EMBO reports

Single cell clonal analysis identifies an AID-dependent pathway of plasma cell differentiation

Carmen Gómez-Escolar, Alvaro Serrano-Navarro, Alberto Benguria, Ana Dopazo, Fatima Sanchez-Cabo, and Almudena Ramiro
DOI: 10.15252/embr.202255000

Corresponding author: Almudena Ramiro (aramiro@cnic.es)

Review Timeline:

Submission Date:	8th Mar 22
Editorial Decision:	4th Apr 22
Revision Received:	25th Jul 22
Editorial Decision:	6th Sep 22
Revision Received:	13th Sep 22
Accepted:	15th Sep 22

Editor: Achim Breiling

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. Depending on transfer agreements, referee reports obtained elsewhere may or may not be included in this compilation. Referee reports are anonymous unless the Referee chooses to sign their reports.)

1st Editorial Decision **4th Apr 2022**

Dear Dr. Ramiro,

Thank you for the submission of your research manuscript to EMBO reports. I have now received the reports from the three referees that were asked to evaluate your study, which can be found at the end of this email.

As you will see, the referees think that these findings are of interest. However, all referees have several comments, concerns, and suggestions, indicating that a major revision of the manuscript is necessary to allow publication of the study in EMBO reports. As the reports are below, and all their points need to be addressed, I will not detail them here.

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 or in the 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. Please contact me to discuss the revision (also by video chat) if you have questions or comments, or should you need additional time.

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. 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

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

See also the guidelines for figure legend preparation:
https://www.embopress.org/page/journal/14693178/authorguide#figureformat

3) 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

4) that primary datasets produced in this study (e.g. RNA-seq, ChIP-seq, structural and array data) are deposited in an appropriate public database. If no primary datasets have been deposited, please also state this in a dedicated section (e.g. 'No primary datasets have been generated and deposited'), see below.

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. This is now mandatory (like the COI statement). Please note that the Data Availability Section is restricted to new primary data that are part of this study. This section is mandatory. As indicated above, if no primary datasets have been deposited, please state this in this section

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, 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 for potential EV figures and all those in the final Appendix). Please also check that all the p-values are explained in the legend, and that these fit to those shown in the figure. Please provide statistical testing where applicable. Please avoid phrases like 'independent experiment', but clearly state if these were biological or technical replicates. See also:
http://www.embopress.org/page/journal/14693178/authorguide#statisticalanalysis

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

10) We updated our journal's competing interests policy in January 2022 and request authors to consider both actual and perceived competing interests. Please review the policy https://www.embopress.org/competing-interests and update your competing interests if necessary. Please name this section 'Disclosure and Competing Interests Statement' and put it after the Acknowledgements section.

11) Please also add up to 5 key words to the title page.

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.

Please use this link to submit your revision: https://embor.msubmit.net/cgi-bin/main.plex

Yours sincerely,

Achim Breiling
Senior Editor
EMBO Reports

Referee #1:

Gomez-Escolar and Ramiro and their colleagues employ single cell RNA-sequencing and clonal analysis to determine the biological consequences of functional AID-ablation in the T-dependent antigen-activated B cell response which includes the germinal center (GC) B cell response, a cell population with high AID expression. While in this study, the scRNA-seq analysis of immunized wild-type mice serves as the control for the analysis of the AID-ablated B cells, the results both confirm previous publications and add new information. It is a previously not described plasma cell-precursor population that is demonstrated to have characteristics of an early plasma cell population which is significantly reduced, along with plasma cells, in the AID-deficient immune response. This observation, and the conclusion based on the finding that memory B-cell and memory precursor B-cell populations in the GC are largely unchanged in terms of frequency and gene expression, provide important new information for the field of adaptive immunity.

A few points may require attention:

This is another paper among several analyzing the GC response by scRNA-seq, and the results of the normal GC response which are largely, but not entirely consistent with published work, clearly demonstrate that there is a need for independent analyses towards setting up a coherent picture of the GC response. The prePB subset identified here is different from those reported in previous analyses, and specifically mentioned is the Ise et al. 2018 work. Conversely, it seems that the subset identified by Ise et al. is not identified in the present analysis. Here one would like to see more discussion on the possible reasons for the different findings beyond the sentence in lines 361/362 "...suggesting the existence of alternative pathways leading to the differentiation of PB/PC." What was different among the analyses reg. antigen or immunization conditions or animal model? Can this account for the different findings or could differences in the biostatistical analysis lead to differing results?

Similarly, the finding of a reduction in the prePB subset in mice with AID-deficient activated B cells is intriguing. But one may like to see more discussion on why the plasmablast/plasma cell differentiation arm is specifically affected than that given in lines 374-376 "...probably reflecting changes in antibody affinity maturation and Ig isotype-dependent checkpoints." The authors should speculate more how one can envision that absence of AID can change affinity maturation in the plasmablast/plasma cell differentiation arm only.

Figure 4 and section starting line 246 on the transitional clusters in AID deficient mice. A fold-difference, and the actual medium values, should be provided for Fig. 4B (and 4I) in the text. It seems the difference is two-fold compared to the control mice? With this in mind, Fig. 4E finds the same overall percentage of plasma cells (PB), and an about 2.5-fold difference for the prePB, meaning the corresponding cells are reduced but are not absent. The following figures convincingly support the conclusions drawn reg. the consequences of AID deletion, especially the branching such as in Fig. 5B to D. But the addition of some actual values reg. Fig. 4 seems appropriate.

Referee #2:

Gómez-Escolar and colleagues use single-cell sequencing of immunized mice to study the relationship of B cell and plamsa cells and how this is modulated by AID deficiency.

Strength: This study uses advanced bioinformatic analysis to identify B cell subsets in an unbiased way. By using mice reporting the previous expression of AID, the authors can identify previously activated B cells. By comparing AID deficient to AID sufficient mice, they re-assess the role of AID, and thus class-switch-recommendation and antibody class switching, for MBC and PC differentiation.

Major Limitations:

1. A key weakness of the study is the lack of antigen specificity of the cells analysed. It is thus entirely unclear what relationship GC B cells have to PC and MBC and I would expect that while GC B cells may reflect predominantly cells responding to OVA, most MBC and PC were present before the immunisation.

2. One key finding, the identified prePB stage, is poorly characterised and the characteristics of these cells are not compared to existing literature characterising PC differentiation from GC. Rather, the authors speculate about an "alternative route to PB/PC differentiation without providing any molecular basis, such as changes in transcription factor expression, required for PC/PB differentiation. Thus, little evidence is provided that the prePB subset is indeed giving rise to PC. Also, it is unclear why the authors choose to analyse CD38+ GL7- cells for pre-PB cells (Figure 2c). CD38+ GL7- cells represent resting naïve and memory B cells unlikely to be on the way towards PC differentiation. Rather, PC precursors should be contained among the activated/GC (GL7 and/or FAS+) B cells. Based on the data presented, it appears more likely that what is labelled as "prePB" rather reflects Memory B cells. To confirm prePB identity, known genes expressed upon PC differentiation (such as IRF4 and PRDM1 expression) should be analysed.

To the best of my knowledge, FcεRIɣ is not known to be expressed in B cells. While it is possible that a small subset indeed expresses FcεRIɣ, staining with anti-CD19 and/or anti-Ig should be used to confirm B cell identity of FcεRIɣ expressing cells as not all B220+ cells (Nikolic et al., EJI 2002) are B cells. Along these lines, it is apparent in Figure 2 A that a small, contaminating cluster of FcεRIɣ expressing cells exists in addition to the B and PC clusters, highlighting the possibility that not all cells identified as Tom+ are B or PC. While such contamination is commonly occurring and at a major issue, any conclusions about small cell populations need to be thoroughly validated.

3. Little to no comparison of proposed MBC clusters with previously described genes expressed in MBC subsets is made.

4. The authors suggest using AID-based genetic tracer allows them to isolate GC-derived cells, while at the same time acknowledging that extrafollicular cells will also express AID. It remains unclear what advantage the use of the AID reporter has beyond excluding naive B cells for the analysis performed here.

4. Effects of AID deficiency are analysed in the absence of considering BCR isotype and affinity for antigen, both dependent on AID and well known to predispose to fate choices in GC responses. GC response kinetics, magnitude, duration and output are all changed in the absence of AID, making any conclusion about cause and consequence difficult. It is unclear to me what question the authors wanted to answer with their experiments.

Other comments:
- It is unclear why OTII T cell transfer and OVA immunisation was chosen as a model antigen system. Numerous other antigens exist which do not require the transfer of T cells. Can the authors provide a rationale for their choice of experimental system?
- FceR1g is typically referred to as FcRg (see Brandsma et al., Immunity, 2016)
- Figure 1C: Unclear what % cells refers to.
- Is the clustering of DZ and LZ cells predominantly based on cell cycle genes? While proliferation is associated with DZ, it is typically also seen in a fraction of LZ cells.

Summary: This manuscript provides little to no advance in our understanding of B cell activation and differentiation. While some of the findings may be interesting, they would require significant further validation to allow for definite conclusions about the relationships of identified B cell subsets. As such, the inclusion of antigen specificity and the longitudinal analysis over the course of an immune response are key missing components. Beyond identifying the transcriptional clusters of B cells, the sole conclusion the authors make is that AID shapes plasma cell differentiation, a well known fact given its role in affinity maturation and antibody class switching.

Referee #3:

Gomez-Escolar and colleagues use an AID-driven reporter to investigate AID-experienced B cells during secondary immune responses to the ovalbumin (OVA) protein in wild-type and AID-deficient mice. Single-cell transcriptomic and VDJ analyses on cells expressing the reporter fluorescent protein Tomato (Tom) identified distinct transcriptional cluster representative of germinal center (GC) dark (DZ) and light zone (LZ) B cells, plasmablasts/plasma cells (PB), precursor plasmablasts (prePB) and memory B cells (including 4 distinct subgroups). The authors further characterized the prePB cells that displayed expression of several markers including Fcer1g. The expression of Fcer1g was confirmed at protein level and it was used as a marker to identify the prePB population by cytofluorimetric analysis of Tom+ cells. Mice lacking AID displayed a reduced proportion of prePB and, as expected, GC expansion and reduced DZ/LZ ratio. Although it is known that AID-deficiency in mice leads to a reduction in plasma cells, the data provided here suggest that the plasma cell impairment is related to the reduced prePB population.

Overall, this work provides a transcriptomic characterization at single-cell level of AID-experienced B cell populations and of the effects of AID deficiency on these populations. The identification of a previously unrecognized precursor plasmablast population that is related also to memory B cells and is affected by AID loss, provides the basis to further explore the interaction between different post-GC subpopulations in wild-type and AID-deficient models.

