nature portfolio

Peer Review File

Early life microbial succession in the gut follows common patterns in humans across the globe

Corresponding Author: Professor Vanja Klepac-Ceraj

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

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This manuscript provides a valuable contribution to our understanding of early-life gut microbial succession. The authors have assembled an impressive collection of metagenomes and identified robust age-associated colonization patterns that hold across different geographical regions. This large-scale analysis reinforces prior research and offers new insights into the dynamics of early-life microbiome development.

The multivariate age-prediction model developed in this study is particularly promising and has the potential to be a useful tool for future research into pediatric health and disease. However, further validation is needed to fully establish its accuracy and utility.

Specific Comments

Model Validation: The current validation strategy, while internally consistent, would benefit from a more standard train-test split and, ideally, external validation using an independent cohort. This would provide a more rigorous assessment of the model's performance and generalizability. I request the authors source additional data that were not used to train this model and report on accuracy and performance in a new cohort.

Study Focus: The manuscript could be strengthened by a clearer articulation of its primary objective. Is it to primarily describe colonization patterns, or is the age-prediction model itself a key deliverable?

-If the age-prediction model is a central goal, applying the model to disease cohorts and exploring age-discordance (as the authors suggest in the introduction) will provide valuable additional insights into the model's utility and encourage researchers to apply it in new ways.
-If the goal was to describe geographically robust developmentally trajectories, then the authors should be more explicit about the limitations of their model in the discussion, and clearer about the intent of the study in the introduction.

Key Strengths:

Large and diverse dataset
Robust identification of age-associated colonization patterns
Development of a promising age-prediction model

Key Areas for Improvement:
Strengthen model validation
Clarify the primary study objective
Consider additional analyses to explore the model's applications and limitations

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
- What are the noteworthy results?

Within this manuscript, Bottino et al. uses a random forest model to estimate child age from metagenomic-derived gut microbial taxonomic relative abundance data of 1,827 infants from 12 different countries throughout the first one and a half years of life. They report important taxa and functional features while identifying universal microbial succession patterns across varying populations, proposed as a normative microbiome age for comparison across pediatric outcomes.

- Will the work be of significance to the field and related fields? How does it compare to the established literature? If the work is not original, please provide relevant references.

No major or minor concerns.

This work incorporates a large and geographically diverse population to generate a global-scale predicted age of infant and young children microbiome, which can serve as a comprehensive reference for the development of gut microbiomes of infants and young children and potentially inspire further research and promising new clinical applications. Compared to individual cohorts, this work includes multiple cross-geographic datasets with large age ranges for infant and young children microbiome.

- Does the work support the conclusions and claims, or is additional evidence needed?

No major concerns.

Minor concerns:

The conclusions and claims are supported, although the authors suggest that the model can be used alongside other measures of child development. However, the authors do not provide examples of these comparisons within the main text. As this is one of their points of utilization and one of the main conclusions the authors present as a value of their work, we recommend including comparisons between the variables that can be normalized between the cohorts or in subcohort with reduced sample size. It is understandable to exclude these for the general build of the model, but using the metadata, albeit in a reduced sample size, to compare across samples would be of great interest to questions pertaining to public health and the global health effects or results of differing microbial biodiversity. Although the general patterns of microbial changes observed in other cohorts are recapitulated here, it would be interesting to see whether the early life exposure differences withstand scrutiny with the impressive dataset included here.

- Are there any flaws in the data analysis, interpretation and conclusions?

No major concerns.

Minor concerns:

It is unclear why the author changed their approach from a continuous, temporal consideration of the microbiome age and the corresponding features to the approach of looking at the 3-month and 12-month time points in Figure 4. Please clarify this within the text or by using different methods.

In addition, within the final figure, please ensure that you indicate the meaning of the red and blue triangles within the figure itself so that the reader does not need to refer to the figure legend.

The discussion primarily highlights that other studies used 16S rRNA sequencing. However, we believe that the main strength of this study is the multi-site and geographical diversity of the different studies incorporated, as a growing number of other studies (e.g. Stokholm et al. and Hoskinson et al.) have indeed begun using shotgun metagenomics data to produce microbiome-derived ages. The authors would benefit from focusing on this aspect of the study within the discussion.

