EMBO reports

Aire-dependent transcripts escape Raver2-induced splice-event inclusion in the thymic epithelium

Francine Padonou, Virginie Gonzalez, Nathan Provin, Sümeyye Yayilkan, Nada Jmari, Julia Maslovskaja, Kai Kisand, Pärt Peterson, Magali Irla, and Matthieu Giraud
DOI: 10.15252/embr.202153576

Corresponding author: Matthieu Giraud (matthieu.giraud@inserm.fr)

Review Timeline:

Submission Date:	6th Jul 21
Editorial Decision:	7th Jul 21
Revision Received:	12th Oct 21
Editorial Decision:	9th Nov 21
Revision Received:	14th Dec 21
Accepted:	21st Dec 21

Editor: Achim Breiling

Transaction Report: This manuscript was transferred to EMBO reports following peer review at The EMBO Journal.

Dear Dr. Giraud,

Thank you for transferring your manuscript to EMBO reports. I now went through your manuscript, the referee reports from The EMBO Journal (attached again below) and your revision plan. Both referees acknowledge that the findings are of interest. Nevertheless, they have raised a number of concerns and suggestions to improve the manuscript, or to strengthen the data and the conclusions drawn. It will be required to address the points of referee #1 experimentally in full, whereas we feel it is out of scope of the manuscript to provide the in vivo data requested by referee #2. Nevertheless, please provide a response to these points of referee #2 in your rebuttal letter.

Given the constructive referee comments, we would like to invite you to revise your manuscript with the understanding that all referee concerns must be addressed in the revised manuscript and/or in a detailed point-by-point response. Acceptance of your manuscript will depend on a positive outcome of a second round of review. It is EMBO reports policy to allow a single round of revision only and acceptance of the manuscript will therefore depend on the completeness of your responses included in the next, final version of the manuscript.

Revised manuscripts should be submitted within three months of a request for revision. We are aware that many laboratories cannot function at full efficiency during the current COVID-19/SARS-CoV-2 pandemic and we have therefore extended our 'scooping protection policy' to cover the period required for full revision. Please contact me to discuss the revision should you need additional time, and also if you see a paper with related content published elsewhere.

When submitting your revised manuscript, please also carefully review the instructions that follow below.

PLEASE NOTE THAT upon resubmission revised manuscripts are subjected to an initial quality control prior to exposition to re-review. Upon failure in the initial quality control, the manuscripts are sent back to the authors, which may lead to delays. Frequent reasons for such a failure are the lack of the data availability section (please see below) and the presence of statistics based on n=2 (the authors are then asked to present scatter plots or provide more data points).

When submitting your revised manuscript, we will require:

1) a .docx formatted version of the final manuscript text (including legends for main figures, EV figures and tables), but without the figures included. Please make sure that changes are highlighted to be clearly visible. Figure legends should be compiled at the end of the manuscript text.

2) individual production quality figure files as .eps, .tif, .jpg (one file per figure), of main figures and EV figures. Please upload these as separate, individual files upon re-submission.

The Expanded View format, which will be displayed in the main HTML of the paper in a collapsible format, has replaced the Supplementary information. You can submit up to 5 images as Expanded View. Please follow the nomenclature Figure EV1, Figure EV2 etc. The figure legend for these should be included in the main manuscript document file in a section called Expanded View Figure Legends after the main Figure Legends section. Additional Supplementary material should be supplied as a single pdf file labeled Appendix. The Appendix should have page numbers and needs to include a table of content on the first page (with page numbers) and legends for all content. Please follow the nomenclature Appendix Figure Sx, Appendix Table Sx etc. throughout the text, and also label the figures and tables according to this nomenclature.

For more details please refer to our guide to authors:
http://www.embopress.org/page/journal/14693178/authorguide#manuscriptpreparation

See also our guide for figure preparation:
http://wol-prod-cdn.literatumonline.com/pb-assets/embo-site/EMBOPress_Figure_Guidelines_061115-1561436025777.pdf

3) a .docx formatted letter INCLUDING the reviewers' reports and your detailed point-by-point responses to their comments. As part of the EMBO Press transparent editorial process, the point-by-point response is part of the Review Process File (RPF), which will be published alongside your paper.

4) a complete author checklist, which you can download from our author guidelines (https://www.embopress.org/page/journal/14693178/authorguide). Please insert page numbers in the checklist to indicate where the requested information can be found in the manuscript. The completed author checklist will also be part of the RPF.

Please also follow our guidelines for the use of living organisms, and the respective reporting guidelines:
http://www.embopress.org/page/journal/14693178/authorguide#livingorganisms

5) that primary datasets produced in this study (e.g. RNA-seq, ChIP-seq and array data) are deposited in an appropriate public

database. This is now mandatory (like the COI statement). If no primary datasets have been deposited in any database, please state this in this section (e.g. 'No primary datasets have been generated and deposited').

See also: http://embor.embopress.org/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 & Methods) that follows the model below. Please note that the Data Availability Section is restricted to new primary data that are part of this study.

