THE EMBO JOURNAL

Unbiased assessment of genome integrity upon gene editing with Cas9 and a donor template

Daniele Canarutto, Claudia Asperti, Valentina Vavassori, Simona Porcellini, Elisabetta Rovelli, Marianna PAULIS, Samuele Ferrari, Angelica Varesi, Martina Fiumara, Aurelien Jacob, Lucia Sergi Sergi, Ilaria Visigalli, Francesca Ferrua, Luis González-Granado, Vassilios Lougaris, Andrea Finocchi, Anna Villa, Marina Radrizzani, and Luigi Naldini
DOI: 10.15252/embj.2023114188

Corresponding author: Luigi Naldini (naldini.luigi@hsr.it)

Review Timeline:

Submission Date:	4th Apr 23
Editorial Decision:	12th May 23
Revision Received:	18th Jul 23
Editorial Decision:	8th Sep 23
Revision Received:	15th Sep 23
Accepted:	25th Sep 23

Editor: Kelly Anderson

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 **12th May 2023**

Dear Prof. Naldini,

Thank you for submitting your manuscript for consideration by the EMBO Journal. It has now been seen by three referees whose comments are shown below.

Given the referees' positive recommendations, I would like to invite you to submit a revised version of the manuscript, addressing the comments of all three reviewers. I should add that it is EMBO Journal policy to allow only a single round of revision, and acceptance of your manuscript will therefore depend on the completeness of your responses in this revised version. It would be good to discuss your plan to address the referee concerns and I am available to do so by zoom or email in the coming weeks. I've also attached a revision guide for your convenience.

When preparing your letter of response to the referees' comments, please bear in mind that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process, please visit our website: https://www.embopress.org/page/journal/14602075/authorguide#transparentprocess

We generally allow three months as standard revision time. As a matter of policy, competing manuscripts published during this period will not negatively impact on our assessment of the conceptual advance presented by your study. However, we request that you contact the editor as soon as possible upon publication of any related work, to discuss how to proceed. Should you foresee a problem in meeting this three-month deadline, please let us know in advance and we may be able to grant an extension.

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

Yours sincerely,

Kelly M Anderson, PhD
Editor
The EMBO Journal
k.anderson@embojournal.org

Further information is available in our Guide For Authors: https://www.embopress.org/page/journal/14602075/authorguide

We realize that it is difficult to revise to a specific deadline. In the interest of protecting the conceptual advance provided by the work, we recommend a revision within 3 months (10th Aug 2023). Please discuss the revision progress ahead of this time with the editor if you require more time to complete the revisions.

Referee #1:

This study addressed the potential treatment of hyper-IgM1, an X-linked combined immunodeficiency disease caused by CD40LG mutations, using Cas9 and a DNA template to induce corrective CD4+ T-cell genome editing. To study potential genomic alterations following long-range genome editing, the authors used digital droplet PCR (ddPCR) assays for unbiased detection of large on-target deletions. This analysis revealed an unexpectedly high frequency of such deletions that occurred upon editing with an adenoviral-associated donor template. Large deletions were similarly detected after editing different loci and cell types and using an alternative Cas9 delivery method. In CD40LG edited T cells, on-target deletions could be counter-selected in culture and eliminated by enrichment for edited cells using a selector coupled to gene correction. The accuracy of the ddPCR detection method was cross-checked using optical genome mapping for unbiased detection of genome wide rearrangements, revealing occasional on-target integration of one or more vector copies. Notably, these integrations don't alter cell function when editing was done using an integrase defective lentiviral donor template. Importantly, genome-edited cells from patients had correctly reconstituted CD40LG gene expression and function, likely due to the inviability of cells that experienced large on-target genetic deletions. The authors conclude that although large deletions and donor template integrations should be expected in these types of genome editing procedures, they can be detected and mitigated using a combined detection and cell outgrowth protocol.