The experimental approaches and data analyses are adequate to address the biological question that is of interest for the field of antibody-mediated immunity. However, some aspects should be clarified and discussed more in detail (see below). In particular, the authors should convey more clearly the information that relate the prePB population not only to plasma cells but also to the memory compartment. In addition, an unbiased analysis of the transcriptomic data from the AID-deficient cells should be considered.

Specific Comments:
• The number of analyzed cells and their distribution in the different subpopulations should be reported for each of the 4 specimens. It should also be clarified if all analyses are performed only on cells for which good quality data are available for both RNA and VDJ-C.
• Figure 1D: please clarify which clusters are identified initially when analyzing all cells together and which ones are identified

after sub-clusterization of the memory group. In particular, it is unclear whether the prePB cluster is identified only after sub-clusterization.
• Figure 1D: the principal component analysis clearly places the prePB closer to the memory clusters than to the PB cluster. The authors are analyzing a secondary immune response (2 weeks after a second antigen exposure): are these prePB coming from memory B cells generated during the primary immune response rather than being the products of recent GC reactions? This aspect needs to be discussed more in detail. The authors emphasize the relationship between prePB and PB, however the relationship with memory should also be detailed more clearly especially based on: i) the relatedness highlighted in the UMAP (Fig. 1D); ii) the large fraction of markers shared with memory in particular Mem1 (Fig. 1I); iii) the IgH V mutation pattern that resembles both PB and Mem1 (Fig. 3C); and iv) the expression of mostly unswitched IgH (Fig. 3E).
• Figure 1D: a small cluster of cells located in the top right quadrant of the UMAP should be checked for quality or other factors driving the distinct transcriptome.
• Figure 1E: It should be clarified how the specific populations are labeled. For example, the authors refer to previously published DZ/LZ signatures: however the DZ- and LZ-signature scores are >0 only in a fraction of the cells that are called DZ/LZ (Fig. S2A, D).
• Figure 3B, D: do the cells used for the analysis carry complete information about heavy and light immunoglobulin chains? Usually not all cells with good RNA expression data have equally informative VDJ and constant region data. The number of cells with suitable VDJ data should be reported in the legends, results and/or in the methods.
• Figure 3A-E: PrePB and PB show about 50% of cells completely unmutated and a very high proportions of unswitched cells is observed in GC (>50%) and in PB (>75%). These values appear low. Please comment.
• Figure 4C-E: The DZ/LZ ratio reduction is shown by FACS analysis but it doesn't seem to be captured by the sc-transcriptomic data.
• No UMAP plots are shown for the AicdaCre/- mouse sc-transcriptomic data. Did this analysis identified the same subpopulations as in the normal mice? The authors mention only "transfer of the transcriptome data into the clusters defined for wild type B cells" (Page 11, lines 267-268). An independent analysis of AicdaCre/- mouse sc-transcriptomic data should be provided.
• The circus plot obtained from the wild-type mice and displayed in Figure 3H is different from the one showed in Figure 7A (left panel). See for example at the sharing between aMem and GC. Please clarify.

Minor details:
• Fig. 1E: why the clusters are ordered by size (I guess) rather than biology or relatedness? Number of cells in each cluster should be provided.
• Fig. 1 legend: Panel E refers to clusters identified in D not C.
• Fig. 2A legend: unclear the reference to Fig. S2A.
• Fig. 4C: the percentage of cells in each gate should be shown.
• Fig. 4F legend: it mentions a booster with OVA but the scheme shows NP-CGG.
• Fig. 4 legend: Panel H is not described and panel I is not listed. The ANOVA refers to a not existing panel.
• Fig. 5D legend: no left and right panels.
• Fig. 7 legend: the text mention p-values as asterisks but no asterisk is displayed in the figure.
• Fig. S3A legend: not clear to which data/figure these p-values refer to.
• Some panels in Figure 4 (F-I) could be moved to Supplemental and the rest could be merged with Figure 5 to have only one figure on the population changes in AID-deficient mice. Similarly, Figure 6 and 7 could be merged displaying only one figure on clonal features and the transitions inferred from them.

Point-by-point Response to Reviewers
Gomez-Escolar et al. EMBOR-2022-55000V1

Response to Reviewer #1
We thank the reviewer for his/her thorough revision and constructive comments. We are very pleased that she/he considers our manuscript to provide important new information for the field of adaptive immunity. We address below the specific point raised by the reviewer.

A few points may require attention:

This is another paper among several analyzing the GC response by scRNA-seq, and the results of the normal GC response which are largely, but not entirely consistent with published work, clearly demonstrate that there is a need for independent analyses towards setting up a coherent picture of the GC response. The prePB subset identified here is different from those reported in previous analyses, and specifically mentioned is the Ise et al. 2018 work. Conversely, it seems that the subset identified by Ise et al. is not identified in the present analysis. Here one would like to see more discussion on the possible reasons for the different findings beyond the sentence in lines 361/362 "...suggesting the existence of alternative pathways leading to the differentiation of PB/PC." What was different among the analyses reg. antigen or immunization conditions or animal model? Can this account for the different findings or could differences in the biostatistical analysis lead to differing results?

We thank the reviewer for this comment. This is indeed a very important point that we have now addressed in depth and that has improved the quality of our manuscript by providing additional information of prePB and aMem subsets, as will be explained below.

As suggested by the reviewer, the approach in the Ise et al study is different from ours in many respects. However, as we will explain here, the data from both studies are compatible and nicely complement each other. First of all, the Ise et al study (PMID: 29669250) is specifically designed to identify a prePB subset (Fraction 1) based on the premise that they should be contained within the LZ population and should express IRF4. They however use Bcl6 low expression as a surrogate to detect IRF4+ cells. As noted in their article: *"Based on the requirement for IRF4 in post-GC plasma cell differentiation (De Silva et al., 2012), we hypothesized that IRF4+ LZ GC cells might exist as plasma cell precursors. Since Irf4 reporter mice are currently not available, in order to circumvent this limitation, we used Bcl6 reporter mice as an alternative. We reasoned that a Bcl6low LZ GC B cell subset might exist and could contain the IRF4+ cells, because previous experiments using transformed GC B cell lines show downregulation of Bcl6 by IRF4 overexpression"*. In contrast, our identification of prePB cells was completely hypothesis free: we analyzed all cells labeled by the expression of AID and identified prePB cells as an independent transcriptional cluster which share many clones with PB cells; this is a key finding that unmistakingly allows their assignment as prePB. Although other experimental differences exist between the two studies (immunogen, kinetics, etc), we believe that those differences have no impact our conclusions, as we explain below.

Given the design of our study, we reasoned that the Ise prePB cells should also be present in our data set. To test this idea, we transferred the top 100 genes defining the Fraction 1 prePB

cells (Bcl6lowCD69high) -extracted from the RNAseq experiment in the Ise study- to the clusters in our study. We found that Fraction 1 prePB cells shared the highest transcriptional similarities with two of our clusters (new Fig. 2D). One of them is indeed within the LZ population, just like it was described in the Ise paper, which very nicely shows that both studies are compatible and comparable. The other identity is with the cluster that we originally designated aMem. In other words, the Fraction 1 population described by Ise et al shares the transcriptional prePB identity with a non-GC cluster named aMem in our original manuscript.

Indeed, aMem had been an elusive population for us. It was distinct but distally related to MBCs, which made us designate them as aMem. We had also been intrigued by the finding that, in the absence of AID, clonal sharing between aMem and PB was greatly increased. Our identification of aMem as part of the previously reported Fraction 1 prePB cells by the Kurosaki lab, nicely puts these pieces together and indicates that two of our clusters represent two distinct prePB subsets. Therefore, we have changed the aMem designation into early prePB (E-prePB) and prePB into late prePB (L-prePB) cells. We summarize here the main features that we have uncovered regarding these subsets:

- E-prePB cells represent a slightly earlier state than L-prePB according to pseudotime analysis (new Fig. 2F). This agrees with E-prePB harboring fewer mutations than L-prePB (Fig. 4A-C, new Fig. 4J), and with their showing fewer diversification steps than L-prePB cells (new Fig. 4M). In turn, this is consistent with milder expansion of E-prePB compared to L-prePB (Fig. 6C).
- In AID deficient mice, the proportion of E-prePB cells is increased, while the proportion of L-prePB cells is decreased (Fig. 5E). E-prePB cells are more clonally expanded in AID deficient mice (Fig. 6C).
- While in AID proficient mice PB show the highest clonal sharing with L-prePB (Fig. 4G, Fig. 6E), in the absence of AID the highest sharing of PB is with E-prePB (Fig. 6E).

Together, these results indicate that there are two alternative pathways of PB differentiation that involve two distinct prePB subsets, which differ in their differentiation/diversification state. L-prePB cells are part of a pathway that involves a higher degree of diversification and is therefore AID dependent.

Similarly, the finding of a reduction in the prePB subset in mice with AID-deficient activated B cells is intriguing. But one may like to see more discussion on why the plasmablast/plasma cell differentiation arm is specifically affected than that given in lines 374-376 "...probably reflecting changes in antibody affinity maturation and Ig isotype-dependent checkpoints." The authors should speculate more how one can envision that absence of AID can change affinity maturation in the plasmablast/plasma cell differentiation arm only.

We agree that this is an important point that deserves further discussion. As the reviewer knows, it is very well established that high affinity for antigen favors the differentiation towards the PB/PC fate, in a process that very likely involves both Tfh help and BCR signaling. MBCs typically harbor fewer mutations than PB/PC cells, but selection of GC B cells into the MBC fate is far less well understood. It is not clear whether MBCs arise stochastically from the pool of B cells that failed to be positively selected, or whether they are actively instructed to differentiate by weak positive selection signals. These events could correlate with an earlier

emergence of MBC precursors, as compared with PB precursors. These possibilities are very nicely discussed in a recent review article by Victora and Nussenzweig (PMID: 35113731). In this context, our findings agree well with the idea that selection into the PB/PC fate is crippled in the absence of AID. This can be explained both by the absence of somatic hypermutation, which precludes the generation of high affinity variants, and by the absence of class switch recombination, as IgG switching favors B cell competition in the GC (PMID: 33857421) and PC differentiation (PMID: 26944202).

We have now expanded our discussion on this point.