Data availability

The datasets produced in this study are available in the following databases:

- RNA-Seq data: Gene Expression Omnibus GSE46843 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46843)
- [data type]: [name of the resource] [accession number/identifier/doi] ([URL or identifiers.org/DATABASE:ACCESSION])

*** Note - All links should resolve to a page where the data can be accessed. ***

Moreover, I have these editorial requests:

6) We strongly encourage the publication of original source data with the aim of making primary data more accessible and transparent to the reader. The source data will be published in a separate source data file online along with the accepted manuscript and will be linked to the relevant figure. If you would like to use this opportunity, please submit the source data (for example scans of entire gels or blots, data points of graphs in an excel sheet, additional images, etc.) of your key experiments together with the revised manuscript. If you want to provide source data, please include size markers for scans of entire gels, label the scans with figure and panel number, and send one PDF file per figure.

7) Our journal 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: http://www.embopress.org/page/journal/14693178/authorguide#referencesformat

8) Regarding data quantification and statistics, can you please specify, where applicable, the number "n" for how many independent experiments (biological replicates) were performed, the bars and error bars (e.g. SEM, SD) and the test used to calculate p-values in the respective figure legends. Please provide statistical testing where applicable, and also add a paragraph detailing this to the methods section. See:
http://www.embopress.org/page/journal/14693178/authorguide#statisticalanalysis

9) Please note our new reference format: http://www.embopress.org/page/journal/14693178/authorguide#referencesformat

10) Please add a paragraph detailing the author contributions to the manuscript. Please order the manuscript sections like this: Title page - Abstract - Introduction - Results - Discussion - Materials and Methods - DAS - Acknowledgements - Author contributions - Conflict of interest - References - Figure legends - Expanded View Figure legends.

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

Kind regards,

Achim

Achim Breiling
Editor
EMBO Reports

Referee #1:

The manuscript of Padonou et al examines intrathymic mechanisms that are relevant to T-cell tolerance. Specifically, they examine the role of Aire expression in medullary thymic epithelial cells (mTEC) in generating intrathymic expression of self antigens that are important in thymic selection. This area of research is important yet remains poorly understood. As such, the subject area is of significant interest.

The main findings are that genes controlled by Aire expression in mTEC are under-represented in relation to the production of splice variants. Furthermore, the authors show that this is because Aire-dependent genes escape the action of Raver2 as a result of differences in post-translational histone modifications.

Overall, the experiments are well performed and provide interesting new information on control of self-antigen expression in the thymus. However, a number of points need to be addressed.

1. The approaches used to isolate mTEC are not typical of many others studies. For example, mTEC are typically sorted as CD45-Ly51-UEA1+ cells. Here, the authors don't use UEA1 as a positive mTEC marker, and rely on Ly51 negativity, why is this? In addition, while mTEClo and mTEChi are usually sorted on MHCIIhiCD80hi, the authors here rely on MHCII only. Why is this? No examples of purity/phenotype are included, and the authors must include this data (see below for example).

2. The authors use a 3D culture system to knockdown Raver2 expression in mTEC, and analyse downstream consequences. They state that total mTECs are seeded into cultures, then recovered 3-5d later. It is important that the authors show data of the phenotype of the mTEC added to the cultures, and retrieved from the cultures. Also, what is the purity at sort (important as beads and not FACS has been used). Also, as RANKL is added to the cultures, this will shift mTEC towards mTEChi. The authors need to show data for expression of Ly51/EpCAM/MHCII/Aire in mTEC at the beginning and end of the culture period.

3. Fig 5B shows expression of Raver2, and the authors conclude that 'Raver2 is over-represented in mTEClo and mTEChi'. More data is needed to support this point. For example, comparison is made to RNA from total peripheral tissues, while the mTEC samples are purified epithelial cells. The authors need to make a fairer comparison by looking for Raver2 in epithelial cells from peripheral tissues, rather than whole tissue. Also, to complete the intrathymic expression analysis, Raver2 should be examined in cTEC.

Referee #2:

The work by Padonou et al. 2021 studied the regulation of alternative splicing in medullary thymic epithelial cells, and considers how it may contribute to T cell tolerance. The authors build on previous studies that showed that mTEChi cells possess relatively high rates of alternative splicing, which increases the diversity of peptide self-antigens presented by MHC molecules on the cell surface. In this report, the authors measured the extent of alternative splicing among Aire-dependent and Aire-independent genes in mTEChi cells. They report the unexpected result that alternative splicing events were higher in Aire-independent transcripts than in Aire-dependent ones. The authors also identified the alternative splicing regulator Raver2 and used shRNA in 3D organotypic cultures to show that it is a crucial factor for the incorporation of alternative splicing events in Aire-independent genes. Furthermore, they found that Aire-dependent genes have lower levels of H3K36 methylation, implying that alternative splicing was lower in these genes owing to a lack of Raver2 recruitment. This work is novel and interesting, with appropriate controls and statistical tools employed. However, the physiological relevance of this finding to central tolerance is inferred rather than directly shown, which limits its overall impact.