Overall this is a careful and well-conducted study that demonstrates the value of assessing genome editing outcomes using ddPCR with primer sets designed to detect loss of sequences flanking but distal from the site targeted by CRISPR-Cas9 cleavage. I have just two minor points for the authors' consideration:

1. Please comment on the utility of the ddPCR and optical mapping approaches for detecting chromosomal deletions or aberrant integrations in other cell types; is the protocol outlined here applicable to the analysis of non-immune cells or tissue samples? Also, should such a protocol become standard in the field due to the relatively common occurrence of such deletions?

2. Do the results reported here imply that in vivo editing of T cells may be safe enough for clinical implementation given the apparent elimination of cells that experience a large genetic deletion? This is a possible conclusion based on the results of this study, and is worth discussing briefly given the potential future direction of the CAR T field.

Referee #2:

* General comments:
CRISPR-Cas nucleases basically aim to target generate DNA double strand breaks (DSBs) at the sites of interest. But recent studies have revealed that such DSBs also cause long-range deletion or genomic translocation as well as small insertion/deletion (indel), giving some warning to the research field. In this manuscript, Canarutto et al. conducted gene editing with Cas9 ribonucleoproteins (RNPs) and viral vector as a donor template in CD4 T-cells, and tried to find genome integrity such as large deletions via unbiased droplet digital PCR assays (ddPCR) or optical genome mapping (OGM). The topic is interesting and important especially for therapeutic applications, but the main text is not easy to follow up and quantifications of outcomes seem insufficient. Furthermore, I am not sure whether this study could satisfy the creteria of this journal, mainly because of the low novelty of the concept. I would like to raise several issues to strengthen this study.

* Specific comments:
1) The authors used Cas9 RNPs with a donor template for knock-in. The results would also generate NHEJ-mediated small indels or HDR-mediated knock-ins. First of all, the quantitative frequencies of them should be provided to validate the editing efficiency and purity.

1-2) In this experiment, I am curious that it is also necessary to discuss the NHEJ-mediated small indels, in terms of safety issue. It might be ignored because the target site is positioned within intron, but relatively large deletions or incomplete knock-in of the donor DNA might cause the CD40LG gene knock-out. Does such gene KO affect the therapy, or not? It should be discussed.

2) In the abstract and discussion parts, the authors emphasized the development of the unbiased ddPCR method, compared to the conventional ddPCR. However, I could not find the detailed mechanism and advantages of the unbiased ddPCR method. It should be provided in the main text.

2-2) The authors used probes for the ddPCR located in neighboring genes including LNC00892 and ARHGEF6, and showed whether those genes were disrupted or not. I think that it is a kind of rough (not exact) method.

3) In Figure 1E, it seems that (RNP only) did not cause CNV differences but (RNP+AAV6) caused meaningful differences. It is thought that long-range deletions are relative to Cas-mediated DSBs. Does the existence of donor templates also affect the long-range deletions? Or does AAV affect? It should be discussed.

3-2) In this regard, other paper [PMID: 34416913] reported that the presence of the donor templates reduced the long-range deletions, which is contradictory to the present study. It should be discussed.

4) In Figures 1-2, the authors argued that the long-range deletions are commonly occurred by Cas nucleases, but such events were rarely occurred in donor cells as shown in Figure 4. The detailed reason should be discussed.

4-2) In addition, do the authors have any idea to reduce the CRISPR-mediated genome integrity? If then, it is necessary to be discussed in the discussion section.

(Minor)
- Figure 1D is not mentioned in the main text.
- The term of 'T cell', 'T-cell' should be unified.

Referee #3:

CRISPR-Cas9-mediated genome editing can introduce undesired on-target effects, such as large deletions, chromothripsis, chromosomal translocations and other complex chromosomal rearrangements, in addition to the well-studied off-target effects. Engineered T cells have shown huge potential for the treatment of multiple cancers and immunodeficiencies such as Hyper-

IgM1 as described in Canarutto et. al. studies here. In this manuscript, authors detected rare structural variations in the edited T cells via optical genome mapping (OGM), a revolutionary tool for "next generation cytogenomics analysis". This approach evaluates the safety and efficacy of the genome engineering platform for T-cell editing which represents a major advance for the field. These results are important and will be of interest to a wide audience in the fields of both genome editing and T cells. This reviewer only have a few questions, which should be addressed or clarified prior to publication.