To evaluate the generalizability of the model across different data sources, the author performed the leave-one-datasource-out cross-validation (LOOCV) experiment and showed that the average RMSEA increased from 2.6 (RMSEACV) to 3 months for predicting a 18 months range. However, to further prove the generalizability, we suggest to check the associations between the predicted age based original k-fold RF model and the predicted age based on the LOOCV RF model or to check the consistence of the importances of species based on these models.

The root mean square error in abstract is 2.61, but 2.56 in results section, please confirm which one is the right number or if they are based on different analyses.

Please state what EC stands for.

The name of data source in supplementary figure 3 is not consistent with other figures and tables.

The hierarchical clustering analysis reveals one large universal cluster of species and succession patterns containing 18 (53%) of the top 34 taxonomic predictors. It would be nice to be able to view it in Figure 3, which is ranked based on the hierarchical clustering analysis, by adding the dendrogram.

- Do these prohibit publication or require revision?

No major concerns.

Minor Concerns:

The manuscript will be fit for publication following minor revisions, including incorporating any metadata that can be normalized across datasets and adjusting Figure 4, whether by clarifying its description and their reason for depicting the data in this manner, altering the figure itself to reflect similar approaches to the rest of the paper, or changing their underlying methodology to use the continuous consideration of age as compared to timepoint (or perhaps by incorporating both).

- Is the methodology sound? Does the work meet the expected standards in your field?

No major concerns.

Minor Concerns:

The methodology is sound. No major or minor concerns.

- Is there enough detail provided in the methods for the work to be reproduced?

No major concerns. However, I believe the reader would appreciate a more detailed description of the random forest model (i.e. is predicted age based on the k-fold random forest model the average of the predicted value across 100 runs, how the parameters in the random forest model are selected, such as the number of trees) and Transition Score in the main text, which could help the readers to replicate this work.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have addressed my comments, no further revisions requested.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
On behalf of myself and co-reviewers, we thank the authors for carefully considering our comments and for integrating their responses into their manuscript. The authors clarified our points of interest and/or adjusted their manuscript to better reflect our concerns and their work. As of now, we are satisfied with the revisions and modifications to the manuscript. We would be supportive of an acceptance of this manuscript.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Response to the Reviewer's comments

We thank the reviewers for their careful reading our manuscript #NCOMMS-24-45749: *Early life microbial succession in the gut follows common patterns in humans across the globe*,and their insightful comments and questions. In response to their comments, we have include analyses using a completely new-to-us cohort, and we better rationalized the aims of the study. We have also carefully reviewed and addressed all other comments and responded to them in detail.

We have structured our responses by first including the reviewer's comment and then presenting our response. The page and line numbers following each specific response indicate the location of these changes in our revise, marked manuscript. Changes made in response to the comments are indicated in bolded font.

We would like to thank you and all reviewers for dedicating time and effort to review our manuscript. We believe their insightful comments and constructive feedback have helped us improve it substantially.

Reviewer #1

This manuscript provides a valuable contribution to our understanding of early-life gut microbial succession. The authors have assembled an impressive collection of metagenomes and identified robust age-associated colonization patterns that hold across different geographical regions. This large-scale analysis reinforces prior research and offers new insights into the dynamics of early-life microbiome development.

The multivariate age-prediction model developed in this study is particularly promising and has the potential to be a useful tool for future research into pediatric health and disease. However, further validation is needed to fully establish its accuracy and utility.

> **We thank Reviewer 1 for their positive feedback and careful and constructive reviews of our manuscript.**

Specific Comments

(1) Model Validation: The current validation strategy, while internally consistent, would benefit from a more standard train-test split and, ideally, external validation using an independent cohort. This would provide a more rigorous assessment of the model's performance and generalizability. I request the authors source additional data that were not used to train this model and report on accuracy and performance in a new cohort.

