nature portfolio

Peer Review File

In Vivo Mitochondrial Base Editing Restores Genotype and Visual Function in a Mouse Model of LHON

Corresponding Author: Professor Hyunji Lee

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Nature Communications.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have addressed my concerns

Reviewer #3

(Remarks to the Author)
I have been asked to review the revised manuscript in the place of Reviewer 2, who has been unable to continue the peer review process. The journal editors have requested that I provide feedback on the responses to Reviewer 2's original points, which I have aimed to do as best I can, taking into account the initial reviewer's opinions and my own. I have done my best to assess the revised manuscript, bearing in mind that I did not read the original version.

Overall I found the points raised by Reviewer 2 to be generally valid, and I would echo their suggestion to avoid overstating results. I would encourage the authors to present the data in this study in a balanced matter, which I believe they have gone some way to achieving with the revised manuscript. I have added any further thoughts below on a point-by-point basis:

1) The authors are correct to point out that a proportion of LHON patients carry heteroplasmic, rather than homoplasmic, mutations (for m.11778G>A this is generally reported to be around 15%), but the fact remains that the pathogenicity of the mutation in the mouse model presented differs significantly to that seen in human patients. The discussion of species-specific thresholds in the revised manuscript addresses this to some extent, however, I feel it would be helpful to explicitly discuss the known pathogenesis of LHON in the introduction (i.e. majority of cases homoplasmic, incomplete penetrance, more common in males, known environmental triggers) so that the differences seen in the mouse model can be framed in the correct context. This is crucial, because whilst the model may be helpful for studying the molecular pathology of LHON in RGCs, it seems less likely to be useful for study of potential genetic triggers of pathology in asymptomatic carriers.

2) This is a significant point and it is unfortunate that the authors chose to use the BL6N strain rather than BL6J. To fully address this the authors should acknowledge in the materials and methods in the 'Mice' section that the BL6N strain carries the known Rd8 mutation and briefly discuss how they have mitigated for the potential impact of this on their results (i.e. include what they have outlined in their response letter) - it would be pertinent to mention the comparable results seen in the BL6J model, although this does not necessarily need to be included in the figures.

3) The original comment was very complex, and would require a lot of work to untangle... it is indeed interesting that there doesn't appear to be a worsening of phenotype with mutation burden, but I feel that it is probably beyond the scope of this study to address that finding, especially when the heteroplasmy level of the RGCs appears to differ so much from the tissues commonly used for genotyping. For this point I believe that adding the thorough description of human LHON pathogenesis to the introduction, suggested in point 1, along with the discussion of species-specific thresholds and the need for further

characterisation of this model, which have already been added, will be sufficient.

4) The additional data included by the authors has helped to clarify this point, but the authors should be more rigorous in how they present this data - based on Fig 2 there appear to be two frequent bystander mutations highlighted blue (m.G11182A & m.C11188T) that are present at high levels alongside the desired m.G11185A mutation in all of the blastocysts and pups shown. It would be relevant to describe these in the results, rather than just using the term 'bystander', describe which amino acid substitutions they result in (unless they are non-synonymous) and assess the potential pathogenicity of the substitution, e.g. by using mutpred. This can be presented alongside the data added to Extended Data Fig 9 to add transparency about what the bystander mutations do and strengthen the claim that they are not contributing to the observed pathology.

As a minor additional point - I do not understand the colouring used in Fig 2e, it is not clear to me what the gray portion of each bar represents, and from the data presented in 2c & 2d I would expect the dark blue portion of the bar to be much larger in the majority of cases (e.g. blastocyst 1 in 2c has 36.2% m.G11185A only, 41.4% m.G11185A plus bystander(s) and 4.8% bystander only). Perhaps the light blue and dark blue bars are overlapping? Overall the plot is confusing, and I would suggest displaying this as a stacked bar chart.