Major concerns:

1. Fig 1E, the authors developed a droplet digital PCR-based copy number variation assay to detect the large on-target deletions. While there was no significance between untreated cells and cells electroporated with Cas9 RNP, authors found lost up to 40% of amplicons with Cas9 RNP + AAV6 donor template group. Does this result indicate that the high frequency of large on-target deletions is specifically attributed to the AAV donor template? Do the authors observe similar results upon editing on the CD40LG gene with Cas9 RNP + IDLV donor template?
2. Fig. 1J, do the authors observe an increase of large on-target deletions or CNV difference upon editing on the AAVS1 or B2M site when Cas9 was delivered as RNA or RNP in bulk-edited CB or MPB cells?
3. In Fig 2-4, the authors validated the optical genome mapping technology for unbiased detection of genome-wide rearrangements. They uncovered on-target integration of one or more vector copies upon editing using the IDLV donor template. It would be great if the author could also use other technologies, such as PacBio sequencing or NGS, to independently verify the on-target integration of LDLV donor sequences upon editing in one of the donors.

Minor concerns:

1. Missing a citation for Fig. 1D in the main text.
2. Authors are using "Hyper IgM1" in the title, but also using "Hyper-IgM1" in the abstract and main text. It would be nice to keep the terms consistent.
3. Some of the figures are generated from screenshot of analysis software, for example: 2D, 2F, 2G and the font size is too small to read.

We would like to thank the Editor and the Reviewers for their careful evaluation of our manuscript and helpful comments, which prompted us to improve the quality of the manuscript. We now provide a revised version of the manuscript for consideration in The EMBO Journal according to your requests, including a substantial amount of relevant new data, additional clarifications on previously reported findings and changes in the manuscript flow to improve readability and optimally embed the new findings in the whole study. As requested by the reviewers, our experimental efforts were mostly directed towards, but not limited to:

- Further investigating deletions and culture dynamics upon HDR editing with AAV6
- Addressing the presence of deletions when using IDLV as donor template
- Corroborating our optical genome mapping findings with PCR-free nanopore sequencing

Most relevant additions and changes in the revised manuscript are highlighted in red to facilitate Reviewers' and Editor's assessment.

Referee #1:

This study addressed the potential treatment of hyper-IgM1, an X-linked combined immunodeficiency disease caused by CD40LG mutations, using Cas9 and a DNA template to induce corrective CD4+ T-cell genome editing. To study potential genomic alterations following long-range genome editing, the authors used digital droplet PCR (ddPCR) assays for unbiased detection of large on-target deletions. This analysis revealed an unexpectedly high frequency of such deletions that occurred upon editing with an adenoviral-associated donor template. Large deletions were similarly detected after editing different loci and cell types and using an alternative Cas9 delivery method. In CD40LG edited T cells, on-target deletions could be counter-selected in culture and eliminated by enrichment for edited cells using a selector coupled to gene correction. The accuracy of the ddPCR detection method was cross-checked using optical genome mapping for unbiased detection of genome wide rearrangements, revealing occasional on-target integration of one or more vector copies. Notably, these integrations don't alter cell function when editing was done using an integrase defective lentiviral donor template. Importantly, genome-edited cells from patients had correctly reconstituted CD40LG gene expression and function, likely due to the inviability of cells that experienced large on-target genetic deletions. The authors conclude that although large deletions and donor template integrations should be expected in these types of genome editing procedures, they can be detected and mitigated using a combined detection and cell outgrowth protocol.

Overall this is a careful and well-conducted study that demonstrates the value of assessing genome editing outcomes using ddPCR with primer sets designed to detect loss of sequences flanking but distal from the site targeted by CRISPR-Cas9 cleavage. I have just two minor points for the authors' consideration:

We thank the Reviewer for his/her appreciation of our work and finding our demonstration of genomic integrity analyses of value for the gene editing field.