Figure 4 and section starting line 246 on the transitional clusters in AID deficient mice. A fold-difference, and the actual medium values, should be provided for Fig. 4B (and 4I) in the text. It seems the difference is two-fold compared to the control mice? With this in mind, Fig. 4E finds the same overall percentage of plasma cells (PB), and an about 2.5-fold difference for the prePB, meaning the corresponding cells are reduced but are not absent. The following figures convincingly support the conclusions drawn reg. the consequences of AID deletion, especially the branching such as in Fig. 5B to D. But the addition of some actual values reg. Fig. 4 seems appropriate.

We thank the reviewer for making this point; we have now included actual values in the text, as requested.

Response to Reviewer #2

We thank the reviewer for her/his insightful comments, which we address below.

Gómez-Escolar and colleagues use single-cell sequencing of immunized mice to study the relationship of B cell and plamsa cells and how this is modulated by AID deficiency.
Strength: This study uses advanced bioinformatic analysis to identify B cell subsets in an unbiased way. By using mice reporting the previous expression of AID, the authors can identify previously activated B cells. By comparing AID deficient to AID sufficient mice, they re-assess the role of AID, and thus class-switch-recommendation and antibody class switching, for MBC and PC differentiation.
Major Limitations:
1. A key weakness of the study is the lack of antigen specificity of the cells analysed. It is thus entirely unclear what relationship GC B cells have to PC and MBC and I would expect that while GC B cells may reflect predominantly cells responding to OVA, most MBC and PC were present before the immunisation.

We agree with the reviewer that analysis of antigen specificity can be very informative for a variety of purposes. We also agree that prior to immunization, wild type mice can have a bystander presence of MBC, PC, and also GC cells. Regarding this, we would like to point out that titers of anti-OVA antibodies are clearly increased upon immunization (we now include these data as new Fig. EV1C).

However, please note that the strength of our study is the establishment of clonal relationships across different subsets, regardless of their specificity. Finding a clone that contains prePB cells and PB, immediately establishes that they originate from the same precursor, regardless of their antigen specificity. Thus, while we are aware that the clones we analyze must contain cells that are not OVA specific, we need to stress that this does not alter our conclusions at all. In fact, we believe ours is a truly unbiased manner to establish clonal relationships in a large number of clones so that the findings are robust. In this sense, it is worth noting that selection of antigen specific B cells is strongly dependent on affinity (PMID: 33125897) (PMID: 34106207). Therefore, using antigen selection to analyze antigen specific B cells would inevitably 1) skew the analysis towards high affinity B cells, and 2) dramatically reduce the number of cells subject to the clonal analysis.

To sum up, while we agree that using an antigen specific approach is critical to ascertain a number of points, we must disagree that that is particularly relevant to our analysis. We have now included a sentence in the discussion that further clarifies this point.

2. One key finding, the identified prePB stage, is poorly characterised and the characteristics of these cells are not compared to existing literature characterising PC differentiation from GC. Rather, the authors speculate about an "alternative route to PB/PC differentiation without providing any molecular basis, such as changes in transcription factor expression, required for PC/PB differentiation. Thus, little evidence is provided that the prePB subset is indeed giving rise to PC.

We thank the reviewer for this suggestion. We had previously identified prePB cells as a distinct transcriptional entity and confirmed its existence by flow cytometry. We had provided information on somatic hypermutation and class switch recombination for this subset. We had further performed clonal analysis which showed that prePB is the subset that shares more clones with PB, indicating that prePB represent the most common precursor for PB cells. Finally, we showed that in AID deficient mice, this prePB→PB relationship is crippled. We now have expanded these results by comparing the prePB cells identified in our study with the Fraction 1 prePB cells reported by the Kurosaki lab (Ise et al PMID: 29669250).

Please note that the Ise study searched for such prePB population specifically within LZ Bcl6low cells, while we have performed our study on all cells with AID expression history. Transfer of the Ise prePB signature (100 most expressed genes) into our clusters identified two Kurosaki prePB subsets. One of them is indeed part of the GC.LZ cluster, in agreement with their report; the other subset corresponds to the cluster originally designated by us as aMem. This analysis thus indicates that aMem actually represents a part of the Kurosaki Fraction 1 prePB cells, which is distinct from our prePB cluster. Therefore, we have changed the aMem designation into early prePB (E-prePB) and prePB into late prePB (L-prePB) cells. We summarize here the main features that we have uncovered regarding these subsets:

- E-prePB cells represent a slightly earlier state than L-prePB according to pseudotime analysis (new Fig. 2F). This agrees with E-prePB harboring fewer mutations than L-prePB (Fig. 4A-C, new Fig. 4J), and with their showing fewer diversification steps than L-prePB cells (new Fig. 4M). In turn, this is consistent with milder expansion of E-prePB compared to L-prePB (Fig. 6C).
- In AID deficient mice, the proportion of E-prePB cells is increased, while the proportion of L-prePB cells is decreased (Fig. 5E). E-prePB cells are more clonally expanded in AID deficient mice (Fig. 6C).
- While in AID proficient mice PB show the highest clonal sharing with L-prePB (Fig. 4G, Fig. 6E), in the absence of AID the highest sharing of PB is with E-prePB (Fig. 6E).

This information is now included in our manuscript in the abovementioned figures and in the discussion section. Please also refer to our answer to Reviewer#1 on page 1 for additional details.

Also, it is unclear why the authors choose to analyse CD38+ GL7- cells for pre-PB cells (Figure 2c). CD38+ GL7- cells represent resting naïve and memory B cells unlikely to be on the way towards PC differentiation. Rather, PC precursors should be contained among the activated/GC (GL7 and/or FAS+) B cells. Based on the data presented, it appears more likely that what is labelled as "prePB" rather reflects Memory B cells.

We agree that this is an unfortunate choice because it may create confusion, given that CD38 is sometimes used as an MBC marker. However, to the best of our knowledge, CD38 does not define MBC identity, since it has no known role in MBC cell function and is actually expressed in naïve B cells as well. Please note that, within Tom+ cells, virtually all B220+ cells that are not GC (GL7+) or PB/PC (CD138+) are CD38+. This does not mean that all CD38+ are MBC, but rather, that those markers alone are insufficient to identify additional subsets. Indeed, identification of novel functional subsets invariably relies on new markers or new

combination of markers. After these considerations and for the sake of clarity, we now use the following gates for L-prePB definition: Tom+, B220+Cd138-,GL7-, FcRg+ (new Fig 3C). These cells are CD19+. For a more detailed discussion on the relatedness of prePB with MBC, please see our answer to the next point.

To confirm prePB identity, known genes expressed upon PC differentiation (such as IRF4 and PRDM1 expression) should be analysed.

We thank the reviewer for this comment. Please note, however, that the key point to define the prePB cluster identified in our study (now L-prePB) is that many of the PB/PC cells belong to the same clone as L-prePB cells (Fig. 4G-H, 6E-F). Given that PB/PC represent a final differentiation stage, the only possible explanation for clones containing L-prePB and PB cells, is that L-prePB are (or can be) PB precursors. In addition, L-prePB share transcriptional similarities with PB (Fig. 2B-C).

features.plot (a)	avg.exp (b)	pct.exp (c)	cluster	avg.exp.scaled
Bcl6	0,48	38,17	GC.DZ	1,10
Bcl6	0,72	30,80	GC.LZ	2,00
Bcl6	0,05	1,88	E-prePB	-0,56
Bcl6	0,15	4,71	Mem.Act	-0,18
Bcl6	0,02	1,13	Mem.GC	-0,68
Bcl6	0,04	1,37	Mem.Act2	-0,61
Bcl6	0,08	4,35	L-prePB	-0,43
Bcl6	0,03	3,23	PB	-0,64
Irf4	0,11	8,90	GC.DZ	-0,58
Irf4	0,09	4,73	GC.LZ	-0,61
Irf4	0,40	18,96	E-prePB	-0,01
Irf4	0,22	6,31	Mem.Act	-0,35
Irf4	0,35	13,84	Mem.GC	-0,10
Irf4	0,17	5,48	Mem.Act2	-0,46
Irf4	0,25	13,62	L-prePB	-0,30
Irf4	1,66	69,35	PB	2,42
Prdm1	0,02	1,95	GC.DZ	-0,37
Prdm1	0,04	2,19	GC.LZ	-0,25
Prdm1	0,05	2,50	E-prePB	-0,22
Prdm1	0,00	0,10	Mem.Act	-0,44
Prdm1	0,02	0,85	Mem.GC	-0,35
Prdm1	0,00	0,00	Mem.Act2	-0,46
Prdm1	0,01	1,45	L-prePB	-0,38
Prdm1	0,58	36,29	PB	2,47

Table RI. Expression of individual genes by 10x Genomics
(a) Gene name.
(b) Average expression level (normalized counts)
(c) Percent of positive cells

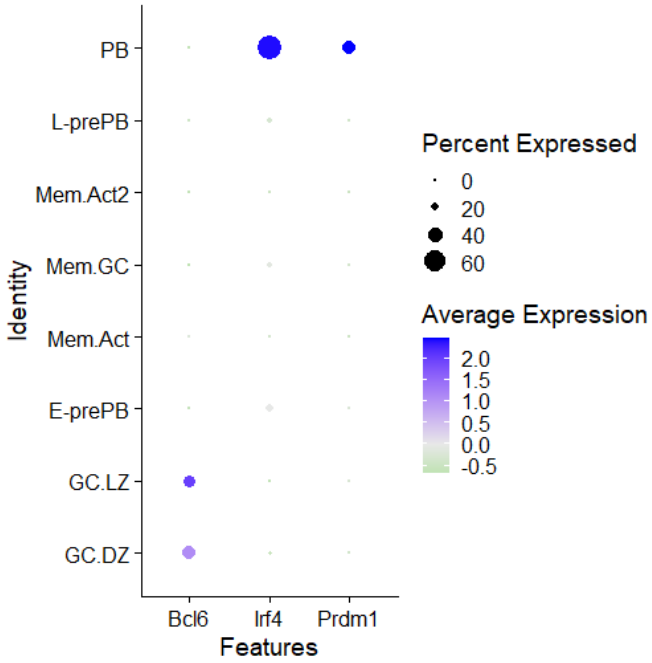

Figure R1. Dot plot representation of Bcl6, Irf4 and Prdm1 genes. Plot represents the percent of positive cells and the scaled averaged expression values shown in Table RI.