1) This study reports that nearly a quarter of alternative splicing events present in Aire-sensitive genes in peripheral tissues, were excluded in mTEChi cells. This exclusion could allow the release of autoreactive T-cells from the thymus, enhancing the risk of autoimmunity. But the authors do not show any evidence for this. It would be extremely interesting to see whether the T-cells specific for any peptides arising from 'thymus-excluded' alternative splicing events in Aire-dependent genes, are present in peripheral lymphoid organs or tissues. This would require the identification of such peptides, and the creation of pMHC tetramers to identify T cells specific to them, and to study the T-cell response in the periphery.

2) As suggested by the authors and others (Danan-Gotthold et al, 2016; Keane et al, 2014) mTEChi cells show high levels of alternative splicing events. But the importance of alternative splicing to mechanisms of central tolerance, such as T-cell deletion or Tregs selection was not directly tested. Thus, the identification of Raver2 as a key factor driving alternative splicing events in mTEChi cells brings important progress to our understanding of this phenomenon. Since the Raver2 KO mice are available (Jax stock# 51111-JAX) it would be interesting to study the function of this protein on the mechanisms of T-cell selection in an in vivo model.
a) Are ASEs in Aire-neutral genes reduced in Raver2 KO mice?
b) Are there any changes in the selection of Tconv and Tregs in those mice?
c) Are the Raver2 KO mice more susceptible to the development of autoimmune disease?

Dear Dr Breiling,

We appreciate the opportunity to submit a revised version of our manuscript entitled "Aire-dependent transcripts escape Raver2-induced splice-event inclusion in medullary thymic epithelial cells", by Padonou et al. (EMBOR-2021-53576-T). We thank both reviewers for their valuable suggestions, which have substantially improved our study. We believe we have addressed all the comments in the revised manuscript, as detailed bellow.

All modifications are shown in track changes. Text amendments related to reviewer 2 concerns are highlighted to be more easily distinguishable.

We provide below responses to all issues raised by the two reviewers.

Referee #1:
The manuscript of Padonou et al examines intrathymic mechanisms that are relevant to T-cell tolerance. Specifically, they examine the role of Aire expression in medullary thymic epithelial cells (mTEC) in generating intrathymic expression of self antigens that are important in thymic selection. This area of research is important yet remains poorly understood. As such, the subject area is of significant interest.

The main findings are that genes controlled by Aire expression in mTEC are under-represented in relation to the production of splice variants. Furthermore, the authors show that this is because Aire-dependent genes escape the action of Raver2 as a result of differences in post-translational histone modifications.

Overall, the experiments are well performed and provide interesting new information on control of self-antigen expression in the thymus. However, a number of points need to be addressed.

1. The approaches used to isolate mTEC are not typical of many others studies. For example, mTEC are typically sorted as CD45-Ly51-UEA1+ cells. Here, the authors don't use UEA1 as a positive mTEC marker, and rely on Ly51 negativity, why is this? In addition, while mTEClo and mTEChi are usually sorted on MHCIIhiCD80hi, the authors here rely on MHCII only. Why is this? No examples of purity/phenotype are included, and the authors must include this data (see below for example).

We thank reviewer 1 for highlighting the need to provide further details on the mTEChi/ mTEClo sorting strategy that we followed, notably with regard to the use of additional markers, i.e., EpCAM, UEA-1 and CD80.

* We provide original data showing details of the mTEChi/ lo sorts that we carried out for RNA sequencing. **Please see new Appendix Figure S1 and pg.5 of the revised MS.**

This simplified mTEChi/ lo sorting strategy was originally set-up based on the observations that in 4-wk old mice:

1- virtually all CD45- cells showing a high or low MHCII expression are EpCAM+ (Gray, JIM 2002, Fig 3B)

2- the vast majority of CD45- EpCAM+ cells that don't express Ly51, are positive for UEA-1 (Seach, J Immunol 2008, Fig 3C)

3- CD45- MHCII high and low cells are UEA-1+ (i.e. mTECs) (Gray, JIM 2008, Fig 3B first two panels)

4- CD45- Ly51- MHCII high cells express high levels of CD80 and Aire; CD45- Ly51- MHCII low cells express low levels of CD80 and are almost entirely negative for Aire expression (Gray, Blood 2006, Fig 6D, E).

We also note that the tight correlation between CD80 and MHCII was reported in CD45- Ly51- TECs (Desanti, J Immunol 2012, Fig 3A).

We would like to bring forward that this simplified mTEChi/ lo strategy was followed in a number of important publications such as: (Gray, JEM 2007), (Abramson, Cell 2010), (Koh, PNAS 2010), (Giraud, PNAS 2012), (Meredith, Nat Immunol 2015), (Bansal, Nat Immunol 2017) or more recently (Guyon, Elife 2020).