1. Please comment on the utility of the ddPCR and optical mapping approaches for detecting chromosomal deletions or aberrant integrations in other cell types; is the protocol outlined here applicable to the analysis of non-immune cells or tissue samples? Also, should such a protocol become standard in the field due to the relatively common occurrence of such deletions?

As discussed in text (pag 14) this protocol is in principle applicable to other cell types and tissue samples, as long as sufficient high molecular weight DNA can be extracted for optical genome mapping (and possibly nanopore sequencing). As elaborated in the discussion, we believe that this (or similar approaches) should be used in view of any prospective new clinical application of editing technologies.

2. Do the results reported here imply that in vivo editing of T cells may be safe enough for clinical implementation given the apparent elimination of cells that experience a large genetic deletion? This is a possible conclusion based on the results of this study, and is worth discussing briefly given the potential future direction of the CAR T field.

We thank for the reviewers for raising this interesting point which we have now addressed at pag 14 in the discussion. Briefly, while it is likely that large losses of DNA impair cellular fitness, it would be inapporpriate to generalize this assumption to all targeted loci. For instance, deletion of a tumor suppressor gene might confer a selective growth advantage and promote transformation. These events should also be assessed in the setting of in-vivo editing of T-cells and incorporated into a comprehensive risk-benefit analysis.

Referee #2:

* General comments:
CRISPR-Cas nucleases basically aim to target generate DNA double strand breaks (DSBs) at the sites of interest. But recent studies have revealed that such DSBs also cause long-range deletion or genomic translocation as well as small insertion/deletion (indel), giving some warning to the research field. In this manuscript, Canarutto et al. conducted gene editing with Cas9 ribonucleoproteins (RNPs) and viral vector as a donor template in CD4 T-cells, and tried to find genome integrity such as large deletions via unbiased droplet digital PCR assays (ddPCR) or optical genome mapping (OGM). The topic is interesting and important especially for therapeutic applications, but the main text is not easy to follow up and quantifications of outcomes seem insufficient. Furthermore, I am not sure whether this study could satisfy the creteria of this journal, mainly because of the low novelty of the concept. I would like to raise several issues to strengthen this study.

We thank the Reviewer for providing his/her perspective on the topic. We believe that the main point of our work, i.e. that genome integrity upon gene editing must be addressed with the least degree of bias is highly relevant. Novelty resides in the validation and application of optical genome mapping, whereby we identify on-target concatemers of template delivery vector (both yet unreported to the best of our knowledge). On the other hand, we do not claim that detection of on-target deletions per se is a novel finding; we rather argue for unbiased assessment of deletions not assuming the conservation of primer binding sites, as demonstrated with our tiling drop-off ddPCR assays. Furthermore, we value incorporation of this part in the manuscript as it highlights the need to complement different technologies for comprehensive genome integrity evaluations.
Concerning quantifications of outcomes we have added a substantial amount of new data to support our claim with more robust statistics and throughout different experimental variables, as detailed below.

* Specific comments:
1) The authors used Cas9 RNPs with a donor template for knock-in. The results would also generate

NHEJ-mediated small indels or HDR-mediated knock-ins. First of all, the quantitative frequencies of them should be provided to validate the editing efficiency and purity.

We now provide HDR efficiency estimates by ddPCR for all samples (or flow cytometry in Fig 3; note that we have always found comparable results in terms of ddPCR and flow cytometry as also shown in our previous work: Vavassori*, Mercuri* et al., Embo Mol Med 2021).

We also included quantification of indels in Fig 1I and 1P. In general, we found quantification of indels by T7 assay not to be very informative, as only alleles with the conserved amplicon can be amplified. For instance, in Fig 1P there is little difference between RNP and IDLV edited samples, which is somewhat misleading considering that approximately 25% of cells are HDR-edited (more so in male LNGFR enriched samples from Fig1O-Q indel frequency was 0.57-0.64 while HDR was 0.68-0.81).