We have now followed the reviewer suggestion and assessed the expression of Irf4 and Prdm1 in the different clusters of our study (Table RI and Fig. R1). As a reference, we also include the data on Bcl6 expression. We detected Irf4 and Prdm1 expression in only 13,6% and 2,50% of L-prePB cells, respectively. However, we believe this result is hardly informative, because the expression levels of both genes are very low even in PB/PC cells (1,66 for Irf4 and 0,58 for Prdm1), which leads to the detection of only 69% Irf4+ cells and 36% Prdm1+ cells in this PB/PC cluster. Therefore, the sensitivity is not good enough to detect cells that presumably will have lower expression of these genes. Please note that Bcl6 is detected in less than 40% of GC B cells. This high proportion of negative cells for individual genes is due to a phenomenon called “dropout”, where a gene is observed at a low or moderate expression level in one cell but is not detected in another cell of the same cell type. These dropout events occur due to the low amounts of mRNA in individual cells and inefficient mRNA capture, as well as the stochasticity of mRNA expression. The excessive zero counts cause the data to be zero-inflated, only capturing a small fraction of the transcriptome of each cell drop-out events (PMID: 32127540). This is particularly important with genes that are expressed at relatively low levels and to mRNAs with particularly low stability. Thus, signature analyses, rather than individual genes, are the most reliable way to assign transcriptional identities to clusters obtained from single cell data.

This being said, and in connection with the previous point raised by the reviewer, we agree that L-prePB also share some transcriptional similarities with MBCs. These similarities were pointed out before and are further stressed in the revised manuscript. Whether those similarities entail a true MBC identity cannot be ascertain within the scope of this revision. We however favor the notion that L-prePB are in a transitional differentiation state between the GC and the PB stages, and that their transcriptional overlap with MBCs reflects the lack of strong GC or PB signatures. The exact same situation would apply to E-prePB, which we formerly named aMem based on their distal similarity to MBC, but which we now know are part of the Fraction 1 prePB cells described by the Kurosaki lab. We must stress here that indeed both E-prePB and L-prePB are only distally related to MBCs, because they appeared as

distinct clusters, separated from MBCs in our first clusterization. Finally, we would like to point out that even if L-prePB were true MBC -which we think is highly unlikely, as expressed above-, this would not invalidate our designation as prePB, in that it would be equally feasible that they were the precursors of PB.

We have now expanded our discussion to include these issues.

To the best of my knowledge, FcεRIγ is not known to be expressed in B cells. While it is possible that a small subset indeed expresses FcεRIγ, staining with anti-CD19 and/or anti-Ig should be used to confirm B cell identity of FcεRIγ expressing cells as not all B220+ cells (Nikolic et al., EJI 2002) are B cells.

We have included a plot showing that L-prePB, FceRIg+ (FcRγ+ in the revised version) cells are CD19+ (new panel in Fig. 3C)

Along these lines, it is apparent in Figure 2 A that a small, contaminating cluster of FcεRIγ expressing cells exists in addition to the B and PC clusters, highlighting the possibility that not all cells identified as Tom+ are B or PC. While such contamination is commonly occurring and at a major issue, any conclusions about small cell populations need to be thoroughly validated.

The reviewer refers to a small group of 10 cells at the top right corner of UMAPs (Fig. R2, labeled in black for more clarity). For comparison, L-prePB cluster contains 345 cells. These 10 cells are heterogeneous and are not assigned as an independent cluster, but instead Seurat assign them into the GC, E and L-prePB and Mem.GC clusters. Therefore, their appearance as a group most likely is an artifact of the UMAP representation. Indeed, we have verified that if we represent the data using UMAP projection in three dimensions (instead of 2 dimensions), the cells are correctly integrated into their respective clusters (Fig. R2). Thus, we believe that this inexactitude in the plot results from the reduction of multidimensional data into two dimensions. However, since this representation has no consequence in cluster assignation it does not affect at all our analysis.

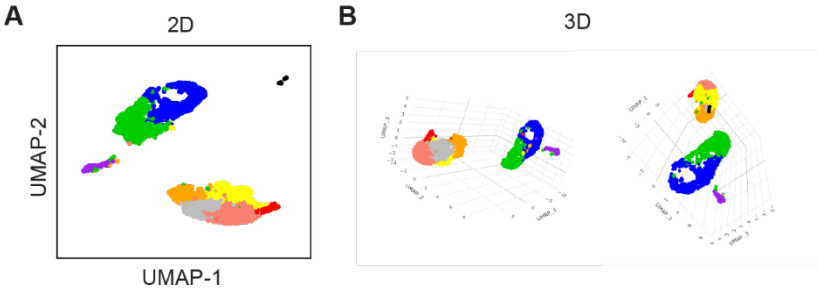

Figure R2. A. UMAP plot in 2D. **B.** Same cells in A represented with components UMAP-1, UMAP-2 and UMAP-3.

3. Little to no comparison of proposed MBC clusters with previously described genes expressed in MBC subsets is made.

We thank the reviewer for this suggestion. To address this point we have used recently published transcriptional data sets. We found that globally, our Mem clusters are enriched in the MBC signature published by Glaros et al (PMID: 34525339) (new Fig. 1G). In another recent study, Viant et al showed, making use of a lineage-tracking approach, two types of MBCs, i.e., activated B cell-derived and GC-derived, which differ in isotype, overall gene expression, somatic hypermutation and antigen affinity (PMID: 34106207). We have found that our Mem.GC is enriched in the transcriptional signature of Viants's GC-derived MBCs, while Mem.Act and the minor cluster Mem.Act2 are transcriptionally similar to activated B cell-derived MBCs (new Fig. 2A). In addition, Mem.GC harbor higher mutational load and CSR than Mem.Act cells. Thus, these new data provide strong additional support on the Mem clusters identified in our study.

4. The authors suggest using AID-based genetic tracer allows them to isolate GC-derived cells, while at the same time acknowledging that extrafollicular cells will also express AID. It remains unclear what advantage the use of the AID reporter has beyond excluding naive B cells for the analysis performed here.

Throughout the text, we profusely make the point that we are using a mouse model that labels cells that have expressed AID. It is well known, and we state so in the text, that precise labeling of all and only GC B cells is not possible with this model; it is probably not possible with any other mouse model available. We find the *Aicda-Cre*$^{+/ki}$*;R26tdTom*$^{+/ki}$ model extremely useful, for instance, as mentioned by the reviewer, to exclude naïve B cells from the analysis, which are not always easy to tell apart from MBCs. The model also allows us to detect GC-independent MBCs. But most importantly, the *Aicda-Cre*$^{-/ki}$*;R26tdTom*$^{+/ki}$ model allows us to precisely address the loss of AID in those cells that have been programmed for AID expression. We have further detailed this in the discussion.

4. Effects of AID deficiency are analysed in the absence of considering BCR isotype and affinity for antigen, both dependent on AID and well known to predispose to fate choices in GC responses. GC response kinetics, magnitude, duration and output are all changed in the absence of AID, making any conclusion about cause and consequence difficult. It is unclear to me what question the authors wanted to answer with their experiments.

As mentioned by the reviewer and in our introduction, AID, as the initiator of somatic hypermutation and class switch recombination, is obviously a critical player in the GC reaction. However, the role of AID in regulating GC fate choices is not well understood. We approached this issue by performing a detailed single cell analysis of clonal relationships among different transcriptional transcripts, which has led us to the identification of an AID-dependent pathway for PB/PC differentiation. We believe this is a novel and valuable piece of information that helps understanding GC dynamics.

Other comments:
- It is unclear why OTII T cell transfer and OVA immunisation was chosen as a model antigen system. Numerous other antigens exist which do not require the transfer of T cells. Can the authors provide a rationale for their choice of experimental system?

We believe that this immunization protocol is relatively standard and yields robust GC responses. In addition, we also confirmed the identification of prePB cells using an alternative, NP-CGG immunization protocol (these data have now been moved to Fig. EV4, as requested by R#3)

- FceR1g is typically referred to as FcRg (see Brandsma et al., Immunity, 2016)

We now refer to FceR1g as FcRg, as suggested by the reviewer.

- Figure 1C: Unclear what % cells refers to.

Figure 1C depicts the % of Tom+ cells, germinal center B cells (GC; Tom+ CD138- GL7+), plasma cells/plasmablasts (PB; Tom+ CD138+) and putative memory B cells (pMem; Tom+ CD138- GL7- CD38+) within live cells. The label on the Y axis has now been changed to clarify this point.

- Is the clustering of DZ and LZ cells predominantly based on cell cycle genes? While proliferation is associated with DZ, it is typically also seen in a fraction of LZ cells.

Seurat analysis of our scRNA seq data identified clusters 1 and 2, both of which expressed a number of GC-associated genes (Fig. 1F, new Fig. 1G, Table S1). Cluster 1 was labeled as GC.LZ and cluster 2 was labeled as GC.DZ based on their respective enrichment in LZ and DZ genes, as defined by Victora et al (PMID: 21074050). These gene sets include cell cycle genes, as well as many other genes. We have performed an additional analysis which shows that regressing out the effect of cell cycle does not have a mayor impact on the cluster analysis, as shown in Figure R3.

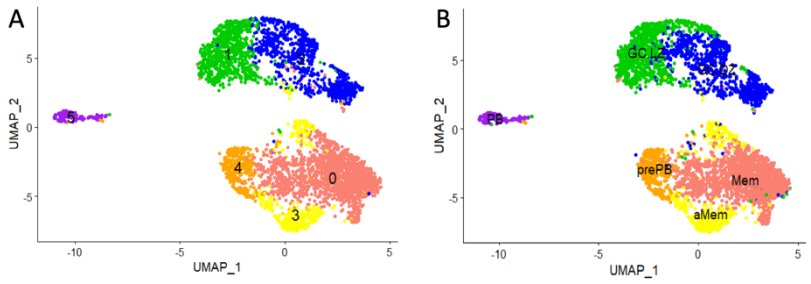

Figure R3. Regressing out cell cycle effect does not have a major effect on clusterization. **A.** UMAP plot showing clusters identified after regressing out cell cycle effect. **B.** UMAP plot in A colored by original cell identities (without cell cycle regression).

Further details on GC.LZ and GC.DZ clusterization are now included in the manuscript.

Response to Reviewer #3

We thank the reviewer for her/his constructive comments and suggestions, which have considerably help us clarify several points.

