of IN1 and compare this to the integration profiles of Ty1 expressing the same mutant (and wild type) versions of IN1, and also show that Ty1 integrates at Pol I transcribed genes and that this also depends on the same interactions between AC40 (which is also a Pol I subunit) and IN1 as for Pol III targeted integration. I would accordingly recommend that the authors revise their abstract to do better justice to the findings they report. (Even the keywords do not give any indication that genome-wide experiments have been done to link IN1 association to integration sites.)
One important point is that the sequence deposition numbers appear to be incorrect. The ChIP-seq deposition at ArrayExpress cited contains data from work on interaction between Mediator and TFIIB, while a search for the SRA number returned no found items. The authors need to provide correct information for this, and should state the number of replicate experiments performed. Although standards are beginning to dictate that replicates be performed for all ChIP-seq experiments, I would not demand that in this instance (if it has not been done), as the large amount of corroborative data supports the conclusions derived from ChIP-seq. Similarly, although technically it is best to swap bait and prey in 2-hybrid experiments, I do not think that is necessary here given the amount of follow-up, independent experiments that support results of 2-hybrid experiments.

We would like to warmly thank Reviewer #2 for her/his interest in our study and appreciate her/his concern about the abstract. We have modified it in the hope of making it more attractive, as recommended. We have also modified the keywords to highlight our genomic approaches.

We are very sorry for the mis-annotation of our sequence deposition. The correct annotations are now given in the text (**Lines 675-678).**

For ChIP-seq analyses, the immunopurified DNA from at least three independent cultures were validated independently by quantitative real-time PCR of *SCR1* and *GAL1* genes and then pooled for sequencing.

For genome-wide Ty1 insertion analyses, total genomic DNA was extracted from 70,000 to 100,000 His^+ colonies recovered from seven to ten independent cultures. For each condition (WT or mutant Ty1his3AI), all genomic DNA samples were validated independently by selective PCR (shown in **Figure 4A**) and were subsequently pooled together to prepare sequencing libraries.

We did not swap prey and bait in the two-hybrid experiments because we confirmed the role of specific residues of IN1 in the interaction with AC40 by complementary approaches, as pointed by the reviewer.

Other minor points:

Line 142, "conserved" should be "conservation"

Corrected (**Line 143**).

Lines 175-6: "mutants that prevent the interaction with AC40"; this is too strong. In fact, at later points in the manuscript the authors recognize that expression level is likely to affect the outcome of the various experiments used to assess interactions (lines 230 ff). The authors should temporize conclusions regarding interactions at each relevant point (e.g. they should recognize that the expression levels in the 2-hybrid experiment are much higher than physiological).

We agree with the reviewer and have changed **prevent** by **strongly affect** in the text (**Line 176**).

In the two-hybrid and coIP experiments, IN1 constructs are ectopically expressed form strong promoters (*ADH1* and *GAL1* respectively) on a two-micron plasmid, while in the ChIP experiments they are expressed from a weaker promoter (*CYC1* tetracycline-off promoter) on a centromeric plasmid. Thus, it is true that IN1 proteins are expressed at higher levels in two-hybrid and coIP experiments. Therefore, we have also indicated in the text that the K_{617}, S_{621} or L_{622} Ala substitutions may have a residual level of interaction with AC40 that may not be detected under conditions where IN1 proteins are highly expressed i.e. in two-hybrid or coIP assay (**Lines 240-241).**

The association of IN1 relative to tDNAs appears possibly asymmetric (Fig. 3E). Is this correct? It could be that there is some association with the transcribing Pol III that gives a low level of "downstream" association, but that the predominant association occurs when Pol III is recruited but has not begun to transcribe. Some comment, perhaps?

We thank the reviewer for raising this interesting point, which is related to point 11 of reviewer #1. A slight asymmetry for the upstream region of Pol III transcribed genes was recently described in the recruitment of the integrase of Ty3 (Patterson *et al*., Genome Res, 2019). However, the tiling array of *SCR1* shown now in **Figure 3G**, did not show an

asymmetry of IN1 occupancy towards the 5' upstream region, where Ty1 integrates. Structural studies of IN1/Pol III interaction will be necessary to determine whether a specific form of Pol III interacts with IN1.

Line 951, "biniding"

Corrected (**Line 1054**).

Lines 467-68 says SC-LEU-TRP for both growth and interaction; should be SC-LEU-TRP-HIS for interaction.

Corrected (**Line 525**).

Line 562 ff: solubilized chromatin was recovered as the supernatant, correct? It reads as though the pellet might have been used.

The description of the chromatin preparation step was indeed not sufficiently clear. We have corrected the text **(Line 623-629).**

Line 582: I believe this should read "Reads per Kilobase per Millions of Mapped reads".

This is correct. The text has been modified (**Line 646**).

In Figure 3C and D **(now Figures 3D and 3F)**, it is unclear how the enrichment was calculated, and what region was used in the calculation-was it just the transcribed portion for Pol III-transcribed genes or a larger region? For Pol II genes, was it a specific upstream region, or the ORF? Was the data normalized for ORF length if ORFs were used? This should be explained in the Methods and/or the figure legend.

