EMBO reports

Increased Alu RNA processing in Alzheimer's brains is linked to gene expression changes

Yubo Cheng, Luke Saville, Babita Gollen, Ana Alvarez Veronesi, Majid Mohajerani, Jeffrey Joseph, and Athanasios Zovoilis

DOI: 10.15252/embr.20 2052255

Corresponding author(s): Athanasios Zovoilis (athanasios.zovoilis@uleth.ca)

Review Timeline:

Transfer from Review Commons:	10th Dec 20
Editorial Decision:	19th Jan 21
Revision Received:	23rd Jan 21
Accepted:	1st Feb 21

Editor: Esther Schnapp

Review COMMONS

Transaction Report: This manuscript was transferred to EMBO reports following peer review at Review Commons

(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.)

Review #1

1. How much time do you estimate the authors will need to complete the suggested revisions:

Estimated time to Complete Revisions (Required)

(Decision Recommendation)

Between 3 and 6 months

2. Evidence, reproducibility and clarity:

Evidence, reproducibility and clarity (Required)

Summary:

The work reports about the finding of a correlation between SINEB2 lncRNA and AD. These observations extend previous and partially unpublished work from the same lab about the role of mouse Alu RNA in Pol II regulation. Thorough transcriptiomics analysis is presented in support of global SINEB2 processing in post-mortem brain tissues from a cohort of AD patients. These findings are novel and may have implications for the comprehension and treatment of this multifactorial disease.

Major comments:

The main message of the work is that brain specimen from AD patients show a significant correlation between the SINEB2 processing and the disease. Although the work does not contain any direct mechanistic analysis that would prove a link with the disease, it sheds light on a potentially key, previously uncharacterized molecular aspect of the pathology. Indeed, aberrant SINEB2 processing would lead to global defect in RNA Pol II regulation, thus a major cause of loss of transcriptome homeostasis, perhaps contributing either to the onset or the progression of the disease.

-Should the authors qualify some of their claims as preliminary or speculative, or remove them altogether?

AD remains a multifactorial, complex disease. The discovery of a new pathway may open a novel line of research towards a potential therapeutic approach. In this context, based on the current analysis, and previous seminal work, authors claims appear fairly balanced.

-Would additional experiments be essential to support the claims of the paper?

Cost/Timewise (3-5months ?) Pol II Chipseq or TagChIP may be a fair, feasible request as this would substantiate the predicted defect and its possible relationship with deregulation of AD relevant genes.

-Are the data and the methods presented in such a way that they can be reproduced?

Experiments are well presented and methods sufficiently described. However, based also on our experience, I suspect that several bioinformatics "tricks" will need direct inquiries to the authors.

Minor comments:

-Authors should improve the manuscript by discussing and quoting current state of the art in the field of transposable elements and ncRNA. Surprisingly this important, critical aspect element is missing in the

conceptual framework of the manuscript.

-Overall text and figures are clear and accurate.

3. Significance:

Significance (Required)

The work is an important contribution to the largely unexplored functional role of the repeat part of the genome and in particular its lncRNA component. The evidence of a strong correlation between SINEB2 processing, deregulated transcription and a high impact multifactorial disease (AD) will be of interest to a broad readership spanning from fundamental biology of retrotransposons, to lncRNA and mechanisms of gene regulation. The implications for potential applications may be relevant also for a medical audience.

Our lab is interested in mechanisms of epigenome regulation, in particular the interplay between ncRNA, chromatin remodeling and transcription. These include also retrotransposons and disease models.

Review #2

1. How much time do you estimate the authors will need to complete the suggested revisions:

Estimated time to Complete Revisions (Required)

(Decision Recommendation)

Between 1 and 3 months

2. Evidence, reproducibility and clarity:

Evidence, reproducibility and clarity (Required)

In the manuscript entitled 'Transcriptome-wide deregulation in processing of SINE Alu RNAs in Alzheimer's disease reveals a novel connection of SINE RNAs with brain molecular pathology' by Cheng et al, the authors employ sRNA-seq and polyA-RNAseq data derived from human AD and non AD brain in order to profile changes in Alu RNAs induced by AD pathology. The role of Alu elements in physiological brain function and neurodegeneration has only recently started receiving attention and novel insights suggest a widespread impact of Alu RNAs over gene expression, especially under pathological conditions. Along these lines, A-to-I editing levels were recently found reduced in the brain of AD patients, further indicating a possible link between Alu-mediated transcriptional regulation and pathology.

Hence, the hypothesis that Alu elements may contribute to AD pathology is intriguing and further studies are required to explore the mechanisms underlying this putative crosstalk. However, the work reported here for the most part remains observational and the authors often (mistakenly) overestimate and misinterpret correlation as causation. The direct (in vitro) functional evidence provided in the manuscript is not adequately controlled and the derived conclusions are therefore not accurate. While the reported datasets are of great value per se, the authors will need to tone down their interpretation.

More specifically:

-All graphs presented as boxplots should be replaced by dot plots and p-values (or q-values where applicable) should always be indicated.

-The authors should discuss the limitations of analyzing RNAseq data from bulk tissue preparations and/or include in silico deconvolution approaches to infer cell type proportion in the sample cohorts used.

-Related to the previous point: The correlation between p53 levels and Alu processing is not informative as it is now discussed in the manuscript. Would the authors expect a change in p53 levels as a function of microgliosis/astrogliosis/neuronal death observed in the AD samples? In that sense, how can the authors prove (or strengthen) the specificity of this observation? The authors should discuss.

-Are the trends observed in Fig. 5A & B significant?

-In Fig. 5C, there seems to be a biphasic distribution of the samples according to ApoE genotype, which does not seem to correlate with the distribution of the differentially processed Alu RNAs. The authors should explain.

-The 'no protein' control is not a valid one for the in vitro experiments presented in Fig. 7, hence these data cannot be appropriately interpreted unless these experiments are performed again including the required controls. Also, which was the internal normalizer used in this case?

-In the MAP cohort, the authors found no correlation between HSF1 levels and full length Alu RNAs. However, the in vitro results seem to contradict this observation. The authors should explain the discrepancy.

-Can any conclusions be made with regard to Pol III-transcribed Alu RNAs from the datasets obtained here?

-Non peer-reviewed preprints cited in the text should be clearly cited as such.

-There are some typos in the figures that need to be corrected.

3. Significance:

Significance (Required)

The role of Alu elements in physiological brain function and neurodegeneration has only recently started receiving attention and novel insights suggest a widespread impact of Alu RNAs over gene expression, especially under pathological conditions. Along these lines, A-to-I editing levels were recently found reduced in the brain of AD patients, further indicating a possible link between Alu-mediated transcriptional regulation and pathology.