Gomez-Escolar and colleagues use an AID-driven reporter to investigate AID-experienced B cells during secondary immune responses to the ovalbumin (OVA) protein in wild-type and AID-deficient mice. Single-cell transcriptomic and VDJ analyses on cells expressing the reporter fluorescent protein Tomato (Tom) identified distinct transcriptional cluster representative of germinal center (GC) dark (DZ) and light zone (LZ) B cells, plasmablasts/plasma cells (PB), precursor plasmablasts (prePB) and memory B cells (including 4 distinct subgroups). The authors further characterized the prePB cells that displayed expression of several markers including Fcer1g. The expression of Fcer1g was confirmed at protein level and it was used as a marker to identify the prePB population by cytofluorimetric analysis of Tom+ cells. Mice lacking AID displayed a reduced proportion of prePB and, as expected, GC expansion and reduced DZ/LZ ratio. Although it is known that AID-deficiency in mice leads to a reduction in plasma cells, the data provided here suggest that the plasma cell impairment is related to the reduced prePB population.
Overall, this work provides a transcriptomic characterization at single-cell level of AID-experienced B cell populations and of the effects of AID deficiency on these populations. The identification of a previously unrecognized precursor plasmablast population that is related also to memory B cells and is affected by AID loss, provides the basis to further explore the interaction between different post-GC subpopulations in wild-type and AID-deficient models.
The experimental approaches and data analyses are adequate to address the biological question that is of interest for the field of antibody-mediated immunity. However, some aspects should be clarified and discussed more in detail (see below). In particular, the authors should convey more clearly the information that relate the prePB population not only to plasma cells but also to the memory compartment. In addition, an unbiased analysis of the transcriptomic data from the AID-deficient cells should be considered.

These points are fully addressed under specific comments below.

Specific Comments:

• The number of analyzed cells and their distribution in the different subpopulations should be reported for each of the 4 specimens. It should also be clarified if all analyses are performed only on cells for which good quality data are available for both RNA and VDJ-C.

We thank the reviewer for bringing up this point on quality checks. Transcriptomic analyses were performed with all the cells that passed the Seurat quality filters for gene expression data. We have now included the number of cells sequenced for each of the samples both in the Materials and methods section and in the corresponding figure legends. For VDJ analysis, we used only those events that passed the corresponding Immcantation quality filter. This information has been included in the Materials and methods section. Specifically, from a total of 8116 cells with high quality transcript information, 5856 cells (72%) also yielded high quality VDJ information.

• Figure 1D: please clarify which clusters are identified initially when analyzing all cells together and which ones are identified after sub-clusterization of the memory group. In particular, it is unclear whether the prePB cluster is identified only after sub-clusterization.

We very much appreciate this comment. We have now very minutely explained the process of clusterization, subclusterization and cluster identity and we believe that this has considerably clarified this section; briefly:

- To answer the specific question of the reviewer, we show that initial clusterization includes clusters 0 to 5, where 4 are prePB cells (Fig. EV2A). Therefore, the prePB cluster was an independent subset from this initial clusterization, as was cluster 3 of aMem (now renamed as E-prePB cells).
- Subclusterization only affected cluster 0. To clarify this point, Fig. 1D and Fig. 1E have been modified and now show the numbers on the clusters and cluster 0 is divided into 0a, 0b and 0c, which correspond to MBCs.
- The identity of the different clusters is more extensively detailed and justified in the text in a stepwise manner. New Fig. 1G has been included to help explain the assignation of identities.

• Figure 1D: the principal component analysis clearly places the prePB closer to the memory clusters than to the PB cluster. The authors are analyzing a secondary immune response (2 weeks after a second antigen exposure): are these prePB coming from memory B cells generated during the primary immune response rather than being the products of recent GC reactions? This aspect needs to be discussed more in detail.

We specifically approached the reviewer's concern by performing new immunization experiments and measuring L-prePB cells after a single antigen exposure or after the second exposure. We find that L-prePB cells can already be detected after the first immunization (new Fig. EV2E), in agreement with our previous findings. While this observation does not formally exclude that L-prePB represent MBCs, it is compatible with the idea that they are a distinct subset, as further discussed below.

The authors emphasize the relationship between prePB and PB, however the relationship with memory should also be detailed more clearly especially based on: i) the relatedness highlighted in the UMAP (Fig. 1D); ii) the large fraction of markers shared with memory in particular Mem1 (Fig. 1I); iii) the IgH V mutation pattern that resembles both PB and Mem1 (Fig. 3C); and iv) the expression of mostly unswitched IgH (Fig. 3E).

prePB cells (now L-prePB) share transcriptional similarities with PB/PC, but also with MBC cells (Fig. 2B), as pointed out by the reviewer. These similarities were pointed out before and are further stressed in the revised manuscript. Whether those similarities entail a true MBC identity cannot be ascertain within the scope of this revision. We however favor the notion that L-prePB are in a transitional differentiation state between the GC and the PB stages, and that their transcriptional overlap with MBCs reflects the lack of strong GC or PB signatures

(we kindly refer the reviewer to a more detailed discussion in our response to Reviewer#2 on pages 6 and 7 above). The exact same situation would apply to E-prePB, which we formerly named aMem based on their distal similarity to MBC and on their expression of Fcrl5, but which we now know are part of the Fraction 1 prePB cells described by the Kurosaki lab. We must stress here that indeed both E-prePB and L-prePB are only distally related to MBCs, because they appeared as distinct clusters, separated from MBCs in our first clusterization. Finally, we would like to point out that even if L-prePB were true MBC -which we think is highly unlikely, as expressed above-, this would not invalidate our designation as prePB, in that it would be equally feasible that they were the precursors of PB.

• Figure 1D: a small cluster of cells located in the top right quadrant of the UMAP should be checked for quality or other factors driving the distinct transcriptome.

The reviewer refers to a small group of 10 cells at the top right corner of UMAPs (Fig. R2, labeled in black for more clarity). For comparison, L-prePB cluster contains 345 cells. These 10 cells are heterogeneous and are not assigned as an independent cluster, but instead Seurat assign them into the GC, E and L-prePB and Mem.GC clusters. Therefore, their appearance as a group most likely reflects an artifact the UMAP representation. Indeed, we have verified that if we represent the data using UMAP projection in three dimensions (instead of 2 dimensions), the cells appear correctly integrated into their respective clusters (Fig. R2). Thus, we believe that this inexactitude in the plot results from the reduction of multidimensional data into two dimensions. However, since this representation has no consequence in cluster assignation it does not affect at all our analysis.

• Figure 1E: It should be clarified how the specific populations are labeled. For example, the authors refer to previously published DZ/LZ signatures: however the DZ- and LZ-signature scores are >0 only in a fraction of the cells that are called DZ/LZ (Fig. S2A, D).

The clustering and labeling of the different populations is now explain stepwise as described above. In the specific case of the GC.LZ and GC.DZ, they were obtained as two distinct clusters (1 and 2). We show in new Fig. 1G that both clusters are enriched in a GC signature. As for LZ and DZ signatures (obtained from the Victora dataset PMID: 21074050), we find that the scores for LZ and DZ are highest at the ends of the corresponding cluster and progressively decrease towards the proximity to the other cluster (new Fig. 1G). We think that this fading of the signatures a nice indication of the iterative transitions between the LZ and the DZ states.

• Figure 3B, D: do the cells used for the analysis carry complete information about heavy and light immunoglobulin chains? Usually not all cells with good RNA expression data have equally informative VDJ and constant region data. The number of cells with suitable VDJ data should be reported in the legends, results and/or in the methods.

From a total of 8116 cells with high quality transcript information, 5856 cells (72%) also yielded high quality VDJ information. These numbers are now included in figure legends and in materials and methods.

• Figure 3A-E: PrePB and PB show about 50% of cells completely unmutated and a very high proportions of unswitched cells is observed in GC (>50%) and in PB (>75%). These values appear low. Please comment.

We agree with the reviewer that these values are possibly in the low range for SHM and CSR, but not dramatically lower than in other immunization experiments. In addition, we would like to point out that our 10x genomics data roughly correlate with flow cytometry data (Figure S1B) and that the rate of CSR generally correlates with SHM load (Figure S3B-D), indicating that the experiment is consistent and robust.

• Figure 4C-E: The DZ/LZ ratio reduction is shown by FACS analysis but it doesn't seem to be captured by the sc-transcriptomic data.

The sc-transcriptomic data also captures an increased proportion of LZ B cells in AID deficient mice (Fig. 5E, star shows significant difference for GC.LZ). However, we agree with the reviewer that we observed a bigger difference by FACS than by sc 10x genomics. We think that one plausible explanation for this difference is that the definition by FACS is based only on gates on the CD86 and CXCR4 markers and may not be as precise as the transcriptional multiparametric clusterization used by Seurat.

• No UMAP plots are shown for the AicdaCre/- mouse sc-transcriptomic data. Did this analysis identify the same subpopulations as in the normal mice? The authors mention only "transfer of the transcriptome data into the clusters defined for wild type B cells" (Page 11, lines 267-268). An independent analysis of AicdaCre/- mouse sc-transcriptomic data should be provided.

We find that the standard way of comparing two datasets is the transfer of cell type labels from a reference dataset onto a new query dataset (standard Seurat workflow). However, we very much appreciate the reviewer's suggestion and have now performed independent clustering of AicdaCre/- samples, which again yielded clusters 0-5 (Fig. R4A). Transferring AicdaCre/+ transcriptome cell labels intro the AicdaCre/- transcriptomic data yielded the labeling shown in Fig R4B. We find an almost perfect correlation with the clusters obtained before.

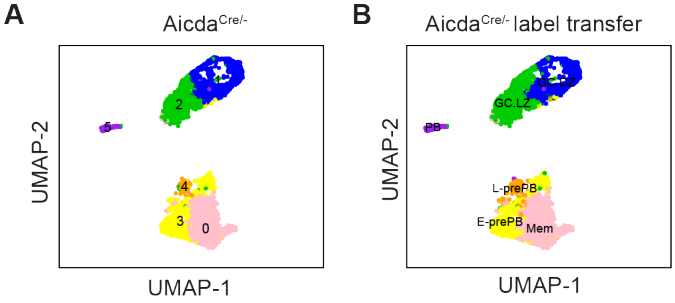

Figure R4. A. UMAP plot showing the 6 clusters obtained in Aicda Cre/- mice. **B.** Label transfer of Aicda Cre/+ data to Aicda Cre/- cells in A.

• The circus plot obtained from the wild-type mice and displayed in Figure 3H is different from the one showed in Figure 7A (left panel). See for example at the sharing between aMem and GC. Please clarify.

