EMBO Molecular Medicine

Multichannel optical cochlear implants enable spectrally distinct auditory activity

Niels Albrecht, Elisabeth Koert, Anna Vavakou, Lennart Roos, Lukasz Jablonski, Juan Marcoleta, Josep Cardona Audi, Jette Alfken, Eric Klein, Patrick Ruther, Thomas Mager, Kathrin Kusch, and Tobias Moser

Corresponding author(s): Tobias Moser (tmoser@gwdg.de)

Review Timeline:

Submission Date:	13th May 26
Editorial Decision:	9th Jul 26
Revision Received:	29th Aug 26
Editorial Decision:	19th Sep 26
Revision Received:	25th Sep 26
Accepted:	29th Sep 26

Editor: Jingyi Hou

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.)

9th Jul 2026

Dear Tobias,

Thank you again for submitting your work to EMBO Molecular Medicine. I apologize for the delay in the review process. This is because we invited the same referees to evaluate the side-by-side submissions, which required additional time for them to assess both manuscripts. We have now received reports from the three referees who agreed to review the current manuscript. As you will see from the comments below, the referees found the study both interesting and relevant and expressed strong support for its publication. They also raised a number of relatively minor issues, which we ask you to address in your revised manuscript.

The referees' recommendations are clear and need not be repeated here. All of the issues raised need to be carefully addressed. As you may already know, our editorial policy allows in principle only one round of major revision, so it is essential that your response fully addresses the referees' comments. Please feel free to contact me if you would like to discuss any of the issues raised by the referees in further detail.

EMBO Molecular Medicine has a "scooping protection" policy, whereby similar findings that are published by others during review or revision are not a criterion for rejection. Should you decide to submit a revised version, I do ask that you get in touch after three months if you have not completed it, to update us on the status.

Please also contact us as soon as possible if similar work is published elsewhere. If other work is published, we may not be able to extend the revision period beyond three months.

Please read below for important editorial formatting and consult our author's guidelines for proper formatting of your revised article for EMBO Molecular Medicine.

I look forward to receiving your revised manuscript.

Kind regards,
Jingyi

Jingyi Hou
Senior Editor
EMBO Molecular Medicine

*****

When preparing your revised manuscript, please refer to our guidelines: https://link.springer.com/journal/44321/submission-guidelines#cms-Revised-submissions. We perform an initial quality control of all revised manuscripts before re-review; failure to include requested items will delay the evaluation of your revision.

We require:

1) A .docx formatted version of the manuscript text (including legends for main figures, EV figures and tables). Please make sure that the changes are highlighted to be clearly visible.

2) Individual production quality figure files as .eps, .tif, .jpg (one file per figure). For guidance, download the 'Figure Guide PDF': https://media.springernature.com/original/springer-cms/rest/v1/content/27825798/data/v1.

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

4) A complete author checklist, which you can download from our author guidelines. Please insert information in the checklist that is also reflected in the manuscript. The completed author checklist will also be part of the RPF.

5) Please note that all corresponding authors are required to supply an ORCID ID for their name upon submission of a revised

manuscript.

6) It is mandatory to include a 'Data Availability' section after the Materials and Methods. Before submitting your revision, primary datasets produced in this study need to be deposited in an appropriate public database, and the accession numbers and database listed under 'Data Availability'. Please remember to provide a reviewer password if the datasets are not yet public.

In case you have no data that requires deposition in a public database, please state so in this section. Note that the Data Availability Section is restricted to new primary data that are part of this study.

7) For data quantification: please specify the name of the statistical test used to generate error bars and P values, the number (n) of independent experiments (specify technical or biological replicates) underlying each data point and the test used to calculate p-values in each figure legend. The figure legends should contain a basic description of n, P and the test applied. Graphs must include a description of the bars and the error bars (s.d., s.e.m.).

8) At EMBO Press we ask authors to provide source data for the main manuscript figures. You will receive a separate email with instructions for providing source data with your revised manuscript, including how to upload and organize the files.

9) 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.

10) We replaced Supplementary Information with Expanded View (EV) Figures and Tables that are collapsible/expandable online. A maximum of 5 EV Figures can be typeset. EV Figures should be cited as 'Figure EV1, Figure EV2" etc... in the text and their respective legends should be included in the main text after the legends of regular figures.

- For the figures that you do NOT wish to display as Expanded View figures, they should be bundled together with their legends in a single PDF file called *Appendix*, which should start with a short Table of Content. Appendix figures should be referred to in the main text as: "Appendix Figure S1, Appendix Figure S2" etc.

- Additional Tables/Datasets should be labeled and referred to as Table EV1, Dataset EV1, etc. Legends have to be provided in a separate tab in case of .xls files. Alternatively, the legend can be supplied as a separate text file (README) and zipped together with the Table/Dataset file.

11) The paper explained: EMBO Molecular Medicine articles are accompanied by a summary of the articles to emphasize the major findings in the paper and their medical implications for the non-specialist reader. Please provide a draft summary of your article highlighting

- the medical issue you are addressing,

- the results obtained and

- their clinical impact.

This may be edited to ensure that readers understand the significance and context of the research. Please refer to any of our published articles for an example.

12) Author contributions: You will be asked to provide CRediT (Contributor Role Taxonomy) terms in the submission system. These replace a narrative author contribution section in the manuscript.

13) A Conflict of Interest statement should be provided in the main text.

14) Please provide a 'Synopsis' to further enhance discoverability. Synopses are displayed on the journal webpage and are freely accessible to all readers. They include a short stand first (maximum of 300 characters, including space) as well as 2-5 one-sentences bullet points that summarizes the paper. Please write the bullet points to summarize the key NEW findings. They should be designed to be complementary to the abstract - i.e. not repeat the same text. We encourage inclusion of key acronyms and quantitative information (maximum of 30 words / bullet point). Please use the passive voice. Please attach these in a separate file or send them by email, we will incorporate them accordingly.

Please also provide a visual abstract as a separate PNG file with dimensions of exactly 550 px (width) × 300 px (height) to highlight the paper on our homepage.

Please refer to published papers for examples.

15) All Materials and Methods need to be described in the main text using our 'Structured Methods' format. According to this format, the Methods section includes a Reagents and Tools Table (listing key reagents, experimental models, software and relevant equipment and including their sources and relevant identifiers) followed by a Methods and Protocols section describing the methods, ideally using a step-by-step protocol format. The aim is to facilitate adoption of the methodologies across labs.

Please download and fill our Reagents and Tools Table template (.docx), which you can find in our author guidelines: https://link.springer.com/journal/44320/submission-guidelines#structuredmethods.

When submitting your revised manuscript, please do not include the Reagents and Tools Table in the Methods section of the manuscript but upload it as a separate file choosing the file type "Reagent Table".

EMBO Molecular Medicine has a "scooping protection" policy, whereby similar findings that are published by others during review or revision are not a criterion for rejection. Should you decide to submit a revised version, I do ask that you get in touch after three months if you have not completed it, to update us on the status.

***** Reviewer's comments *****

Referee #1 (Comments on Novelty/Model System for Author):

the optogenetic actuator used is likely too slow to support hearing in a clinically meaningful way. still data provide and important step forward towards the development of a treatment that can achieve that.

Referee #1 (Remarks for Author):

In their manuscript Albrecht et al. present their data on optogenetic tonotopic simulation of spiral ganglion neurons (SGN) using a Multi(up to ten)channel linear LED array implanted into the cochlea in the sense of a hearing-restoring cochlear implant (CI) using gerbils as a model system. The optogenetic actuator used is ChReef.
The authors clearly show that stimulus energies needed to achieve sufficient optical SGN stimulation are within the translationally feasible range. Though slightly less definite (because only data from a previous study were used as reference) they also demonstrate that the optogenetic ChReef-CI outperforms electrical CIs particularly in terms of tonotopic discrimination accuracy. They also state that tonotopic discrimination accuracy observed for the optogenetic CI was indistinguishable form native/acoustic stimulation. While this wording is technically correct, the indistinguishability is probably partially due to the limited resolution their outcome measure - a 32 channel linear electrode array recording from the inferior colliculus. Hence, the "indistinguishable form acoustic stimulation" statements should appear more in context.
(Minor: it would be interesting to understand if/how the authors think the long inactivation time constant of ChReef could support functional hearing.)
Specific comments (in order of appearance in ms)
- Fig 1C: the symbols for LED and Laser stim are not intuitive here (only in this fig)
- The electronic CI reference data stem form a paper the authors published roughly 7 years ago. Could surgeon / or surgeon-training related-effects explain the poor performance of the electronic CIs?
- Fig 2 D-F and related areas in the ms: the numbers above the error bars refer to "n"s, I understand.
o Given that there were only 8,8, and 12 animals included in D-F, resp. I assume these "n" arise from repeated recordings?
o "n" varies across SPLs tested. So the average dots shown refer to different populations at each SPL. I would argue that all average values should be calculated from the same reference population (i.e. only form the data where values exist for each SPL level)
o Analog issue in data for figure 6
- Fig 3D: are the 4 emitters distributed across the entire cochlea?
- Page 9 and Fig 4: I agree to the conclusion that there seems to be a spread of light to the opposite side of the cochlea - explaining the faint-ish second peak indicated by the black arrow but also seen in several other panels. Possibly because this reviewer is a retinologist, the logic for the sharp secondary peak (the authors call it broader spread of activation with central minimum) is not becoming clear to me. Potentially also other readers might have this probelm
- Abbreviation RC is not introduced
- Fig 5 C has n unusual aspect ratio that might faint variability actually present in the data
- The logic for the "LED main peak" analysis is not becoming fully clear. This would be to simulate how optogenetic hearing could perform if higher density LED arrays were used?
- Figure 5G Cf should be subscript f
- Fig 6 H and 8 J: A LMM was used to probe for statistically significant differences between cohort at the individual ERs
o 1) data are skewed. LMM can handle that but needs to be checked that residuals are not skewed
o In fig 8J: effects are clearly not linear in native/acoustic hearing, so an LMM wouldn't be able to describe the data appropriately.
- The discrimination spacing at the neurometric threshold floors at 0 mm (i.e. same electrode on the recording array) for the

acoustic stimulation. It is logically not truly 0, just the resolution of the measurement method is insufficient - that fair and likely unavoidable, but should be accounted for in the interpretation
Overall nice work. Thanks.

Referee #2 (Remarks for Author):

This study by Albrecht et al. tests the efficacy of a new channelrhodopsin ChReef in the context of optical cochlear stimulation with a readout of multi-unit activity in the inferior colliculus. This is part of an ongoing program of work by this group to establish the clinical feasibility of optical cochlear implants. This is very important and exciting work and the present study is a worthy addition. The work is admirably thorough, the key conclusions are well supported, and the writing is clear. I have only minor concerns around language and emphasis that should be addressed before publication.

In general, the language used in describing the results seems overly favorable toward optical stimulation (i.e., it overstates the degree to which the current results support the similarity of optical and acoustic stimulation). This is in part due to an overemphasis of statistical significance rather than effect size. If two distributions have well separated centroids but fail to pass a statistical significance test, that should not be taken as evidence that the centroids are comparable. Think about the confidence intervals on the difference of the centroids: example A is 50% (-1%,101%) and example B is 0% (-1%, 1%). Both of these would fail the significance test. Example B is strong evidence that the centroids are comparable. Example A is not. The interpretation of Example A is that, given the data, the best guess of the centroid difference is 50% but in fact it might well be 0% or 100%.

The discussion of spectral spread on page 13 is a good example. There is much talk of statistical significance, and of the fact that optical stimulation is closer to acoustic than to electric, but at no point is the magnitude of the difference in spread between acoustic and optical for a given evoked rate ever stated. If one looks at figure 6H, the difference between the lines for acoustic and LED Main Peak is sizable. If one chooses a particular evoked rate, say 200 Hz, and then looks at the y axis values for each of those lines, the difference is probably on the order of 30 or 40%. The same can be said of Figure 8I. Choose a particular evoked firing rate, again say 200 Hz, and look at the y axis values. They are rather different. It may be that they are not statistically significantly different, but if you had to guess whether the difference was 0 or a factor of 2, the evidence would support a guess of a factor of 2.

Another example of overly favorable language is that around the reduction of SGN density due to injection. It is described as mild in the main text. But looking at figure EV2C, the average results in the right panel appear to have an effect size on the order of 33%, meaning the injection results in a 33% loss of SGNs relative to control. I wouldn't describe that as mild. And if we look at the left panel for the apex, the difference looks rather extreme.

Let me be clear: I don't see any major problems with the actual data or analyses. It is simply the language used to describe the results and choice of which aspects of the statistics to focus on. These problems are easily fixed.

Regarding the comparison of light pulses to sustained tones: I think the approach used of limiting the analysis to the onset time window is perfectly fine, but going forward it might be good to also test tone bursts of, say, 10 ms with the five ms onset and offset ramps that are currently used and/or to try to elicit sustained activity with the light stimulus. The latter is presumably what is being worked toward in order to elicit the most physiologically realistic patterns of activity in response to complex sounds.

The analysis that results in the establishment of the maximum light evoked activity being equivalent to the activity evoked by a 46 dB SPL pure tone is excellent. This type of benchmarking makes the comparison very clear. That said, this is also reported to be approximately 10 dB above the acoustic threshold, which is quite limited. For the purposes of this study, it's fine because the focus is on sharp spectral tuning, which is most evident for stimuli close to threshold. But going forward, it's going to be important for optical stimulation to be able to illicit activity that is comparable to that elicited by real-world sounds, which are of course typically at much higher levels. I think this should be included in the limitations section of the Discussion. Related to this, I don't understand the statement in the Discussion about the response range up to 200 hertz being most relevant for optical coding strategies. Is the hope that central plasticity following implantation would adopt the dynamic range downstream to make use of the limited dynamic range peripherally?

Referee #3 (Remarks for Author):

This manuscript characterizes intensity and spectral encoding by a multichannel optical cochlear implant (CI), which employed green-LED stimulation of spiral ganglion cells expressing the low-energy-threshold channelrhodopsin, ChReef. Optical stimulation was compared to acoustic and electrical cochlear stimulation using extracellular recordings along the tonotopic axis of the gerbil inferior colliculus (IC). Optical stimulation using ChReef replicated sub-μJ optical thresholds and showed intensity coding at least as high as that by 46-dB SPL pure tones. It provided spectral specificity comparable to acoustic stimulation for low-to-moderate IC activation levels but also revealed activation of off-target tonotopic regions. Discriminability between optical

channels was more similar to that by acoustic than by electric stimulation, even at relatively high levels of IC activation.