Hence, the hypothesis that Alu elements may contribute to AD pathology is intriguing and further studies are required to explore the mechanisms underlying this putative crosstalk. However, the work reported here for the most part remains observational and the authors often (mistakenly) overestimate and misinterpret correlation as causation. The direct (in vitro) functional evidence provided in the manuscript is not adequately controlled and the derived conclusions are therefore not accurate. While the reported datasets are of great value per se, the authors will need to tone down their interpretation.

Referees cross-commenting

Regarding Reviewer #1's comments:

Pol II Chipseq or TagChIP may not be absolutely required.

Regarding Reviewer #3's comments:

Agree with the majority of the remarks.
Addressing point #8 would be fundamental to the main conceptual message of the manuscript if indeed prior literature cannot support the link.

Overall:

The authors would primarily need to better clarify/discuss/interpret their data.
Small experimental additions will benefit the manuscript.

Review #3

1. How much time do you estimate the authors will need to complete the suggested revisions:

Estimated time to Complete Revisions (Required)

(Decision Recommendation)

Between 1 and 3 months

2. Evidence, reproducibility and clarity:

Evidence, reproducibility and clarity (Required)

The authors present smal- and total-RNA seq data indicating increased cleavage of Alu elements in the brain of AD patients. Using the sequencing data they identified the cleavage sites within Alu elements and proposed a model whereby the precessing ratios of Alus regulates gene expression. The study is interesting; however, most of the presented data are correlative, which makes the conclusions/model highly speculative and not strongly supported by the data.

Major comments:

1)In the Methods section, the authors utilized Qiagen MinElute Cleanup kit to clean phosphorylated RNA. The column of this kit binds only RNAs longer than 200 nt, while shorter RNAs go to the flow through. Which fraction did the authors use? and If the size cut off value of the column is 200 nt, how did the authors combine full length Alus (300 nt) Alu with degradation fractions (less than 200 nt) in one pool for sequencing?

2)To calculate Alu processing ratio, the authors calculated the read counts in the TSS as full-length Alu. In fact, degradation products can also generate reads at the TSS: if an Alu is cleaved at XR1 site, the 5' end of the 3' cleavage product will generate reads at the XR1 site, while the 5' end of the 5' cleavage product will generate reads at the TSS region. How did the authors exclude the possibility that the 5' cleavage product might be stable enough to generate sequencing reads?

3)The authors mentioned that post-mortem delay was comparable between AD and no AD samples. Given that this is difficult to control, the RIN values (average and SD) of isolated RNA need to be mentioned for both groups to ensure comparable RNA integrity prior to library preparation.

4)The authors explained higher levels of Alus fragments in AD patients as higher processing and destabilization of Alu RNAs. They overlooked the possibility that processing rate might be the same and the observed results is due to pathologic accumulation of the degradation products due to, e.g., AD-associated defect in some nucleases. This possibility need to be addressed or at least discussed.

5)If the authors are able to detect intact Alus (please refer to comment 2 above), they need to provide direct comparison between the level of intact Alus in AD and no AD groups. This will help understand better the relationship between intact Alu levels and Alu processing ratios.

5)In Fig. 2B and Fig. 4B, around 20% of Alus appear to have higher processing ratios in the no-AD group. Is there any thing specific to this group of Alus like specific sequences or motifs?

6)In Fig. 7A, it is highly recommended to increase the amount of RNA so as to be able to detect both full-length Alu as well as the cleave product of degradation intermediate. Does the size of the in-vitro cleavage product match the expected size upon cleavage at XL!, XR1, and/or XR2?

7)In Fig. 4C, D, if HSF1 increases the processing of Alus as shown in Fig 7C, it is expected to reduce the level of full-length Alus. However, it does not (Fig. 7D). Does this reinforce the notion that the increased processing ratio is due to accumulation of degradation products rather than enhanced processing (see point 4 above)? The authors need to discuss and explain this point better.

8)The model of Alu binding to Pol II is based on previous studies (mainly on mouse B2). The physical/direct binding between Pol II and Alu elements (intact and precessed) need to be experimentally shown through, e.g., RIP experiments.

Minor comments:

1)The detailed description of published results on how intact SINEs inhibit Pol II while processed SINEs activate Pol II is better to be moved from page 11 to the introduction to avoid any confusion.

2)Methods for short-RNA-seq (and how they are modified from previously published methods to include the 300-nt full-length Alus) need to be better explained.

3)The discussion lacks references in many parts.

3. Significance:

Significance (Required)

This study extends what we know about mouse B2 SINE to human Alus and utilizes brain samples from normal individuals as well as patients to generate the presented data.

I am an RNA biologist and have research lines on Alu elements, and can see that the manuscript (after addressing the comments) will be significant and of interest to RNA biologists, researchers in the field of retrotransposons, as well as researchers interested in AD.

Referees Cross-commenting

Reviewer #1 comments:

I agree that Pol II Chipseq will be a very good addition to the data set, especially when combined with the total-RNA seq data.

Reviewer #3:

I agree with most of the comments, and would like to emphasize the importance of the reviewer's comments that pertain to the correlation between p53 levels and Alu processing and the in-vitro experiments.

REVIEWER 1

Reviewer #1 (Evidence, reproducibility and clarity (Required)):

Summary:

The work reports about the finding of a correlation between SINEB2 lncRNA and AD. These observations extend previous and partially unpublished work from the same lab about the role of mouse Alu RNA in Pol II regulation. Through transcriptomics analysis is presented in support of global SINEB2 processing in post-mortem brain tissues from a cohort of AD patients. These findings are novel and may have implications for the comprehension and treatment of this multifactorial disease.

Major comments:

This reviewer generally remarks that "*The main message of the work is that brain specimen from AD patients show a significant correlation between the SINEB2 processing and the disease. Although the work does not contain any direct mechanistic analysis that would prove a link with the disease, it sheds light on*

a potentially key, previously uncharacterized molecular aspect of the pathology. Indeed, aberrant SINEB2 processing would lead to global defect in RNA Pol II regulation, thus a major cause of loss of transcriptome homeostasis, perhaps contributing either to the onset or the progression of the disease."

We appreciate the encouraging comments made by this reviewer.

-Should the authors qualify some of their claims as preliminary or speculative, or remove them altogether?

AD remains a multifactorial, complex disease. The discovery of a new pathway may open a novel line of research towards a potential therapeutic approach. In this context, based on the current analysis, and previous seminal work, authors claims appear fairly balanced.

-Would additional experiments be essential to support the claims of the paper?