* To address the concerns raised about the purity/phenotype of mTEChi and mTEClo sorted following this strategy, we now include FACS data showing that these cells are all positive for EpCAM and UEA-1, therefore confirming their mTEC phenotype. In addition, we confirm that our sorted mTEChi and mTEClo are CD80high and CD80low, respectively. **Please see new Appendix Figure S2 of the revised MS.**

* To compare mTEChi/ lo obtained by the simplified sorting strategy and the strategy including the additional EpCAM, UEA-1 and CD80 markers, we took advantage of a public RNA-seq dataset of thymic epithelial cells isolated from the same mouse strain than the one we used (B6.129S2-Airetm1.1Doi/J from Drs Mathis and Benoist - ref:004743 at The Jackson Laboratory) and sequenced as paired-end data (2x100bp) (St-Pierre, J Immunol 2015). This dataset includes reads of mTEChi and mTEClo (three replicates) sorted from WT littermates as CD45- EpCAM+ Ly51- UEA-1+ CD80hi MHCIIhi and CD45- EpCAM+ Ly51- UEA-1+ CD80lo MHCIIlo, respectively. After mapping the reads to the mm10 genome, we calculated the level of gene expression for each of these samples and compared the expression of CD45, EpCAM, Ly51, CD80 and MHCII with our data. We added to this comparison highly specific mTEChi and lo genes that we obtained by re-analyzing a public scRNA-seq dataset (Bornstein, Nature 2018). Remarkably, we observe exact same levels of gene expression for mTEChi in both datasets and very close levels for mTEClo, showing that the sorted mTEChi populations are identical between the two strategies and that mTEClo are very similar. **Please see new Appendix Figures S4A and B, and pgs.8 and 9 of the revised MS.**

* In addition, we calculated the percent splicing inclusion (PSI) of alternative splicing events (ASEs) of Aire-sensitive and neutral genes in the mTEChi and mTEClo datasets sorted with the full list of markers and found very similar values to those that we obtained with our own mTEChi/ lo datasets. **Please see new Appendix Figures S4C-E, and pg.9 of the revised MS**. These new observations show that the simplified and the full-marker sorting strategies to isolate mTEChi/lo result in very similar datasets with an almost perfect match for mTEChi.

Gray DHD, Chidgey AP, Boyd RL. Analysis of thymic stromal cell populations using flow cytometry. J Immunol Methods. 2002; 260(1-2):15–28.
Seach N, Ueno T, Fletcher AL, Lowen T, Mattesich M, Engwerda CR, et al. The lymphotoxin pathway regulates Aire-independent expression of ectopic genes and chemokines in thymic stromal cells. JI. 2008; 180(8):5384–92.
Gray DHD, Fletcher AL, Hammett M, Seach N, Ueno T, Young LF, et al. Unbiased analysis, enrichment and purification of thymic stromal cells. J Immunol Methods. 2008; 329(1-2):56–66.
Gray DHD, Seach N, Ueno T, Milton MK, Liston A, Lew AM, et al. Developmental kinetics, turnover, and stimulatory capacity of thymic epithelial cells. Blood. 2006; 108(12):3777–85.
Desanti GE, Cowan JE, Baik S, Parnell SM, White AJ, Penninger JM, et al. Developmentally regulated availability of RANKL and CD40 ligand reveals distinct mechanisms of fetal and adult cross-talk in the thymus medulla. J Immunol. 2012; 189(12):5519–26.
Gray D, Abramson J, Benoist C, Mathis D. Proliferative arrest and rapid turnover of thymic epithelial cells expressing Aire. Journal of Experimental Medicine. 2007; 204(11):2521–8.
Abramson J, Giraud M, Benoist C, Mathis D. Aire's partners in the molecular control of immunological tolerance. Cell. 2010; 140(1):123–35.
Koh AS, Kingston RE, Benoist C, Mathis D. Global relevance of Aire binding to hypomethylated lysine-4 of histone-3. Proc Natl Acad Sci USA. 2010; 107(29):13016–21.
Giraud M, Yoshida H, Abramson J, Rahl PB, Young RA, Mathis D, et al. Aire unleashes stalled RNA polymerase to induce ectopic gene expression in thymic epithelial cells. Proc Natl Acad Sci USA. 2012; 109(2):535–40.
Meredith M, Zemmour D, Mathis D, Benoist C. Aire controls gene expression in the thymic epithelium with ordered stochasticity. Nat Immunol. 2015; 16(9):942–9.
Bansal K, Yoshida H, Benoist C, Mathis D. The transcriptional regulator Aire binds to and activates super-enhancers. Nat Immunol. 2017; 18(3):263–73.
Guyon C, Jmari N, Padonou F, Li Y-C, Ucar O, Fujikado N, et al. Aire-dependent genes undergo Clp1-mediated 3'UTR shortening associated with higher transcript stability in the thymus. Elife. 2020; 9:2078.
St-Pierre C, Trofimov A, Brochu S, Lemieux S, Perreault C. Differential Features of AIRE-Induced and AIRE-Independent Promiscuous Gene Expression in Thymic Epithelial Cells. J Immunol. 2015; 195(2):498–506.
Bornstein C, Nevo S, Giladi A, Kadouri N, Pouzolles M, Gerbe F, et al. Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells. Nature. 2018; 559(7715):622–6.