We agree that these different representations can be confusing and we thank the reviewer for pointing it out. Both types of circos representations depict only pairwise relationships for the sake of clarity, i.e. they only show clones shared by two clusters. In the first circos plot (now as Fig. 4G) GC.LZ and GC.DZ are shown separately, which very nicely shows the rich clonal relationship between both clusters, representing more than 50% of all the clonal interactions; this can be obviously expected from the known iteration of cells between LZ and DZ. However, this dominance of LZ/DZ clones, and the fact that only clones shared by two clusters are shown, hinders the visualization of clones shared by the rest of the clusters. Thus, in the circos plots in Fig. 6E we decided to plot together GC/DZ and GC.LZ as a compound cluster (i.e. we do not show their clonal relationships) so that we could show a better view of the rest of the clusters, including the separate Mem clusters. In the example provided by the reviewer, the circus in 6E shows a big proportion of clones shared by E-prePB (previously aMem) and GC because some of them are shared both by GC.LZ and GC.DZ (i.e., among 3 clusters), and were therefore not shown in Fig. 4G. Therefore, both types of representations are simplifications that provide slightly different perspectives to the data. We have now explained this in more detail in the text. Please also note that complete information of all clones is provided in Table S3 and in Fig. EV5A-B.

Minor details:

• Fig. 1E: why the clusters are ordered by size (I guess) rather than biology or relatedness? Number of cells in each cluster should be provided

Clusters are by default ordered by size. These numbers are used to initially label clusters in Fig. 1D and E. The new stepwise description of clusterization and labeling, also further clarifies this point. Number of cells in each cluster are now included in figure legends.

• Fig. 1 legend: Panel E refers to clusters identified in D not C.
This mistake has been amended.

• Fig. 2A legend: unclear the reference to Fig. S2A.
We have now clarified the reference to Fig S2A (Fig EV3A in the revised version).

• Fig. 4C: the percentage of cells in each gate should be shown.
Percentages are now shown (note that this is now Fig. 5C)

• Fig. 4F legend: it mentions a booster with OVA but the scheme shows NP-CGG.
This mistake has been amended.

• Fig. 4 legend: Panel H is not described and panel I is not listed. The ANOVA refers to a not existing panel.
This mistake has been amended

• Fig. 5D legend: no left and right panels.
This mistake has been amended

• Fig. 7 legend: the text mention p-values as asterisks but no asterisk is displayed in the figure.
This mistake has been amended.

• Fig. S3A legend: not clear to which data/figure these p-values refer to.
Reference to Fig 4A is shown in S3A legend.

• Some panels in Figure 4 (F-I) could be moved to Supplemental and the rest could be merged with Figure 5 to have only one figure on the population changes in AID-deficient mice. Similarly, Figure 6 and 7 could be merged displaying only one figure on clonal features and the transitions inferred from them.
We have followed all these suggestions

Dear Dr. Ramiro,

Thank you for the submission of your revised manuscript to our editorial offices. I have received the reports from the three referees that were asked to re-evaluate your study that I have already forwarded to you (you will also find again below). As you know, referees #1 and #3 now fully support the publication of your study in EMBO reports, whereas referee #2 remains critical and indicates remaining concerns. Looking through the point-by-point response (revision plan) you sent, and after cross-commenting with referees #1 and #3, I think that the remaining points referee #2 will be adequately addressed during a further revision as indicated in your letter. I thus ask you to address these comments in a final revised manuscript. Please also take care of the minor points of referee #2. Please also provide a detailed final point-by-point-response to the final referee concerns.

Moreover, I have these editorial requests I also ask you to address:

- We updated our journal's competing interests policy in January 2022 and request authors to consider both actual and perceived competing interests. Please review the policy https://www.embopress.org/competing-interests and update your competing interests if necessary. Please name this section 'Disclosure and Competing Interests Statement' and put it after the Acknowledgements section.

- We now use CRediT to specify the contributions of each author in the journal submission system. CRediT replaces the author contribution section. Please use the free text box to provide more detailed descriptions. Thus, please remove the author contributions section from the manuscript text file. See also guide to authors: https://www.embopress.org/page/journal/14693178/authorguide#authorshipguidelines

- Please make sure that the manuscript sections are ordered like this (using these names):
Title page - Abstract - Keywords - Introduction - Results - Discussion - Materials & Methods - DAS (data availability section) - Acknowledgements - Disclosure and Competing Interests Statement - References - Figure legends - EV Figure legends.

- Please remove the referee token (password) from the data availability section and make sure the data are public latest upon publication of the study.

- 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 (main, EV and Appendix figures), and that statistical testing has been done where applicable. Please avoid phrases like 'independent experiment', but clearly state if these were biological or technical replicates. Please add complete statistical testing to all diagrams (main, EV and Appendix figures). Please also indicate (e.g. with n.s.) if testing was performed, but the differences are not significant. In case n=2, please show the data as separate datapoints without error bars and statistics.

- It seems there is no clout for Fig. 1D. Please check and make sure that all figure panels are called out separately and sequentially.

- Please name the three Tables (S1-3 or EV1-3) Datasets, using the nomenclature Dataset EVx. Please also change the callouts accordingly. Please add a title and a legend for these on the first TAB of the excel file and upload the revised files as datasets. Finally, please remove the legends for Tables EV1-3 from the manuscript main text file.

- Finally, please find attached a word file of the manuscript text (provided by our publisher) with changes we ask you to include in your final manuscript text, and some queries, we ask you to address. Please provide your final manuscript file (using the attached file as basis) 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 (two lines each).
- a schematic summary figure (in jpeg, png 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.

Please use this link to submit your revision: https://embor.msubmit.net/cgi-bin/main.plex

Best,

Achim Breiling

Editor Senior
EMBO Reports

Referee #1:

I have read the response to my criticism and the revised manuscript. The authors have satisfactorily addressed my concerns and, in my view, the manuscript is much stronger now and provides important new information for the field.

Referee #2:

The authors have extensively revised the manuscript which now includes a more thorough integration of the presented data with existing literature. Some aspects of the study are potentially interesting, however, some concerns remain.

1. I remain very sceptical about the validity of the FcgR staining and the use of this molecule to identify a novel B cell subset. I appreciate that the authors have done some validation but all data could still be explained by a combination of contaminating cells and unspecific staining. Given the authors want to conclude strongly, arguing that FcRg is (1) expressed in B cells (which has not been reported before) and (2) allows to identify a new transitional stage, in my opinion, this warrants the need for extensive validation. Reasons for my continued scepticism are:
- a very sizeable fraction of Tom+ cells in Figure 3c is comprised of non-B/PC. This population is part of RNA-seq and flow cytometry data.
- figure EV2D & EV4C shows that a significant proportion of FcRg+ (and FcRg-) cells have a very high FSC-A signal, suggesting these events reflect cell duplets, in particular given no duplet exclusion is performed throughout the study (no complete gating strategy is shown).
- The frequency of the FcRg staining obtained in figure 3c is in the range of what is typically seen as "sticky" staining of IgD/M+ B cells, in particular given the staining was done intracellularly and using a polyclonal antibody.
- FcRg is a cell membrane protein which should not require intracellular staining. What is the rationale for using a fixation method that allows intra-nuclear protein detection to stain FcRg?
- qPCR comparison of total Tom+ vs rare, sorted cells that show FcRg staining could be due to non-exclusion of duplets and thus contamination with other cells that highly express FcRg. The rarer the population, the more likely are contaminations.

I appreciate that unambiguously validating the staining is difficult, but several additional steps should be undertaken to further investigate the identity of the FcRg+ subset:
- co-staining with markers to exclude possible contaminating subsets, such as for example GR-1 and NK1.1, to exclude myeloid and NK cells, respectively.
- control staining with secondary antibody/streptavidin only and/or an isotype control antibody should be shown (it is unclear how exactly FcRg was detected).
- staining with cells known to express/do not express FcRg should be shown alongside B cells. This needs to be done with and without fixation to confirm that fixation retains the specificity of the anti-FcRg antibody. I could not find any other report using this antibody to stain murine cells (although the vendor specifies it should be cross-reactive) and all human/macaque studies I could find perform surface staining.
- Ideally, FcRg ko mice (readily available) should be used to confirm the staining.
- co-staining with a monoclonal anti-FcRg, if available.
- back-gating the L-prePB population in particular analysing their scatter properties, duplet exclusion (fsc-a vs fsc-h; fsc-a vs fsc-w), level of dead-cell staining, autofluorescence level in "empty channels" and expression of B220, CD19, GR-1, NK1.1.
- show that FcRg staining is indeed restricted to GL7- non-GC B cells, as suggested by scRNA-seq.
- show that L-prePB contain class-switched cells (staining for IgD, IgM, IgG, IgA). These are typically less "sticky" and should be contained among L-prePB cells.
- The representation of the data in figure 3D is difficult to interpret. I would suggest showing ddCt values for L-prePB and Tom+ cells side-by-side to allow the reader the see the signal intensity obtained for the individual genes in both subsets. Comparison to cells known to express FcRg would be beneficial.
- show that FcRg transcripts in the RNA-seq data can be used as a "marker" to identify L-prePB cells. The authors mentioned in their response that 345 cells make up the L-prePB cluster. However, looking at Figure 3A, it appears that while the clusters of FcRg, Actn, Dnm3 and Ptpn22 are in the same area (the designated L-prePB cluster) each is only expressed in a fraction of L-prePB cells and the expression is largely non-overlapping. Thus, FcRg+ cells are, if validated, more likely a small subset within the identified L-prePB subset, making it increasingly difficult to assign a specific characteristic to these cells. I would suggest to analyse FcRg+ cells from sc-RNA-Seq separately and compare them to the remainder of the dataset.

2. The interpretation of the data, in particular with regards to AID deficient mice, is sometimes misleading. For example, the final sentence of the abstract states that AID enables PB differentiation from a prePB stage. There is little information in this sentence as PB must, of course, come from a prePB cell, whatever the identity of this cell is. However, to which extent AID directly influences pre-PB to PB differentiation is not investigated by this study. Also, the identity of the L-prePB as "PC precursors" rather than MBC remains largely speculative as it is mostly based on clonal relationship. Other than that, there is no indication

that these cells are indeed on the way towards differentiating PC rather than being resting MBC (and if they are resting MBC, whether they preferentially differentiate into PB upon re-activation). Can the authors show preferential ex-vivo PC differentiation of L-prePB to provide evidence that these cells are really committed towards PB? Such in vitro culture systems are available (see Kurosaki, Kitamura)? This would also be used to further confirm the identity of the FcRg+ population.
It must be expected, and in fact is known, that pretty much all response characteristics are changed in the absence of AID. For example, the reduction in L-prePB and PC could simply reflect the cumulative effect of changes in T cell mediated selection, response duration, time of response magnitude, absence of class switched cells (which can have different propensities for differentiation) and absence of affinity maturation. Along the same lines, the multiple underlying causes are likely responsible for changes in clonal relationships within transitional stages in a system that prevents vs. does not prevent clonal diversification. This is further complicated by the experimental setup as the data reported here is from a single, late time point following secondary antigen challenge which makes drawing conclusions about cause and consequence close to impossible, in particular in a model knockout system to is known to have major consequences for numerous aspects of the response.
Overall, I think this study has too many limitations to provide a significant conceptual advance to what is already known.