<u>Point 1.</u> The reviewer proposes that "*Cost/Timewise (3-5months ?) Pol II Chipseq or TagChIP may be a fair, feasible request as this would substantiate the predicted defect and its possible relationship with deregulation of AD relevant genes.*"

This is an important point and we thank the reviewer for this comment. Indeed, the changes we report in expression levels of AD relevant genes correspond to final mRNA levels in the cell and not to the initial levels transcribed by Pol II. These initially transcribed levels may have been subject to a cascade of potential post transcriptional modifications and processing that may have affected stability and half-life. Thus, final mRNA levels may not correspond completely to the elongation activity of RNA Polymerase II, for which we are mostly interested in the current study. To this end, employing ChIP-seq would help to substantiate the observed transcriptional activation at the chromatin level. A significant limiting factor is availability and/or suitability for such assays of post-mortem tissue from the related biobanks and consortium studies since most of this material has already been approved or earmarked for specific assays. Luckily enough, the ROSMAP study has produced relevant ChIP-seq data for exactly the same patients used in our current study for the identification of the AD deregulated genes in the RNA-seq data. In particular, the related ChIP-seq data we have in hand for these patients include H3K9ac ChIP-seq data. Acetylation on histone H3 lysine 9 is a mark for the switch of RNA Pol II from transcription initiation to elongation and marks recruitment of the super elongation complex (SEC) to chromatin necessary for this stage. Thus, testing this chromatin mark corresponds very well to monitoring of RNA Pol II activity needed in our study. We now provide this data that show that chromatin state at AD upregulated genes is consistent with elevated Pol II elongation and the switch into a more activated Pol II mode. This data is now presented as new Suppl. Figure 5 and described in results section "Changes in Alu RNA expression and processing are associated with changes in gene expression".

-Are the data and the methods presented in such a way that they can be reproduced?

<u>Point 2.</u> The reviewer remarks that "*Experiments are well presented and methods sufficiently described. However, based also on our experience, I suspect that several bioinformatics "tricks" will need direct inquiries to the authors.*"

Upon the deposition of this data to EGA we will provide where applicable also source data of our analysis, including processed files with Alu fragment counts. MAP RNA-seq and H3K9ac data are also available through the AMP-AD Knowledge Portal.

Minor comments:

-Authors should improve the manuscript by discussing and quoting current state of the art in the field of transposable elements and ncRNA. Surprisingly this important, critical aspect element is missing in the conceptual framework of the manuscript.

We now have updated both the introduction and discussion sections with additional information on the topic *including information about transposable elements (page 4), increase in SINE ncRNA levels during cellular stress (page 5), retrotransposons/Alus and neurodegenerative diseases (page 17-18), including their potential role as A to I editing targets.*

-Overall text and figures are clear and accurate.

Reviewer #1 (Significance (Required)):

The work is an important contribution to the largely unexplored functional role of the repeat part of the genome and in particular its lncRNA component. The evidence of a strong correlation between SINEB2 processing, deregulated transcription and a high impact multifactorial disease (AD) will be of interest to a broad readership spanning from fundamental biology of retrotransposons, to lncRNA and mechanisms of gene regulation. The implications for potential applications may be relevant also for a medical audience.

Our lab is interested in mechanisms of epigenome regulation, in particular the interplay between ncRNA, chromatin remodeling and transcription. These include also retrotransposons and disease models.

REVIEWER 2

Reviewer #2 (Evidence, reproducibility and clarity (Required)):

This reviewer generally remarks that "*In the manuscript entitled 'Transcriptome-wide deregulation in processing of SINE Alu RNAs in Alzheimer's disease reveals a novel connection of SINE RNAs with brain*

molecular pathology' by Cheng et al, the authors employ sRNA-seq and polyA-RNAseq data derived from human AD and non AD brain in order to profile changes in Alu RNAs induced by AD pathology. The role of Alu elements in physiological brain function and neurodegeneration has only recently started receiving attention and novel insights suggest a widespread impact of Alu RNAs over gene expression, especially under pathological conditions. Along these lines, A-to-I editing levels were recently found reduced in the brain of AD patients, further indicating a possible link between Alu-mediated transcriptional regulation and pathology.

Hence, the hypothesis that Alu elements may contribute to AD pathology is intriguing and further studies are required to explore the mechanisms underlying this putative crosstalk. However, the work reported here for the most part remains observational and the authors often (mistakenly) overestimate and misinterpret correlation as causation. The direct (in vitro) functional evidence provided in the manuscript is not adequately controlled and the derived conclusions are therefore not accurate. While the reported datasets are of great value per se, the authors will need to tone down their interpretation."

As in other transcriptome studies in human, the primary aim of this study was to confirm the findings that we have already described in an AD-related mouse model (Cheng et al, elife 2020) regarding the increase of SINE RNA processing in human patients. To this end, providing a mechanistic insight of the control of gene expression by Alu RNAs in human neural cells, as it has been previously described in HeLa cells by the Kugel and Goodrich labs, has been beyond our aim. However, since the changes in gene expression in AD patients observed in our study were found to be in alignment with the general mechanism for regulation of gene expression by SINE RNAs suggested by the Kugel lab as well as our own recent findings both in mouse and human (Zovoilis et al, Cell 2016; Hernandez et al, PNAS 2020, Cheng et al, elife 2020) it is likely that many readers will make the same connection that this reviewer made regarding the potential causation between Alu RNAs and gene expression in human brain. Thus, the reviewer's concern that described correlation may be easily misinterpreted as causation is absolutely justified. In our initial submission, we had made a significant effort to prevent this and, in our discussion, we had explicitly stated that "it remains unclear which of the genes found to be upregulated in AD patients are direct targets of Alu RNAs". However, based on the reviewer's comment we see that the possibility of a misinterpretation of the results remains.

For the above reason, and since both this and the next reviewer's comments denote that there is a significant interest in providing some mechanistic data in our context, we decided to expand the focus of the current study and provide such data based on the reviewers' comments. To this end, we now provide functional genomics data from ChIP-seq experiments for the same patients, that provide a clearer picture regarding the Pol II elongation status of the identified AD up-regulated genes (new Suppl. Figure 5). Subsequently, we investigated the correlation co-efficient between expression of these genes and Alu RNA processing (new Suppl. Table 6, new Fig.7A) and we have now separated the AD upregulated genes into two categories: those strongly correlated with Alu RNA processing (new Suppl. Table 8) and those not correlated. By doing so, we provide a clearer picture regarding those AD up-regulated genes that may indeed be subject to regulation by Alu RNAs. Then, we designed LNAs against Alu RNAs, induced the down-regulation of Alu RNAs in a neuronal cell line frequently used in AD research (SHSY5Y cells) (new Fig.7B-C) and measured the expression of a representative set of potentially Alu RNA regulated genes and

the respective negative controls (new Fig.7D-E). In addition, in order to address the concern that "*The direct (in vitro) functional evidence provided in the manuscript is not adequately controlled*", we performed a series of RNA and protein control experiments to substantiate further the validity of our observations regarding HSf1 induced acceleration of Alu RNA self-cleavage activity. We have also modified the model figure (Fig.9) to exclude any assumption regarding the mechanisms leading to the Alu RNA destabilization in our context.