> **We thank Reviewer 1 for raising concerns with model validation.**

Here, we want to clarify how we validated the model's performance and generalizability. The Results section, Learned gut microbial patterns generalize across different sites (line numbers 197-208), describes the Leave-One-Datasource-Out cross-validation experiments. Our validation consisted of 6 separate, independent model training experiments – for each experiment, one left-out cohort acted as a completely blind test set, and internal cross-validation was performed only among the remaining cohorts left in-bag. For example, when the ECHO cohort was left out, a completely new microbiome age model was trained and cross-validated *from scratch* on all pooled samples *excluding any and all* ECHO samples. Then, the ECHO samples were used as a separate test set during evaluation. Given that we included datasets from multiple independent datasets from different research groups, we believe that this procedure demonstrates the generalizability of the model both across the globe and across collection procedures.

Nevertheless, to address the reviewer's concern about generalizability, we identified a new-to-us dataset and used it to validate our final model further to address Reviewer 1's request. We used data from Ennis & Yassour (2024), which contained 66 infant metagenomes with matched biosample collection ages inside the range of our original cross-validated model. After obtaining sequence data and processing through the unified computational pipeline, we reached a test-set RMSE of 1.55 months on this completely new dataset, matching the expected RMSE for the age range of the input data. We included the results (Ln 202-208), amended the Methods section Machine Learning (Ln 475-488), and also added additional language to the description of our initial cross-validation procedure to make it clear that these were separate models with no leakage between training and test data.

Ln 197-208:

To evaluate the generalizability of our model across different data sources and test the predictive ability of each data source toward age, we performed a leave-one-datasource-out cross-validation (LOOCV) experiment. We trained new, independent versions of our original model, excluding each data source completely from the training dataset, and benchmarked on out-of-bag data. LOOCV yielded an average RMSE of leave-one-out cross-validation of 3.03 +- 0.63 months (Supplementary Table 1, Supplementary Fig. 3A). Additionally, we pursued validating our model on an independent test dataset, not used in training or previous validation. This dataset, obtained from Ennis et al.[36], contained 66 infant metagenomes and paired biosample collection ages in the range of our original cross-validated model. After obtaining sequence data and processing through the uniformized pipeline, we reached a test-set RMSE of 1.55 months on this new dataset, matching the expected RMSE for the age range of the input data (Supplementary Fig. 3B-C).

Ln 475-488:

Models were trained on a repeated cross-validation scheme with 5 folds and 100 repetitions. Validation-set predictions for each sample were calculated as average values from the 100 repetitions, considering only the subset of models where each sample was not present on the training fold. Hyperparameter optimization was performed via factorial grid search considering the maximum number of nodes, the minimum number of samples per leaf and the proportion of subfeatures to consider for a split.

Additional validation was performed independently via Leave-One-Datasource-Out Cross Validation (LOOCV) and an external test set. For LOOCV, independent models were trained, validated and optimized with complete holdout of each major data source, which in turn was only used for model benchmarks. For external validation, sequence data and matching age annotations (N=66) were obtained from a study not among the original training data[36] and processed on the same pipeline as the original training data to generate taxonomic profiles. For test data, age predictions were generated as an average from all submodels of the grand model.

(2) Study Focus: The manuscript could be strengthened by a clearer articulation of its primary objective. Is it to primarily describe colonization patterns, or is the age-prediction model itself a key deliverable?

-If the age-prediction model is a central goal, applying the model to disease cohorts and exploring age-discordance (as the authors suggest in the introduction) will provide valuable additional insights into the model's utility and encourage researchers to apply it in new ways.
-If the goal was to describe geographically robust developmentally trajectories, then the authors should be more explicit about the limitations of their model in the discussion, and clearer about the intent of the study in the introduction.