5) The experiments using a PBS control in one eye and AAV in the contralateral eye provide compelling evidence for successful correction of the mutation and reduction of heteroplasmy (although injection of an empty vector AAV may be a better control than PBS), and have also allowed the authors to assess the level of additional off-target edits. Extended Data Fig 14 is somewhat unclear - In (a) it could be made clearer what "OT1" and "OT2" refer to, I assume they were the only two 'off target' sequences found in the nuclear genome, is there a rationale for choosing a two-base mismatch rather than three or more? Where in the genome are OT1 & OT2, are they close to known protein coding genes? Assuming they are not, this information will help address the concerns that the phenotypic changes following correction of m.G11185A are due to off target effects.

I am less concerned about the perceived lack of functional rescue, as long as this is not overstated in the manuscript (i.e. the key point is that the mutation has been corrected and heteroplasmy successfully reduced, rather than that the pathology has been cured), so overall I think the authors have addressed this point sufficiently. I do wonder why the authors did not perform the visual phenotyping measurements pre- and post- AAV injection in the same mice, to get a direct comparison of the rescue following correciton, but perhaps this was not possible and comparing to PBS controls was the only feasible method.

6) Figure 7 now includes data from homoplasmic m.G11185A patient cells, this largely addresses the reviewer's original point, and shows that successful correction is indeed possible in homoplasmic human cells. I would again urge the authors to make sure that this is framed in the correct context - they acknowledge there are issues with bystander edits, and despite the enhancements they introduce, I feel that the final claim "Taken together, these results demonstrate that our enhanced sTALED-V28R catalyzes the desired single A-to-G conversion, while minimizing bystander edits, in human LHON patient-derived cells with the homoplasmic mutation, making it an effective base editing tool for gene therapy." is too strong, as there would clearly need to be significant additional validation and optimisation performed before this specific editor could be considered ready for gene therapy trials. Also, I would suggest that the authors include the data from the Complex I activity assay in the extended data - whilst the result is not significant, any trend seen can be interpreted in this context, and of course it could be hypothesised that the trend may become more significant if the heteroplasmy were reduced further.

As a general note on Fig 7, I would suggest that the authors think carefully about how to present the editing efficiency data, including the nomenclature used - the pathogenic mutation is m.G11185A, so technically the correction would be better described as m.A11185G, it's also worth considering that, up to this point, all mentions of editing efficiency % have given the percentage of the mutant A base, and intuitively it may make more sense to continue to use this metric in Fig 7 (e.g. in 7b the cells are not being edited from 0% to 5%, but rather corrected from 100% to 95%). Personally I feel that this would help keep the distinction between editing to introduce the mutation and editing to correct the mutation, but this can be left to the author's discretion.

7) The original reviewer's concerns regarding power can be addressed by using appropriate hypothesis testing to assess statistical significance - this appears to have been done for most data in Fig 6 (although the type of test used is not described in the figure legend), so assuming appropriate choice of test was made for each dataset, this appears to have been addressed. ERG waves are now included, which addresses the second part of this point, although I'm not sure if there is a suitable way of hypothesis testing this type of data?

8) This point has not been addressed, and I am unsure as to why it has not been possible to submit the sequencing data to an online repository, this is usually easy to achieve, and access keys for editors/reviewers can be generated prior to the data being made publicly accessible, so I don't see a reason for this to not be done.

9) The authors have added the requested reference to mitoBEs

10) The authors have included additional data to assess respiratory capacity of the MEF cells in culture, which I believe addresses the reviewer's suggestion sufficiently.

Overall I congratulate the authors on the additional work they have performed to revise this manuscript. Whilst I feel there are still some additional aspects to work on, as outlined above, I believe that these can be achieved with some careful editing of the manuscript, with a particular focus on presenting the data at face value, avoiding overstatement of the findings. I do not think that any further experiments will be required to achieve the above requests.

Version 1:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
The authors should be commended on addressing all of the outstanding comments thoroughly and carefully. I believe the study is now ready for publication and have no further suggestions to make regarding the revised manuscript.