1-2) In this experiment, I am curious that it is also necessary to discuss the NHEJ-mediated small indels, in terms of safety issue. It might be ignored because the target site is positioned within intron, but relatively large deletions or incomplete knock-in of the donor DNA might cause the CD40LG gene knock-out. Does such gene KO affect the therapy, or not? It should be discussed.

We thank the Reviewer for this comment; we have explained more in detail (pag 5) why in our opinion this may not be the case. The issue is not really that the target is within an intron, but rather that in HIGM1 patients the CD40LG is already not functional. Thus, further disruptions of the gene are not expected to further worsen the phenotype nor emergence of hypothetical dominant negative forms is expected, given its hemizygous condition in the affected male population. This is also the reason why we focused on deletions that involve the closest annotated genes rather than CD40LG itself.

2) In the abstract and discussion parts, the authors emphasized the development of the unbiased ddPCR method, compared to the conventional ddPCR. However, I could not find the detailed mechanism and advantages of the unbiased ddPCR method. It should be provided in the main text.

We thank the Reviewer for this comment. We have further detailed in pag 5, 11-12 why we believe that these assays are unbiased by design, i.e. they do not require assuming the presence of primer binding sites in treated samples.

2-2) The authors used probes for the ddPCR located in neighboring genes including LNC00892 and ARHGEF6, and showed whether those genes were disrupted or not. I think that it is a kind of rough (not exact) method.

These assays are designed with reproducibility and standardization in mind. We now further discuss their limitations in page 12, especially with respect to the length of the deletions and the fact that deletions may also be unbalanced translocations with heterogeneous junctions. We chose neighboring genes as “mileposts”, to make the point that Cas9 deletions should be investigated beyond the target site in the genomic neighborhood.

3) In Figure 1E, it seems that (RNP only) did not cause CNV differences but (RNP+AAV6) caused meaningful differences. It is thought that long-range deletions are relative to Cas-mediated DSBs. Does the existence of donor templates also affect the long-range deletions? Or does AAV affect? It should be discussed.

We thank the Reviewer for noticing this trend and have now addressed this point with new data. Indeed, in our experimental setting the presence of AAV6 significantly increased the frequency of deletions at nadir (Fig1H-L), with a dose-response trend.

3-2) In this regard, other paper [PMID: 34416913] reported that the presence of the donor templates reduced the long-range deletions, which is contradictory to the present study. It should be discussed.

We addressed this point with new data. Even with IDLV, we still observed the presence of deletions in female cells. In pag 15, we argue that presence of a template cannot guarantee the absence of deletions and address differences between our findings and those of Wen et al. that may explain the reported discrepancies.

4) In Figures 1-2, the authors argued that the long-range deletions are commonly occurred by Cas nucleases, but such events were rarely occurred in donor cells as shown in Figure 4. The detailed reason should be discussed.

Cells assessed by optical mapping were analysed at day 14 (i.e. 12 days after editing and selection for LNGFR-positive cells, at the end of the GMP manufacturing process) and were sourced uniquely from males. We show that deletions are mostly detectable early after editing and are counterselected over time and upon LNGFR selection, and less frequent in males. We thus did not expect to detect deletions by OGM at that time point. Indeed, leveraging on both natural and ad-hoc designed counterselection to purge Cas-induced deletions is a “take-home-message” of our study supporting clinical translation of our strategy.

4-2) In addition, do the authors have any idea to reduce the CRISPR-mediated genome integrity? If then, it is necessary to be discussed in the discussion section.

We thank the Reviewer for raising this point. We added (pag 15) references to novel findings which point to culture conditions and cell cycle progression as main drivers of long deletions. While limiting proliferation appears to be an option for NHEJ applications, cell cycle progression is currently required for HDR applications. At present, selection (in our case by LNGFR) is a reasonable mitigation strategy.

(Minor)
- Figure 1D is not mentioned in the main text. Done.
- The term of 'T cell', 'T-cell' should be unified. Done.