In any case, the ability to provide additional mechanistical studies in case of AD should be seen under the light of the limitations that exist for *in vivo* experiments in humans and lack of immortalized human hippocampal or cortical cells that could simulate more effectively the gene expression changes observed in amyloid beta pathology. In fact, these limitations are usually mitigated though the use of mouse models of AD, which in our case has already been done though our previous work in the mouse hippocampus (Cheng et al, elife 2020). Given how nascent this field is, we understand that our study inevitably still leaves more questions open than those ones it answers and, as this reviewer acknowledges, "*further studies are required to explore the mechanisms underlying this putative crosstalk*". To this end, we believe that the additional experimental data we provide here will serve as a good foundation for such future studies and draw further the scientific community's attention to the role of these largely overseen non coding RNAs in AD.

"More specifically:"

Point 1. The reviewer proposes that *"All graphs presented as boxplots should be replaced by dot plots and p-values (or q-values where applicable) should always be indicated.*

We have now employed a hybrid approach that includes both the boxplots and dot plots in order to present better the sample dispersion and any outliers. P values that were previously reported only in the figure legends are now depicted also within the graphs.

Point 2. The reviewer proposes that *"The authors should discuss the limitations of analyzing RNAseq data from bulk tissue preparations and/or include in silico deconvolution approaches to infer cell type proportion in the sample cohorts used."*

This is a great suggestion and we thank the reviewer for this comment. Throughout the text in the initial manuscript we had consistently used the term "neural" instead of "neuronal" for two reasons: Firstly, because, as the reviewer notes, non-single cell RNA sequencing does not distinguish between neuroglia and neurons. Secondly, because it is a reasonable expectation that our findings will apply to both neurons and neuroglia given that the basic apparatus for cellular response to stress shares similarities among multiple neural cell types. We agree that including an *in silico* cellular deconvolution approach would be a nice addition to the study and we now provide this as a new supplementary figure 8 (Suppl. Fig.8D,E). Since as part of the ROSMAP study we also have access to RNA sequencing data from microglia, we also tested this dataset as a positive control for a neuroglia cell type in our analysis. This analysis revealed a neuron: non-neuron proportion of approx. 43%:57% in the sequenced tissue, and as expected a 100% non-neuron percentage in our microglia control. We have also compared our RNA-seq data against a set of neuronal and non-neuronal markers inferred from whole brain single cell sequencing data including both neuronal

and non-neuronal cell types to substantiate further these findings (Suppl. Fig.8A-C) and we confirm that both neuronal and neuroglia cells contributing to the Alu RNA levels observed in our study (Suppl. Fig.8F), with percentages of this contribution likely to be different. We now discuss this new data and possible limitations in the Discussion section.

Point 3. **The reviewer proposes that** *-Related to the previous point: The correlation between p53 levels and Alu processing is not informative as it is now discussed in the manuscript. Would the authors expect a change in p53 levels as a function of microgliosis/astrogliosis/neuronal death observed in the AD samples? In that sense, how can the authors prove (or strengthen) the specificity of this observation? The authors should discuss.*

That's an important point and we are grateful to the reviewer for this comment. The rational for testing p53 levels was that in a similar context in mouse, p53 levels are downstream within pathways of stress response genes regulated by SINE RNAs. Subsequently, our similar findings regarding association of p53 with ALu RNA processing in our study provide the rational for testing further gene expression levels, as we did in Fig 6, to identify potential upstream regulators and pathways that may be also impaired. However, we never included any description of such pathways and as noted by the reviewer this information is not informative as it is presented. We now provide a clearer rational in the results section, and we have moved the p53 graphs at the end of figure 5 and directly before figure 6, to make the above connection clear. To this end, P53 results are now presented in a separate results section preceding the section regarding AD upregulated genes that could be potential upstream regulators in P53 pathways. For this reason, we have now also classified AD upregulated genes based on the correlation co-efficient of their relationship with Alu RNA processing (new Suppl. Table 6, new Fig.7A) and provide a full Gene Ontology and pathway analysis for the subset of the strongly correlated genes (new Suppl. Table 8, new Suppl. Fig.6 and 7). Based on this analysis we present the pathways upstream of p53 that include genes strongly correlated with Alu RNA processing and confirm increase of P53 after targeting of ALu RNAs in new Figure 7D. As noted in point 2 though we are not in the position to dissect the exact contribution of microgliosis, astrogliosis and neuronal death to our findings, since our RNA-seq is not single cell specific and all the identified pathways are universal cellular stress response and signalling pathways. We now discuss also this limitation in Discussion.

Point 4. **The reviewer asks "***Are the trends observed in Fig. 5A & B significant?"*

The answer is yes. We had included this information in the legend but we now realize that this information can be easily missed so we have now included this information also within the figure.

Point 5. **The reviewer notes that** *"In Fig. 5C, there seems to be a biphasic distribution of the samples according to ApoE genotype, which does not seem to correlate with the distribution of the differentially processed Alu RNAs. The authors should explain.*

This figure is now Suppl.Figure 2A. There are current no genetic markers for late onset AD with a high correlation to the observed phenotype. Even in case of ApoE, which until today is one of the few reliable genetic markers shown to confer some degree of higher susceptibility to AD, that connection is not absolute. The vast majority of people with these genotypes will not develop AD. For this reason, we also do not

expect the correlation between Alu RNA processing and this genotype to be a 1:1 match, especially since Alu RNA processing is a continuous variable while Apo E genotype being a discrete one. Nevertheless, we believe that presenting this data would benefit the readers as it shows that, as processing ratio climbs higher, it is more likely to identify an individual with this phenotype. The reason why within the higher processing ratio, distribution of genotypes seems to be biphasic (with one density area in the middle of the range, and one towards the right end of the range) remains unknown and likely should be attributed to the large heterogeneity among AD patients and the multi-factorial nature of this disease. If the reviewer or the editor think that presenting this would still be confusing we would be happy to omit it. We anyway had to move it to supplement in order to improve the flow of our presentation as mentioned at point 3 of this reviewer. He have now included this explanation in the legend of this figure.

Point 6. The reviewer remarks that *"The 'no protein' control is not a valid one for the in vitro experiments presented in Fig. 7, hence these data cannot be appropriately interpreted unless these experiments are performed again including the required controls. Also, which was the internal normalizer used in this case?"*.

Figure 7 is now Figure 8 in the revised manuscript. We are grateful to the reviewer for this comment. The reason we include the "no protein control" (i.e. incubation in the same buffer but in the absence of HSF1) is in order to account also for non-specific Alu RNA destabilization due to degradation, hydrolysis or Alu RNA endogenous self-cleavage.
Regarding necessary controls, in our recent PNAS study, in which we have substantiated the self-cleaving properties of Alu RNAs, we had included a number of protein control samples. However, we agree that including such controls here would also benefit the current study. Thus, we are now providing the following controls: i) simultaneous incubations of control RNAs (RNAs other than Alu RNAs) with HSF1, to provide their comparison with Alu RNA and show that in contrast to Alu RNA, these RNAs are not destabilized beyond the standard non-specific decay observed during such incubations (new Figure 8C), and ii) simultaneous incubations of Alu RNA with control proteins (including denaturated HSF1) , to provide their comparison with HSF1 and show that acceleration of Alu RNA self-cleavage is specific to HSF1 (new Figure 8B).