Point-by-point response Response for Reviewer #3:

I have been asked to review the revised manuscript in the place of Reviewer 2, who has been unable to continue the peer review process. The journal editors have requested that I provide feedback on the responses to Reviewer 2's original points, which I have aimed to do as best I can, taking into account the initial reviewer's opinions and my own. I have done my best to assess the revised manuscript, bearing in mind that I did not read the original version.

Overall I found the points raised by Reviewer 2 to be generally valid, and I would echo their suggestion to avoid overstating results. I would encourage the authors to present the data in this study in a balanced matter, which I believe they have gone some way to achieving with the revised manuscript. I have added any further thoughts below on a point-by-point basis:

Response)

We would like to sincerely thank the reviewer for the careful, thorough, and insightful evaluation of our manuscript. We truly appreciate the time and effort dedicated to reviewing our work, as well as the constructive comments and suggestions that have significantly helped us to improve the quality, clarity, and scientific rigor of the study. The reviewer's comments demonstrated a deep understanding of the topic and raised important points regarding the experimental design, interpretation of the data, and contextualization of the findings within the broader field of LHON research. We found the feedback to be both fair and highly valuable, and we have revised the manuscript accordingly to address each point raised.

In particular, we have incorporated the reviewer’s suggestions by [e.g., expanding the introduction to include a more detailed description of LHON pathogenesis, clarifying the use of mouse strains in the methods section, improving data presentation in the figures, and refining the discussion to avoid overstating the results, as outlined in our point-by-point responses.

Once again, we appreciate to reviewer for constructive input and thoughtful guidance, which we believe have contributed greatly to strengthening the manuscript.

1) The authors are correct to point out that a proportion of LHON patients carry heteroplasmic, rather than homoplasmic, mutations (for m.11778G>A this is generally

reported to be around 15%), but the fact remains that the pathogenicity of the mutation in the mouse model presented differs significantly to that seen in human patients. The discussion of species-specific thresholds in the revised manuscript addresses this to some extent, however, I feel it would be helpful to explicitly discuss the known pathogenesis of LHON in the introduction (i.e. majority of cases homoplasmic, incomplete penetrance, more common in males, known environmental triggers) so that the differences seen in the mouse model can be framed in the correct context. This is crucial, because whilst the model may be helpful for studying the molecular pathology of LHON in RGCs, it seems less likely to be useful for study of potential genetic triggers of pathology in asymptomatic carriers.

Response1)

In response to the reviewer's insightful feedback regarding the need for a clearer explanation of the pathogenic mechanisms of LHON, we have supplemented the manuscript with a more detailed description of the currently known pathogenic processes. This addition has been incorporated into **lines 76–81 of the manuscript** and highlighted in yellow.

In most patients, the pathogenic mutation is present in a homoplasmic state and symptoms in patients appear in young adulthood, with around 80% of patients being male. Furthermore, some individuals exhibit incomplete penetrance, whereby the absence of disease symptoms despite carrying the mutation may, in part, result from a lack of environmental factors such as smoking and alcohol consumption, which are known to contribute to the pathogenic mechanism.

2) This is a significant point and it is unfortunate that the authors chose to use the BL6N strain rather than BL6J. To fully address this the authors should acknowledge in the materials and methods in the 'Mice' section that the BL6N strain carries the known Rd8 mutation and briefly discuss how they have mitigated for the potential impact of this on their results (i.e. include what they have outlined in their response letter) - it would be pertinent to mention the comparable results seen in the BL6J model, although this does not necessarily need to be included in the figures.

Response 2)

We agree with the reviewer's comment that it is important to clearly state the background of the mouse strain used. In the revised **Materials and Methods** section, we have addressed the mutations and visual disorders associated with the B6N strain and further included our perspective on how these issues were considered and resolved in the context of our study. The additional description can be found in the manuscript at **lines 736-741.**

We used C57BL/6N mice, which carry the rd8 mutation in the Crb1 gene. Since rd8 affects the outer retina, whereas LHON primarily involves the ganglion cell layer and retinal nerve fiber layer, we distinguished LHON-specific phenotypes from background effects. ND4 mutant mice generated on the C57BL/6N background exhibited RGC abnormalities consistent with LHON. To confirm that these findings were not strain-specific, we also generated m.G11185A mutant mice on the C57BL/6J background and observed the same pathological features (data not shown)