We thank the reviewer for noting this lack of clarity. In **Figures 3D and 3F**, the enrichment was calculated using the RPKM formula (i.e. number of mapping Reads /(Region Length/1000 * total Number Reads/1,000,000)) and shown in Log2. For ORFs and RNA Pol III-transcribed genes, the regions used for calculation are gene bodies, since the peaks in ChIP-seq data have been detected mainly on gene bodies. This is now explained in the legend of panels 3D and 3F (**Lines 1028 and 1033**).

Referee #3:

Summary
In this work, Asif-Laidin et al. provide convincing evidence that the IN1 targeting determinant recognized by Pol III AC40 is located within the bipartite IN1 NLS. Importantly, the IN1 tethering determinant is necessary and sufficient to mediate preferential integration. The authors bring together other Pol III proteins that interact with IN1 as well as the recently reported NPC involvement in targeting into a coherent testable model for how Ty1 preferentially integrates. The default integrations at subtelomeric regions are also interesting as other features of chromatin organization may be revealed by further study. However, the

author's should determine whether insertions at Pol II genes still occur when tethering to AC40 is defective. Lastly, the authors speculate that the AC40/Ty1 interaction could be incorporated into novel retroviral vectors that may minimize deleterious off-targeting in gene therapy applications. Although a long-shot as AC40sp cannot replace AC40sc for Ty1 targeting, it is worth testing.

Major concern
Line 300. It is important to know if an IN1 tethering mutant alters the Ty1 fraction and promoter bias among mutants at well-studied target genes such as CAN1. This simple straightforward experiment based on analyzing Ty1-induced canavanine resistant mutants would directly address insertions affecting a target gene transcribed by RNA Pol II.

We thank this reviewer for raising this important point. We have analyzed WT and mutant Ty1 insertions at *CAN1 (YER063C)* in our high-throughput sequencing data. The gene is located on Chr V, adjacent to the left subtelomere and is represented in **Figure 5F** between YEL064C and YEL062W. As shown on **figure 5F**, there is no WT or mutant Ty1 insertions at this locus. The information has been added in the text (Lines **341-343**).

We also wanted to determine the insertion frequency of Ty1 (WT and $K_{617}A$) at the *CAN1* locus by selecting Canavanine resistant clones as proposed by the reviewer. We have transformed a *spt3Δ rad52Δ CAN1* strain by pGAL1Ty1-*his3*AI (WT or $K_{617}A$) and then induced Ty1 retrotransposition in the presence of galactose at 20°C. The experiment was initiated just before the containment of the French population and could not be completed. We can only answer that under these conditions the frequency of WT and mutant Ty1 insertions was as expected (about 10^3 His^+/cell) and that no canavanine resistant colony could be detected among 10^8 cells. These data have not been reproduced in independent experiments but seem to confirm the high-throughput sequencing data.

Minor concerns
Line 70. To continue to unify the transposon and retrovirus integration fields, please use intasome instead of PIC throughout.

We have changed PIC in intasome (**4 changes**).

Line 80. Please specify that Ty1 is a predominant retroelement in certain strains such as the S288c reference strain. Little is known about Ty1 copy number dynamics in other human associated strains or natural isolates.

We have added the information (**Lines 77-78**).

Line 97. Typo, replace "to" with "in".

Corrected (**Line 94**).

Line 145-147. Soften, use "strongly suggest". Although unlikely, E. coli may contain a bridging protein. The most convincing evidence is an in vitro interaction with purified components. Alternative is proteomic analysis of pull downs.

We agree that we cannot formally rule out the possibility that an *E. coli* protein may bridge AC40 and IN1, although this is highly unlikely as our data would suggest that the three IN1 mutations (K_{617}A, S_{621}A and L_{622}A) all disrupt this interaction.
We have softened our conclusion as suggested **(Line 145)**.

Line 167. Incorrect citation. A more informative reference is Curcio and Garfinkel Mol Cell Biol 1992.

Corrected (**Line 168**).

Line 215. The authors claim that cDNA is not necessary for IN1 recruitment at Pol III-transcribed genes. However, they do not determine relative cDNA levels in their system at 30°C. Soften interpretation as growth at 30d decreases but does not eliminate transposition, and the extent of temperature sensitivity depends on whether Ty1 is overexpressed or not (see Lawler et al. JVirol 2002 and Garfinkel et al. Genetics 2005).

The reviewer is right and we agree that the transposition levels of Ty1 (over-expressed or not) at 30°C are not negligible and that there might be some residual Ty1 cDNA in the cell under our experimental conditions. However, in the ChIP-seq experiments HA-IN is ectopically expressed and thus probably not present in the VLPs, where the cDNA is produced. Consequently, we have clarified our point and softened our conclusion. We think this information is now more appropriate to the discussion, so we have moved it in the revised version **(Lines 417-424)**.