Point 7. The reviewer remarks that *"In the MAP cohort, the authors found no correlation between HSF1 levels and full length Alu RNAs. However, the in vitro results seem to contradict this observation. The authors should explain the discrepancy."*

This is an interesting point and we thank the reviewer for this comment. We have included the HSF1 vs. full-length plot as a control to the HSF1 vs. processing ratio plot. In that way we wanted to exclude the possibility that Alu RNAs may be under the direct transcriptional control of HSF1, which would confound our findings by causing an increase in the denominator of our estimated Alu RNA processing ratio. Our results show that HSF1 is unlikely to be an upstream direct regulator of Alu RNA transcription, as there is practically no correlation with Alu RNA full length levels and rather exerts its action on Alu RNAs by increasing the proportion of fragmented Alu RNAs. As with many gene circuits and pathways involved in cellular response to stress, there are usually a number of compensatory cellular homeostasis mechanisms in

place that through positive and negative feedbacks regulate RNA levels. In our case, it would be reasonable to expect that in vivo there are compensatory pathways that would respond to the reduction of Alu RNA levels by increased processing through an increase of Pol III Alu RNA transcription. This could even result in an increase of total Alu RNA levels after chronic exposure to such stimulus as observed in Fig 4D. In fact, we are currently preparing a manuscript on this mechanism.
In an in vitro setting though, during the incubation of Alu RNA with HSF1, there is a finite amount of RNA, that during processing is not replaced by any physiological process, so inevitably Alu RNA levels will decrease in time. We now include this potential explanation in the discussion.

Point 8. **The reviewer asks** *"Can any conclusions be made with regard to Pol III-transcribed Alu RNAs from the datasets obtained here?"*

That's a very interesting point, as it raises the question how many of the Alu sequences tested here originate from PolIII transcripts and how many from transcripts embedded into mRNAs (likely nascent ones) that may be also under the same endogenous ribozyme activity of the Alu sequence, and be processed in response to stimuli. Due to the repetitive nature of these RNAs, it is difficult to answer this question as the same sequence could be located simultaneously at regions between genes as well as regions within gene introns. This was also the reason why we have employed here mapping against the ALUome instead of against the genome. Nevertheless, there are ways to indirectly approach this question. In particular, to test this, we have repeated our mapping, performing it against the genome, and separated the Alu elements, against which we map the RNA fragments, into two categories: i) Alu elements that fall within gene regions, and (ii) Alu elements outside of gene regions. Despite multiple mapping a level of spatial specificity is expected to be maintained. Therefore, if the mapped Alu RNAs originated exclusively from either only Pol III Alu elements or mRNA embedded Alu elements, we would expect at least some difference in the distribution of fragments between Alu elements of these two categories, as the genic ones overlaps with mRNAs. As shown in the new Suppl.Fig.9, the fact that distribution models are very similar between the two categories indeed supports the hypothesis that both types of Alu elements may contribute to Alu RNA processing. However, given the limitations posed by the repetitive nature of Alu RNAs, it remains difficult to provide an exact number regarding the portion of B2 RNA fragments produced by each category and this is clearly noted in our revised discussion part. However, even the indication that Alu RNAs embedded in mRNAs may also play an important role in our model provides a new perspective that should be investigated further in future studies.

Point 9. **The reviewer remarks that** *"Non peer-reviewed preprints cited in the text should be clearly cited as such."*

Our paper has been now published in eLife and the respective citations have been updated accordingly.

Point 10. **The reviewer notes that** *"There are some typos in the figures that need to be corrected."*.

We have now gone through the figures and corrected them.

Reviewer #2 (Significance (Required)):

The role of Alu elements in physiological brain function and neurodegeneration has only recently started receiving attention and novel insights suggest a widespread impact of Alu RNAs over gene expression, especially under pathological conditions. Along these lines, A-to-I editing levels were recently found reduced in the brain of AD patients, further indicating a possible link between Alu-mediated transcriptional regulation and pathology.

Hence, the hypothesis that Alu elements may contribute to AD pathology is intriguing and further studies are required to explore the mechanisms underlying this putative crosstalk. However, the work reported here for the most part remains observational and the authors often (mistakenly) overestimate and misinterpret correlation as causation. The direct (in vitro) functional evidence provided in the manuscript is not adequately controlled and the derived conclusions are therefore not accurate. While the reported datasets are of great value per se, the authors will need to tone down their interpretation.

Referees cross-commenting

Regarding Reviewer #1's comments:

Pol II Chipseq or TagChIP may not be absolutely required.

Please see response to Reviewer 1. We do have equivalent data from the ROSMAP study and we now present them at new Suppl.Figure 5.

Regarding Reviewer #3's comments:

Agree with the majority of the remarks.
Addressing point #8 would be fundamental to the main conceptual message of the manuscript if indeed prior literature cannot support the link.

Please see response to point 8 of reviewer 3. Suppression of the physical/direct binding between Pol II and Alu elements (intact and processed) has been shown by already shown by the Goodrich/Kugel labs, including binding and Pol II suppression assays for specific Alu RNA fragments. Nevertheless, ff necessary, we would be happy to perform Pol II RIP-seq and provide this data in the final manuscript.

Overall:

The authors would primarily need to better clarify/discuss/interpret their data.
Small experimental additions will benefit the manuscript.

REVIEWER 3

Reviewer #3 (Evidence, reproducibility and clarity (Required)):

This reviewer generally remarks that "*The authors present smal- and total-RNA seq data indicating increased cleavage of Alu elements in the brain of AD patients. Using the sequencing data they identified the cleavage sites within Alu elements and proposed a model whereby the precessing ratios of Alus regulates gene expression. The study is interesting; however, most of the presented data are correlative, which makes the conclusions/model highly speculative and not strongly supported by the data.*"

Please see also our response to the general comment of Reviewer II. Our current results confirm previous *in vivo* findings from a related mouse model (Cheng et al, eLife 2020) and are in alignment with studies from the Kugel and Goodrich labs on the suppressive effect of Alu RNAs in transcription in human cells and our previous study on the self-cleaving properties of SINE RNAs, including Alu RNAs. Per reviewers' suggestions, we have now expanded the scope of the current study including ChIP-seq data of MAP patients (new Suppl.Fig.5) and a functional assay (Alu RNA KD) in a human neural cell line (new Figure 7).