We agree with the reviewer that the goals of this study could be more clearly articulated. The primary objective of the study was to describe early-life microbial colonization patterns that hold across different geographical regions. To do so, we developed and validated an age-prediction model (based on these patterns). Hence, the predictive model itself is both a mechanism of phenomenological understanding and a deliverable of the work, but not the main motivation in itself. Given previous work on the pediatric microbiome and its links to child development, we certainly believe that this model may have broader applicability but have endeavored to make the distinction between our goals for this manuscript and future work more clear. To that end, we have revised the introduction and discussion sections to clarify the nature of this dual focus, highlighting the comprehension of the underlying biology as the work's priority and positioning the robust nature of the model as well as its potential applications in future research, including disease cohorts (Introduction, Ln 123-132; Discussion, Ln 364-377).

Ln 123-132:

In this study, we sought to characterize generalizable early-life microbial colonization patterns across different geographical regions. To achieve this, we developed an age-prediction model based on gut microbial taxonomic relative abundances, with high temporal resolution for the first 1.5 years of life. Using 3,154 shotgun-sequenced samples from 1,827 infants across 12 countries spanning Africa, Europe, Asia and America, the model provides a robust tool for understanding microbial succession patterns and serves as a foundational benchmark for future studies linking gut microbiome maturation to child development.

Ln 364-377:

Directly testing the relationship between microbiome development and other normative measures of child development is beyond the scope of the present work, a limitation stemming from the nature of the data we used. Our study pooled data from a combination of observational and interventional studies with heterogeneous annotations, many of which lacked the detailed metadata necessary for such analyses. Despite this, the microbiome-age model provided here, built from a diverse and global population of human children, provides a foundational resource for future studies exploring these relationships. For example, previous targeted work by Subramanian et al.[17] demonstrated the utility of microbiome-age models in understanding malnutrition, while Shenhav et al.[41] linked alterations in microbiome maturation to the onset of asthma. These examples highlight the potential of microbiome-age models to advance public health research and shed light on the global health impacts of differing microbiomes. Future research incorporating harmonized metadata and targeted subcohort analyses will be crucial to fully realize the potential of microbiome-age models in these contexts.

Reviewer #2

Within this manuscript, Bottino et al. uses a random forest model to estimate child age from metagenomic-derived gut microbial taxonomic relative abundance data of 1,827 infants from 12 different countries throughout the first one and a half years of life. They report important taxa and functional features while identifying universal microbial succession patterns across varying populations, proposed as a normative microbiome age for comparison across pediatric outcomes.

We thank Reviewer 2 for their constructive and thoughtful feedback and address their comments point-by-point below.

- Will the work be of significance to the field and related fields? How does it compare to the established literature? If the work is not original, please provide relevant references.

No major or minor concerns.

This work incorporates a large and geographically diverse population to generate a global-scale predicted age of infant and young children microbiome, which can serve as a comprehensive reference for the development of gut microbiomes of infants and young children and potentially inspire further research and promising new clinical applications. Compared to individual cohorts, this work includes multiple cross-geographic datasets with large age ranges for infant and young children microbiome.

- Does the work support the conclusions and claims, or is additional evidence needed?
No major concerns.

Minor concerns:
(1) The conclusions and claims are supported, although the authors suggest that the model can be used alongside other measures of child development. However, the authors do not provide examples of these comparisons within the main text. As this is one of their points of utilization and one of the main conclusions the authors present as a value of their work, we recommend including comparisons between the variables that can be normalized between the cohorts or in subcohort with reduced sample size. It is understandable to exclude these for the general build of the model, but using the metadata, albeit in a reduced sample size, to compare across samples would be of great interest to questions pertaining to public health and the global health effects or results of differing microbial biodiversity. Although the general patterns of microbial changes observed in other cohorts are recapitulated here, it would be interesting to see whether the early life exposure differences withstand scrutiny with the impressive dataset included here.

We thank Reviewer 2 for this comment and agree that comparing across samples would be of great interest to many. Our newly expanded discussion in the last paragraph of the Discussion section (Ln 358-376) addresses this matter in more detail, while also incorporating references to studies where cohort-specialized age models were linked to outcomes of interest.