Referee #3:

CRISPR-Cas9-mediated genome editing can introduce undesired on-target effects, such as large deletions, chromothripsis, chromosomal translocations and other complex chromosomal rearrangements, in addition to the well-studied off-target effects. Engineered T cells have shown huge potential for the treatment of multiple cancers and immunodeficiencies such as Hyper-IgM1 as described in Canarutto et. al. studies here. In this manuscript, authors detected rare structural variations in the edited T cells via optical genome mapping (OGM), a revolutionary tool for "next generation cytogenomics analysis". This approach evaluates the safety and efficacy of the genome engineering platform for T-cell editing which represents a major advance for the field. These results

are important and will be of interest to a wide audience in the fields of both genome editing and T cells. This reviewer only have a few questions, which should be addressed or clarified prior to publication.

We thank the Reviewer for his/her appreciation of our work and finding our demonstration of genomic integrity analyses of importance and general value for the gene editing and T-cell therapy fields.

Major concerns:
1. Fig 1E, the authors developed a droplet digital PCR-based copy number variation assay to detect the large on-target deletions. While there was no significance between untreated cells and cells electroporated with Cas9 RNP, authors found lost up to 40% of amplicons with Cas9 RNP + AAV6 donor template group. Does this result indicate that the high frequency of large on-target deletions is specifically attributed to the AAV donor template? Do the authors observe similar results upon editing on the CD40LG gene with Cas9 RNP + IDLV donor template?

We thank the Reviewer for raising this point. We have added new data, confirming these findings. Indeed, in our experimental setting the presence of AAV6 significantly increased the frequency of deletions at nadir (Fig1H-L), with a dose-response trend. This is not to say that there is no deletion without AAV, as now specifically shown in new panels O-Q of Fig. 1.
Concerning the use of IDLV, we did detect deletions also in its presence, albeit to lower extent than when using AAV6 and with a trend toward lower frequency than with RNP alone.
We argue in pag 14-15 that while others have reported mitigation in the proportion of deletions in presence of an AAV6 template this is not a universal finding, and that the presence of deletions should always be investigated when using Cas9 (and possibly nickases).

2. Fig. 1J, do the authors observe an increase of large on-target deletions or CNV difference upon editing on the AAVS1 or B2M site when Cas9 was delivered as RNA or RNP in bulk-edited CB or MPB cells?

We have provided new data addressing this interesting point (Fig 1S). Comparing RNA and RNP delivery at same levels of cutting efficiency we observed similar levels of deletions.

3. In Fig 2-4, the authors validated the optical genome mapping technology for unbiased detection of genome-wide rearrangements. They uncovered on-target integration of one or more vector copies upon editing using the IDLV donor template. It would be great if the author could also use other technologies, such as PacBio sequencing or NGS, to independently verify the on-target integration of LDLV donor sequences upon editing in one of the donors.

We thank the Reviewer for raising this challenging point. We complemented our findings with on-target, PCR-free nanopore sequencing. As now reported in the new Fig. 4E,F and EV5, albeit at very low coverage, we were able to sequence, beyond a majority of intended edits with precise bilateral HDR, a 9.8 kb insert consisting of a concatemer of 2 IDLVs joined at an intervening concatemer, with a single LTR, that had been integrated by HDR. Interestingly, its size reproduced the most frequent size of insert recovered by OGM.

Minor concerns:
1. Missing a citation for Fig. 1D in the main text. Done
2. Authors are using "Hyper IgM1" in the title, but also using "Hyper-IgM1" in the abstract and main text. It would be nice to keep the terms consistent. Done

3. Some of the figures are generated from screenshot of analysis software, for example: 2D, 2F, 2G and the font size is too small to read. Done

Dear Luigi,

Congratulations on a great revision! Overall, the referees have been positive. However, referee 2 has pointed out some non-experimental concerns that we ask you to please address in a revised version. When you submit your revised version, please also take care of the following editorial items and add this also to your point-by-point response:

1. Please ensure that the author names in EJP are exactly matching the manuscript, particularly: Samuele Ferrari in ms vs. Sanuele Ferrari in eJP; Angelica Varesi in ms vs. Angelica Vaersi in eJP; Aurelien Jacob in ms, vs. Jacob Aurelien in eJP.