Major comments:

<u>Point 1.</u> The reviewer asks *"In the Methods section, the authors utilized Qiagen MinElute Cleanup kit to clean phosphorylated RNA. The column of this kit binds only RNAs longer than 200 nt, while shorter RNAs go to the flow through. Which fraction did the authors use? and If the size cut off value of the column is 200 nt, how did the authors combine full length Alus (300 nt) Alu with degradation fractions (less than 200 nt) in one pool for sequencing?"*

Indeed, as in most silica-based columns used for RNA purification the standard protocol omits short fragments. However, a modification of this protocol proposed by many manufacturers and used in our previous short-RNA-seq studies allows to retain the shorter fragments in the column through the use of double amount of ethanol in the binding buffer. In particular, instead of the 100ul-350ul-250ul sample-binding buffer-100%ethanol proportions we used 525ul 100% ethanol. The modification from our previous short RNA seq protocol (Cheng et al, 2020) was that, instead of separating the short and long RNA fraction, we have combined them, applied ribodepletion and then small RNA-seq kit. Although the yield of short RNAs in the sequence run is reduced due to the presence of longer RNAs, this approach enables us to include both short and longer than 200 nt RNAs. We have now added this additional information in the methods. We also provide the respective Bioanalyzer electropherogram of one of the libraries that shows a fragment range of 150-650 nt that corresponds to RNA inserts of approx. 30-530 nt upon removal of the 120-nt adapter (new Suppl. Fig.10A).

<u>Point 2.</u> The reviewer remarks that *"To calculate Alu processing ratio, the authors calculated the read counts in the TSS as full-length Alu. In fact, degradation products can also generate reads at the TSS: if an Alu is cleaved at XR1 site, the 5' end of the 3' cleavage product will generate reads at the XR1 site, while*

the 5' end of the 5' cleavage product will generate reads at the TSS region. How did the authors exclude the possibility that the 5' cleavage product might be stable enough to generate sequencing reads?"

This is an important point and we are grateful to the reviewer for this comment. Indeed, in case of RNA-seq (ROSMAP data) we can exclude this possibility based on the way the library was constructed, which includes only poly(A)-selected RNA fragments. To this extent the 5' end fragment of the XR1 cleavage is not included in the RNA pool to be sequenced. As described below, we now provide a confirmation of this in new Suppl. Figure10C (upper panel). Thus, Figure 3 and onwards include data where the possibility described by the reviewer can be excluded.

In contrast, the library construction for short-RNA-seq of the Calgary Brain Bank samples (now Fig.2D) does not exclude non-poly(A) fragments and, thus, these samples may include both the 5' and 3' cleavage fragments in the RNA pool that was sequenced. To test this possibility we aligned the 3'end of the sequenced fragments. As we show in new Suppl.Fig 10C, in contrast to RNA-seq (MAP patients), in short-RNA-seq data (CBB patients) we did identify a significant portion of truncated 5' end fragments in our data, whose 3' end maps arounds XR1. This finding should be taken into account in our calculations of the Alu RNA processing ratio of these samples and we now describe this in detail in our methods section. To calculate the processing ratio in short-RNA-seq (no poly(A) selection) we are now using the same approach we applied in our previous papers where we estimated the total SINE RNA read coverage per base levels and normalized the results with a Pol III transcript that serves as a housekeeping RNA. The results remain unchanged, and processing ratio of AD patients remains significantly higher than in no AD controls. Thus, although we had initially applied for the short-RNA-seq data the same processing ratio analysis approach as with RNA-seq data to provide more homogeneity in the applied methods, we now see that the reviewers' concerns were indeed valid and we reverse to our standard approach for estimation of processing ratio in short RNA-seq data.

<u>Point 3.</u> The reviewer remarks that "*The authors mentioned that post-mortem delay was comparable between AD and no AD samples. Given that this is difficult to control, the RIN values (average and SD) of isolated RNA need to be mentioned for both groups to ensure comparable RNA integrity prior to library preparation."*

We now provide a comparison of the RIN scores in new Suppl.Figure 10B that shows no difference between the two groups.

<u>Point 4.</u> The reviewer notes that "*The authors explained higher levels of Alus fragments in AD patients as higher processing and destabilization of Alu RNAs. They overlooked the possibility that processing rate might be the same and the observed results is due to pathologic accumulation of the degradation products due to, e.g., AD-associated defect in some nucleases. This possibility need to be addressed or at least discussed."*

This is an important note and we thank the reviewer for this comment. Indeed, acceleration of Alu self-cleavage is likely only one of the ways through which the cell may control Alu RNA levels. Other pathways may also be at play including nucleases or protection through A-I editing. In fact, as noted below, new data

we present in New Fig.8D imply an additional Alu RNA processing step *in vivo* beyond its self-cleaving activity. To account partially for those potential defects in nucleases that are non-specific to Alu RNAs, in the analysis of short RNA-seq data we also normalize our data with another Pol III transcript that would be subject to similar non-specific effects. For nuclease defects though specific to Alu RNAs we currently don't have a way to correct for this. This limitation is now discussed in the revised discussion section.

Point 5. **The reviewer remarks that** "*If the authors are able to detect intact Alus (please refer to comment 2 above), they need to provide direct comparison between the level of intact Alus in AD and no AD groups. This will help understand better the relationship between intact Alu levels and Alu processing ratios.*"

We now include the full length Alu levels (5' end counts at TSS) between AD and no AD patients in Fig 4D (right panel). As noted on point 2 and shown in Suppl. Fig.10C, these are poly(A) selected RNAs and, thus, those fragments correspond to the intact full length RNAs. To maintain coherence in the way we present the results we also provide the same for CBB samples in Fig 2.D, though in that case, the levels correspond to total Alu RNA levels that include both full length and truncated products.

The reviewer also remarks that "*In Fig. 2B and Fig. 4B, around 20% of Alus appear to have higher processing ratios in the no-AD group. Is there any thing specific to this group of Alus like specific sequences or motifs?*"

That's an interesting point and, indeed, we have checked this and did not find any specific pattern. A similar finding is also observed when comparing Alu RNA processing ratio with RNA-seq data at Figure 6B, where a small subcategory of no AD patients shows a profile more similar to the AD ones. We are compelled to assume here that these findings may have to do with the great clinical heterogeneity observed in AD disease that results in misclassifying some patients with AD pathology into the no AD group. Such patients, despite having the underlying AD molecular pathology, may possess higher cognitive reserves that result in delayed symptom manifestation and classification into the no AD group.

Point 6. **The reviewer remarks that** "*In Fig. 7A, it is highly recommended to increase the amount of RNA so as to be able to detect both full-length Alu as well as the cleave product of degradation intermediate. Does the size of the in-vitro cleavage product match the expected size upon cleavage at XL!, XR1, and/or XR2?*".

Fig.7 is now Fig.8 in the revised manuscript. In the new control experiments we provide in Fig.7B and 7C we have now doubled the Alu RNA concentration as suggested by the reviewer as well as the exposure time of the gels in order to provide a more detailed picture of degradation products and also show the degree of non-specific RNA degradation in our experiments. The full respective gels are now presented in Suppl. Figure 11A and B. We have also performed sequencing of the *in vitro* processed product and present it in new Fig.8D.