In this work, our primary aim was to build a microbiome-age model to characterize the generalizable early-life microbial colonization patterns across different geographical regions by adding publicly available data to our data. We also sought to motivate hypothetical use cases for future studies where such metadata is available. Due to the heterogeneous nature of publicly available annotations and the observational character of several of the studies employed, we were unable to identify properly-powered matched metadata across cohorts (e.g., malnutrition, asthma or autism). We furthermore understand that obtaining matched, harmonized clinical metadata falls outside the scope of our work. However, our model can be used in future studies with data across different samples and cohorts.

(2) It is unclear why the author changed their approach from a continuous, temporal consideration of the microbiome age and the corresponding features to the approach of looking at the 3-month and 12-month time points in Figure 4. Please clarify this within the text or by using different methods.

In addition, within the final figure, please ensure that you indicate the meaning of the red and blue triangles within the figure itself so that the reader does not need to refer to the figure legend.

We thank Reviewer 2 for this careful attention to our methods and figures. As suggested, we modified the Figure 4 legend to incorporate the visual meaning of the up/down triangles on the x-axis.

Our initial motivation for using only the 3-month and 12-month time points was that this analysis, in addition to exploring functional changes over time, also reproduced previous results obtained by Vatanen et al. (2018). In an effort to perform a true confirmation, we reproduced the original methodology from Vatanen et al. (2018), which discretized data according to the collection time point. By doing that, we observed that 25% of the ECs previously flagged as the largest abundance transitions in that original publication were also top-ranked in our modern, non-westernized cohort.

That said, we agree that the discrepancy between continuous and discrete analyses appears discordant, so to address Reviewer 2's comment in more depth, we examined the continuous nature of the longitudinal data as it would be a better match with the age model architecture. These results are presented in Supplementary Figure 5. 40% of the top ECs flagged by either a discretized analysis presented in Vatanen et al. (2018) or our original Figure 4 were also top hits on the continuous analysis, reinforcing the consistency of the results.

(3) The discussion primarily highlights that other studies used 16S rRNA sequencing. However, we believe that the main strength of this study is the multi-site and geographical diversity of the different studies incorporated, as a growing number of other studies (e.g. Stokholm et al. and Hoskinson et al.) have indeed begun using shotgun metagenomics data to produce microbiome-derived ages. The authors would benefit from focusing on this aspect of the study within the discussion.

We thank Reviewer 2 for raising this point and agree with it. Therefore, we adapted the Discussion section about the sequencing strategy (Discussion, Ln 291-309) to focus more positively on the rising prevalence of metagenomics in infant gut microbiome studies and how our model responds to and matches this modern reality, rather than solely focusing on the limitations of 16S amplicon sequencing.

Ln 291-309:

Most studies to date have characterized microbiome age using taxonomic classifications from amplicon sequencing of the 16S rRNA gene. However, an increasing number of studies are now adopting shotgun metagenomics, recognizing the greater resolution this approach brings to understanding microbial communities[16,40,41]. Accordingly, we developed our model using shotgun metagenomics, not only to address the well-known limitations of amplicon sequencing[23–25], and to align with the emerging standards in the field. Leveraging the ability of the metagenomic approach to sample all genes in a complex sample, we maximize taxonomic resolution and the breadth of identified taxa. In addition to taxa, this comprehensive approach enabled us to also explore the functional pathways associated with microbial genes—offering a deeper view into how the functional repertoire of the microbial communities evolved with age.

(4) To evaluate the generalizability of the model across different data sources, the author performed the leave-one-datasource-out cross-validation (LOOCV) experiment and showed that the average RMSEA increased from 2.6 (RMSEACV) to 3 months for predicting a 18 months range. However, to further prove the generalizability, we suggest to check the associations between the predicted age based original k-fold RF model and the predicted age based on the LOOCV RF model or to check the consistence of the importances of species based on these models.

We thank the reviewer for this suggestion and agree that this analysis provides helpful context. An in-depth view of the consistency of variable importances between the independent LOOCV experiments can be found in the newly added Supplementary Figure 4.

The root mean square error in abstract is 2.61, but 2.56 in results section, please confirm which one is the right number or if they are based on different analyses.