Minor comments/suggestions:
Introduction:
- The abbreviation "L-prePB" is not explained in the abstract.
- Line 22: Consider mentioning short-lived plasmablasts as additional fate choice.
- Line 33: ... deaminates the region encoding the antigen-recognizing, variable part of immunoglobulin genes, generating mutations which can give rise to variants with altered affinity for antigen.
- Line 39: ... exchange of IgM/IgD isotypes with IgG, IgE or....
- Line 44-46: I would suggest to re-word this as all B cells, not only those with increased BCR affinity are undergoing selection by Tfh. What differs is their fate.
- Line 56: Consider changing to: Accordingly, the average frequency of SHM is higher in PCs than in MBCs.
- Line 62: consider changing "looser" to something like "a broader range of affinities". "of B cells" could be deleted as it is contained within "MBC".
Results:
- Complete flow cytometry/FACS gating strategies should be shown for all data.
A mix of present and past tense is used in the results section, in particular in the first part. I would suggest to consistently use past tense.
- Line 93: consider changing "coding for" to "encoding"

Referee #3:

The authors have addressed my major concerns.

Point-by-point Response to Reviewers
Gomez-Escolar et al. EMBOR-2022-55000V2

Referee #1
I have read the response to my criticism and the revised manuscript. The authors have satisfactorily addressed my concerns and, in my view, the manuscript is much stronger now and provides important new information for the field.

AU RESPONSE: We are grateful to Referee #1's constructive suggestions and positive assessment of our manuscript.

Referee #2
The authors have extensively revised the manuscript which now includes a more thorough integration of the presented data with existing literature. Some aspects of the study are potentially interesting, however, some concerns remain.

1. I remain very sceptical about the validity of the FcgR staining and the use of this molecule to identify a novel B cell subset. I appreciate that the authors have done some validation but all data could still be explained by a combination of contaminating cells and unspecific staining. Given the authors want to conclude strongly, arguing that FcRg is (1) expressed in B cells (which has not been reported before) and (2) allows to identify a new transitional stage, in my opinion, this warrants the need for extensive validation. Reasons for my continued scepticism are:
- a very sizeable fraction of Tom+ cells in Figure 3c is comprised of non-B/PC. This population is part of RNA-seq and flow cytometry data.
- figure EV2D & EV4C shows that a significant proportion of FcRg+ (and FcRg-) cells have a very high FSC-A signal, suggesting these events reflect cell duplets, in particular given no duplet exclusion is performed throughout the study (no complete gating strategy is shown).
- The frequency of the FcRg staining obtained in figure 3c is in the range of what is typically seen as "sticky" staining of IgD/M+ B cells, in particular given the staining was done intracellularly and using a polyclonal antibody.
- FcRg is a cell membrane protein which should not require intracellular staining. What is the rationale for using a fixation method that allows intra-nuclear protein detection to stain FcRg?
- qPCR comparison of total Tom+ vs rare, sorted cells that show FcRg staining could be due to non-exclusion of duplets and thus contamination with other cells that highly express FcRg. The rarer the population, the more likely are contaminations.

I appreciate that unambiguously validating the staining is difficult, but several additional steps should be undertaken to further investigate the identity of the FcRg+ subset:
- co-staining with markers to exclude possible contaminating subsets, such as for example GR-1 and NK1.1, to exclude myeloid and NK cells, respectively.
- control staining with secondary antibody/streptavidin only and/or an isotype control antibody should be shown (it is unclear how exactly FcRg was detected).
- staining with cells known to express/do not express FcRg should be shown alongside B cells. This needs to be done with and without fixation to confirm that fixation retains the specificity of the anti-FcRg antibody. I could not find any other report using this antibody to stain murine cells (although the vendor specifies it should be cross-reactive) and all human/macaque studies I could find perform surface staining.
- Ideally, FcRg ko mice (readily available) should be used to confirm the staining.
- co-staining with a monoclonal anti-FcRg, if available.

- back-gating the L-prePB population in particular analysing their scatter properties, duplet exclusion (fsc-a vs fsc-h; fsc-a vs fsc-w), level of dead-cell staining, autofluorescence level in "empty channels" and expression of B220, CD19, GR-1, NK1.1.
- show that FcRg staining is indeed restricted to GL7- non-GC B cells, as suggested by scRNA-seq.
- show that L-prePB contain class-switched cells (staining for IgD, IgM, IgG, IgA). These are typically less "sticky" and should be contained among L-prePB cells.
- The representation of the data in figure 3D is difficult to interpret. I would suggest showing ddCt values for L-prePB and Tom+ cells side-by-side to allow the reader the see the signal intensity obtained for the individual genes in both subsets. Comparison to cells known to express FcRg would be beneficial.
- show that FcRg transcripts in the RNA-seq data can be used as a "marker" to identify L-prePB cells. The authors mentioned in their response that 345 cells make up the L-prePB cluster. However, looking at Figure 3A, it appears that while the clusters of FcRg, Actn, Dnm3 and Ptpn22 are in the same area (the designated L-prePB cluster) each is only expressed in a fraction of L-prePB cells and the expression is largely non-overlapping. Thus, FcRg+ cells are, if validated, more likely a small subset within the identified L-prePB subset, making it increasingly difficult to assign a specific characteristic to these cells. I would suggest to analyse FcRg+ cells from sc-RNA-Seq separately and compare them to the remainder of the dataset.

AU RESPONSE: We would like to note that R2 had already touched on this question in his/her review of our original manuscript and specifically asked for *"staining with anti-CD19 and/or anti-Ig should be used to confirm B cell identity of FcεRIɣ expressing cells as not all B220+ cells (Nikolic et al., EJI 2002) are B cells"*. We precisely followed R2's suggestion and had included in our revised manuscript the requested plot showing that L-prePB, FceRIg+ (FcRɣ+ in the revised version) cells are CD19+ (new panel in Fig. 3C). Upon R2 revisiting this point, we have now included broad additional information on the "*the validity of the FcgR staining and the use of this molecule to identify a novel B cell subset*", as follows:

Regarding the technical considerations on the use of an intracellular staining, the staining protocol used, and its validation, please note that although indeed FcRγ is a transmembrane protein, it has a very short (5aa) extracellular tail. We did not find any antibody raised specifically against those 5 extracellular residues. To validate FcRγ expression, we tested two different antibodies by flow cytometry. One of them was an antibody from Biorbyt raised against the complete FcRγ protein, the other one was a Sigma-Aldrich antibody raised against an intracellular peptide (aa80-86). Both antibodies were tested in B3Z cells transfected with an FcRγ-CD2 fusion protein. Following the same fixation/permeabilization protocol as used in the manuscript, we found that FcRγ-CD2 B3Z cells are detected by the anti-FcRγ antibody from Sigma-Aldrich as well as with an anti-CD2 antibody, while untransfected cells are negative for both markers (this staining is now included in **new Figure EV2D**). In contrast, the Biorbyt antibody failed to stain FcRγ-CD2 transfected B3Z cells, neither fresh, nor after fixation/permeabilization (not shown), indicating that, at least under these conditions, the Biorbyt antibody is not useful for flow cytometry analysis of FcRγ expression. Therefore, we did our experiments with the Sigma-Aldrich antibody against the intracellular part of the protein, after permeabilization and fixation. **This control demonstrates anti-FcRγ antibody specificity**.

R2 brings up the possibility that duplets were not excluded from our analysis. **Duplets were indeed excluded from the analysis, so they cannot possibly affect the interpretation of our stainings**. We apologize for not having specified this routine practice in the materials and methods section. Gating strategy is now included in **new Figure EV1B**.

R2 is concerned that our FcγR staining may not be specific for B cells. This is now shown by our control experiments in FcRγ-CD2 transfected B3Z cells showing that antibody staining is specific (Figure EV2D) and further supported by our clarification that duplets were carefully excluded from all our gates (Figure EV1B). In addition, we provide the following additional clarifications as per the R2's suggestions:

- We have performed **back-gating analysis,** showing that FcγR+ cells are: GL7-, B220+, Tom+, live and singlets. In addition, their size by FSC parameter is compatible with that of a B cell. This is now shown in **new Figure EV2E.**
- R2 raises the concern that "*a very sizeable fraction of Tom+ cells in Figure 3c is comprised of non-B/PC. This population is part of RNA-seq and flow cytometry data*". Regarding this, we want to clarify that we detect only a minor fraction of Tom+ that cannot be identified as B cells or PC, about 6% in our single cell data and about 3-5% of all Tom+ cells in our FACS data. Importantly, and contrary to R2's remark, these cells were carefully excluded from the scRNA analysis and the FACS analysis. In the case of scRNA analysis we used the automatic feature of SingleR, which uses different databases for annotation (including Immgen) and filtered out all non-B non-PC cells. This has been now more explicitly explained in the Methods section. Likewise, these cells are not included in our FACS analysis of FcRγ+ cells: all our FACS analyses are based in a gating strategy that selects only Tom+, B220+ or CD138+ (see above and new Figure EV1B and EV2E). **Thus, Tom+ non-B/PC were not included in any of the analyses shown in the manuscript.**
- Along the same lines, R2 also asks to *"show that FcRg staining is indeed restricted to GL7- non-GC B cells, as suggested by scRNA-seq*". This information was already contained in Figure EV2. Now we show FcRγ expression, as requested by R2, specifically in the GL7- and GL7+ fractions of live, singlet, Tom+ CD138- and B220+ gated cells (**new Figure EV2F**). The analysis shows that **FcRγ staining is very much restricted to GL7-, non-GC B cells**.
- In addition, R2 suggests "*to analyse FcRg+ cells from sc-RNA-Seq separately and compare them to the remainder of the dataset*". We now provide here the scRNAseq expression values for the indicated genes in FcRg+ (pct.1) cells and the rest of cells in the dataset (pct. 2), shown in **new Figure EV2G**.