Overall, the study was thorough and technically well executed. The demonstration of an efficient multichannel oCI is an important advancement for the clinical translation of optogenetic stimulation. A nice strength of the work was its mapping of frequency into cochlear place/distance, and vice versa, for more direct comparisons between stimulation modalities. My concerns were mostly around issues of clarity and presentation in the manuscript. Please see below some general comments as well as several specific points regarding the text and figures.

General:
1.) For readers who are not fully up to-date on optogenetic literature, the abstract and introduction could do a better job of highlighting what is novel and significant about the study. Higher spectral selectivity with oCIs (relative to eCIs) has been shown, of course. My sense, however, is that this is one of the first demonstrations of a clinically feasible multichannel oCI (via the efficiency of LED-ChReef pairing). Is that the main takeaway? I see that the introduction says: "This challenged the utility of implantable microscale emitters for preclinical characterization of multichannel oCI (Dieter et al., 2020; Keppeler et al., 2020)". But a little more context would give better emphasis on the new contribution; for instance, example(s) of those past challenges with multichannel oCIs and what has been shown with ReefCh oCIs so far, and what hasn't (the gaps to filled).

2.) Based on the title "multichannel optical cochlear implants..." and terms like channel discriminability, in the abstract and introduction, I had an initial impression that multiple oCI channels would be stimulated together (concurrently) and the resulting ICC patterns analyzed. That was likely my own bias going into it. Nevertheless, I think this is worth considering this in the discussion. Whereas electric CIs necessitate interleaved stimulation to avoid charge interaction, optical CI presumably will not. Is that correct? Therefore, future assessments of simultaneous channel stimulation might be more clinically relevant (e.g., Middlebrooks & Synder, 2007; 10.1007/s10162-007-0070-2). The interactions of those channels are potentially more important than discrimination between different channels alone. This also seems relevant considering the secondary ICC activation peaks, shown in this study, which might interact (i.e., get enhanced).

3.) Supplemental Figures S3 and S7 don't seem to be referenced in the text. These should be described or excluded.

Specific Points:
Abstract:
1.) Following the sentence, "We characterized the oCI encoding...", you might add that: oCI encoding was compared to that achieved by acoustic and electrical CI stimulation.
2.) 46 dB tones, should be 46 dB SPL pure tones.
3.) "moderate activity levels" wasn't clear without reading the main text; slightly more meaningful might be "moderate levels of ICC activation"

The paper Explained:
1.) "high levels of midbrain activity" - again, I wasn't clear on this until reading the text. Maybe qualify in some way, e.g., "...and, compared with acoustic stimulation, relatively high levels of midbrain activation with only low micro-joule energy". Using the full spelling of micro- to mirror that of nano- in the same sentence.

Introduction:
1.) "noisy backgrounds" - plural. You might add "or competing sound sources".
2.) Consider: "sound-frequency spatial (tonotopic) organization"
3.) "multipolar stimulation" - since this is not contrasted to monopolar, consider more general terms: e.g., "current focusing techniques"
4.) 2nd paragraph: For first sentence, I suggest changing to: "However, the power requirement of optogenetic SGN stimulation has remained a challenge for oCIs to become...". Using the present perfect tense throughout this section (e.g., has shown) might be preferrable to the simple past (e.g., showed) to indicate the work is still ongoing.
5.) Last paragraph: consider changing "for up to moderate evoked ICC activities" to "for stimulus intensities eliciting low-to-moderate levels of ICC activities." This clarifies the relation between stimulus intensity and ICC activation.
6.) Last paragraph: the sentence beginning: "The tonotopic organization..." was a bit unclear. More closely than electric stimulation? Does it mean that considering the previous sentences, tonotopic organization was overall more like acoustic than eCI responses? Or that tonotopic organization for oCI activity was appropriate to the intracochlear placement of the emitters? If the later, consider putting this sentence after the sentence about micro-CT.
7. Last paragraph: I suggest: "...single emitters reached that elicited by 46 dB SPL pure tones"

Results
1.) The first paragraph could better delineate the use of laser stimulation for the preliminary assessment, which is followed by LEDs for the main experiment. Something like: "After a minimum of 8 weeks, we first assessed auditory activation using a single intracochlear optical fiber and recordings of the optically evoked auditory brainstem response (oABR). Of the 15 gerbils, 13 showed oABR responses to sub-uJ light pulses."
2.) I was a bit confused by the animal numbers. 15 received AAV injections, and the first paragraph says 13 showed oABR sub-uJ thresholds (by extension <10 uJ). The results then jump to "out of eight animals..." without explaining the reduction. The

methods somewhat clarify that only animals with ICC thresholds <10mJ were used, but it says that was 9 (not 8) out of 13 AAV-injected animals. Is that because one had the mid-turn cochleostomy and was excluded? I would suggest a clearer breakdown of these numbers, so the reader doesn't have to jump around to understand. Also, how many animals had the 10- vs. 5-LED arrays?
3.) It's unclear why "for acoustic stimulation" is included in the parentheses (for acoustic stimulation, Fig. 1A). Also, what does "ipsilateral" for oCI animals and "contralateral" for acoustic animals mean? Ipsi and contra to the ICC recording hemisphere? "Acoustic animals" should be non-injected animals, I assume, since all animals had intact hearing. It's probably best to state here that implanted animals were not deafened.
4.) I suggest specifying the analysis duration, 13 ms, in the paragraph starting, "We first computed evoked firing rates (evoked rate, ER)..."
5.) The dynamic range paragraphs needs to be examined for clarity. The sentence "This is why the stimulus range..." could say, "The dynamic range was quantified as 10-90% of the max ER across 32-recording sites (Fig 2G)"; then explain how saturation played a role. Otherwise, it was difficult initially to understand why optical DRs in Fig. 2G were so much smaller than electrical just from comparing rate-level functions in Fig. 2E & F. The following bit about calculating 10% max firing rate (Fig S4): does this mean the 10-100% range? I suggest putting the last sentence about expressing dB values relative to threshold as the 2nd sentence following the introduction of rate-level functions. Please also state this in the captions, since I had trouble interpreting 0 dB without revisiting the text. Lastly, it would be useful to have the maximum stimulus intensities listed in this text, or the figure caption, as it took some time to find it in the methods.
6.) Fig. 7: As a colorblind reader, I wasn't sure if the mean rate-level functions in Fig. 2E & F are different colors (red/green)? Please consider that color choice throughout the manuscript.
7.) I didn't see how 46- and 42-dB SPL are represented in Fig 2H-J. Fig 2J show's distributions of max firing rates with "max equivalent SPL" medians of ~10 dB. Is this because dB is expressed relative to pure-tone threshold, rather than standard dB SPL? Also, it was unclear what the thin colored horizontal to vertical lines in Fig 2I represent. Are those mean values from Fig 2E and F? Does the Fig 2K-L caption mean to say, "The data shown in E-F was mapped onto the rate-level function for acoustic stimulation as shown in I"? Please examine the caption for correct panel references. Lastly, I don't see that Fig 2K & L are described anywhere in the text.
8.) Reference to Fig. S4 is mentioned before S2 or 3; consider rearranging those figures to match their order of appearance in the text.
9.) I suggest indicating the directions of dorsal and ventral recording positions in Fig 3, or at least in the caption.
10.) The arrowhead in Fig 3B somewhat contradicts the statement in the text that "additional peak(s) [were] positioned toward the lower frequency range relative to the main peak." That extra peak is in the high-frequency direction.
11.) Add "from" in: "This analysis indicated that secondary peak of ICC activity likely resulted [from] the spread of light..."
12.) "...with a peculiar central activity minimum": this wasn't clear. Does this refer to the bimodal shape of the STCs for firing rates of >=200 Hz (dotted lines)?
13.) Fig. 4: Does the caption for 4A mean to say, "...oriented towards the RC as depicted in Fig. 1B", not Fig. 1C? Also, the caption doesn't describe the thin colored line, the Greenwood function.
14.) 0.60 [0.14, 0.89] mW; I suggest indicating these and other values given this way are median and 25th, 75th quartiles.
15.) For the paragraph describing Fig. 5, I found the terms confusing when referring to "single spectral peaks". There is an initial contrast between a "single spectral peak" and "two peaks." For Fig. 5D, all about the two peak responses, the text describes the intensity range of the "main spectral peak," but Fig5A and 5D use the term "single peak." Then, for giving the maximal ER, the text uses "single peak" in referring to STCs without a secondary peak (based on the Fig 5E caption). Please evaluate and maybe only use the term "main peak" when referring to two-peak STCs.
16.) The sentence, "Optogenetic stimulation was tonotopically organized..." - this was established in the previous text/figures. Consider something like, "We quantified the tonotopic organization of 10-LED optogenetic stimulation and the corresponding variations in ICC threshold"
17.) Fig. 5: What is the red line in 5C? I don't think I found this in the text or caption. Does this have something to do with the positive-going outlier data for the filled circles? That outlier should probably be mentioned. Also, I would suggest the caption to say, "Each combination of marker type and color corresponds to ICC recordings for an individual animal". Instead of "Markers indicate ICC recordings from the same animal"
18.) In the first paragraph describing Fig 6: It was not immediately clear from the text or caption how SoE was computed. Because this is important, it's probably best to define clearly without needing to refer to the methods. Perhaps add to the following sentence: "Specifically, the width of the STC was defined as the distance between [ICC threshold contours (d' = 1)] at the dorsal-most and ventral-most activated electrode, [measured at the lowest intensity level that elicited each ICC activity level.] In Fig. 6A, I suggest adding examples of the SoE measurement for 200 Hz and 400 Hz, in addition to the current 300-Hz example, for greater clarity of the method.
19.) The first paragraph for the discriminability analysis: could be slightly more descriptive, since readers will presumably not have read the methods in detail or may not be that familiar with RSA. I would suggest stating first in more general terms what it is that RSA captures; ...correlation in the overall geometry between ICC response maps (?). As written, it's also somewhat unclear whether RSA is the ultimate comparison between modalities, or the pairwise comparisons within each modality. An example could provide a more intuitive sense of how it works. For instance, after the sentence ending "..., measured as correlation distance" it could say, "For example, Fig 7B shows how... [so forth]". This could also mention the matching of excitation levels in the analysis. Lastly, the term representational dissimilarity (RD) is not well defined here. Is it the same as the "correlation distance" mentioned above?
20.) The following paragraph for the multidimensional scaling plots (Fig 7C-E): It is somewhat difficult to pick out the patterns

described for each modality. Here again, describing an example or two from the figure would be helpful, maybe assisted by annotation in the figures (e.g., arrows).
21.) Fig. 8F-H: consider making the triangle markers the same light shading as the individual data lines. The darker markers, matching the mean data, initially seemed as if that was the spacing boundary across all animals for each modality, which is certainly not the case for optical.
22.) The sentence, "In contrast, optical thresholds did not differ significantly from acoustic thresholds...": Consider adding something like "despite being ~0.5-mm higher for optical than acoustic across those activity levels"; That offset is still stands out, so it' s probably worth noting.

Discussion:
1.) In the sentence starting: "Increasing stimulation energy of individual oCI channels up to 3 µJ..." Does it later mean to say, "...approached [the maximum] evoked firing rates achieved by pure tone acoustic stimulation"?
2.) In the same paragraph, the phrase "and/or multi-channel stimulation present important avenues": does multi-channel stimulation in this context (and the following sentence) refer to combined (i.e., simultaneous) channel stimulation?
3.) The meaning of the sentence, "This will also allow charting the range of cumulative ICC activity..." was not clear. What is "This" referring to exactly? Please examine for clarity.
4.) In the spread of excitation (SoE) paragraph, is it possible to relate these firing rate values (>=150 Hz for electrical stimulation, and <100 Hz for acoustic stimulation) back to the equivalent acoustic sound pressure levels? - the same way that 46 dB SPL is derived as the highest SPL by the oCI? That could provide a more meaningful interpretation of relevant stimulus ranges. Also, later in this paragraph, why is up to 200 Hz the most relevant range for optical coding strategies?
5.) For the secondary peak paragraph: consider adding before the 3rd sentence something like, "The secondary peak could account for much of the broadening of SoE at higher optical intensities (i.e., for FR >= 150 Hz). On the other hand, when focusing on the main spectral peak (i.e., excluding the secondary peaks), the SoE remained much lower and closely..."
6.) In the sentence, "A key question was therefore whether the relationships among ICC responses..." consider changing "distinct responses" to "distinct tonotopic populations"
7.) It may be worthwhile to highlight and discuss further why discriminability of oCI is more similar to acoustic discriminability at higher activity levels, >=200Hz, which contrasts sharply with the SoE analysis wherein oCI and acoustic are more similar at lower levels, <200 Hz. As mentioned briefly in the results, this could mean oCI channels remain meaningfully distinct (i.e., maybe perceptually...) at higher intensities, despite the broader spread of excitation, which is not the case for electric stimulation.

Methods:
1.) Postnatal intracochlear injection: what was the route of cochlear injection?
2.) Detection of multi-unit activity: I suggest re-wording this sentence as follows, "After applying a band-pass filter (0.6 - 6 kHz, 4th order Butterworth filter), the threshold for detection of firing activity was defined as four times the mean absolute deviation within a pre-stimulus baseline (estimated as mean/0.675 for data 100 to 2 ms before stimulation)."
3.) Quantification of activity level: is 100-µs biphasic pulses 100 µs per phase (200 µs total)? What was the rate of electrical pulses?
4.) Analysis of Spectral Peaks: is the "d'-array" the threshold contour (d'=1)?
5.) Spectral spread of excitation: The sentence "For each activity level..." was confusing. Consider simplifying: "For each modality, we first found the lowest stimulus intensity that elicited a given activity level (e.g., 200, 300, 400 Hz)."
6.) Some method sections point to relevant figures throughout the paper, and other do not. I would suggest always doing so where appropriate; e.g., Quantification of activity level = Fig 2A-C; Mapping of CF = Fig S1; particularly, Spectral spread of excitation = Fig 6.

Point-by-Point Responses to Reviewer Comments:

Reviewer comments are colored in blue with our responses shown in black. Minor changes beyond reviewer comments occurred due to the repeated review of the analysis and are highlighted in the manuscript text.

Referee: 1

Referee #1 (Comments on Novelty/Model System for Author):

the optogenetic actuator used is likely too slow to support hearing in a clinically meaningful way. still data provide and important step forward towards the development of a treatment that can achieve that.