3) The original comment was very complex, and would require a lot of work to untangle... it is indeed interesting that there doesn't appear to be a worsening of phenotype with mutation burden, but I feel that it is probably beyond the scope of this study to address that finding, especially when the heteroplasmy level of the RGCs appears to differ so much from the tissues commonly used for genotyping. For this point I believe that adding the thorough description of human LHON pathogenesis to the introduction, suggested in point 1, along with the discussion of species-specific thresholds and the need for further characterisation of this model, which have already been added, will be sufficient.

Response 3)

We sincerely thank you for taking the time to examine our data in such detail and for approaching it from our perspective. We truly appreciate your thoughtful and careful consideration, which has been very helpful in refining our work. In addition, we have made a concerted effort to incorporate the revisions you suggested, so as to better address and clarify the description of our data.

4) The additional data included by the authors has helped to clarify this point, but the authors should be more rigorous in how they present this data - based on Fig 2 there appear to be two frequent bystander mutations highlighted blue (m.G11182A & m.C11188T) that are present at high levels alongside the desired m.G11185A mutation in all of the blastocysts and pups shown. It would be relevant to describe these in the results, rather than just using the term 'bystander', describe which amino acid substitutions they result in (unless they are non-synonymous) and assess the potential pathogenicity of the substitution, e.g. by using mutpred. This can be presented alongside the data added to Extended Data Fig 9 to add transparency about what the bystander mutations do and strengthen the claim that they are not contributing to the observed pathology.

As a minor additional point - I do not understand the colouring used in Fig 2e, it is not clear to me what the gray portion of each bar represents, and from the data presented in 2c & 2d I would expect the dark blue portion of the bar to be much larger in the majority of cases (e.g. blastocyst 1 in 2c has 36.2% m.G11185A only, 41.4% m.G11185A plus bystander(s) and 4.8% bystander only). Perhaps the light blue and dark blue bars are overlapping? Overall the plot is confusing, and I would suggest displaying this as a stacked bar chart.

Response 4)

We fully agree with the reviewer's comment that it is not sufficient to consider the off-target mutations observed within the editing range merely as "bystanders," but that precise descriptions of these mutations are required. First, the m.G11182A mutation results in a serine-to-asparagine substitution at position 399 (S399N) in both human and mouse genes. When analyzed using the MutPred program, as suggested by the reviewer, this mutation received a score of 0.462, indicating a minimal predicted impact. In contrast, the m.C11188T mutation showed a relatively higher MutPred score of 0.67. Interestingly, while the 341st amino acid of the *ND4* gene is isoleucine in humans, it is threonine in mice. Therefore, this mutation in the mouse *ND4* gene leads to a threonine-to-isoleucine substitution. Given the highly conserved role of protein-coding genes between mice and humans, the impact of this mutation is predicted to be minor. Taken together with the results shown in Extended Data Fig. 9, we believe that this mutation is unlikely to raise pathological concerns.

We considered that the description of this issue would be more appropriate in the Discussion section. Accordingly, in the Results section, we deleted the term "bystander" and replaced it

with *undesired*. In addition, mutations previously referred to as “bystanders” are now specifically described as the **m.G11182A** and **m.C11188T** mutations. **These mutations are first mentioned at line 134 of the manuscript**, and thereafter, all references to undesired editing are described directly in terms of either the corresponding amino acid changes or the nucleotide substitutions. And additional description about those undesired mutations has been included in the discussion section (**manuscript lines 382–396**). Furthermore, we also change the mentions on Main figure and Extended data figure.

We appreciate the reviewer’s comment that the presentation of the graph in Figure 2e could make interpretation difficult. In response, we have modified the graph to improve clarity. In the revised figure, each colored bar represents proportion of independent genotype in blastocysts and pups: the light blue bar indicates the genotype with the m.G11185A mutation as a single mutation allele; the navy bar represents the genotype carrying both the m.G11185A, m.G11182A, and m.C11188T mutation within the same allele; the gray bar denotes the genotype with no m.G11185A mutation but only the m.G11182A and m.C11188T editing, and the white bar represents the genotype of WT. The updated graph is attached below.