Point 7. The reviewer remarks that "*In Fig. 4C, D, if HSF1 increases the processing of Alus as shown in Fig 7C, it is expected to reduce the level of full-length Alus. However, it does not (Fig. 7D). Does this reinforce the notion that the increased processing ratio is due to accumulation of degradation products rather than enhanced processing (see point 4 above)? The authors need to discuss and explain this point better.*"

Please see response to Reviewer 2, point 7. This is an important point and it is indeed the focus on a manuscript we currently have in preparation that investigates potential positive feedback loops that lead to enhancement of Alu RNA transcription during increased Alu RNA processing. We now discuss the possibility of such a regulation mode as well as the one proposed by the reviewer in our revised discussion section.

Point 8. The reviewer notes that "*The model of Alu binding to Pol II is based on previous studies (mainly on mouse B2). The physical/direct binding between Pol II and Alu elements (intact and precessed) need to be experimentally shown through, e.g., RIP experiments.*"

The suppression and physical direct binding of the full length Alu to Pol II has been shown in detail in human in the following study by the Kugel and Goodrich labs (Mariner, P. D., et al (2008). Mol Cell, 29(4), 499-509. doi:10.1016/j.molcel.2007.12.013). The study shows that each arm of Alu RNA binds Pol II, and that the right arm is a potent transcriptional repressor, including the A-rich linker and a domain in the right arm adjacent to the processing points we have identified. In that study, multiple fragments of Alu RNA have been tested to confirm their ability or inability to bind Pol II and the fragments identified in our study correspond well to those fragments. In fact, in our study we selected the region beyond the 135 nt for estimating Alu RNA fragments based on the Kugel/Goodrich study as this is the region that has already been shown to bind and suppress Pol II before. Nevertheless, if necessary, we would be happy to discuss with the editor the need for performing additionally the Pol II RIP seq experiment and provide it in our final manuscript. However, there is no guarantee that existing Pol II antibodies will work well for this assay.

Minor comments:

1)The detailed description of published results on how intact SINEs inhibit Pol II while processed SINEs activate Pol II is better to be moved from page 11 to the introduction to avoid any confusion.

We have now deleted this section from the results section and first mention the model in the Suppl. Figure 1 in the introduction.

2)Methods for short-RNA-seq (and how they are modified from previously published methods to include the 300-nt full-length Alus) need to be better explained.

We have now updated the methods section and also added the respective Bioanalyzer results that show the desired fragment lengths (new Suppl. Fig.10A).

3)The discussion lacks references in many parts.
We have added further references in this part, please see also response to the minor comment of Reviewer 1.

Reviewer #3 (Significance (Required)):

This study extends what we know about mouse B2 SINE to human Alus and utilizes brain samples from normal individuals as well as patients to generate the presented data.

I am an RNA biologist and have research lines on Alu elements, and can see that the manuscript (after addressing the comments) will be significant and of interest to RNA biologists, researchers in the field of retrotransposons, as well as researchers interested in AD.

Referees Cross-commenting

Reviewer #1 comments:

I agree that Pol II Chipseq will be a very good addition to the data set, especially when combined with the total-RNA seq data.

Reviewer #3:

I agree with most of the comments, and would like to emphasize the importance of the reviewer's comments that pertain to the correlation between p53 levels and Alu processing and the in-vitro experiments.

1st Editorial Decision 19th Jan 2021

Dear Dr. Zovoilis,

Thank you for the submission of your revised manuscript. We have now received the comments from the referees and all support its publication now. Only a few minor editorial changes will be required before we can proceed with the official acceptance of your study.

- Please add up to 5 keywords to your manuscript file.

- Please add a DAS (data availability section) to the end of the method section that lists the URL to access your deposited data. The data must be freely accessible upon the online publication of your paper.

- Our reference style has changed to Harvard style and lists up to 10 authors before "et al", please correct. https://www.embopress.org/page/journal/14693178/authorguide#referencesformat

- Please send us a completed author checklist that can be found here: <https://www.embopress.org/page/journal/14693178/authorguide> The completed author checklist will also be part of the transparent peer-review process file.

- Please upload all figures and tables as individual files.

- All supplementary tables need to be uploaded as individual files. Tables 1-8 should be called Dataset EV1-EV8 and uploaded as excel files, Table 9 should be called Table EV1.

- The supplemental figures can be called Figure EV1-EV5 and uploaded as individual files. Their legends need to be listed after the main figure legends. If more than 5 extra figures will be added, the remaining figures should be moved to an Appendix file with a table of content and page numbers. Alternatively, all supplemental figures can be moved to the Appendix file that should be uploaded as a single pdf file. You can find more information about our file types in our guide to authors.

- In Fig S11 the weight markers need lines around them to indicate that they were spliced on top of the gels. Please explain what was done and why. If Fig S11 is the source data for Fig 8 it should be uploaded as Source data file.

- I attach to this email a related manuscript file with comments by our data editors. Please address all comments in the final manuscript file.

- I add more info from our regular decision letters below, just for your information.

I would like to suggest a few minor changes to the title and abstract. Please let me know whether you agree with the following and whether the text accurately reflects your data (I am not 100% certain about the inverse correlation) :

Deregulation of human SINE Alu RNA processing in Alzheimer's disease brains

Despite significant steps in our understanding of Alzheimer's Disease (AD) pathogenesis, AD remains a complex disease, with many of the molecular processes underlying its pathogenesis

largely unknown. Here we focus on a mechanism that involves processing of a class of non-coding RNAs produced by repetitive Small Interspersed Nuclear Elements (SINEs). RNAs from SINE B2 repeats in mouse and SINE Alu repeats in human, long regarded as "junk" DNA, control gene expression by binding RNA polymerase II and suppressing transcription. They also possess self-cleaving activity that is accelerated through their interaction with certain proteins and disables the suppression of gene transcription, thus serving as transcriptional switch. We have recently shown that amyloid beta pathology in the brain of mouse models of Alzheimer's disease correlates with accelerated SINE B2 RNA processing. Here we show that similar to mouse brain, human SINE RNAs, and particularly SINE Alu RNAs, are processed, and the processing rate is increased in brains of AD patients compared with aging individuals that are cognitively healthy. In agreement with our previous work, increased processing of Alu RNAs correlates with the activation of genes upregulated in AD patients, while intact Alu RNA levels inversely correlate with the expression of genes downregulated in AD. In vitro assays show that processing of human Alu RNAs is accelerated by HSF1, a key stress response factor. Overall, our data show that RNAs from SINE Alu elements in the human brain are processed similar to SINE B2 RNAs in mouse brains, and show a similar pattern of deregulation during amyloid beta pathology, suggesting a role for SINE RNAs in human brain molecular pathology.