We would like to thank the reviewer for the thorough review and appreciation of our manuscript. We agree that the temporal properties of ChReef are unlikely to be sufficiently fast for clinical translation. We have revised the relevant passages to clarify this limitation early on in the MS and to emphasize that the aim of the present study was to use ChReef as a model for investigating the spectral and intensity-coding performance achievable with an opsin that has the light sensitivity required for clinical translation.

> Introduction, 3rd paragraph: "Although the temporal properties of ChReef (τ_{off} ~30 ms at physiological temperature) challenge its use for clinical hearing restoration, the high light sensitivity it conveys to SGNs paved the way for preclinical characterization of the utility of multichannel oCI at a power budget more compatible with clinical translation.
>
> Limitations, 2nd paragraph. "…the kinetics of ChReef challenges its use for future clinical hearing restoration."

Referee #1 (Remarks for Author):

In their manuscript Albrecht et al. present their data on optogenetic tonotopic simulation of spiral ganglion neurons (SGN) using a Multi(up to ten)channel linear LED array implanted into the cochlea in the sense of a hearing-restoring cochlear implant (CI) using gerbils as a model system. The optogenetic actuator used is ChReef.
The authors clearly show that stimulus energies needed to achieve sufficient optical SGN stimulation are within the translationally feasible range.

Though slightly less definite (because only data from a previous study were used as reference) they also demonstrate that the optogenetic ChReef-CI outperforms electrical CIs particularly in terms of tonotopic discrimination accuracy.

We thank the reviewer for pointing out this limitation. In response to the comment, we clarified the comparability of the recent data to the previous dataset (eCI data from Dieter et al, 2019). Specifically, the present recordings were obtained using the same experimental setup, and both datasets were processed using the same analysis pipeline.

> Results, 3rd paragraph: "To maximize comparability, the present recordings were obtained using the same experimental setup, and both datasets were processed using the same analysis pipeline."

Nevertheless, because the electrical-stimulation data were acquired in a previous study, potential differences related to the individual experimenter cannot be excluded. We have now explicitly acknowledged this limitation in the manuscript.

> Limitations, final paragraph: "Finally, the electrical stimulation data were acquired in a previous study using the same experimental setup, protocol, and analysis pipeline. Yet, the comparison of these datasets remains a limitation, as subtle experimenter-related differences cannot be fully excluded."

They also state that tonotopic discrimination accuracy observed for the optogenetic CI was indistinguishable form native/acoustic stimulation. While this wording is technically correct, the indistinguishability is probably partially due to the limited resolution their outcome measure - a 32 channel linear electrode array recording from the inferior colliculus. Hence, the "indistinguishable form acoustic stimulation" statements should appear more in context.

We thank the reviewer for highlighting this limitation. In line with comments by reviewer 2, we have carefully adjusted the wording from "indistinguishable from acoustic stimulation" to emphasize improvements over state-of-the-art electrical stimulation to not overclaim equivalence to acoustic stimulation in the relevant passages. Additionally, we directly stated the neurometric threshold results within the framework of these limitations:

> Results, second to last paragraph: "Pure tones were sometimes discriminable at the smallest tested spacing (Table EV3), limiting the estimation of their true thresholds, and the 32-channel recording array may not fully capture the spatial representation of acoustic responses in the ICC. Therefore, the difference of e/oCI neurometric thresholds to acoustic thresholds might be underestimated."

(Minor: it would be interesting to understand if/how the authors think the long inactivation time constant of ChReef could support functional hearing.)

In the presence of relatively high spectral resolution, speech perception may require only limited temporal-envelope resolution. Xu et al. (2005, DOI: 10.1121/1.1886405) reported consonant recognition above 80% at a 16-Hz envelope cutoff in quiet for acoustic hearing, and Xu and Zheng (2007, DOI: 10.1121/1.428325) found performance plateaus at low-pass filter cutoffs between 8–16 Hz for consonants and approximately 4 Hz for vowels at both +6 and 0 dB Signal-to-Noise-

Ratio. Although modifying acoustic signals with an envelope cutoff is not directly equivalent to stimulation rate, Fu and Shannon (2000) also observed approximately 35% vowel and consonant recognition at 50 pulses/s/channel. These findings support the possibility of useful speech perception at lower stimulation rates, while faster opsins may remain advantageous for robust performance and auditory functions not captured by these phoneme tasks.

> Limitations, 2nd paragraph: “Although studies using low-pass-filtered acoustic or vocoded speech suggest that speech perception is possible with limited temporal-envelope information (Xu et al 2005, Xu and Zheng 2007), particularly at high spectral resolution, the kinetics of ChReef challenges its use for future clinical hearing restoration.”

Specific comments (in order of appearance in ms)
- Fig 1C: the symbols for LED and Laser stim are not intuitive here (only in this fig)

We thank the reviewer for this observation and have changed the symbols to text for clarity.

- The electronic CI reference data stem form a paper the authors published roughly 7 years ago. Could surgeon / or surgeon-training related-effects explain the poor performance of the electronic CIs?

We thank the reviewer for indicating this possibility. As described above, surgeon- or training-related effects cannot be entirely excluded and were included in the limitations. We consider them unlikely to account for the observed difference between electrical and optical stimulation. The eCI reference data were acquired using the same experimental setup and protocol as the present recordings. Moreover, the broader spread of excitation observed with eCI stimulation was also evident in a previous comparison with an optical dataset acquired by the same experimenter (Dieter et al., 2019; Keppeler et al., 2020). Importantly, the oCI used by Keppeler et al. (2020) differed from that used in the present study only in its emission spectrum, yet that study likewise found a significantly lower spread of excitation for oCI than for eCI stimulation. Together, these findings suggest that the difference observed here is unlikely to be primarily attributable to the individual experimenter or their level of training.

- Fig 2 D-F and related areas in the ms: the numbers above the error bars refer to "n"s, I understand.

We have included “n=” in the respective figure for clarity and revised the section in the figure legend:

> Fig. 2 Legend: “Because multiple frequencies or emitters were tested within individual animals, the number above each error bar indicates the number of recordings contributing to that data point.”

o Given that there were only 8,8, and 12 animals included in D-F, resp. I assume these "n" arise from repeated recordings?

We thank the reviewer for raising this point. Multiple recordings were obtained from each animal because different pure tone frequencies or emitters were tested and analyzed here. We have revised the figure legend to clarify this distinction.

> Fig. 2 Legend: "(**D-F**) Evoked firing rate quantified for each pure tone frequency or optical/electrical emitter using the maximally responding electrode at each stimulus intensity relative to threshold (mean ± SD) for acoustic (**D**), optical (**E**), and electrical (**F**) stimulation. Traces for individual recordings are plotted in the background. Because multiple frequencies or emitters were tested within individual animals, the number above each error bar indicates the number of recordings contributing to that data point."

Furthermore, to account for multiple recordings originating from the same animal, recordings are now averaged within animals in all other panels before plotting and statistical comparison. This increased the p-values and altered the absolute estimates, although most previously significant comparisons remained significant. Interestingly, the difference in equivalent SPL between optical and electrical stimulation, no longer reached statistical significance. This change does not affect the conclusions of the manuscript. Following animal-level averaging, the estimated equivalent SPL changed from 11 to 16 dB SPL above threshold for optical stimulation and from 8 to 15 dB SPL above threshold for electrical stimulation.

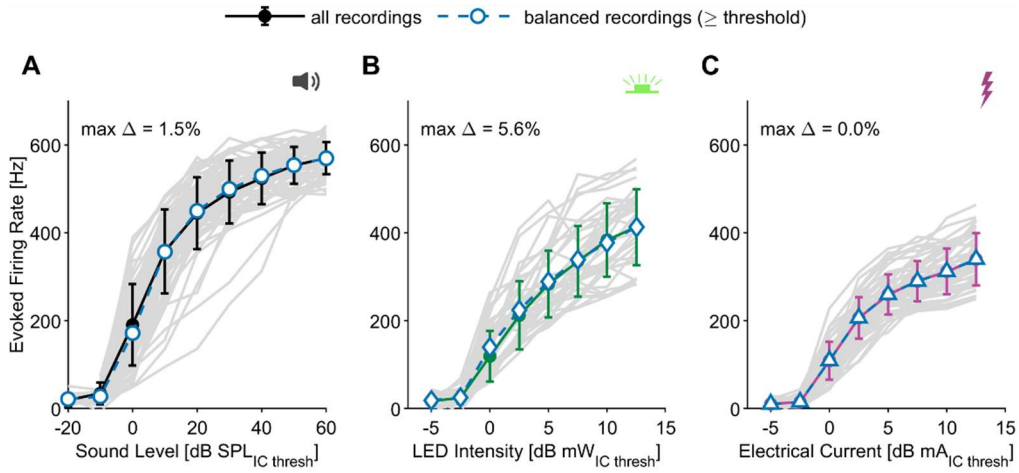
o "n" varies across SPLs tested. So the average dots shown refer to different populations at each SPL. I would argue that all average values should be calculated from the same reference population (i.e. only form the data where values exist for each SPL level)

We agree that the varying number of recordings across SPLs means that the descriptive averages are not based on the same set of recordings. However, restricting the analysis to recordings for which data were available at every SPL would preferentially retain those stimulus frequencies and emitters that showed the lowest threshold as intensity was quantified relative to threshold. Specifically, the maximum intensity was limited by the speaker or implant (i.e. 90 dB for acoustic, 3 mW for LED). Therefore, in case of higher thresholds stimuli did not reach as high intensities relative to threshold.

We also re-ran the analysis exactly as the reviewer suggested, restricting each average to a single reference population, namely the recordings for which a value was available at every intensity above threshold. The resulting rate-level functions were similar to those based on all available recordings, with a maximum deviation across supra-threshold intensities of 5.6% (optogenetic), 1.5% (acoustic) and 0% (electric; for the electric data every recording already covered all intensities, so the two analyses are identical). Restricting to this balanced population did not strongly alter the results but discarded a substantial fraction of the data and as noted above,

could bias the sample toward low-threshold stimuli. Therefore, we prefer to keep the full dataset in the main figure and provide the balanced comparison suggested by the reviewer here (**Response Fig. 1**) and in the Appendix (**Appendix Fig. S3**).

> Results, 5th paragraph: "Importantly, recalculating the rate-level functions using only recordings with data available at every plotted suprathreshold intensity yielded similar rate-level function shapes across in modalities (**Appendix Fig. S3**)."



Response Figure 1. Balanced vs. unbalanced averaging of the rate-level functions. Mean onset-evoked firing rate (± SD) vs. intensity relative to ICC threshold for (**A**) acoustic, (**B**) LED, and (**C**) electric (eCI) stimulation. "All recordings" (colored) uses every recording available in each bin, whereas "balanced recordings" (blue dashed) includes only recordings with data available in every plotted suprathreshold bin. Grey lines indicate individual recordings included in the balanced subset. Max Δ, maximum difference between the two averages across supra-threshold bins. Number of recordings: acoustic, total n = 236, balanced n = 73; LED, total n = 59, balanced n = 34; electrical, total n = 47, balanced n = 47.

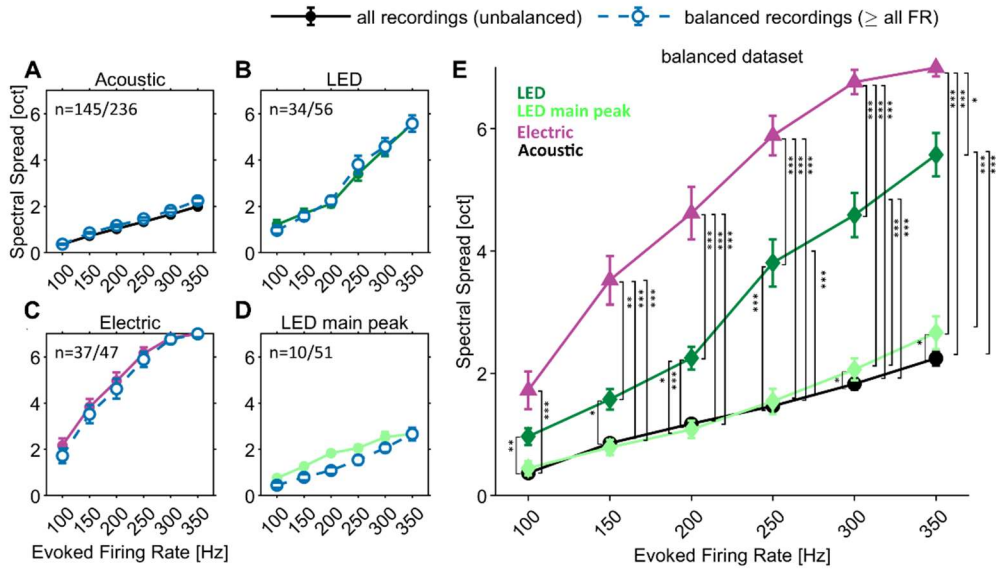
o Analog issue in data for figure 6

The reason for the unbalanced data is the same as in Fig. 2. Stimuli showed different thresholds in the ICC and therefore reached different firing rates in the limited range of intensities. We decided to not exclude stimuli that did not reach the highest firing rates to not selectively exclude stimuli with weaker or higher-threshold responses. Repeated measures within animals as well as the unbalanced design were accounted for by linear mixed effects models and were therefore not revised here.

We have revised the figure legend for clarity:

Fig. 6 legend: “Because multiple frequencies or emitters were tested within individual animals, the numbers above each panel indicate the number of contributing recordings.”

Again, we also tested the suggested approach by comparing the current analysis to a balanced dataset where only recordings were used that reached every tested activity level (referenced as balanced dataset):



Response Figure 2. Balanced vs. unbalanced analysis of Spread of Excitation.
(**A-D**) Mean spectral spread of Excitation (SoE, ± SEM, octaves) as a function of activity level (evoked firing rate [Hz]) for (**A**) acoustic, (**B**) LED, (**C**) electric and (**D**) the LED main spectral peak. "All recordings" (colored) averages every recording available at each firing rate (unbalanced); "balanced" (blue dashed) averages only the recordings that provide an SoE value at every firing rate (n = balanced/total per panel).
(**E**) Comparative overview of the balanced dataset for all modalities. Statistical comparisons were conducted using a linear mixed-effects model, with animal treated as a random factor and evoked firing rate and modality treated as categorical fixed effects, (*p<0.05, **p<0.01, **p<0.001). Evoked firing rate was modeled as a categorical factor to avoid assuming a linear relationship between firing rate and the outcome. Only significant effects are indicated. The number of recordings contributing to each activity level is shown in **A-D** and was obtained from n = 8/8/12 gerbils for acoustic/optical/electrical stimulation, respectively.