However, one important consideration in the design and execution of this study was that our mutant mice harbored undesired mutations mediated by Hifi-DdCBEs. Undesired mutations were m.G11182A and m.C11188T, which induce S339N, T341I amino acid substitutions in the *ND4* gene. We predicted the pathogenic score via Mutpred2[29], a computational tool for predicting protein structures and assessing the likelihood of disease-causing effects from genetic mutations. The results revealed that the pathogenicity score of the S339N mutation was 0.462 in both mice and humans, suggesting it is likely benign. Although the T341I mutation in the mouse ND4 gene yields a relatively high MutPred score of 0.67, it is unlikely to significantly impact protein function. This is because threonine is the native residue at position 341 in the mouse ND4 gene, whereas isoleucine is the wild-type residue at the corresponding position in the human ND4 gene. This suggests functional conservation between the two species. Furthermore, no functional defects resulting from undesired editing were observed (Extended Data Fig. 9). These results indicate that the observed defects in the mutant mice were solely attributable to the m.G11185A mutation. Nevertheless, precise introduction of only pathogenic mutations is essential for developing reliable gene-edited animal models.

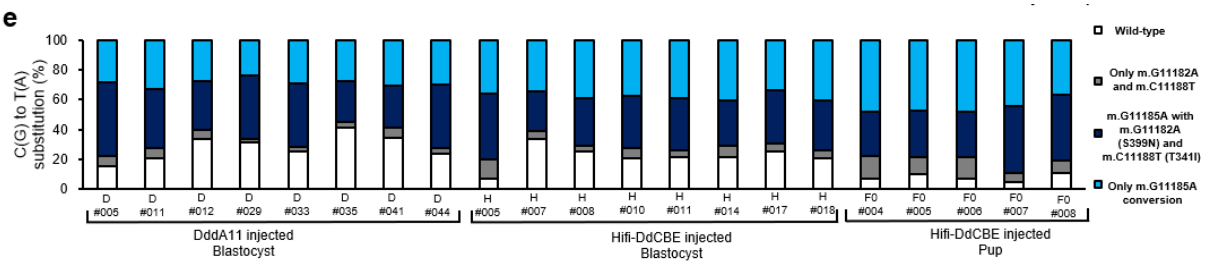


Figure 2

5) The experiments using a PBS control in one eye and AAV in the contralateral eye provide compelling evidence for successful correction of the mutation and reduction of heteroplasmy (although injection of an empty vector AAV may be a better control than PBS), and have also allowed the authors to assess the level of additional off-target edits. Extended Data Fig 14 is somewhat unclear - In (a) it could be made clearer what "OT1" and "OT2" refer to, I assume they were the only two 'off target' sequences found in the nuclear genome, is there a rationale for choosing a two-base mismatch rather than three or more? Where in the genome are OT1 & OT2, are they close to known protein coding genes? Assuming they are not, this information will help address the concerns that the phenotypic changes following correction of m.G11185A are due to off target effects.

I am less concerned about the perceived lack of functional rescue, as long as this is not overstated in the manuscript (i.e. the key point is that the mutation has been corrected and heteroplasmy successfully reduced, rather than that the pathology has been cured), so overall I think the authors have addressed this point sufficiently. I do wonder why the authors did not perform the visual phenotyping measurements pre- and post- AAV injection in the same mice, to get a direct comparison of the rescue following correciton, but perhaps this was not possible and comparing to PBS controls was the only feasible method.