2. The authors use a 3D culture system to knockdown Raver2 expression in mTEC, and analyse downstream consequences. They state that total mTECs are seeded into cultures, then recovered 3-5d later. It is important that the authors show data of the phenotype of the mTEC added to the cultures, and retrieved from the cultures. Also, what is the purity at sort (important as beads and not FACS has been used). Also, as RANKL is added to the cultures, this will shift mTEC towards mTEChi. The authors need to show data for expression of Ly51/EpCAM/MHCII/Aire in mTEC at the beginning and end of the culture period.

* We chose to culture enriched rather than pure mTECs in using magnetic bead separation to preserve cell viability and improve the yield of the retrieved mTECs at the end of the culture period. The benefit of using magnetic bead was reported by Dr Pinto in her original paper describing the 3D culture system (Pinto, J Immunol 2013).

* To specify the phenotype of mTECs seeded to the 3D culture, we provide data showing the gates used to sort CD45- EpCAM+ TECs from total CD45 depleted thymic cells and the proportion of mTEChi and mTEClo following Ly51 depletion and MHCII enrichment by magnetic bead separation. We observe a strong reduction of Ly51+ cells (cTECs) following Ly51 depletion and a marked increase of MHCII+ cells (mTECs) coupled with a surfeit of mTEChi after MHCII enrichment. As a result, the cells that we seed onto the 3D culture system contain about 62% of mTEChi and 7% of mTEClo, corresponding to a 9 to 1 ratio. **Please see new Appendix Figures S7, and pg.12 of the revised MS.**

* To determine the proportion of mTEChi and mTEClo after the culture period, we provide data showing the gates that we used for cell sorting and the level of expression of EpCAM, Ly51 and MHCII. We show that almost all retrieved cells are positive for EpCAM (TECs) - no fibroblast contamination -, and negative for Ly51 (no cTECs). We also observe a reduced mTEChi to mTEClo ratio (3 to 1), with about 75% of mTEChi and 25% of mTEClo. **Please see new Appendix Figure S8A, and pg.12 of the revised MS.** We show that this shift towards mTEClo is dramatically enhanced in absence of RankL, with a mTEChi to mTEClo ratio (1 to 6) which is inverted in comparison to the condition with RankL. Hence, these new data show that RankL, added to the 3D culture system, helps mTEChi survive and preserve their mature phenotype. **Please see new Appendix Figure S8B, and pg.12 of the revised MS.**

* To complete the characterization of the cells that we isolate from the 3D culture, we provide qPCR data of *Aire* and *Ins2* (a prototypic Aire-dependent self-antigen gene) expression in the seeded and retrieved cells. **Please see new Appendix Figure S9, and pg.12 of the revised MS**. As expected, we observe a strong expression of *Aire* and *Ins2* in both samples, with higher levels (2-fold) at the beginning of the culture. Although the proportions of mTEChi in the seeded and retrieved cells are relatively similar (62% and 75%, respectively), the decrease of *Aire* expression is consistent with the transitory nature of its expression and the fact that it is not balanced out by new Aire-expressing mTEChi arising from differentiation/maturation of mTEClo in the 3D culture system.

Pinto S, Schmidt K, Egle S, Stark H-J, Boukamp P, Kyewski B. An organotypic coculture model supporting proliferation and differentiation of medullary thymic epithelial cells and promiscuous gene expression. J Immunol. 2013; 190(3):1085–93.

3. Fig 5B shows expression of Raver2, and the authors conclude that 'Raver2 is over-represented in mTEClo and mTEChi'. More data is needed to support this point. For example, comparison is made to RNA from total peripheral tissues, while the mTEC samples are purified epithelial cells. The authors need to make a fairer comparison by looking for Raver2 in epithelial cells from peripheral tissues, rather than whole tissue. Also, to complete the intrathymic expression analysis, Raver2 should be examined in cTEC.