The main modality comparisons were qualitatively unchanged: the eCI-monopolar SoE remained significantly larger than the other modalities, and the LED SoE remained intermediate, exceeding the acoustic SoE. The only comparisons affected in a significant way were the acoustic vs. LED-main-peak difference (significant at all activity levels in the full dataset, but only at 300-350 Hz in the balanced set). This likely reflects selection of LED recordings with lower thresholds and clearly

separable main peaks across the full analyzed firing-rate range, thereby focusing the balanced subset to more spatially confined responses. In addition, balancing substantially reduced the sample size of the LED main-peak analysis, as only 10 recordings contributed data at every firing-rate level. The main conclusions of the MS are unchanged by this, but the balanced analysis may bias the sample toward low-threshold stimuli. Therefore, we prefer to keep the full dataset in the main figure and provide the balanced comparison here and in the Appendix:

> Results, second to last SoE paragraph: "Because fewer recordings reached the highest activity levels, we repeated the spectral-spread analysis using only recordings contributing across the full tested activity range (**Appendix Fig. S7**). The main conclusions were unchanged, except that the difference between acoustic spread and the LED main peak was no longer significant for activity levels below 300 Hz ($p > 0.05$, LME with Holm correction)."

Importantly, this comparison included the changes made to the linear mixed effects models in response to another comment (see below).

- Fig 3D: are the 4 emitters distributed across the entire cochlea?

We thank the reviewer for raising this question. The emitters in Figure 3D have an inter-emitter spacing of 300 µm and the oCIs were inserted from a base-turn cochleostomy. We estimate that they reach from the basal turn to the early midturn. We have updated the approximated emitter locations in the figure caption:

> Fig. 3 legend: "Columns in **B** show STCs for LED emitters positioned approximately 3600, 2400, 1200, and 0 µm from the round window (emitters 1, 4, 7, and 10, respectively). Columns in **C** show STCs for emitters positioned approximately 2700, 1800, 900, and 0 µm from the round window (emitters 1, 4, 7, and 10, respectively). Columns in **D** show STCs for emitters positioned approximately 1200, 900, 600, and 300 µm from the basal-turn cochleostomy (emitters 1–4, respectively)."

- Page 9 and Fig 4: I agree to the conclusion that there seems to be a spread of light to the opposite side of the cochlea - explaining the faint-ish second peak indicated by the black arrow but also seen in several other panels. Possibly because this reviewer is a retinologist, the logic for the sharp secondary peak (the authors call it broader spread of activation with central minimum) is not becoming clear to me. Potentially also other readers might have this probelm

We thank the reviewer for raising this point and have revised the logic in the corresponding section. To make the phenomenon easier to identify, we annotated the pronounced central activity minimum in the corresponding example (star in the fourth panel of Fig. 4D). We additionally tested whether this pattern could be explained by partial obstruction of the light path by the basilar membrane by segmenting the basilar membrane in the µCT reconstruction.

However, the reconstructed membrane crossed the implant near LED 7, which did not show such an activity minimum, arguing against this interpretation (Fig. EV3). We have therefore revised the text accordingly and now present the origin of the central minimum as uncertain, while noting that a local recording-related effect, such as increased electrode impedance, could provide an alternative explanation.

> Results, 9th paragraph: "Moreover, the lateral position and more tangential orientation of the LEDs relative to Rosenthal's canal within the wider scala tympani at the cochlear base may account for the broader activation observed in several recordings (e.g. 3rd to 5th panels of Fig. 4C, D; Fig. 3C, emitter 10). We initially hypothesized that the pronounced central activity minimum observed in some responses (e.g., star in the 4th panel of Fig. 4D) resulted from partial obstruction of the light path by the basilar membrane. However, µCT reconstruction localized penetration of the basilar membrane near LED 7, which did not show such an activity minimum (Fig. EV3), arguing against this explanation. Thus, the origin of the minimum remains uncertain with potential reasons including a locally increased impedance of the ICC recording electrode."

- Abbreviation RC is not introduced

RC is now introduced as Rosenthal's canal in the legend for Fig. 4A.

- Fig 5 C has n unusual aspect ratio that might faint variability actually present in the data

We thank the reviewer for this comment. Because Fig. 5G presents supplementary results, we moved it to the Appendix as Fig. S6. This allowed us to adjust the layout of the remaining panels and present Fig. 5C with a more balanced aspect ratio, thereby providing a clearer visualization of the variability in the data.

- The logic for the "LED main peak" analysis is not becoming fully clear. This would be to simulate how optogenetic hearing could perform if higher density LED arrays were used?

We thank the reviewer for this question. The "LED main peak" analysis was not primarily intended to simulate the performance of higher-density LED arrays. The occurrence of secondary spectral peaks, likely caused by cross-turn illumination of SGNs, has not been frequently observed in previous studies of cochlear optogenetics, likely because of lower light sensitivity conveyed by ChReef, the ChR used there. In the present MS, we reason that the secondary peaks contribute to the increased spectral spread observed for LED stimulation at higher activity levels. Therefore, we assessed their contribution by quantifying the spectral spread after excluding the secondary peaks from the analysis, i.e., by measuring the spread of the remaining main peak. The difference between the "LED" and "LED main peak" data can thus be interpreted as the additional spectral

spread attributable to secondary peaks. We have revised the relevant section of the manuscript to clarify this rationale.

> Results, 3rd to last paragraph on spectral spread: "To determine whether secondary, off-target spectral peaks contributed to the increased spectral spread at higher activity levels, we repeated the analysis after excluding these peaks and quantifying the spectral spread of the main spectral peak of activation."

- Figure 5G Cf should be subscript f

Updated.

- Fig 6 H and 8 J: A LMM was used to probe for statistically significant differences between cohort at the individual ERs
o 1) data are skewed. LMM can handle that but needs to be checked that residuals are not skewed

We thank the reviewer for this important comment. We found that the residuals of the model used for Fig. 6H showed substantial deviations from normality. We therefore revised the analysis by applying a cube-root transformation to the dependent variable, the SoE. In addition, we expanded the random-effects structure to include random slopes and intercepts, rather than random intercepts alone, because a likelihood-ratio test indicated a significantly improved model fit. The residuals of the revised model no longer showed pronounced deviations from normality in the Q–Q plot (**Response Fig. 3A**). Therefore, we included this revised model as the primary analysis in the MS.

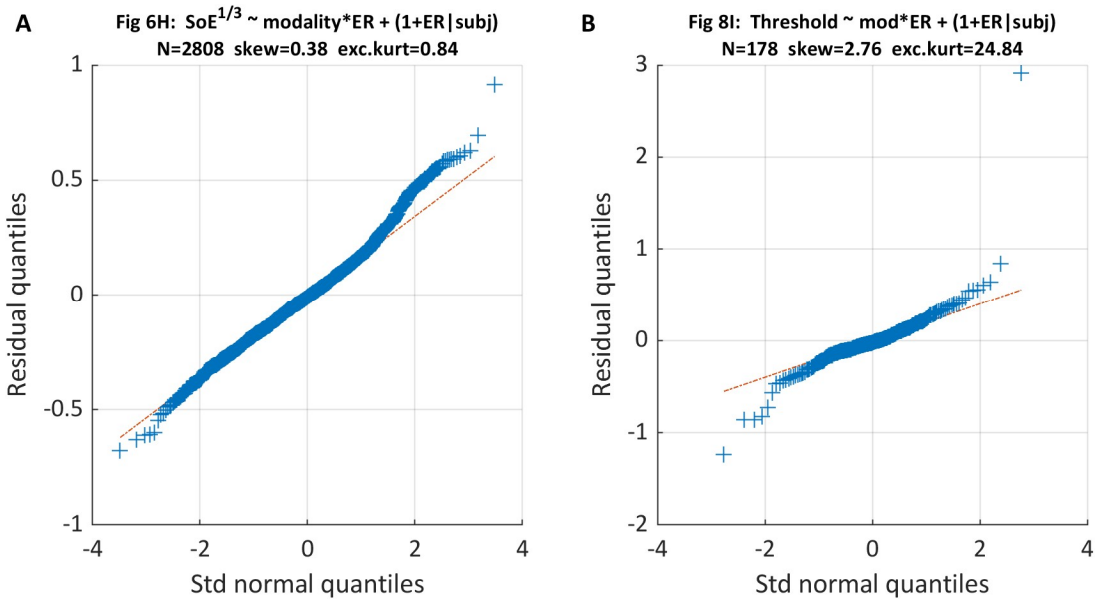
The significant findings from the previous analysis were retained. In addition, the revised analysis identified a significant difference between LED and acoustic stimulation at 100 Hz and between the LED main-peak and acoustic conditions at 100 and 150 Hz. We have revised the corresponding sections of the manuscript accordingly. In response to this comment and a related comment from Reviewer 2, we also removed statements implying that the spectral spread or tonotopic discriminability of LED and acoustic stimulation was indistinguishable.

For Fig. 8I, the residual deviations from normality were driven primarily by a small number of extreme observations at an activity level of 100 Hz rather than by pronounced overall skewness (**Response Fig. 3B**). As for Fig. 6H, we expanded the random-effects structure to include random slopes and intercepts, because a likelihood-ratio test indicated improved model fit ($p = 0.008$). We did not apply a transformation to the dependent variable.

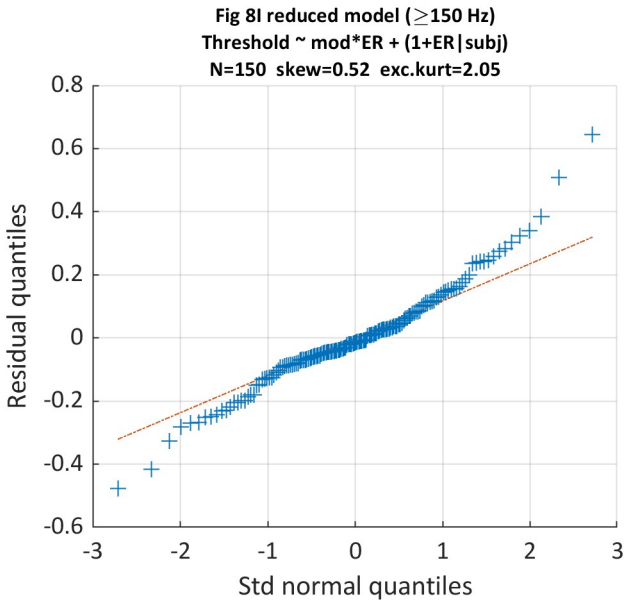
As a sensitivity analysis, we refitted the model after excluding the 100-Hz activity level. This reduced residual skewness from 2.76 to 0.52 and excess kurtosis from 24.84 to 2.05 (**Response Fig. 4**). The reduced model retained all statistically significant differences identified by the full model. In addition, it identified significant differences between acoustic and LED stimulation at

200 and 250 Hz and between LED and electrical stimulation at 250 and 300 Hz (**Response Table 1**). These additional findings may reflect changes in the estimated variance structure after removal of the highly variable 100-Hz observations.

For the primary analysis in the manuscript, however, we retained the full model to avoid excluding the low-activity data. Consistent with the comments by Reviewer 2, we do not interpret nonsignificant differences in this model between LED and acoustic stimulation as evidence of indistinguishability. This is particularly important because the corresponding contrast became significant in the reduced sensitivity model, further illustrating that non-significance in the full model should not be interpreted as evidence of equivalence. Additionally, we also tested the robustness of the conclusions for the 100 Hz condition using Kruskal-Wallis test with Holm's post-hoc correction. This did not reproduce the statistical difference between LED and electric to acoustic stimulation. The difference should therefore be interpreted with caution which we included in the manuscript. The significance stars in Fig. 8I at 100 Hz are now highlighted in gray to indicate this limitation.



Response Figure 3. Residual normality of the revised mixed-effects models for Fig. 6H and Fig. 8I. Quantile–quantile plots of the model residuals against normal quantiles for Fig 6H (A) and Fig. 8I (B); the line marks the expectation for normally distributed residuals. Abbreviations: SoE, Spread of Excitation; ER, evoked firing rate; subj, subject (animal); exc. kurt, excess kurtosis (kurtosis-3).



Response Figure 4. Residual normality of the reduced mixed-effects models for Fig. 8I. Quantile–quantile plots of the model residuals against normal quantiles for Fig. 8I excluding an activity level of 100 Hz where outliers might violate assumptions of normality; the line marks the expectation for normally distributed residuals. Abbreviations: ER, evoked firing rate; subj, subject (animal); exc. kurt, excess kurtosis (kurtosis-3).

Fig. 8I — Neurometric threshold

Model: threshold ~ modality × activity level + (1 + activity level | animal).
Full: all levels (primary); ≥150 : 100 Hz condition excluded.

Activity Level (Hz)	Ac vs LED		Ac vs eCI		LED vs eCI	
	Full	≥150	Full	≥150	Full	≥150
100 Hz	**0.080**	—	**0.011**	—	0.491	—
150 Hz	0.313	0.077	0.275	0.038	0.927	0.876
200 Hz	0.121	**0.017**	**0.006**	**<.001**	0.333	0.165
250 Hz	0.101	**0.021**	**<.001**	**<.001**	0.101	**0.021**
300 Hz	0.097	0.044	**<.001**	**<.001**	0.065	**0.011**
350 Hz	0.106	0.067	**<.001**	**<.001**	**0.005**	**<.001**
400 Hz	0.149	0.112	**<.001**	**<.001**	**0.007**	**<.001**

Response Table 1. p-Values for Modality comparison in Fig. 8I using a full and a reduced LMM excluding an activity level of 100 Hz.
p-Values from the linear mixed model, Holm-corrected within each activity level. 'Full' = primary model (all activity levels). '≥150' = the same model refit with the 100 Hz condition excluded.

Shading marks significance: light = p<.05, medium = p<.01, dark = p<.001; unshaded (grey) = not significant. eCI = electrical (monopolar) stimulation.

o In fig 8J: effects are clearly not linear in native/acoustic hearing, so an LMM wouldn't be able to describe the data appropriately.