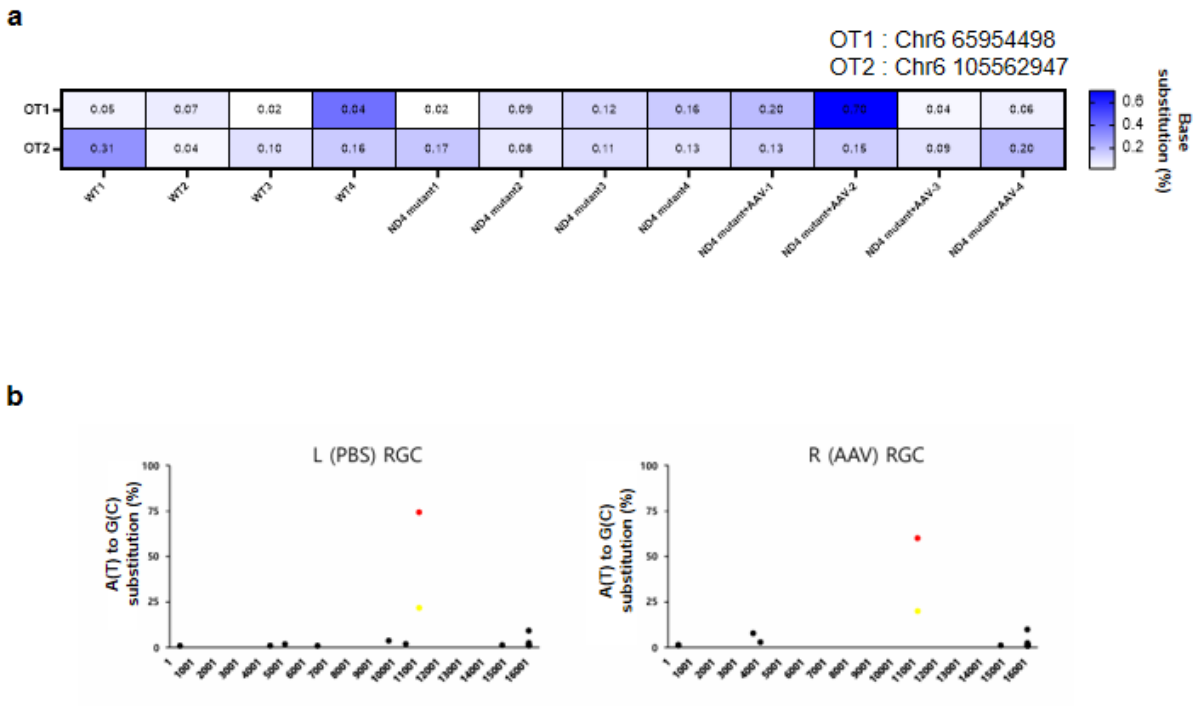
Response 5)

We sincerely thank the reviewer for pointing out the lack of detailed information regarding the off-target sites. In response to the reviewer's comments, we would like to clarify the following.

First, mitochondrial genome editors function through TALE proteins binding to DNA and recognizing the target sequence. Therefore, the analysis of potential off-target activity in the nuclear genome was conducted based on the TALE sequence. Specifically, we used the

Cas-OFFinder program to identify potential off-target sites with ≤2 mismatches to our designed TALE sequence, and subsequently validated off-target activity at these sites by defining the editing range and performing deep sequencing. The reason for restricting our analysis to sites with ≤2 mismatches is that TALEs are highly specific DNA-binding proteins; even a single mismatch significantly reduces binding probability, and activity at sites with ≥2 mismatches is rarely detectable (PMID: 23408851, 24531420, 24569350). Based on these references, we therefore focused on sites with ≤2 mismatches. Information about the identified off-target sites has now been added to Extended Data Figure 14 and Supplementary Table 3 and is also provided below. These sites were all located on Chromosome 6 and were confirmed not to be in proximity to protein-coding genes.

Regarding the reviewer's concern about the lack of visual function assessment before and after AAV injection, we would like to clarify that intravitreal injections were performed at postnatal week 2 to maximize editing efficiency before the maturation of retinal ganglion cells (RGCs). At this stage, the animals are not yet fully matured, and it is technically challenging to obtain reliable measurements of visual function. Phenotypic analyses of visual function were then conducted at postnatal week 26. Since there are considerable developmental differences in vision between 2 and 26 weeks of age, we anticipated that intra-individual variability in vision would be too high to provide reliable data. Therefore, we considered PBS-injected mice as the most appropriate control group. We hope this explanation addresses the reviewer's concern.