* To address reviewer 1's concerns about the level of *Raver2* expression in mTEChi and mTEClo, we followed the proposed recommendations and collected public RNA-seq datasets of epithelial cells isolated from a variety of peripheral tissues of B6 mice. We included a dataset of skin epithelial cells that was generated by the group of Dr C Perreault and used as reference for mTEC and cTEC transcriptome analyses in (St-Pierre, Sci Rep 2013) and (Danan-Gotthold, Genome Biol 2016). In addition, we added a comprehensive set of epithelial cells isolated form a variety of peripheral tissues (i.e. brain, caecum, large intestine, small intestine, heart, kidney, liver, lung, lymph node, spleen) (Krausgruber, Nature 2020), and completed this collection with other epithelium public datasets isolated from neural tube (ENCODEProjectConsortium, Nature 2012), mammary glands (Pal, 2021), olfactory system (Tan, Elife 2020), esophage (Wiles, Cell Mol Gastroenterol Hepatol 2021) and colon (Marincola Smith, Am J Physiol Gastrointest Liver Physiol 2021). For each dataset, we mapped the reads to the mm10 genome, calculated the levels of gene expression and compared *Raver2* expression between these datasets and the mTEChi/ lo datasets. We found that the level of *Raver2* expression is significantly higher in mTEChi or mTEClo,

therefore showing a preferential expression in the thymic epithelium. **Please see new Figure EV3, Mat/meth section "Multi-tissue comparison analysis" and pg.11 of the revised MS**.

In addition, we note that *Raver2* expression in the skin epithelium is the one that is the closest to mTECs, probably reflecting similarities between the types of epithelium. Indeed, Foxn1 has been reported to be expressed in skin keratocytes (Palamaro, Int Rev Immunol 2014) and Aire in skin tumor keratinocytes (Hobbs, Nat Genet 2015).

* We thank reviewer 1 for calling our attention to the need to examine the expression of *Raver2* in cTECs. Given the close similarities between our mTEChi/ lo data and the mTEC data in (St-Pierre, J Immunol 2015), we selected and analyzed the cTEC dataset that was also generated in this study. We found that the level of *Raver2* expression in cTECs is 1) significantly higher than in peripheral tissues or epithelial cells from a variety of organs, and 2) slightly lower than in mTEChi or mTEClo. **Please see modified Fig 5B, new Figure EV3, and pg.11 of the revised MS.**

* We then asked about the levels of ASE inclusion in cTECs. We calculated the PSI values of ASEs of Aire-sensitive and neutral genes expressed in cTECs, and found a reduced inclusion of ASEs of Aire-sensitive genes. This new finding shows that the weak inclusion of ASEs of Aire-sensitive genes that we found to be a common feature of mTECs is actually conserved between mTECs and cTECs, suggesting that it could stem from a common thymic epithelial progenitor. **Please see new Figs 3G-I and pg.9 of the revised MS.**

St-Pierre C, Brochu S, Vanegas JR, Dumont-Lagacé M, Lemieux S, Perreault C. Transcriptome sequencing of neonatal thymic epithelial cells. Sci Rep. 2013; 3(1):1860–10.

Danan-Gotthold M, Guyon C, Giraud M, Levanon EY, Abramson J. Extensive RNA editing and splicing increase immune self-representation diversity in medullary thymic epithelial cells. Genome Biol. 2016; 17(1):219.

Krausgruber T, Fortelny N, Fife-Gernedl V, Senekowitsch M, Schuster LC, Lercher A, et al. Structural cells are key regulators of organ-specific immune responses. Nature. 2020; 583(7815):296–302.

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489(7414):57–74.

Pal B, Chen Y, Milevskiy MJG, Vaillant F, Prokopuk L, Dawson CA, et al. Single cell transcriptome atlas of mouse mammary epithelial cells across development. Breast Cancer Res. 2021; 23(1):69–19.

Tan K, Jones SH, Lake BB, Dumdie JN, Shum EY, Zhang L, et al. The role of the NMD factor UPF3B in olfactory sensory neurons. Elife. 2020; 9.

Wiles KN, Alioto CM, Hodge NB, Clevenger MH, Tsikretsis LE, Lin FTJ, et al. IκB Kinase-β Regulates Neutrophil Recruitment Through Activation of STAT3 Signaling in the Esophagus. Cell Mol Gastroenterol Hepatol. 2021.

Marincola Smith P, Choksi YA, Markham NO, Hanna DN, Zi J, Weaver CJ, et al. Colon epithelial cell TGFβ signaling modulates the expression of tight junction proteins and barrier function in mice. Am J Physiol Gastrointest Liver Physiol. 2021; 320(6):G936–57.

Palamaro L, Romano R, Fusco A, Giardino G, Gallo V, Pignata C. FOXN1 in organ development and human diseases. Int Rev Immunol. 2014; 33: 83–93.

Hobbs RP, DePianto DJ, Jacob JT, Han MC, Chung B-M, Batazzi AS, et al. Keratin-dependent regulation of Aire and gene expression in skin tumor keratinocytes. Nat Genet. 2015; 47: 933–938.