We thank the reviewer for raising this point. We agree that the relationship between evoked firing rate and the discrimination spacing is not linear. However, evoked firing rate was included in the linear mixed-effects model as a categorical factor, with 100 Hz as the reference level, rather than as a continuous numerical predictor. The model therefore did not assume a linear relationship across firing-rate levels. Instead, it estimated separate effects for each firing-rate level and separate firing-rate-by-modality interactions, allowing the non-monotonic response observed for acoustic stimulation to be represented. We have specifically added this point in the figure legends for Fig. 8J and Fig. 6H. We have also revised the supplementary tables for the LME fits to include the type of variable for the fixed effects and referenced this in the methods.

> Fig. 6 and Fig. 8 legend: "Evoked firing rate was modeled as a categorical factor to avoid assuming a linear relationship between firing rate and the outcome."

- The discrimination spacing at the neurometric threshold floors at 0 mm (i.e. same electrode on the recording array) for the acoustic stimulation. It is logically not truly 0, just the resolution of the measurement method is insufficient - that fair and likely unavoidable, but should be accounted for in the interpretation.
Overall nice work. Thanks.

We would like to thank the reviewer for the comment. The discrimination spacing at the neurometric threshold for acoustic stimulation floors close to 0 which corresponds to smallest stimulus spacing tested (0.03 mm = 0.042 octaves). This is what is documented as downward censored values in **Table EV3**. Of course, this must be accounted for in the interpretation, and we have framed the results considering this limitation in the main text (see also above):

> Results, second to last paragraph: "Pure tones were sometimes discriminable at the smallest tested spacing (Table EV3), limiting the estimation of their true thresholds, and the 32-channel recording array may not fully capture the spatial representation of acoustic responses in the ICC. Therefore, the difference of e/oCI neurometric thresholds to acoustic thresholds might be underestimated."

Referee: 2

Referee #2 (Remarks for Author):

This study by Albrecht et al. tests the efficacy of a new channelrhodopsin ChReef in the context of optical cochlear stimulation with a readout of multi-unit activity in the inferior colliculus. This is part of an ongoing program of work by this group to establish the clinical feasibility of optical cochlear implants. This is very important and exciting work and the present study is a worthy addition. The work is admirably thorough, the key conclusions are well supported, and the writing is clear. I have only minor concerns around language and emphasis that should be addressed before publication.

We would like to thank the reviewer for the thorough review and appreciation of our manuscript. We have carefully revised the manuscript in response to the comments.

In general, the language used in describing the results seems overly favorable toward optical stimulation (i.e., it overstates the degree to which the current results support the similarity of optical and acoustic stimulation). This is in part due to an overemphasis of statistical significance rather than effect size. If two distributions have well separated centroids but fail to pass a statistical significance test, that should not be taken as evidence that the centroids are comparable. Think about the confidence intervals on the difference of the centroids: example A is 50% (-1%,101%) and example B is 0% (-1%, 1%). Both of these would fail the significance test. Example B is strong evidence that the centroids are comparable. Example A is not. The interpretation of Example A is that, given the data, the best guess of the centroid difference is 50% but in fact it might well be 0% or 100%.

We thank the reviewer for raising this central point. We have carefully adjusted the phrasing of the results especially in terms of the remaining differences in spectral spread and channel discriminability between optical and acoustic stimulation. We stress that while the difference in spectral spread and neurometric thresholds between acoustic and optical stimulation failed to pass the statistical significance test at certain activity levels, this should not be taken to reflect equivalence. Moreover, the difference in spectral spread between acoustic and optical stimulation is statistically significant at every activity level after implementing the changes in response to comments by Reviewer 1. We have addressed this in two ways: firstly, the sections describing these results now include the magnitude of the observed difference alongside the statistical significance indicating the certainty of this difference. Secondly, we conclude an advantage of optical over electrical stimulation in terms of the spectral properties while a gap between acoustic and optical stimulation remains. This is now reflected in all relevant paragraphs during the MS:

> Introduction, last paragraph: “Consistent with previous studies, we demonstrate that optogenetic stimulation substantially improved spectral selectivity compared with electrical stimulation but did not achieve that of acoustic stimulation.”
>
> Result, 3rd paragraph on spectral spread: ”Across the tested activity levels, the estimated spectral spread was largest for electrical stimulation, smallest for acoustic stimulation, and intermediate for optical stimulation.”

This consideration contributed to an additional paragraph in the discussion elaborating on closing the remaining gap between optical and acoustic stimulation.

> Discussion, last paragraph before limitations: “Despite this improvement in channel discriminability, optical stimulation still produced a larger spread of excitation […].”

The discussion of spectral spread on page 13 is a good example. There is much talk of statistical significance, and of the fact that optical stimulation is closer to acoustic than to electric, but at no point is the magnitude of the difference in spread between acoustic and optical for a given evoked rate ever stated. If one looks at figure 6H, the difference between the lines for acoustic and LED Main Peak is sizable. If one chooses a particular evoked rate, say 200 Hz, and then looks at the y axis values for each of those lines, the difference is probably on the order of 30 or 40%. The same can be said of Figure 8I. Choose a particular evoked firing rate, again say 200 Hz, and look at the y axis values. They are rather different. It may be that they are not statistically significantly different, but if you had to guess whether the difference was 0 or a factor of 2, the evidence would support a guess of a factor of 2.

We would like to thank the reviewer for the comment. We now explicitly state the difference in spectral spread between optical and acoustical and optical and electrical stimulation as well as their development with increasing activity level. Importantly, as stated above, after revision in response to the comments by reviewer 1, the difference between optical and acoustic stimulation became significant for all activity levels. For the convenience of the reviewer, we paste further modifications in response to the comment below:

> Results, 3rd paragraph on spectral spread: “At low-to-moderate activity levels of up to 200 Hz, the average difference between optical and electrical stimulation increased, reaching a maximum of 2.85 octaves, or 57% relative to electrical stimulation, at 200 Hz.”

The same point also applies to the description of the results derived from the LDA and the difference in neurometric threshold across intensities. We have applied the same rationale there:

> Results, 2nd to last paragraph: “Averaged across this activity range, optical thresholds were 0.44 mm lower than electrical thresholds (range 0.03–0.75 mm across activity levels) and 0.47 mm higher than acoustic thresholds (range 0.4–0.51 mm across activity levels). In relative terms, the electrical threshold was on average about 1.7-fold the optical threshold, increasing from near-parity at 150 Hz to roughly 1.9-fold at 400 Hz, whereas

the optical threshold was on average about 3.4-fold the acoustic threshold for higher activity levels (≥ 150 Hz), reaching approximately four-fold at 200 Hz (0.66 vs 0.16 mm)."

As stated above, the conclusions derived from these revisions were also included in the discussion leading to a new paragraph exploring potential solutions for the remaining gap between optical and acoustic stimulation.

Another example of overly favorable language is that around the reduction of SGN density due to injection. It is described as mild in the main text. But looking at figure EV2C, the average results in the right panel appear to have an effect size on the order of 33%, meaning the injection results in a 33% loss of SGNs relative to control. I wouldn't describe that as mild. And if we look at the left panel for the apex, the difference looks rather extreme.

We would like to thank the reviewer for raising this point. The reduction of SGN density averaged across turns amounted to approximately 30 %. We have now specifically included this result in the main text and addressed it in the discussion.

Discussion, 4th paragraph: "The reduced SGN density observed after transduction may also have contributed to the limited ICC activity by reducing auditory nerve input. As previous studies suggest that this loss may be related to pressure during cochlear injection (Bali et al., 2022), alternative approaches for viral delivery may help preserve SGN density and further increase the achievable neural response range (Thirumalai et al., 2025)."

Let me be clear: I don't see any major problems with the actual data or analyses. It is simply the language used to describe the results and choice of which aspects of the statistics to focus on. These problems are easily fixed.

Regarding the comparison of light pulses to sustained tones: I think the approach used of limiting the analysis to the onset time window is perfectly fine, but going forward it might be good to also test tone bursts of, say, 10 ms with the five ms onset and offset ramps that are currently used and/or to try to elicit sustained activity with the light stimulus. The latter is presumably what is being worked toward in order to elicit the most physiologically realistic patterns of activity in response to complex sounds.

We thank the reviewer for this comment. For clinical translation, it will likely be favorable for hearing restoration to elicit response patterns that are as similar as possible to responses to real world acoustic signals. This is one of the reasons we include acoustic stimulation as a control group for optical stimulation. By restricting our analysis to the onset response window, we isolate the spectral properties of the response, while its temporal properties are addressed in the limitations section and in the companion paper. As suggested, a second option would be to match

the modalities at the stimulus level by using brief tone bursts. Because opsins with faster kinetics will be required to reproduce patterns closer to the sustained response evoked by pure tones, we have now added both points to the limitations of the manuscript where temporal properties are discussed.

> Limitations, 1st paragraph: "This matches the different modalities at the level of the analysis window. Future studies could instead match them at the level of the stimulus, for example by comparing single light pulses with brief tone bursts (e.g., 10 ms duration with 5-ms onset and offset ramps). As it will likely be advantageous for clinical hearing restoration, future studies should also focus on faithfully reproducing the sustained temporal pattern of acoustic stimulation, which will require opsins with faster kinetics (Koert et al., 2026). "

The analysis that results in the establishment of the maximum light evoked activity being equivalent to the activity evoked by a 46 dB SPL pure tone is excellent. This type of benchmarking makes the comparison very clear. That said, this is also reported to be approximately 10 dB above the acoustic threshold, which is quite limited. For the purposes of this study, it's fine because the focus is on sharp spectral tuning, which is most evident for stimuli close to threshold. But going forward, it's going to be important for optical stimulation to be able to illicit activity that is comparable to that elicited by real-world sounds, which are of course typically at much higher levels. I think this should be included in the limitations section of the Discussion.

We would like to thank the reviewer for raising this point that was previously not adequately included in the discussion.

In response to Reviewer 1, we revised the equivalent-SPL analysis by averaging recordings obtained from the same animal before estimating the group-level value. This changed the estimated equivalent SPL of maximal optical stimulation from 46 to 51 dB SPL. Although this value is higher than in the original analysis, it remains below the sound levels typically encountered during normal conversation (~65 dB SPL). We have therefore added this limitation after the discussion of the equivalent SPL results. Future studies should focus on approaching the activity levels observed in real world sounds and speech by further increasing opsin sensitivity, optical emitter output, and multi-channels stimulation.

> Discussion, 4th paragraph: "The equivalent sound pressure level approximated for the maximal firing rate achieved by a single LED was 51 dB (SPL). Although this represents substantial neural recruitment by a single optical channel, it remains below sound levels typically encountered during conversational speech (~65 dB SPL). Importantly, such a comparison is necessarily approximate because the energy of natural sounds is distributed across frequency bands, whereas the present estimate reflects stimulation by

a single optical channel. Future studies should address this by multichannel stimulation, as well as by extending the range of neural activity for single oCI channels."

Related to this, I don't understand the statement in the Discussion about the response range up to 200 hertz being most relevant for optical coding strategies. Is the hope that central plasticity following implantation would adopt the dynamic range downstream to make use of the limited dynamic range peripherally?

We thank the reviewer for raising this point. We do not intend to imply that central plasticity would compensate for a limited peripheral dynamic range after implantation. Rather, in the present model, activity levels up to approximately 200 Hz of MU firing appear most relevant for exploring optical coding strategies because spectral spread increases more strongly above this range, in part due to the emergence of secondary spectral peaks. We therefore consider this range to represent the most favorable operating regime for the single channels of the current oCI implementation in the small gerbil cochlea. As discussed above, however, higher activity levels and a broader effective dynamic range will be advantageous for clinical use and will require further improvements in opsin sensitivity, optical output, and multi-channel coding strategies. Nevertheless, the activity range available in the current model provides a useful basis for developing and evaluating multichannel optical coding strategies. We have clarified this point in the corresponding section of the discussion.

Discussion, 8th paragraph: "In the current oCI implementation in the gerbil cochlea, spectral selectivity was best preserved up to approximately 200 Hz, above which off-target tonotopic regions were recruited (secondary spectral peaks). This range therefore appears most relevant for developing coding strategies with the presently available stimulation levels, although higher activity levels will likely be advantageous for clinical use of optogenetic SGN stimulation."

Referee: 3

Referee #3 (Remarks for Author):

This manuscript characterizes intensity and spectral encoding by a multichannel optical cochlear implant (CI), which employed green-LED stimulation of spiral ganglion cells expressing the low-energy-threshold channelrhodopsin, ChReef. Optical stimulation was compared to acoustic and electrical cochlear stimulation using extracellular recordings along the tonotopic axis of the gerbil inferior colliculus (IC). Optical stimulation using ChReef replicated sub-µJ optical thresholds and showed intensity coding at least as high as that by 46-dB SPL pure tones. It provided spectral specificity comparable to acoustic stimulation for low-to-moderate IC activation levels but also revealed activation of off-target tonotopic regions. Discriminability between optical channels was more similar to that by acoustic than by electric stimulation, even at relatively high levels of IC activation.

Overall, the study was thorough and technically well executed. The demonstration of an efficient multichannel oCI is an important advancement for the clinical translation of optogenetic stimulation. A nice strength of the work was its mapping of frequency into cochlear place/distance, and vice versa, for more direct comparisons between stimulation modalities. My concerns were mostly around issues of clarity and presentation in the manuscript. Please see below some general comments as well as several specific points regarding the text and figures.

We would like to thank the reviewer for the thorough review and for the positive assessment of our manuscript. We carefully considered all points for strengthening and clarifying the conclusions of the study and have addressed all comments in the revised manuscript.

General:

1.) For readers who are not fully up to-date on optogenetic literature, the abstract and introduction could do a better job of highlighting what is novel and significant about the study. Higher spectral selectivity with oCIs (relative to eCIs) has been shown, of course. My sense, however, is that this is one of the first demonstrations of a clinically feasible multichannel oCI (via the efficiency of LED-ChReef pairing). Is that the main takeaway? I see that the introduction says: "This challenged the utility of implantable microscale emitters for preclinical characterization of multichannel oCI (Dieter et al., 2020; Keppeler et al., 2020)". But a little more context would give better emphasis on the new contribution; for instance, example(s) of those past challenges with multichannel oCIs and what has been shown with ReefCh oCIs so far, and what hasn't (the gaps to filled).