Extended Data Figure 14

Supplementary Table 3. Off-target site in nuclear genome

Name	Location	Mismatch	Direction	Sequence
OT1	Chr 6 65954498	2	+	TTTGAAGTCCTaGGGCaAT
OT2	Chr 6 105562947	2	+	TTgGAAGTCCTCGGGaCAT

Supplementary Table 3

6) Figure 7 now includes data from homoplasmic m.G11185A patient cells, this largely addresses the reviewer's original point, and shows that successful correction is indeed possible in homoplasmic human cells. I would again urge the authors to make sure that this is framed in the correct context - they acknowledge there are issues with bystander edits, and despite the enhancements they introduce, I feel that the final claim "Taken

together, these results demonstrate that our enhanced sTALED-V28R catalyzes the desired single A-to-G conversion, while minimizing bystander edits, in human LHON patient-derived cells with the homoplasmic mutation, making it an effective base editing tool for gene therapy." is too strong, as there would clearly need to be significant additional validation and optimisation performed before this specific editor could be considered ready for gene therapy trials. Also, I would suggest that the authors include the data from the Complex I activity assay in the extended data - whilst the result is not significant, any trend seen can be interpreted in this context, and of course it could be hypothesised that the trend may become more significant if the heteroplasmy were reduced further.

As a general note on Fig 7, I would suggest that the authors think carefully about how to present the editing efficiency data, including the nomenclature used - the pathogenic mutation is m.G11185A, so technically the correction would be better described as m.A11185G, it's also worth considering that, up to this point, all mentions of editing efficiency % have given the percentage of the mutant A base, and intuitively it may make more sense to continue to use this metric in Fig 7 (e.g. in 7b the cells are not being edited from 0% to 5%, but rather corrected from 100% to 95%). Personally I feel that this would help keep the distinction between editing to introduce the mutation and editing to correct the mutation, but this can be left to the author's discretion.

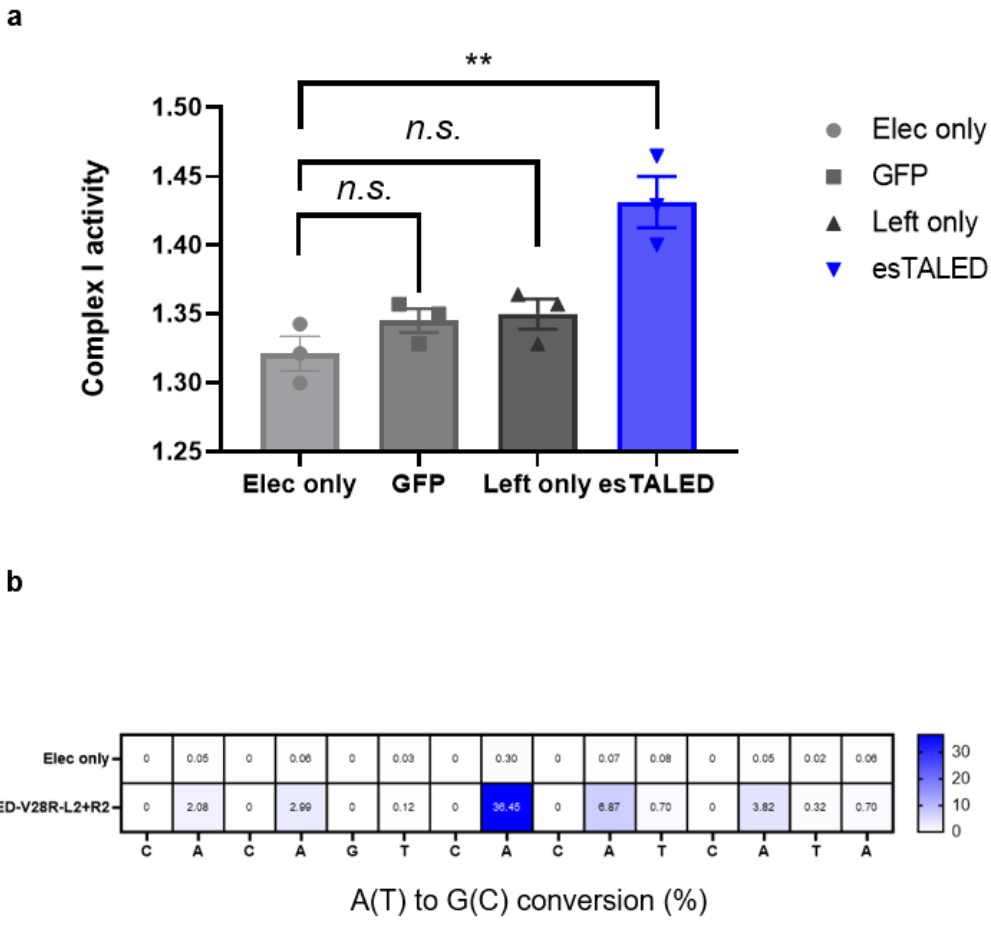
Response 6)