We thank the reviewer for catching this inconsistency. We have now reviewed all numbers and ensured that they are all correct. The correct value is 2.56 months (Ln 69).

Please state what EC stands for.

Done. Added on line 242.

The name of data source in supplementary figure 3 is not consistent with other figures and tables.

We would again like to thank the reviewers for their careful attention to detail. The names of data sources in Supplementary Figure 3 have been corrected as suggested. (Supplementary Figure 3).

The hierarchical clustering analysis reveals one large universal cluster of species and succession patterns containing 18 (53%) of the top 34 taxonomic predictors. It would be nice to be able to view it in Figure 3, which is ranked based on the hierarchical clustering analysis, by adding the dendrogram.

We agree that this visual representation aids the interpretability of the figure and have added the dendrogram and auxiliary notation to Figure 3 as requested. This also allowed us to review numbers and fix typos in the results section about this figure (Ln 197-208).

Ln 197-208:

To evaluate the generalizability of our model across different data sources and test the predictive ability of each data source toward age, we performed a leave-one-datasource-out cross-validation (LOOCV) experiment. We trained new, independent versions of our original model, excluding each data source completely from the training dataset, and benchmarked on out-of-bag data. LOOCV yielded an average RMSE of leave-one-out cross-validation of 3.03 +- 0.63 months (Supplementary Table 1, Supplementary Fig. 3A). Additionally, we pursued validating our model on an independent test dataset, not used in training or previous validation. This dataset, obtained from Ennis et al.[36], contained 66 infant metagenomes and paired biosample collection ages in the range of our original cross-validated model. After obtaining sequence data and processing through the uniformized pipeline, we reached a test-set RMSE of 1.55 months on this new dataset, matching the expected RMSE for the age range of the input data (Supplementary Fig. 3B-C).

- Do these prohibit publication or require revision?

No major concerns.

Minor Concerns:

The manuscript will be fit for publication following minor revisions, including incorporating any metadata that can be normalized across datasets and adjusting Figure 4, whether by clarifying its description and their reason for depicting the data in this manner, altering the figure itself to reflect similar approaches to the rest of the paper, or changing their underlying methodology to use the continuous consideration of age as compared to timepoint (or perhaps by incorporating both).

We addressed the comments in more detail above.

- Is the methodology sound? Does the work meet the expected standards in your field?

No major concerns.

Minor Concerns:

The methodology is sound. No major or minor concerns.

- Is there enough detail provided in the methods for the work to be reproduced?

No major concerns. However, I believe the reader would appreciate a more detailed description of the random forest model (i.e. is predicted age based on the k-fold random forest model the average of the predicted value across 100 runs, how the parameters in the random forest model are selected, such as the number of trees) and Transition Score in the main text, which could help the readers to replicate this work.

We appreciate Reviewer 2's thorough exploration of the Machine Learning methodology and agree that the information described would help the reader better understand the model without having to read the code provided. We added all the suggested information to the corresponding subsection *Machine Learning* in the Methods (Ln 475-488).

Ln 475-488:

Models were trained on a repeated cross-validation scheme with 5 folds and 100 repetitions. Validation-set predictions for each sample were calculated as average values from the 100 repetitions, considering only the subset of models where each sample was not present on the training fold. Hyperparameter optimization was performed via factorial grid search considering the maximum number of nodes, the minimum number of samples per leaf and the proportion of subfeatures to consider for a split.

Additional validation was performed independently via Leave-One-Datasource-Out Cross Validation (LOOCV) and an external test set. For LOOCV, independent models were trained, validated and optimized with complete holdout of each major data source, which in turn was only used for model benchmarks. For external validation, sequence data and matching age annotations (N=66) were obtained from a study not among the original training data[36] and processed on the same pipeline as the original training data to generate taxonomic profiles. For test data, age predictions were generated as an average from all submodels of the grand model.

Reviewer #3 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We sincerely thank Reviewers 3 and 4 for co-reviewing our manuscript and providing valuable feedback through the Nature Communications initiative for Early Career Researchers. Your comments have been carefully considered, and we are grateful for your contribution to improving the quality of our manuscript.