The key advance is indeed the systematic characterization of spectral and intensity coding by multichannel oCIs at energy levels more compatible with clinical use. Previous studies demonstrated spectrally specific activation of the auditory system with LED-based optical

stimulation, but stimulation thresholds were much higher than those reported for clinical electrical CIs. In addition, preclinical characterization of multichannel oCIs was limited because individual emitters did not reliably activate the auditory pathway. More recent work with ChReef demonstrated substantially improved light sensitivity of SGNs and a better match to the output of LED-based oCIs, but systematic characterization of spectral and intensity coding with multichannel LED-based oCIs had yet to be performed. We have now made these previous limitations and the resulting knowledge gap more explicit in the revised introduction and abstract.

> Abstract: "This combination enabled systematic comparison of encoding intensity and spectral information by individual oCI channels to acoustic and electrical stimulation within a translationally feasible energy range."
>
> Introduction, 2nd paragraph: "This challenged the utility of implantable microscale emitters for preclinical characterization of multichannel oCI (Dieter et al., 2020; Keppeler et al., 2020). With previous µLED-based approaches, the light sensitivity of CatCh was frequently insufficient for activation by individual LEDs, requiring simultaneous operation of multiple LEDs (Dieter et al., 2020). Moreover, reliable responses from all channels implemented on multichannel oCI prototypes were not consistently observed within the same animal (Keppeler et al., 2020). Thus, achieving energy-efficient and tonotopic activation of the spiral ganglion by individual oCI channels remained an important objective for characterizing optogenetic stimulation by multichannel oCIs."
>
> Introduction, last paragraph: "Going beyond previous work, reliable activation by individual emitters also enabled us to assess channel discriminability at the level of ICC activity. ICC responses elicited by different oCI channels were more discriminable than those elicited by eCI channels, particularly for stimulus intensities eliciting high activity levels."

2.) Based on the title "multichannel optical cochlear implants..." and terms like channel discriminability, in the abstract and introduction, I had an initial impression that multiple oCI channels would be stimulated together (concurrently) and the resulting ICC patterns analyzed. That was likely my own bias going into it. Nevertheless, I think this is worth considering this in the discussion. Whereas electric CIs necessitate interleaved stimulation to avoid charge interaction, optical CI presumably will not. Is that correct? Therefore, future assessments of simultaneous channel stimulation might be more clinically relevant (e.g., Middlebrooks & Synder, 2007; 10.1007/s10162-007-0070-2). The interactions of those channels are potentially more important than discrimination between different channels alone. This also seems relevant considering the secondary ICC activation peaks, shown in this study, which might interact (i.e., get enhanced).

We would like to thank the reviewer for raising this important point. In the present study, we characterized individually addressed LEDs, which allowed us to assess intensity and spectral encoding by individual oCI channels. In future clinical use, however, sound-coding strategies will activate multiple channels for generating optical stimulation patterns that may include simultaneous or interleaved stimulation. Future characterization of simultaneous multichannel stimulation will therefore provide additional information relevant to clinical coding strategies.

Regarding the need for interleaved optical stimulation, channel interactions are expected to be lower for optical than for monopolar electrical. Interleaved stimulation may therefore be less necessary for oCIs than for conventional eCIs. However, this cannot be concluded from the present single-channel experiments. Studies directly investigating simultaneous optical stimulation at energy requirements relevant for clinical translation will be needed to determine the extent of channel interactions and whether interleaved stimulation provides an advantage. As pointed out by the reviewer, this will also be important given the secondary activation peaks observed at higher stimulation levels in the present study of the small rodent cochlea, which could aggravate channel interaction in oCI if off-target stimulation also occurred in the larger human cochlea.

We have therefore addressed this comment in two ways. First, we now clarify early in the abstract and introduction that we characterized responses to individually stimulated LEDs of a multichannel oCI. Second, we added a paragraph to the limitation section of the discussion that states the need for future oCI studies to address simultaneous multichannel stimulation and channel interactions.

> Abstract: “This combination enabled systematic comparison of encoding intensity and spectral information by **individual** oCI channels to acoustic and electrical stimulation within a translationally feasible energy range.”
>
> Introduction, last paragraph: “Here, we used AAV-based expression of ChReef in SGNs and stimulation by **individual** green LEDs to characterize preclinical multichannel oCI with recordings from the central nucleus of the inferior colliculus (ICC) in Mongolian gerbils.”
>
> Discussion, Limitations, 3rd paragraph: “Another limitation is that we characterized stimulation by individual oCI channels, whereas clinical sound coding relies on multichannel stimulation. Although interactions between simultaneously activated channels are expected to be lower for optical than for monopolar electrical stimulation (Azees et al., 2023), this cannot be inferred from the present experiments. Addressing channel interaction is relevant given the secondary activation peaks observed at higher stimulation levels in the present study of the small rodent cochlea, which could enhance channel interaction in oCI if off-target stimulation also occurred in the larger human cochlea. Future studies should therefore directly investigate simultaneous multichannel stimulation (Middlebrooks and Snyder, 2007).”

3.) Supplemental Figures S3 and S7 don't seem to be referenced in the text. These should be described or excluded.

We thank the reviewer for pointing this out. Both figures are now referenced in the text.

The supplementary figure previously numbered S3 shows the spontaneous firing rate for the three stimulation modalities. Spontaneous rates differed between modalities, being highest for acoustic, intermediate for optical and lowest for electrical stimulation. This is one of the reasons we used evoked firing rates rather than absolute firing rates as the readout of activity level throughout the manuscript, and the figure is now referenced where that choice is described:

> Results, 4th paragraph: "Spontaneous firing rates measured before stimulation were subtracted from the firing rate within a 13-ms window following response onset to obtain evoked firing rates. This baseline correction was particularly important because spontaneous rates differed between modalities (Appendix Fig. S2)."

Fig. S8 (previously S7) shows the quantification of spectral spread based on the cumulative d' analysis. This was included to facilitate comparison with previous publications (Dieter et al., 2019; Middlebrooks and Snyder, 2007), which expressed spectral spread as a function of d-Prime-Level rather than of evoked firing rate. It is now referenced in the results section:

> Results, second to last paragraph on spectral spread: "To facilitate comparison with previous publications (Dieter et al., 2019; Middlebrooks and Snyder, 2007), we additionally quantified spectral spread as a function of d', a signal detection theory index to compare neural activity levels."

For the reason of comparability, we have revised the figure to better match the analysis used in Dieter et al, 2019. It now compares spectral spread at the same d'-levels used in the previous publication and uses repeated measures ANOVA for statistical comparisons.

Specific Points:

We would like to thank the reviewer for the specific points and the suggestions for possible solutions. We have implemented most of the points as suggested, for details see below.

Abstract:

1.) Following the sentence, "We characterized the oCI encoding...", you might add that: oCI encoding was compared to that achieved by acoustic and electrical CI stimulation.

This point was embedded as suggested to clarify the scope of the study in the abstract: "systematic comparison of encoding intensity and spectral information by individual oCI channels to acoustic and electrical stimulation within a translationally feasible energy range."

2.) 46 dB tones, should be 46 dB SPL pure tones.

Agreed and revised as suggested. Note, that following the revisions in response to reviewer 1 the estimate changed from 46 to 51 dB SPL: “51 dB SPL pure tones”.

3.) "moderate activity levels" wasn't clear without reading the main text; slightly more meaningful might be "moderate levels of ICC activation"

As part of the revisions in response to reviewer 2 this point was removed from the abstract. It was replaced by a stronger focus on the benefit compared to electrical stimulation instead of similarity to acoustic stimulation.

The paper Explained:

1.) "high levels of midbrain activity" - again, I wasn't clear on this until reading the text. Maybe qualify in some way, e.g., "...and, compared with acoustic stimulation, relatively high levels of midbrain activation with only low micro-joule energy". Using the full spelling of micro- to mirror that of nano- in the same sentence.

We thank the reviewer for raising this point. We have adjusted the paragraph to include the specific estimate of LED-evoked midbrain activity being equivalent to that evoked by 51 dB SPL pure tones. Note, that following the revisions in response to reviewer 1 the estimate changed from 46 to 51 dB SPL: “51 dB SPL pure tones”. For consistency within the sentence but also the rest of the paper, we have changed nano-joule to nJ.

> “ChReef enabled activation of the auditory pathway with nJ thresholds and midbrain activity equivalent to that achieved with 51 dB SPL pure tones with low µJ radiant energy.”

Introduction:

1.) "noisy backgrounds" - plural. You might add "or competing sound sources".

We agree and have adjusted the wording accordingly to “noisy backgrounds or competing sound sources.”

2.) Consider: "sound-frequency spatial (tonotopic) organization"

We agree that this helps understand the concept for readers and have adjusted the wording accordingly.

3.) "multipolar stimulation" - since this is not contrasted to monopolar, consider more general terms: e.g., "current focusing techniques"

We agree that “current focusing techniques” is more intuitive for readers than “multipolar” and have adjusted the wording accordingly.

4.) 2nd paragraph: For first sentence, I suggest changing to: "However, the power requirement

of optogenetic SGN stimulation has remained a challenge for oCIs to become...". Using the present perfect tense throughout this section (e.g., has shown) might be preferrable to the simple past (e.g., showed) to indicate the work is still ongoing.

We agree and have adjusted the wording to “has remained a challenge”.

5.) Last paragraph: consider changing "for up to moderate evoked ICC activities" to "for stimulus intensities eliciting low-to-moderate levels of ICC activities." This clarifies the relation between stimulus intensity and ICC activation.

As indicated above, we have revised the conclusions to focus on optical stimulation narrowing the gap between electrical and acoustic stimulation rather than implying equivalence between optical and acoustic stimulation. Therefore, the specific wording highlighted by the reviewer is no longer present in this paragraph. We did, however, adopt the reviewer’s suggested formulation in another sentence, which now reads: “particularly for stimulus intensities eliciting high activity levels.”

6.) Last paragraph: the sentence beginning: "The tonotopic organization..." was a bit unclear. More closely than electric stimulation? Does it mean that considering the previous sentences, tonotopic organization was overall more like acoustic than eCI responses? Or that tonotopic organization for oCI activity was appropriate to the intracochlear placement of the emitters? If the later, consider putting this sentence after the sentence about micro-CT.

We thank the reviewer or the comment. “Tonotopic organization” indeed refers to the relationship between intracochlear emitter position and response location in the ICC. We have therefore added a sentence to clarify this definition and placed it right after the sentence about the µCT analysis:

“The tonotopic locations of ICC responses closely matched those expected from the intracochlear emitter positions based on the Greenwood function.”

7. Last paragraph: I suggest: "...single emitters reached that elicited by 46 dB SPL pure tones"

We agree and changed the wording accordingly to “elicited by 51 dB SPL pure tones”.

Results

1.) The first paragraph could better delineate the use of laser stimulation for the preliminary assessment, which is followed by LEDs for the main experiment. Something like: "After a minimum of 8 weeks, we first assessed auditory activation using a single intracochlear optical fiber and recordings of the optically evoked auditory brainstem response (oABR). Of the 15 gerbils, 13 showed oABR responses to sub-uJ light pulses."

We thank the reviewer for the helpful comment and have adjusted the wording close to the reviewer’s suggestion:

“After a minimum of 8 weeks, we first assessed successful transduction using optically evoked auditory brainstem responses (oABRs) elicited by a single optical fiber placed at the round window. oABRs to sub-µJ light pulses were observed in 13 out of 15 gerbils.”

2.) I was a bit confused by the animal numbers. 15 received AAV injections, and the first paragraph says 13 showed oABR sub-uJ thresholds (by extension <10 uJ). The results then jump to "out of eight animals..." without explaining the reduction. The methods somewhat clarify that only animals with ICC thresholds <10mJ were used, but it says that was 9 (not 8) out of 13 AAV-injected animals. Is that because one had the mid-turn cochleostomy and was excluded? I would suggest a clearer breakdown of these numbers, so the reader doesn't have to jump around to understand. Also, how many animals had the 10- vs. 5-LED arrays?

We thank the reviewer for raising this issue. The animal numbers are, of course, an important part of the results. Because of the long duration and complexity of the experiment, recording of ICC activity did not work in all animals where it was attempted. ICC recordings were attempted in all the 13 animals with oABR thresholds <10 µJ but worked in 9. One more animal was excluded due to the mid-turn cochleostomy leading to 8. Of these 8 animals 3 received a 10-channel insertion, and 6 a 5-channel insertion. One animal received both, a 5- and afterwards a 10-channel implant, adding up to 8 animals included in total. We have clarified these numbers in the corresponding paragraph.

> Results Paragraph 2: “Complete ICC recordings were obtained from 9 of 13 animals with oABR thresholds <10 µJ (loss due to the long duration and complexity of the experiments) and from 8 non-injected control animals.”
> Paragraph 3: “Of the eight included animals, three received a full 10-channel insertion and five received a 5-channel insertion; one additional 5-channel insertion was performed in one of the animals that also received a 10-channel insertion.”

Note, that initially the number of non-injected animals with successful ICC recordings was reported to be 9 in a methods paragraph. This was inadvertently wrong, and is now corrected to 8.

3.) It's unclear why "for acoustic stimulation" is included in the parentheses (for acoustic stimulation, Fig. 1A). Also, what does "ipsilateral" for oCI animals and "contralateral" for acoustic animals mean? Ipsi and contra to the ICC recording hemisphere? "Acoustic animals" should be non-injected animals, I assume, since all animals had intact hearing. It's probably best to state here that implanted animals were not deafened.

The control dataset for acoustic stimulation was recorded in a group of non-injected animals (n = 8) that did not undergo a bullostomy of the left ear. We had intended to indicate the purpose of this group by including “for acoustic stimulation” in parentheses, which was unrelated to Fig.

1A. We agree that this was unclear and have therefore moved this information outside the parentheses: “a group of non-injected animals for acoustic stimulation (Fig. 1A).”

Additionally, we now clarify in the Results that the ChReef-treated animals were not deafened, allowing assessment of multielectrode placement along the tonotopic axis using pure-tone stimulation: “The ChReef-transduced animals were not deafened…” These animals had, however, undergone a left-ear bullostomy for assessment of transduction using oABRs before placement of the ICC recording array, which likely reduced acoustic input from the left ear. We therefore considered this technical detail relevant for interpretation of the acoustic stimulation in these animals. However, it is not central to the reported results, and we no longer emphasize it in the results.