----------Referee #2:

The work by Padonou et al. 2021 studied the regulation of alternative splicing in medullary thymic epithelial cells, and considers how it may contribute to T cell tolerance. The authors build on previous studies that showed that mTEChi cells possess relatively high rates of alternative splicing, which increases the diversity of peptide self-antigens presented by MHC molecules on the cell surface. In this report, the authors measured the extent of alternative splicing among Aire-dependent and Aire-independent genes in mTEChi cells. They report the unexpected result that alternative splicing events were higher in Aire-independent transcripts than in Aire-dependent ones. The authors also identified the alternative splicing regulator Raver2 and used shRNA in 3D organotypic cultures to show that it is a crucial factor for the incorporation of alternative splicing events in Aire-independent genes. Furthermore, they found that Aire-dependent genes have lower levels of H3K36 methylation, implying that alternative splicing was lower in these genes owing to a lack of Raver2 recruitment. This work is novel and interesting, with appropriate controls and statistical tools employed. However, the physiological relevance of this finding to central tolerance is inferred rather than directly shown, which limits its overall impact.

1) This study reports that nearly a quarter of alternative splicing events present in Aire-sensitive genes in peripheral tissues, were excluded in mTEChi cells. This exclusion could allow the release of autoreactive T-cells from the thymus, enhancing the risk of autoimmunity. But the authors do not show any evidence for this. It would be extremely interesting to see whether the T-cells specific for any peptides arising from 'thymus-excluded' alternative splicing events in Aire-dependent genes, are present in peripheral lymphoid organs or tissues. This would require the identification of such peptides, and the creation of pMHC tetramers to identify T cells specific to them, and to study the T-cell response in the periphery.

We thank reviewer 2 for these recommendations that comfort the direction that we have taken to follow up on our discovery. To continue the current study, we are selecting a set of peptides derived from thymic-excluded ASEs of Aire-dependent genes to construct fluorochrome-conjugated pMHC tetramers that will be used to isolate specific circulating and thymic T cells. Their phenotypic and transcriptomic characterization, as well as the capture of the VDJ repertoire will be conducted at the single-cell level and compared to T cells specific to ASE-derived peptides present in the thymus. For this project, we build on a very similar research that we are currently carrying out on autoreactive T CD8 cells in type I diabetes, as part of an international consortium notably composed of tetramers experts (under the leadership of Dr Mallone, Cochin Institute, Paris).

These experiments are very demanding and pretty challenging. To our point of view, they constitute a substantial research project by itself with a high potential impact. We thus plan to conduct this work as a standalone project. Nevertheless, to take into account the reviewer concerns, we added changes in the main text to clearly reflect the assumed nature of the effect of ASE exclusion on central tolerance. We also bring forward the need to carry out research to formally demonstrate an effect on central tolerance, as well as a role for Tregs in controlling the incomplete Aire-dependent negative selection in-vivo. **In this regard, we modified the last sentence of the abstract, a sentence in the last paragraph of the introduction, and added changes in the discussion pgs.16 and 17**.

2) As suggested by the authors and others (Danan-Gotthold et al, 2016; Keane et al, 2014) mTEChi cells show high levels of alternative splicing events. But the importance of alternative splicing to mechanisms of central tolerance, such as T-cell deletion or Tregs selection was not directly tested. Thus, the identification of Raver2 as a key factor driving alternative splicing events in mTEChi cells brings important progress to our understanding of this phenomenon. Since the Raver2 KO mice are available (Jax stock# 51111-JAX) it would be interesting to study the function of this protein on the mechanisms of T-cell selection in an in vivo model.
a) Are ASEs in Aire-neutral genes reduced in Raver2 KO mice?
b) Are there any changes in the selection of Tconv and Tregs in those mice?
c) Are the Raver2 KO mice more susceptible to the development of autoimmune disease?

We are grateful for the information that the Knockout Mouse Project (KOMP) has released *Raver2*-KO mice, available as cryopreserved embryos. The corresponding littermate thus can be likely recovered in approximately 4 months and will help in further deciphering the role of Raver2 in the establishment of central tolerance.
We agree that these mice would allow to further confirm the effect of Raver2 on ASE exclusion and determine its impact on negative selection and the escape of autoimmune T cells captured by pMHC tetramers and more generally on the balance of Tconv versus Tregs or signs of autoimmunity. This research will constitute a great addition to the research depicted in the previous point, enabling the characterization of the role of Raver2 and ASE exclusion on central and peripheral tolerance, as a complete follow up on our discovery. **We added a sentence in the discussion to highlight this point pg.17.**

We believe that we have addressed all the issues raised by the reviewers. We remain very grateful to the editor and the reviewers for the competent revision of our manuscript.

Sincerely yours,

Dr Matthieu Giraud
Corresponding author

Dear Dr. Giraud,

Thank you for the submission of your revised manuscript to our editorial offices. I have now received the reports from the two referees that were asked to re-evaluate your study, you will find below. As you will see, the referees support the publication of your study.

Before we can proceed with formal acceptance, I have these editorial requests I ask you to address in a final revised manuscript:

- Please shorten the title to not more than 100 characters (including spaces).

- Please provide the abstract written in present tense throughout.

- Please add up to 5 keywords to the title page.