2. Please review our new policy on conflict of interests on the EMBO author guide website and update the title of this section to: Disclosure and competing interests statement.

3. Please remove the author contribution section from the main manuscript.

4. We require the publication of source data, particularly for electrophoretic gels and blots and graphs, with the aim of making primary data more accessible and transparent to the reader. It would be great if you could provide me with a PDF file per figure that contains the original, uncropped and unprocessed scans of all or key gels used in the figure or for graphs, an Excel spreadsheet with the original data used to generate the graphs. The PDF files should be labeled with the appropriate figure/panel number, and should have molecular weight marker; further annotation could be useful but is not essential. The PDF files will be published online with the article as supplementary "Source Data" files. If you require assistance regarding which data should be included, please contact: h.sonntag@source-data.org.

5. We include a synopsis of the paper (see http://emboj.embopress.org/). Please provide me with a general summary statement and 3-5 bullet points that capture the key findings of the paper.

6. We also need a summary figure for the synopsis. The size should be 550 wide by 200-440 high (pixels). You can also use something from the figures if that is easier.

7. Please update figure 2. In the current format it spans over two pages and thus it won't be properly processed by the typesetters.

8. Please remove Tables EV1-EV3 and their legends from the manuscript file and upload as expanded view files.

9. Table 1 can remain in the manuscript, however it must be editable and black and white. It should be placed between the main and EV figure legends.

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

Kind regards,

Kelly

Kelly M Anderson, PhD
Editor, The EMBO Journal
k.anderson@embojournal.org

Further information is available in our Guide For Authors: https://www.embopress.org/page/journal/14602075/authorguide

---

Referee #2:

The authors have mostly answered the issues I raised in the earlier review. I think this revised version has been improved largely. I recommend the publication of this revised version.

Referee #3:

In this revised version, majority of the concerns raised in the previous version have been addressed, incorporating new data and revising figures as needed.

However, there are still some minor concerns that should be addressed:

1) Please double-check all gene names throughout the manuscript, especially in the figures. For instance, in Fig. 1F, the Y-axis label should read "LNGFR+" instead of "NGFR+". The same typo appears in Fig. 1G on the X-axis.

2) In Fig. 4E, the Y-axis label is "reads" Could you please clarify whether this refers to the exact numbers of reads or percentages of reads? This clarification should be added to the figure legends.

3)In the new Fig. 1G-J, it's unclear how the pink and blue dots are being used for labeling. I assume that pink dots represent female donors, and blue dots represent male donors. However, this information should be explicitly stated in the figures, figure legends, and the main text to avoid any ambiguity.

We thank the referees for their precious work. We have revised the manuscript as suggested.

Referee #2:

The authors have mostly answered the issues I raised in the earlier review. I think this revised version has been improved largely. I recommend the publication of this revised version.

Referee #3:

In this revised version, majority of the concerns raised in the previous version have been addressed, incorporating new data and revising figures as needed.

However, there are still some minor concerns that should be addressed:

1) Please double-check all gene names throughout the manuscript, especially in the figures. For instance, in Fig. 1F, the Y-axis label should read "LNGFR+" instead of "NGFR+". The same typo appears in Fig. 1G on the X-axis.

2) In Fig. 4E, the Y-axis label is "reads" Could you please clarify whether this refers to the exact numbers of reads or percentages of reads? This clarification should be added to the figure legends.

3)In the new Fig. 1G-J, it's unclear how the pink and blue dots are being used for labeling. I assume that pink dots represent female donors, and blue dots represent male donors. However, this information should be explicitly stated in the figures, figure legends, and the main text to avoid any ambiguity.

We have updated the figure legends for clarity.