In contrast, animals used for the acoustic control dataset received poloxamer hydrogel in the right ear canal to induce conductive hearing loss and thereby predominantly restrict acoustic stimulation to the left ear, matching the side used for optical stimulation. This procedural detail was inadvertently omitted from the original manuscript and has now been added to the methods of acoustic stimulation and results paragraph:

> “In the acoustic control animals, poloxamer hydrogel was applied to the right ear canal and bulla to induce conductive hearing loss and thereby predominantly restrict acoustic stimulation to the left ear, matching the side used for oCI stimulation in injected animals.”

4.) I suggest specifying the analysis duration, 13 ms, in the paragraph starting, "We first computed evoked firing rates (evoked rate, ER)..."

We agree that this technical detail adds more depth. We included it in the last sentence of the paragraph: “we calculated firing rates in a 13 ms window following the response latency”

5.) The dynamic range paragraphs needs to be examined for clarity. The sentence "This is why the stimulus range..." could say, "The dynamic range was quantified as 10-90% of the max ER across 32-recording sites (Fig 2G)"; then explain how saturation played a role. Otherwise, it was difficult initially to understand why optical DRs in Fig. 2G were so much smaller than electrical just from comparing rate-level functions in Fig. 2E & F. The following bit about calculating 10% max firing rate (Fig S4): does this mean the 10-100% range? I suggest putting the last sentence about expressing dB values relative to threshold as the 2nd sentence following the introduction of rate-level functions. Please also state this in the captions, since I had trouble interpreting 0 dB without revisiting the text. Lastly, it would be useful to have the maximum stimulus intensities listed in this text, or the figure caption, as it took some time to find it in the methods.

We thank the reviewer for raising this issue. As suggested, we have restructured the paragraph for clarity and now introduce the expression of stimulus intensity relative to threshold in the

second sentence. We also revised the explanation of the comparatively small estimated optical dynamic range. Specifically, we now state that the maximum light intensity was kept approximately constant across recordings, independent of the individual threshold:

> "Because the same maximum light intensity (~3 mW) was applied across experiments, recordings with higher optical thresholds covered a smaller range of suprathreshold intensities, resulting in fewer recordings at the highest relative intensities (Fig. 2E)."

We also clarified that the additional analysis shown in Fig. S4 refers to the stimulus range spanning 10–100% of the maximum ER. Finally, the figure caption now states both the conversion of stimulus intensity relative to threshold and the maximum applied stimulus intensities (~3 mW for optical, 80–90 dB SPL for acoustic, and ~0.5 mA for electrical stimulation).

6.) Fig. 7: As a colorblind reader, I wasn't sure if the mean rate-level functions in Fig. 2E & F are different colors (red/green)? Please consider that color choice throughout the manuscript.
We would like to apologize for not considering this in the first version of the manuscript. We have now changed the color code from black/green/red to black/green/purple including differing markers for the different modalities.

7.) I didn't see how 46- and 42-dB SPL are represented in Fig 2H-J. Fig 2J show's distributions of max firing rates with "max equivalent SPL" medians of ~10 dB. Is this because dB is expressed relative to pure-tone threshold, rather than standard dB SPL? Also, it was unclear what the thin colored horizontal to vertical lines in Fig 2I represent. Are those mean values from Fig 2E and F? Does the Fig 2K-L caption mean to say, "The data shown in E-F was mapped onto the rate-level function for acoustic stimulation as shown in I"? Please examine the caption for correct panel references. Lastly, I don't see that Fig 2K & L are described anywhere in the text.
We thank the reviewer for the comment which helped us to improve clarity. **Fig. 2J** represents the maximum equivalent SPL relative to the pure tone ICC threshold. This was previously not mentioned in the figure caption but is now clarified:

> "**(I)** Quantification of an equivalent sound pressure level (SPL) relative to the pure tone ICC threshold".

The thin lines are supposed to visualize the process of mapping firing rates achieved with optical or electrical stimulation onto the acoustic RLF. This is also now specifically described in the figure caption:

> "The mapping procedure is illustrated by thin green lines projecting the maximum optical firing rate onto the acoustic rate-level function and from there to the corresponding stimulus intensity in dB SPL."

Accordingly, the panel references for in the caption for Fig. 2K-L were indeed incorrect. The references suggested by the reviewer are correct and were revised in the figure caption:

> "The data shown in **E-F** was mapped onto the rate-level function for acoustic stimulation as shown in **I."** Finally, Fig. 2K-L are now described in the main text: "By mapping the firing rates achieved at each step of the optical and electrical rate-level function onto the acoustic reference, each optical and electrical stimulus intensity relative to ICC threshold was assigned an equivalent SPL relative to the pure tone threshold (**Fig. 2K-L**)."

8.) Reference to Fig. S4 is mentioned before S2 or 3; consider rearranging those figures to match their order of appearance in the text.

We would like to thank the reviewer for pointing out this inconsistency. We have now revised the order of the supplementary figures to match the order of appearance in the main text.

9.) I suggest indicating the directions of dorsal and ventral recording positions in Fig 3, or at least in the caption.

We thank the reviewer for this suggestion. Because of limited space in the figure, we clarified the recording direction in the Fig. 3 caption: "Electrode numbers increase from dorsal to ventral recording positions in all STCs shown in **A-E**."

10.) The arrowhead in Fig 3B somewhat contradicts the statement in the text that "additional peak(s) [were] positioned toward the lower frequency range relative to the main peak." That extra peak is in the high-frequency direction.

We thank the reviewer for raising this issue. Secondary peaks more commonly but not always appeared in the high-frequency direction as can be seen in Fig. 5F. We have clarified this inconsistency in the text: "Moreover, in some cases, we found additional peak(s) positioned toward both the higher, but more commonly the lower frequency range relative to the main peak (see arrowhead in **Fig. 3B, D**)."

11.) Add "from" in: "This analysis indicated that secondary peak of ICC activity likely resulted [from] the spread of light..."

We have adjusted the wording as suggested by the reviewer.

12.) "...with a peculiar central activity minimum": this wasn't clear. Does this refer to the bimodal shape of the STCs for firing rates of >=200 Hz (dotted lines)?

We agree that the description of the shape of the STC in Fig. 4C was unclear. It indeed referred to the bimodal shape of the 200 Hz contour. We have revised the corresponding paragraph for clarity and included an additional expanded view figure showing the segmentation of the basilar membrane in the µCT scan. We initially hypothesized that the bimodal shape is attributable to

the obstruction of the light path by the basilar membrane, which does not seem to be the case:

> "Moreover, the lateral position and more tangential orientation of the LEDs relative to Rosenthal's canal within the wider scala tympani at the cochlear base may account for the broader activation observed in several recordings (e.g. 3rd to 5th panels of Fig. 4C, D; Fig. 3C, emitter 10). We initially hypothesized that the pronounced central activity minimum observed in some responses (e.g., star in the 4th panel of Fig. 4D) resulted from partial obstruction of the light path by the basilar membrane. However, µCT reconstruction localized penetration of the basilar membrane near LED 7, which did not show such an activity minimum (Fig. EV3), arguing against this explanation. Thus, the origin of the minimum remains uncertain with potential reasons including a locally increased impedance of the ICC recording electrode ."

13.) Fig. 4: Does the caption for 4A mean to say, "...oriented towards the RC as depicted in Fig. 1B", not Fig. 1C? Also, the caption doesn't describe the thin colored line, the Greenwood function.

Indeed, the sentence is supposed to be "…oriented towards the RC as depicted in Fig. 1B." This is now corrected. The colored line in Fig. 4B is now accurately described in the figure caption:

> "The colored line shown next to the RC reconstruction in B corresponds to the colorbar in A and represents the expected C_f at each RC position according to the gerbil Greenwood function (Müller, 1996)."

14.) 0.60 [0.14, 0.89] mW; I suggest indicating these and other values given this way are median and 25th, 75th quartiles.

We had previously introduced this convention in the paragraph reporting LED threshold. We have additionally added it to the paragraph describing Fig. 5 for clarity.

15.) For the paragraph describing Fig. 5, I found the terms confusing when referring to "single spectral peaks". There is an initial contrast between a "single spectral peak" and "two peaks." For Fig. 5D, all about the two peak responses, the text describes the intensity range of the "main spectral peak," but Fig5A and 5D use the term "single peak." Then, for giving the maximal ER, the text uses "single peak" in referring to STCs without a secondary peak (based on the Fig 5E caption). Please evaluate and maybe only use the term "main peak" when referring to two-peak STCs.

We agree that the terms were inconsistently used throughout the paragraph. We now directly define the main and secondary spectral peaks at the beginning of the paragraph: "The spectral peak occurring at the lowest threshold was termed the main spectral peak, while any peaks occurring at higher intensities were termed secondary spectral peaks (Fig. 5A)." Subsequently,

those terms are now consistently applied throughout the paragraph and the whole manuscript, e.g.:

> "When determining the number of spectral peaks across all recordings of ICC activity elicited by single LEDs (**Fig. 5B**), we found ~50% of the recordings to contain only the main spectral peak, while one secondary peak, i.e. two peaks in total, were found in ~50% and two secondary peaks, i.e. three peaks in total, in only two recordings."

16.) The sentence, "Optogenetic stimulation was tonotopically organized..." - this was established in the previous text/figures. Consider something like, "We quantified the tonotopic organization of 10-LED optogenetic stimulation and the corresponding variations in ICC threshold"

We agree that the general statement about tonotopic organization was redundant in this paragraph. We have therefore replaced it with a more specific description of the quantification:

> "We quantified the tonotopic organization of oCI stimulation by correlating the intracochlear emitter position and the C_f at the best, i.e. lowest threshold, electrode of the ICC response (**Fig. 5C**)."

17.) Fig. 5: What is the red line in 5C? I don't think I found this in the text or caption. Does this have something to do with the positive-going outlier data for the filled circles? That outlier should probably be mentioned. Also, I would suggest the caption to say, "Each combination of marker type and color corresponds to ICC recordings for an individual animal". Instead of "Markers indicate ICC recordings from the same animal"

We thank the reviewer for raising this issue. The colored line represents the linear fit for the recordings shown in Fig. 4 and was intended to provide a cross-reference between Figs. 4 and 5. This was previously only indicated in the caption by stating that "data corresponding to recordings shown in Fig. 4 are highlighted in green." We agree that this was insufficiently clear and have revised the caption accordingly to: "The green colored line indicates the linear fit to the recordings shown in Fig. 4, with the corresponding raw data overlaid as filled green-colored circles." We also adopted the reviewer's suggested wording to clarify the animal-specific marker coding and now state: "Each combination of marker type and color corresponds to ICC recordings from an individual animal." In addition, we now explicitly mention the positive-going outlier in the text: "In one outlier animal (dark filled circles in **Fig. 5C**), the relationship was reversed, which may reflect atypical emitter placement, as exact emitter positioning is difficult to verify during the experiment."

18.) In the first paragraph describing Fig 6: It was not immediately clear from the text or caption how SoE was computed. Because this is important, it's probably best to define clearly without needing to refer to the methods. Perhaps add to the following sentence: "Specifically, the width of the STC was defined as the distance between [ICC threshold contours (d' = 1)] at the dorsal-

most and ventral-most activated electrode, [measured at the lowest intensity level that elicited each ICC activity level.] In Fig. 6A, I suggest adding examples of the SoE measurement for 200 Hz and 400 Hz, in addition to the current 300-Hz example, for greater clarity of the method.

We thank the reviewer for the comment. We agree that the method for the calculation of SoE should become clear from reading the results section as it is central to the conclusions. We added a sentence very close to the one suggested by the reviewer to clarify this point: “Specifically, the width of the STC was defined as the distance between the dorsal-most and ventral-most electrodes within the ICC threshold contour (d’ = 1), measured at the lowest intensity level that elicited each ICC activity level. “

Additionally, examples for the SoE measurement at 200 and 400 Hz were added to Fig. 6A, as suggested.

19.) The first paragraph for the discriminability analysis: could be slightly more descriptive, since readers will presumably not have read the methods in detail or may not be that familiar with RSA. I would suggest stating first in more general terms what it is that RSA captures; ...correlation in the overall geometry between ICC response maps (?). As written, it's also somewhat unclear whether RSA is the ultimate comparison between modalities, or the pairwise comparisons within each modality. An example could provide a more intuitive sense of how it works. For instance, after the sentence ending "..., measured as correlation distance" it could say, "For example, Fig 7B shows how... [so forth]". This could also mention the matching of excitation levels in the analysis. Lastly, the term representational dissimilarity (RD) is not well defined here. Is it the same as the "correlation distance" mentioned above?

We agree that as for the SoE the RSA method should become clearer from reading the results. We have therefore revised the corresponding paragraph and the example given in Fig. 7B is described in detail. We begin by giving an overall definition of what RSA captures: “This was first quantified using representational similarity analysis (RSA), which captures the overall organization of relationships among ICC response patterns within each stimulation modality.”

In the third sentence, we clearly define RD:

“For each pair of tones or o/eCI channels, representational dissimilarity (RD) was calculated as the correlation distance between their ICC response patterns at matched activity levels.” We then directly describe how RD is derived for a given stimulus pair making use of the example given in Fig. 7B:

“For example, **Fig. 7B** illustrates how the ICC response patterns evoked by two stimuli, i.e. two tones, electrical or optical stimulation channels, are compared to obtain their pairwise RD. First, for each stimulus the lowest intensity eliciting a given firing rate (e.g. 200 Hz in **Fig. 7B**) is identified. RD is then calculated as the correlation distance between the responses across all 32 ICC electrodes at these matched activity levels.”

20.) The following paragraph for the multidimensional scaling plots (Fig 7C-E): It is somewhat

difficult to pick out the patterns described for each modality. Here again, describing an example or two from the figure would be helpful, maybe assisted by annotation in the figures (e.g., arrows).

We agree that the MDS plots were previously not sufficiently intuitive to interpret. We have therefore revised the corresponding paragraph and now first explain how the MDS representation should generally be read:

"In these plots, responses with lower RD are positioned closer together, whereas larger distances indicate more dissimilar ICC response patterns."

Additionally, we include examples for the visualization of the acoustic responses as well as a more detailed description of the pattern observed for electrical stimulation. For acoustic stimulation, we now state:

"For example, responses to 3.3, 4, 4.8, and 5.7 kHz, indicated by pink, yellow, blue, and orange circles, respectively, remained largely separated across activity levels. Activity level is indicated by the circle filling. The separation between 3.3 and 4 kHz became particularly apparent at higher activity levels (stars in **Fig. 7C**)."

For electrical stimulation, we now clarify:

"This is visible as responses to different electrodes, indicated by differently colored circles, shifting together from right to left with increasing activity level rather than remaining grouped by electrode identity."