- Please make sure that the number "n" for how many independent experiments were performed, their nature (biological versus technical replicates), the bars and error bars (e.g. SEM, SD) and the test used to calculate p-values is indicated in the respective figure legends (also of the diagrams in the Appendix), and that statistical testing has been done where applicable. Please avoid phrases like 'independent experiment' or 'independent replicate', but clearly state if these were biological or technical replicates. If statistical testing was done but there is no significant difference, please also mark this in the diagrams (n.s.). It seems presently some diagrams have no stats or only partial stats.

- Please move the legends in the Appendix below each figure. That is much more comprehensible for the reader.

- Finally, please find attached a word file of the manuscript text (provided by our publisher) with a few changes and queries we ask you to include in your final manuscript text. Please provide your final manuscript file with track changes, in order that we can see any modifications done.

In addition, I would need from you:
- a short, two-sentence summary of the manuscript (not more than 35 words).
- two to four short bullet points highlighting the key findings of your study.
- a schematic summary figure (in jpeg or tiff format with the exact width of 550 pixels and a height of not more than 400 pixels) that can be used as a visual synopsis on our website.

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

Best,

Achim Breiling
Editor
EMBO Reports

Referee #1:

The authors assert that addressing the functional consequences of alternative splicing changes in non-AIRE expressing cells or the effect of Raver2 deficiency is beyond the scope of the current report, and I am in agreement.

Referee #2:

The authors have comprehensively addressed the comments raised in relation to their initial manuscript. The additional experimental detail, data and interpretation has made for a stronger study that will be of interest.

The authors performed the requested editorial changes.

Dr. Matthieu Giraud
Université de Nantes, INSERM
Centre de Recherche en Transplantation et Immunologie
30 bd Jean Monnet
Nantes, FRANCE 44000
France

Dear Dr. Giraud,

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." Please note that the author checklist will still be published even if you opt out of the transparent process.

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.

Yours sincerely,

Achim Breiling
Editor
EMBO Reports

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: Matthieu Giraud
Journal Submitted to: EMBO Reports
Manuscript Number: EMBOR-2021-53576-T

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 χ2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

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

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?	WT and Aire-KO RNA-seq experiements were repeated three independent times; Chip-seq experiments two independent times. Human RNA-seq was performed on five individuals to hinder the expected heterogeneity. When applicable, we matched the sample size of the comparator group to the tested one to avoid artificial inflation of the statistical significance due to oversize of one group.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	Each sample corresponds to pools of the minimum number of mice required to obtain the minimal number of cells that are necessary for the planned experiments. Hence pools of 4 mice were needed for RNA-seq experiments whereas pools of more than ten mice were needed for the 3D co-culture system experiments.
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	No samples or animals were excluded from our study.
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.	Animals of the same genotype (i.e. Aire-KO or WT) were randomly chosen to constitute pools of mice required for the different experiements
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.	Littermates were chosen for the WT vs Aire-KO comparisons.
4.b. For animal studies, include a statement about blinding even if no blinding was done	All of the experiments were conducted in a blinded manner, without a priori.
5. For every figure, are statistical tests justified as appropriate?	Yes, we used appropriate statistical tests. Data comparisons were performed using either the non-parametric Wilcoxon test, the parametric Student test or the Chi-squared test when appropriate.
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	Student's tests were performed when the prerequisite of a normal distribution of the data was met. Normality of the distributions was assessed with the Shapiro-Wilk test.

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://jjj.biochem.sun.ac.za
https://osp.od.nih.gov/biosafety-biosecurity-and-emerging-biotechnology/
http://www.selectagents.gov/

Is there an estimate of variation within each group of data?	For analyses of splicing entropy or enrichment of H3K36me3, the 95% confidence interval of the median and the 75th and 25th percentiles for H3K36me3, are shown for each replicate.
Is the variance similar between the groups that are being statistically compared?	For the student test, a statistical test of variance was applied

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).	Information on antibodies is given in the "Materials and Methods" section
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	Primary cells were used in this study

* 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.	Aire-KO and wild-type littermates bred on a C57BL/6 background were obtained from Drs Mathis and Benoist, Harvard Medical School, Boston. They were housed according to guidelines of the French Veterinary Department. Experiments were performed on 4-6 week-old mice.
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	All experiments were approved by the Paris-Descartes Ethical Committee for Animal Experimentation (decision CEEA34.MG.021.11).
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.	Yes we followed the ARRIVE Guidelines

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	The study protocol was approved by the Ethics Review Committee on Human Research of the University of Tartu (permission #170/T-I).
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.	The informed consent was obtained from all subjects and the experiments conformed to the mentionned report.
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	Raw sequencing data have been made publicly available and can be accessed in the GEO database of the NCBI. The R code developped for ASE inclusion analysis has also been depositied as a Git repository available from GitHub.
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).	Raw sequencing data have been made publicly available and can be accessed in the GEO database of the NCBI. The R code developped for ASE inclusion analysis has also been depositied as a Git repository available from GitHub.
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).	Human transcriptomic datasets have been deposited in the GEO database
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.	The R code developped for ASE inclusion analysis has been depositied as a Git repository available from GitHub.

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.	no