EMBO press papers are accompanied online by A) a short (1-2 sentences) summary of the findings and their significance, B) 2-3 bullet points highlighting key results and C) a synopsis image that is exactly 550 pixels wide and 200-600 pixels high (the height is variable). You can either show a model or key data in the synopsis image. Please note that text needs to be readable at the final size. Please send us this information along with the revised manuscript.

We would also welcome the submission of cover suggestions, or motifs to be used by our Graphics Illustrator in designing a cover.

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

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

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

Kind regards,
Esther

Esther Schnapp, PhD
Senior Editor
EMBO reports

Referee #1:

Authors have satisfactorily addressed all my points.

Referee #3:

The authors have adequately addressed all my remarks.

More info from our regular decision letters:

7) Before submitting your revision, primary datasets produced in this study need to be deposited in an appropriate public database (see https://www.embopress.org/page/journal/14693178/authorguide#datadeposition). Please remember to provide a reviewer password if the datasets are not yet public. The accession numbers and database should be listed in a formal "Data Availability" section placed after Materials & Method (see also https://www.embopress.org/page/journal/14693178/authorguide#datadeposition). Please note that the Data Availability Section is restricted to new primary data that are part of this study. * Note - All links should resolve to a page where the data can be accessed. *
If your study has not produced novel datasets, please mention this fact in the Data Availability Section.

8) We would also encourage you to include the source data for figure panels that show essential data. Numerical data should be provided as individual .xls or .csv files (including a tab describing the data). For blots or microscopy, uncropped images should be submitted (using a zip archive if multiple images need to be supplied for one panel). Additional information on source data and instruction on how to label the files are available at <https://www.embopress.org/page/journal/14693178/authorguide#sourcedata>.

9) Our journal also encourages inclusion of *data citations in the reference list* to directly cite datasets that were re-used and obtained from public databases. Data citations in the article text are distinct from normal bibliographical citations and should directly link to the database records from which the data can be accessed. In the main text, data citations are formatted as follows: "Data ref: Smith et al, 2001" or "Data ref: NCBI Sequence Read Archive PRJNA342805, 2017". In the Reference list, data citations must be labeled with "[DATASET]". A data reference must provide the database name, accession number/identifiers and a resolvable link to the landing page from which the data can be accessed at the end of the reference. Further instructions are available at https://www.embopress.org/page/journal/14693178/authorguide#referencesformat

The authors have addressed all minor editorial requests.

1st Revision - Editorial Decision 1st Feb 2021

Dr. Athanasios Zovoilis
University of Lethbridge
Southern Alberta Genome Sciences Centre and Canadian Centre for Behavioral Neuroscience
Canada

Dear Athan,

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.

At the end of this email I include important information about how to proceed. Please ensure that you take the time to read the information and complete and return the necessary forms to allow us to publish your manuscript as quickly as possible.

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

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

Should you be planning a Press Release on your article, please get in contact with emboreports@wiley.com as early as possible, in order to coordinate publication and release dates.

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

Best regards,
Esther

Esther Schnapp, PhD
Senior Editor
EMBO reports

****

THINGS TO DO NOW:

You will receive proofs by e-mail approximately 2-3 weeks after all relevant files have been sent to our Production Office; you should return your corrections within 2 days of receiving the proofs.

Please inform us if there is likely to be any difficulty in reaching you at the above address at that time. Failure to meet our deadlines may result in a delay of publication, or publication without your corrections.

All further communications concerning your paper should quote reference number EMBOR-2020-52255V2 and be addressed to emboreports@wiley.com.

Should you be planning a Press Release on your article, please get in contact with emboreports@wiley.com as early as possible, in order to coordinate publication and release dates.

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: Athanasios Zovoilis
Journal Submitted to: EMBO Reports
Manuscript Number: EMBOR-2020-52255V1

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.

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?	Based on previous published transcriptome projects in post-mortem brain tissues, including the ROSMAP study referenced in the Acknowledgements and dataset sections.
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?	In MAP patients data: Based on our preliminary quality control (generation of 5' end heatmaps) of all MAP samples, we identified a potential batch effect in a subset of these samples. Further examination of these samples revealed common technical characterisitics among them, with the most important one being that all were sequenced within a specific time period (year 2013), supporting a potential batch effect . Thus, these samples were excluded from further analysis.
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, no patient samples have been allocated to treatment.
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.	Such steps are decribed in detail in the ROSMAP study paper referenced in the acknowledgements.
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?	We have used standard tests in the field: student's test for short RNA and mRNA level comparisons of individual genes , statistical tools within the DEseq2 suit for transcriptome wide comparisons and correction for multiple testing, default statistics in the R package for the Pearson correlation in the scatterplots.
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	Yes, the DESeq2 incorporated statistical tests for transcriptome wide comparisons, and the student test for comparisons between samples of normalized read counts for the same selected genomic element are the default for RNA-seq data.
Is there an estimate of variation within each group of data?	Yes, depicted as error bars where applicable.

USEFUL LINKS FOR COMPLETING THIS FORM

http://www.antibodypedia.com
http://1degreebio.org
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-report

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://jij.biochem.sun.ac.za
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html
http://www.selectagents.gov/

Is the variance similar between the groups that are being statistically compared?	Yes, we did not observe any systematic skewing in the tested values between the groups or changes in the distribution of values due to outliers between the AD/noAD 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).	NA
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	SH-SY5Y neuroblastoma cell lines were purchased from ATCC (CRL-2266). ATCC is a tissue consortium that performs genotype and mycoplasma tests for all cell lines before sale. STR profiling reported: Amelogenin: X CSF1PO: 11 D13S317: 11 D16S539: 8,13 D5S818: 12 D7S820: 7,10 TH01: 7,10 TPOX: 8,11 vWA: 14,18

* 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.	University of Lethbridge Human Subject Research Committee (HSRC): Protocol #2019-059. Calgary Brain Bank provisions under the University of Calgary Conjoint Health Research Ethics Board Study ID: REB17-2289. ROSMAP study as described in the papers referenced in the Acknowledgements section.
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.	No experiments on research subjects were performed. Consent for provision of post mortem brain tissue material was provided under the Calgary Brain Bank provisions of study REB17-2289.
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.	CBB: Restricted access through EGA as defined in the respective EGA DAC policy for the EGAD00001006886 dataset.
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	Short RNA-seq raw fastq data from the CBB patients are available under controlled access through the European Genome-Phenome Archive (EGA) under the following dataset accession number, EGAD00001006886, and the following study access number, EGAS00001004973. MAP patient RNA-seq data have been obtained from the AMP-AD Knowledge Portal Synapse ID: syn3219045 (Bennett et al., 2012).
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).	Data have been deposited to EGA (see above).
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).	Data have been deposited to EGA (see above).
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.	Methods section provides description of publicly avaialbe software used in this study.

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