We fully agree with and appreciate your comment that an overly strong tone could lead to overinterpretation of the data. Accordingly, we have rephrased the relevant parts as suggested. The revised text has been highlighted in yellow in the **manuscript file Line 372-373** and is also described below.

making it an effective base editing tool for gene therapy -> providing a basis for its potential application in future gene therapy.

Furthermore, in response to the reviewer's comment that higher editing efficiency might result in a significant increase in complex I activity, we optimized the electroporation conditions and achieved an editing efficiency of up to 36%. Under these conditions, we observed a subtle yet statistically significant increase in complex I activity. However, the increase in activity was not

proportional to the editing efficiency. This revision has been incorporated into the **manuscript** and **Extended Data Fig. 18**, and the corresponding data are attached below. We fully agree with the reviewer's comment regarding the manner of data presentation, and we have therefore revised the notation to A11778G. We are deeply grateful for this careful and thoughtful review.



Extended Data Figure 18

7) The original reviewer's concerns regarding power can be addressed by using appropriate hypothesis testing to assess statistical significance - this appears to have been done for most data in Fig 6 (although the type of test used is not described in the figure legend), so assuming appropriate choice of test was made for each dataset, this

appears to have been addressed. ERG waves are now included, which addresses the second part of this point, although I'm not sure if there is a suitable way of hypothesis testing this type of data?

Response 7) We acknowledge that the omission of the statistical test used in the figure legend was an oversight. The analyses were performed using a two-tailed Student's t-test, and this information has now been added to the legend of Figure 6. For the aspect of statistical significance in ERG waves, there is no dedicated statistical test specifically for ERG waveforms. Instead, statistical significance is typically assessed based on quantitative parameters such as amplitude (a- and b-wave) and implicit time, using standard methods such as the Student's t-test, ANOVA, or mixed-effects models depending on the study design. Thus, we would like to note that we were unable to identify a specific statistical method dedicated to validating ERG waveforms, and we kindly ask for your understanding regarding this point.

8) This point has not been addressed, and I am unsure as to why it has not been possible to submit the sequencing data to an online repository, this is usually easy to achieve, and access keys for editors/reviewers can be generated prior to the data being made publicly accessible, so I don't see a reason for this to not be done.

Response 8) As the dataset contained an extensive amount of sequencing data, we previously faced time constraints in organizing these files. We would like to inform you that this task has now been completed in the current revision, and the data can be accessed through the provided SRA number PRJNA1334167.

9) The authors have added the requested reference to mitoBEs

Response 9) We appreciate your acknowledgment of the references we have added.

10) The authors have included additional data to assess respiratory capacity of the MEF

cells in culture, which I believe addresses the reviewer's suggestion sufficiently.

Response) We would like to express our sincere gratitude for your thorough review of the additional experiments performed in response to your comments, and for your favorable assessment.