We also added stars to Fig. 7C to highlight the acoustic example described in the text.

21.) Fig. 8F-H: consider making the triangle markers the same light shading as the individual data lines. The darker markers, matching the mean data, initially seemed as if that was the spacing boundary across all animals for each modality, which is certainly not the case for optical.

We appreciate the reviewer's suggestion and have adjusted the triangle marker face color to match the lighter color of the individual data lines. We retained the darker edge color from the previous version to preserve marker visibility.

22.) The sentence, "In contrast, optical thresholds did not differ significantly from acoustic thresholds...": Consider adding something like "despite being ~0.5-mm higher for optical than acoustic across those activity levels"; That offset is still stands out, so it' s probably worth noting.

We would like to thank the reviewer for raising this important issue. We agree that the difference in neurometric threshold between optical and acoustic stimulation is worth noting. We therefore now specifically include estimates of this difference in the main text:

> "Averaged across this activity range, optical thresholds were 0.44 mm lower than electrical thresholds (range 0.03–0.75 mm across activity levels) and 0.47 mm higher than acoustic thresholds (range 0.4–0.51 mm across activity levels)."

Discussion:

1.) In the sentence starting: "Increasing stimulation energy of individual oCI channels up to 3 µJ..." Does it later mean to say, "...approached [the maximum] evoked firing rates achieved by pure tone acoustic stimulation"?

Indeed, the intended meaning of the sentence was that evoked firing rates observed with optical stimulation in some cases approached the maximum evoked firing rates observed with pure tone acoustic stimulation. We have adjusted the wording accordingly.

2.) In the same paragraph, the phrase "and/or multi-channel stimulation present important avenues": does multi-channel stimulation in this context (and the following sentence) refer to combined (i.e., simultaneous) channel stimulation?

Correct, multi-channel stimulation in this context refers to simultaneous multi-channel stimulation. We have adjusted the wording accordingly to: "…and/or simultaneous multi-channel stimulation present important avenues for future studies."

3.) The meaning of the sentence, "This will also allow charting the range of cumulative ICC activity..." was not clear. What is "This" referring to exactly? Please examine for clarity.

In this sentence, "This" referred to the use of simultaneous multichannel stimulation schemes in future preclinical studies. By recruiting multiple frequency bands, such stimulation could help characterize the range of cumulative ICC activity and thereby inform expectations for loudness perception with clinically relevant coding strategies. For clarity, we now state this explicitly.

4.) In the spread of excitation (SoE) paragraph, is it possible to relate these firing rate values (>=150 Hz for electrical stimulation, and <100 Hz for acoustic stimulation) back to the equivalent acoustic sound pressure levels? - the same way that 46 dB SPL is derived as the highest SPL by the oCI? That could provide a more meaningful interpretation of relevant stimulus ranges. Also, later in this paragraph, why is up to 200 Hz the most relevant range for optical coding strategies?

We would like to thank the reviewer for this suggestion and the comment. It is possible to derive an equivalent SPL for these evoked firing rates. We have added them for the most important activity levels in the corresponding paragraph of the discussion:

"Relating these activity levels to equivalent acoustic sound levels provides a more intuitive estimate of the operating range over which optical stimulation improved spectral selectivity. The SoE of optical stimulation was significantly smaller than that of electrical stimulation at 150 Hz, corresponding to an equivalent SPL of 33 dB. The difference was largest at 36 dB SPL equivalent (200 Hz) and remained significant up to 45 dB SPL equivalent (350 Hz)."

We agree that the explanation for the calculation of the SoE slope in the range up to 200 Hz had not been clear previously. In the gerbil model of LED-oCI stimulation, ICC activity levels up to approximately 200 Hz appear most relevant for exploring optical coding strategies because spectral spread increases more strongly above this range, in part due to the emergence of

secondary spectral peaks. We therefore consider this range to represent the most favorable operating regime of the current oCI implementation in the small gerbil cochlea. The activity range up to 200 Hz therefore provides a useful basis for developing and evaluating multichannel optical coding strategies. We have clarified this point in the corresponding section of the discussion:
“In the current oCI implementation in the gerbil cochlea, spectral selectivity was best preserved up to approximately 200 Hz, above which off-target tonotopic regions were recruited (secondary spectral peaks). This range therefore appears most relevant for developing coding strategies with the presently available stimulation levels, although higher activity levels will likely be advantageous for clinical use of optogenetic SGN stimulation.”

5.) For the secondary peak paragraph: consider adding before the 3rd sentence something like, "The secondary peak could account for much of the broadening of SoE at higher optical intensities (i.e., for FR >= 150 Hz). On the other hand, when focusing on the main spectral peak (i.e., excluding the secondary peaks), the SoE remained much lower and closely..."
We agree that adding this sentence makes the role of secondary spectral peaks for the broadening spread at higher activity levels very clear. We have therefore included it as suggested.

6.) In the sentence, "A key question was therefore whether the relationships among ICC responses..." consider changing "distinct responses" to "distinct tonotopic populations" .
We would like to thank the reviewer for the suggestion. The wording in this paragraph was adjusted in response to another reviewer as well. We agree with the suggestion and have changed “distinct responses” to “activate sufficiently distinct tonotopic populations.”

7.) It may be worthwhile to highlight and discuss further why discriminability of oCI is more similar to acoustic discriminability at higher activity levels, >=200Hz, which contrasts sharply with the SoE analysis wherein oCI and acoustic are more similar at lower levels, <200 Hz. As mentioned briefly in the results, this could mean oCI channels remain meaningfully distinct (i.e., maybe perceptually...) at higher intensities, despite the broader spread of excitation, which is not the case for electric stimulation.
We thank the reviewer for raising this point. We agree that the improved channel discriminability observed at higher activity levels appears, at first, to contrast with the increasing spectral spread of optical stimulation. We have therefore added this aspect to the discussion:

> “Interestingly, this improvement became significant at higher activity levels (≥350 Hz), although the spectral spread of optical responses also increased at these levels. This suggests that oCI responses can remain meaningfully distinct despite broader overall activation. One possible explanation is that a strong, spatially localized response focus remains embedded within a broader activation pattern (e.g. LED 10 in **Fig. 3C**).”

Methods:

1.) Postnatal intracochlear injection: what was the route of cochlear injection?

We would like to thank the reviewer for the comment. We have provided more technical depth in the corresponding paragraph to clarify the route of injection:

“The cartilaginous bulla was identified using the adjacent facial nerve as a landmark. A glass pipette was then advanced toward the cochlea until perilymph reflux was observed, confirming intracochlear placement before injection of the AAV construct.“

2.) Detection of multi-unit activity: I suggest re-wording this sentence as follows, "After applying a band-pass filter (0.6 - 6 kHz, 4th order Butterworth filter), the threshold for detection of firing activity was defined as four times the mean absolute deviation within a pre-stimulus baseline (estimated as mean/0.675 for data 100 to 2 ms before stimulation)."

We agree that the suggested sentence improves the clarity of the method and have adjusted it as suggested.

3.) Quantification of activity level: is 100-µs biphasic pulses 100 µs per phase (200 µs total)? What was the rate of electrical pulses?

We would like to thank the reviewer for raising this issue. Indeed, the duration of electrical pulses was 100 µs per phase, i.e. a total of 200 µs. Electrical pulses were presented at ~3 Hz. We have updated the methods accordingly:

“200-µs biphasic electrical pulses (100 µs per phase) consistently presented at a stimulation rate of ~3 Hz”

4.) Analysis of Spectral Peaks: is the "d'-array" the threshold contour (d'=1)?

Indeed, “d’-array” should be “d’=1-contour”. We have adjusted it accordingly.

5.) Spectral spread of excitation: The sentence "For each activity level..." was confusing. Consider simplifying: "For each modality, we first found the lowest stimulus intensity that elicited a given activity level (e.g., 200, 300, 400 Hz)."

We agree that the sentence should be improved for clarity and have adjusted it as suggested.

6.) Some method sections point to relevant figures throughout the paper, and other do not. I would suggest always doing so where appropriate; e.g., Quantification of activity level = Fig 2A-C; Mapping of CF = Fig S1; particularly, Spectral spread of excitation = Fig 6.

We agree that the reference to relevant figures was not applied consistently throughout the methods section. We have therefore revised the methods section and included the corresponding figure references where appropriate.

19th Sep 2026

Dear Tobias,

Thank you for sending us your revised manuscript. We have now heard back from the two referees who were asked to re-evaluate your study. As you will see below, the referees are now satisfied with the modifications made. Before we can formally accept your manuscript, we would ask you to address the remaining minor issues noted by Referee #3 as well as the following editorial-level issues:

1. The email address provided for contributing author Juan Marcoleta (JMarcoleta@dpz.eu) appears to be no longer valid. Please provide the author's current email address.

2. Please add the European Innovation Council grant OptoWavePro (101158920) and the HORIZON TMA MSCA Postdoctoral Fellowship (OPTOCODE, grant 10110767) to the funders list in the online submission system.

3. Please limit the number of authors listed for each reference to 10, followed by et al. when there are more than 10 authors. Please also remove the DOIs from all published articles in the reference list.

4. Please rename Table EV2 as "Dataset EV1" and renumber the current Table EV3 accordingly. The legend for Dataset EV1 should be included in a separate tab/sheet within the corresponding Excel file. Please update the corresponding manuscript callouts and ensure that all EV tables are cited in sequential order.

5. Appendix Methods should be moved to the Methods section of the main manuscript.

6. Source data files for Figures 8G and 8H appear to be missing. Please check and provide these files. Source data folders that do not contain source data, for example, folders for schematics or deposited data, should be removed.

7. Please remove the BioRender attribution from the figure legends and add a dedicated subsection to the Methods section using one of the following formats, as applicable:
"Graphics:
(some of the... OR Figure #... OR synopsis) Graphics were created with BioRender.com."

8. Please address the following issues in the figure legends, as identified by our Data Editor:
- Please note that the exact p values are not provided in the legends of figures 2G, H; 6H, I; and 8C.
- Please indicate the statistical test used for data analysis in the legends of figures 5C and 8I.
- Please note that the box plots need to be defined in terms of centre, minima, maxima, whiskers, bounds of box, and percentile in the legends of figures 6I and 7I.
-Please note that information related to n is missing in the legends of figures 6I, EV4 A-H.

I look forward to reading a new revised version of your manuscript as soon as possible.

Kind regards,
Jingyi

Jingyi Hou
Senior Editor
EMBO Molecular Medicine

*** PLEASE NOTE *** As part of the EMBO Publications transparent editorial process initiative (see our Editorial at https://www.embopress.org/doi/pdf/10.1002/emmm.201000094), EMBO Molecular Medicine will publish online a Review Process File to accompany accepted manuscripts.

In the event of acceptance, this file will be published in conjunction with your paper and will include the anonymous 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 at contact@embomolmed.org.

***** Reviewer's comments *****

Referee #1 (Remarks for Author):

id like to thank the authors for so carefully addressing the reviewers' comments. the manuscript has improved substantially and all my concerns were addressed. I have no further comments.

Referee #3 (Comments on Novelty/Model System for Author):

N/A

Referee #3 (Remarks for Author):

The authors have thoroughly addressed each of my concerns/comments with substantial improvement to the presentation and clarity.

A couple minor details were noted:

- To further clarify the derivation of 51 and 50 equivalent dB SPL values in the text, the Fig. 2J caption might add something like "Median equivalent SPL values (16 & 15 dB, oCI and eCI, respectively) were related to acoustic sound pressure levels by adding the averaged pure-tone ICC threshold level of 35 dB SPL, resulting in 51 and 50 dB SPL, respectively."

- For Figure 8, it looks like there was a correction to the threshold criterion, now 83.9%. The Fig 8B caption still says 84.3%.

Referee #3 (Remarks for Author):

The authors have thoroughly addressed each of my concerns/comments with substantial improvement to the presentation and clarity.
We thank the reviewer for their positive assessment.

A couple minor details were noted:
- To further clarify the derivation of 51 and 50 equivalent dB SPL values in the text, the Fig. 2J caption might add something like "Median equivalent SPL values (16 & 15 dB, oCI and eCI, respectively) were related to acoustic sound pressure levels by adding the averaged pure-tone ICC threshold level of 35 dB SPL, resulting in 51 and 50 dB SPL, respectively."
We have clarified the calculation of the 51 and 50 dB SPL values in the Fig. 2J legend.

- For Figure 8, it looks like there was a correction to the threshold criterion, now 83.9%. The Fig 8B caption still says 84.3%.
The Fig. 8B legend now states 83.9%.

29th Sep 2026

Dear Tobias,

Apologies for the delay in sending you our decision. I have been traveling and only returned today.

We are pleased to inform you that your manuscript is accepted for publication and is now being sent to our publisher to be included in the next available issue of EMBO Molecular Medicine.

Your manuscript will be processed for publication by EMBO Press. It will be copy edited and you will receive page proofs prior to publication. Please note that you will be contacted by Springer Nature Author Services to complete licensing and payment information.

You may qualify for financial assistance for your publication charges - either via a Springer Nature fully open access agreement or an EMBO initiative. Check your eligibility: https://link.springer.com/journal/44321/how-to-publish-with-us

Should you be planning a Press Release on your article, please notify the production team via this form in order to coordinate publication and release dates: https://forms.cloud.microsoft/e/eH4A8JVP6g

If you have any questions, please do not hesitate to contact the Editorial Office. Thank you for your contribution to EMBO Molecular Medicine.

Kind regards,
Jingyi

Jingyi Hou
Senior Editor
EMBO Molecular Medicine

>>> Please note that it is EMBO Molecular Medicine 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://link.springer.com/partners/embo-press/editorial-policies#Peer%20review

EMBO Press Author Checklist

Corresponding Author Name: Tobias Moser
Journal Submitted to: EMBO Molecular Medicine
Manuscript Number: EMM-2026-23961

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 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?	Yes	Reagents and Tools Table, Material and Methods

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	Reagents and Tools Table, Material and Methods

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.	Not Applicable	

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.	Not Applicable	
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, Uploaded Dataset (age and sex for each animal)
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

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?	Not Applicable	

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.	Yes	Materials and Methods

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	
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.	Yes	Materials and Methods
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.	Yes	Materials and Methods
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, Figure Legends

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

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 approva**l (IRB or equivalent committee(s), provide reference number for approval. Include a statement of compliance with ethical regulations.	Not Applicable	
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?	Yes	
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.	Yes	
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	
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**.	Not Applicable	