2. The interpretation of the data, in particular with regards to AID deficient mice, is sometimes misleading. For example, the final sentence of the abstract states that AID enables PB differentiation from a prePB stage. There is little information in this sentence as PB must, of course, come from a prePB cell, whatever the identity of this cell is. However, to which extent AID directly influences pre-PB to PB differentiation is not investigated by this study. Also, the identity of the L-prePB as "PC precursors" rather than MBC remains largely speculative as it is mostly based on clonal relationship. Other than that, there is no indication that these cells are indeed on the way towards differentiating PC rather than being resting MBC (and if they are resting MBC, whether they preferentially differentiate into PB upon re-activation). Can the authors show preferential ex-vivo PC differentiation of L-prePB to provide evidence that these cells are really committed towards PB? Such in vitro culture systems are available (see Kurosaki, Kitamura)? This would also be used to further confirm the identity of the FcRg+ population.
It must be expected, and in fact is known, that pretty much all response characteristics are changed in the absence of AID. For example, the reduction in L-prePB and PC could simply reflect the cumulative effect of changes in T cell mediated selection, response duration, time of response magnitude, absence of class switched cells (which can have different propensities for differentiation) and absence of affinity maturation. Along the same lines, the multiple underlying causes are likely responsible for changes in clonal relationships within transitional stages in a system that prevents vs. does not prevent

clonal diversification. This is further complicated by the experimental setup as the data reported here is from a single, late time point following secondary antigen challenge which makes drawing conclusions about cause and consequence close to impossible, in particular in a model knockout system to is known to have major consequences for numerous aspects of the response.
Overall, I think this study has too many limitations to provide a significant conceptual advance to what is already known.

In this point, R2 argues that all kinds of alterations are to be expected in the absence of AID and our study does not "provide significant conceptual advance to what is already known". We respectfully disagree with the reviewer's opinion. One can certainly expect any number of differences in the germinal center program in the absence of AID, but ours is the first study to directly assess and get molecular insights on these differences at the single cell level. In fact, our analysis of clonal relationships is unprecedented. As for the specific question raised in this point: "Can the authors show preferential ex-vivo PC differentiation of L-prePB to provide evidence that these cells are really committed towards PB?", we absolutely agree that this is an important experiment. Unfortunately, *ex vivo* experiments cannot be performed with the reagents available at the moment. As explained above, the antibody used to identify L-prePB requires fixation and permeabilization. Therefore, for the time being, L-prePB identification is not compatible with isolation of live cells for differentiation studies and this kind of experiment will have to await the development of an antibody that allows the selection of live cells.

Minor comments/suggestions:
AU RESPONSE: All minor R#2's comments/suggestions has been addressed:

Introduction:
- The abbreviation "L-prePB" is not explained in the abstract.
Abbreviation has been explained

- Line 22: Consider mentioning short-lived plasmablasts as additional fate choice.
For the sake of simplification, we have removed the specific reference to long-lived PC in the first paragraph of the introduction and refer globally to PC fate

- Line 33: ... deaminates the region encoding the antigen-recognizing, variable part of immunoglobulin genes, generating mutations which can give rise to variants with altered affinity for antigen.
Edited as suggested

- Line 39: ... exchange of IgM/IgD isotypes with IgG, IgE or....
Edited as suggested

- Line 44-46: I would suggest to re-word this as all B cells, not only those with increased BCR affinity are undergoing selection by Tfh. What differs is their fate.
Reworded as suggested

- Line 56: Consider changing to: Accordingly, the average frequency of SHM is higher in PCs than in MBCs.
Edited as suggested

- Line 62: consider changing "looser" to something like "a broader range of affinities". "of B cells" could be deleted as it is contained within "MBC".
Rephrased as suggested

Results:
- Complete flow cytometry/FACS gating strategies should be shown for all data.
Gating strategies are shown in Fig EV1B

-A mix of present and past tense is used in the results section, in particular in the first part. I would suggest to consistently use past tense.
Edited as suggested

- Line 93: consider changing "coding for" to "encoding"
Edited as suggested

(Please note that line numbers as indicated by the reviewer do not correspond with the line numbers of our revised manuscript)

Referee #3:
The authors have addressed my major concerns.

AU RESPONSE: We very much appreciate R#3's thorough revision of our manuscript and her/his positive assessment.

Dr. Almudena Ramiro
Centro Nacional de Investigaciones Cardiovasculares
Vascular Physiopathology
Melchor Fernández Almagro, 3
Madrid 28029
Spain

Dear Dr. Ramiro,

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

****

THINGS TO DO NOW:

Once your article has been received by Wiley for production, the corresponding author will receive an email from Wiley's Author Services system which will ask them to log in and will present them with the appropriate license for completion.

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-2022-55000V3 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 Author Checklist

Corresponding Author Name: Almudena Rodriguez Ramiro
Journal Submitted to: EMBO Reports
Manuscript Number: EMBOR-2022-55000V1

USEFUL LINKS FOR COMPLETING THIS FORM

The EMBO Journal - Author Guidelines
EMBO Reports - Author Guidelines
Molecular Systems Biology - Author Guidelines
EMBO Molecular Medicine - Author Guidelines

Reporting Checklist for Life Science Articles (updated January 2022)

This checklist is adapted from Materials Design Analysis Reporting (MDAR) Checklist for Authors. MDAR establishes a minimum set of requirements in transparent reporting in the life sciences (see Statement of Task: 10.31222/osf.io/9sm4x). Please follow the journal's guidelines in preparing your manuscript.

Please note that a copy of this checklist will be published alongside your article.

Abridged guidelines for 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.
- ideally, figure panels should include only measurements that are directly comparable to each other and obtained with the same assay.
- plots 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. Any statistical test employed should be justified.
- Source Data should be included to report the data underlying figures according to the guidelines set out in the authorship 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.

Please complete ALL of the questions below.
Select "Not Applicable" only when the requested information is not relevant for your study.

Materials

Newly Created Materials	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
New materials and reagents need to be available; do any restrictions apply?	Not Applicable	

Antibodies	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
For **antibodies** provide the following information: - Commercial antibodies: RRID (if possible) or supplier name, catalogue number and or/clone number - Non-commercial: RRID or citation	Yes	Materials and Methods (p19)

DNA and RNA sequences	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Short novel DNA or RNA including primers, probes: provide the sequences.	Yes	Materials and Methods (p19 and p20)

Cell materials	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Cell lines: Provide species information, strain. Provide accession number in repository **OR** supplier name, catalog number, clone number, and/**OR** RRID.	Yes	Materials and Methods (p18)
Primary cultures: Provide species, strain, sex of origin, genetic modification status.	Not Applicable	
Report if the cell lines were recently **authenticated** (e.g., by STR profiling) and tested for mycoplasma contamination.	Not Applicable	

Experimental animals	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Laboratory animals or Model organisms: Provide species, strain, sex, age, genetic modification status. Provide accession number in repository **OR** supplier name, catalog number, clone number, **OR** RRID.	Yes	Materials and Methods (p18)
Animal observed in or captured from the field: Provide species, sex, and age where possible.	Not Applicable	
Please detail **housing and husbandry conditions**.	Yes	Materials and Methods (p18)

Plants and microbes	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Plants: provide species and strain, ecotype and cultivar where relevant, unique accession number if available, and source (including location for collected wild specimens).	Not Applicable	
Microbes: provide species and strain, unique accession number if available, and source.	Not Applicable	

Human research participants	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
If collected and within the bounds of privacy constraints report on age, sex and gender or ethnicity for all study participants.	Not Applicable	

Core facilities	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
If your work benefited from core facilities, was their service mentioned in the acknowledgments section?	Yes	Acknowledgments (p24)

Design

Study protocol	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
If study protocol has been **pre-registered, provide DOI in the manuscript**. For clinical trials, provide the trial registration number **OR** cite DOI.	Not Applicable	
Report the **clinical trial registration number** (at ClinicalTrials.gov or equivalent), where applicable.	Not Applicable	

Laboratory protocol	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Provide DOI OR other citation details if **external detailed step-by-step protocols** are available.	Not Applicable	

Experimental study design and statistics	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Include a statement about **sample size** estimate even if no statistical methods were used.	Yes	Figures
Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. **randomization procedure**)? If yes, have they been described?	Not Applicable	
Include a statement about **blinding** even if no blinding was done.	Not Applicable	
Describe **inclusion/exclusion criteria** if samples or animals were excluded from the analysis. Were the criteria pre-established? If sample or data points were omitted from analysis, report if this was due to attrition or intentional exclusion and provide justification.	Not Applicable	
For every figure, are **statistical tests** justified as appropriate? Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it. Is there an estimate of variation within each group of data? Is the variance similar between the groups that are being statistically compared?	Yes	Materials and Methods (p23)

Sample definition and in-laboratory replication	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
In the figure legends: state number of times the experiment was **replicated** in laboratory.	Yes	Figures
In the figure legends: define whether data describe **technical or biological replicates**.	Yes	Figures

Ethics

Ethics	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Studies involving **human participants**: State details of **authority granting ethics approval** (IRB or equivalent committee(s), provide reference number for approval.	Not Applicable	
Studies involving **human participants**: 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.	Not Applicable	
Studies involving **human participants:** For publication of **patient photos**, include a statement confirming that consent to publish was obtained.	Not Applicable	
Studies involving experimental **animals**: State details of **authority granting ethics approval** (IRB or equivalent committee(s), provide reference number for approval. Include a statement of compliance with ethical regulations.	Yes	Materials and Methods (p18)
Studies involving **specimen and field samples:** State if relevant **permits** obtained, provide details of authority approving study; if none were required, explain why.	Not Applicable	

Dual Use Research of Concern (DURC)	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Could your study fall under dual use research restrictions? Please check biosecurity documents and list of **select agents and toxins** (CDC): https://www.selectagents.gov/sat/list.htm	Not Applicable	
If you used a select agent, is the security level of the lab appropriate and reported in the manuscript?	Not Applicable	
If a study is subject to dual use research of concern regulations, is the name of the **authority granting approval and reference number** for the regulatory approval provided in the manuscript?	Not Applicable	

Reporting

The MDAR framework recommends adoption of discipline-specific guidelines, established and endorsed through community initiatives. Journals have their own policy about requiring specific guidelines and recommendations to complement MDAR.

Adherence to community standards	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
State if relevant guidelines or checklists (e.g., **ICMJE, MIBBI, ARRIVE, PRISMA**) have been followed or provided.	Not Applicable	
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.	Not Applicable	
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.	Not Applicable	

Data Availability

Data availability	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Have **primary datasets** been deposited according to the journal's guidelines (see 'Data Deposition' section) and the respective accession numbers provided in the Data Availability Section?	Yes	Data Availability Section (p23)
Were **human clinical and genomic datasets** deposited in a public access-controlled repository in accordance to ethical obligations to the patients and to the applicable consent agreement?	Not Applicable	
Are **computational models** that are central and integral to a study available without restrictions in a machine-readable form? Were the relevant accession numbers or links provided?	Not Applicable	
If publicly available data were reused, provide the respective **data citations in the reference list**.	Yes	Materials and Methods, References