2nd Revision - Editorial Decision **25th Sep 2023**

Dear Luigi,

Congratulations on an excellent manuscript, I am pleased to inform you that your manuscript has been accepted for publication in the EMBO Journal. Thank you for your comprehensive response to the referee concerns and for providing detailed source data. It has been a pleasure to work with you to get this to the acceptance stage.

I will begin the final checks on your manuscript before submitting to the publisher next week. Once at the publisher, it will take about 3 weeks for you manuscript to be published online. As a reminder, the entire review process, including referee concerns and your point-by-point response, will be available to readers.

I will be in touch throughout the final editorial process until publication. In the meantime, I hope you find time to celebrate!

Kind regards,

Kelly

Kelly M Anderson, PhD
Editor, The EMBO Journal
k.anderson@embojournal.org

Please note that it is EMBO Journal policy for the transcript of the editorial process (containing referee reports and your response letter) to be published as an online supplement to each paper. If you do NOT want this, you will need to inform the Editorial Office via email immediately. More information is available here:
https://www.embopress.org/page/journal/14602075/authorguide#transparentprocess

Your manuscript will be processed for publication in the journal by EMBO Press. Manuscripts in the PDF and electronic editions of The EMBO Journal will be copy edited, and you will be provided with page proofs prior to publication. Please note that supplementary information is not included in the proofs.

You will be contacted by Wiley Author Services to complete licensing and payment information. The required 'Page Charges Authorization Form' is available here: https://www.embopress.org/pb-assets/embo-site/tej_apc.pdf - please download and complete the form and return to embopressproduction@wiley.com

EMBO Press participates in many Publish and Read agreements that allow authors to publish Open Access with reduced/no publication charges. Check your eligibility: https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/affiliation-policies-payments/index.html

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

If you have any questions, please do not hesitate to call or email the Editorial Office. Thank you for your contribution to The EMBO Journal.

EMBO Press Author Checklist

Corresponding Author Name: Luigi Naldini
Journal Submitted to: The EMBO Journal
Manuscript Number: EMBOJ-2023-114188

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
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
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
- 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

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 χ^2 tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

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?	Yes	Material and methods/Extended View Content/Data Availability Section

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	Expanded View Content

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	Expanded View Content

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	Expanded View Content
Primary cultures: Provide species, strain, sex of origin, genetic modification status.	Yes	Methods
Report if the cell lines were recently **authenticated** (e.g., by STR profiling) and tested for mycoplasma contamination.	Yes	Expanded View Content. Cell line was bought as described.

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	Expanded View Content (ref to previous work)
Animal observed in or captured from the field: Provide species, sex, and age where possible**.**	Not Applicable	No animals were captured from the field
Please detail **housing and husbandry conditions**.	Yes	Expanded View Content

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	No plants were used
Microbes: provide species and strain, unique accession number if available, and source.	Not Applicable	No microbes were used

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.	Yes	Methods - Human subjects research

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	Aknowledgements

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	Not a clinical study
Report the **clinical trial registration number** (at ClinicalTrials.gov or equivalent), where applicable.	Not Applicable	Not a clinical study

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.	Yes	Expanded View Content > Methods > Nanopore sequencing

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.	Not Applicable	Not applicable
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	Not applicable
Include a statement about **blinding** even if no blinding was done.	Not Applicable	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	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	Non parametrical tests were used throughout the manuscript

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	Figure legends
In the figure legends: define whether data describe **technical or biological replicates**.	Yes	Figure legends

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.	Yes	Methods
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.	Yes	Methods
Studies involving **human participants:** For publication of **patient photos**, include a statement confirming that consent to publish was obtained.	Not Applicable	Not applicable
Studies involving experimental **animals**: State details of **authority granting ethics approva**l (IRB or equivalent committee(s), provide reference number for approval. Include a statement of compliance with ethical regulations.	Yes	Expanded View Content
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	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	NO
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	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	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	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; zip files with source data uploaded with submission
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?	Yes	
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	Proprietary Bionano Access software and Oxford Nanopore MinKnow were used for data analysis
If publicly available data were reused, provide the respective **data citations in the reference list**.	Not Applicable	Not applicable