Line 224: conflicts with Figure 3 and Figure EV2 and legends: Pearson or Spearman correlation?

The quantification is **Pearson** as indicated in the Figures. We have corrected the main text (**Line 227**).

Figure 3C: label y-axis as fold enrichment. Unclear what the dashed line indicates.

Y-axis has been corrected in “Fold enrichment”.
The dashed line indicates the separation between Pol II and Pol III-transcribed genes. The information has been added in the legend of Figure 3C **(Lines 1023-1024)**.

Figure 3E legend: state that Pearson correlations are with WT.

This has been stated in the figure legend **(Lines 1030-1031)**.

3rd Editorial Decision 29th May 2020

Dr. Pascale LESAGE
INSERM U944, CNRS UMR 7212, Genomes & Cell Biology of Disease Unit, Institut de Recherche Saint-Louis, Université de Paris, Hôpital Saint-Louis
1 avenue Claude Vellefaux
Paris, Paris 75010
France

29th May 2020

Re: EMBOJ-2019-104337R1
A small targeting domain in Ty1 integrase is sufficient to direct retrotransposon integration to tRNA genes

Dear Dr. Lesage,

Thank you for submitting your revised manuscript for our consideration. Please apologize the delay in communicating this decision to you, which was due to the high number of new submission we are currently receiving. We have now received the reports from the original referees (see comments below). I am pleased to say that all referees now support publication. Referee # 1 and referee #3 point out some issues that can be resolved in a final revised version. In this version, I would also ask you to please address a number of editorial issues that are listed in detail below. Please make any changes to the manuscript text in the attached document only using the "track changes" option. Once these remaining issues are resolved, we will be happy to formally accept the manuscript for publication.

Thank you again for giving us the chance to consider your manuscript for The EMBO Journal. I look forward to receiving your final revision. Please feel free to contact me if you have further questions regarding the revision or any of the specific points listed below.

Kind regards,

Stefanie Boehm

REFEREE REPORTS

Referee #1:

The authors are to be commended for the considerable effort taken to address my concerns regarding integration into the RNA pol I transcribed genes. They have also addressed my other concerns including altering the name originally given to the targeting domain. I am also happy with the considerable revisions made to the text. Except for some mistakes I found in the numbering of figures at least as described in the point by point rebuttal. (Figure EV3C is missing and Fig. EV2B is incorrect). I strongly recommend they check each reference to Figures in the text.

Referee #2:

The authors have satisfactorily addressed my comments.

Referee #3:

The revised version is improved and queries have been addressed satisfactorily. The authors should carefully proofread for grammatical errors.

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ➔

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: LESAGE Pascale and ACKER Joël
Journal Submitted to: The EMBO journal
Manuscript Number: EMBOj-2019-104337

Reporting Checklist For Life Sciences Articles (Rev. June 2017)

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 $\chi 2$ 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.

In the pink boxes below, please ensure that the answers to the following questions are reported in the manuscript itself. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable). We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

USEFUL LINKS FOR COMPLETING THIS FORM

http://www.antibodypedia.com
http://1degreebio.org
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repor
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-tumo
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)
1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	There was no pre-specified effect size in our analyses.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	NA
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	NA
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.	NA
For animal studies, include a statement about randomization even if no randomization was used.	NA
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.	NA
4.b. For animal studies, include a statement about blinding even if no blinding was done	NA
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.	For retrotransposition test figure 2B, we have used student t-test, we assume that our samples follow the normal distribution.
Is there an estimate of variation within each group of data?	no

Is the variance similar between the groups that are being statistically compared?	Figure 2B, the variance is assumed to be similar between the groups

C- Reagents

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	Anti-Strep from mouse (Qiagen, Cat n°34850) for western blots. Anti-HA (12CA5) from mouse (Roche, Cat n° 11583816001) for western blots. Anti-Streptavidin-HRP (Thermofisher, Cat n°21130). Anti C34 and anti C53 polyclonal antibodies home made from rabbit (Huet et al.; 1985). CaptureSelect™ Biotin Anti-C-tag conjugate (Thermofisher, Cat:7103252100). anti-Tap polyclonal antibody from rabbit (Thermofisher, Cat n°CAB1001) for western blots.
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	NA

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	NA
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	NA
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	NA

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	NA
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	NA
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	NA
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	NA
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	NA
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	NA
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	NA

F- Data Accessibility

18: Provide a "Data Availability" section at the end of the Materials & Methods, listing the accession codes for data generated in this study and deposited in a public database (e.g. RNA-Seq data: Gene Expression Omnibus GSE39462, Proteomics data: PRIDE PXD000208 etc.) Please refer to our author guidelines for 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	Data accession numberis provided in Materials and methods section ChIP-seq data are deposited to Array Express under accession number E-MTAB-9038. Ty1 de novo insertion data are deposited to Sequence Read Archive under accession number PRJNA597319
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	N/A
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	N/A
21. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	N/A

G- Dual use research of concern

22. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	NA