nature portfolio

Peer Review File

A probabilistic histological atlas of the human brain for MRI segmentation

Corresponding Author: Dr Juan Iglesias

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

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

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)
Casamitjana et al. conducted a study aimed at aligning histological (H&E and LFB) images with MR images of five human hemispheres, in the creation of a data resource named NextBrain, which was intended to serve as a probabilistic atlas of human brains. Additionally, they explored the application of NextBrain in aiding the segmentation of further ex vivo MR images and in the study of brain aging.

1. This study falls into an established territory as similar methodologies of cross-modality registration and information fusion for atlas building have been extensively documented in prior research (see below for some references and examples). Notably, this study fails to exhibit a marked improvement over existing techniques applied to both human and non-human animal brains—a detailed and careful comparison is notably absent. Furthermore, the manuscript lacks evidence to back the alignment and segmentation of Regions of Interest (ROIs); it merely offers an ad-hoc description of the procedure accompanied by elementary quantification. Consequently, it remains uncertain whether the NextBrain atlas will yield any novel insights or knowledge.

2. This study's conceptual foundation lies in the cross-modality, cross-scale registration of imaging data, a field that has seen considerable exploration. Past endeavors have spanned a wide range of techniques, such as the co-registration of specimens' MR images with CT images, PET images, gene expression patterns, spectrometry and various forms of microscopy, including light and electron microscopy, among others. In light of these extensive prior efforts, asserting any form of conceptual innovation within this domain would be challenging. This study overlooks a crucial opportunity by failing to benchmark against any existing approaches, not providing any compelling evidence of superior performance.

3. The cornerstone of this study lies in the alignment, or registration, of images. However, as acknowledged in the manuscript (lines 202-211), the essential methodologies had been previously published. Furthermore, the discussion scarcely addresses how one might achieve a biologically accurate digital reconstruction of human brains, which are renowned for their complexity. From a life sciences viewpoint, a critical omission is the lack of validation for the alignment process to ensure that the resultant NextBrain accurately represents a human brain, especially in comparison to prior efforts. In essence, the procedural description provided in this manuscript fails to instill confidence that the presented data are unequivocally accurate and also offer new knowledge missed in prior work.

Some references for comments above:

a. Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., & Marchal, G. (1995, June). Automated multi-modality image registration based on information theory. In IPMI Vol. 3, No. 6, pp. 263-274.

b. Yang, Q., Li, N., Zhao, Z., Fan, X., Chang, E. I. C., & Xu, Y. (2020). MRI cross-modality image-to-image translation.

Scientific Reports, 10(1), 3753.

c. Kieselmann, J. P., Fuller, C. D., Gurney-Champion, O. J., & Oelfke, U. (2021). Cross-modality deep learning: contouring of MRI data from annotated CT data only. Medical Physics, 48(4), 1673-1684.

d. van Tulder, G., & de Bruijne, M. (2018). Learning cross-modality representations from multi-modal images. IEEE Trans Medical Imaging, 38(2), 638-648.

e. Xu, Z., Luo, J., Yan, J., Pulya, R., Li, X., Wells, W., & Jagadeesan, J. (2020). Adversarial uni-and multi-modal stream networks for multimodal image registration. In MICCAI 2020: 23rd International Conference, Lima, Peru, Proceedings, Part III 23 (pp. 222-232). Springer International Publishing.

f. Wang, H., Rivenson, Y., Jin, Y., Wei, Z., Gao, R., Günaydın, H., ... & Ozcan, A. (2019). Deep learning enables cross-modality super-resolution in fluorescence microscopy. Nature Methods, 16(1), 103-110.

g. Santella, A., Kolotuev, I., Kizilyaprak, C., & Bao, Z. (2022). Cross-modality synthesis of EM time series and live fluorescence imaging. eLife, 11, e77918.

h. Van de Plas, R., Yang, J., Spraggins, J., & Caprioli, R. M. (2015). Image fusion of mass spectrometry and microscopy: a multimodality paradigm for molecular tissue mapping. Nature Methods, 12(4), 366-372.

There are many other major issues:

4. This study utilized five donated brains from elderly individuals aged between 78 and 94 years. Notably, there is an absence of documented information or discussion concerning potential brain atrophy in these tissues compared to normal brains—an issue that is prevalent within this specific age demographic and could significantly affect both registration and atlas construction. Such considerations warrant cautious interpretation of any claims regarding the resulting atlas. Given the uncertainties and the limited sample size (5) over the wide range of ages, and two genders (2 males and 2 females and another female with tumor), it's unclear about the validity of the study's findings on aging as presented in this paper.

5. It is somewhat surprising that the authors frequently mention their use of state-of-the-art algorithms without introducing any novel aspects to the algorithms employed, nor do they provide comparisons with other methods. This approach resembles a preliminary exercise characterized by the production of colorful but superficial plots without a thorough explanation of the biological significance underlying the data.

6. I must emphasize that Figures 1 and 3 could be combined, yet they require augmentation with rigorously conducted significance tests. Furthermore, Figure 2 demands substantial improvement through quantitative analysis, even if the work were to be submitted to a standard medical image analysis journal. For consideration in a leading Results or Discovery journal, the presentation of genuine, significant biological discoveries is essential to justify publication.

7. Figure 4 ought to be relocated to the supplementary materials. Figure 5, which currently seems to lack relevance (see below as well), requires significant enhancement to both elucidate the application of NextBrain and affirm the validity of the underlying data (e.g. age-related brain deformation etc.).

8. It is notable that the manuscript lacks in-depth discussion on aligning histological data with MR images, despite the inevitable presence of deformations, along with missing and damaged sections. Additionally, the alignment challenge seems to be framed as a 2D to 3D problem, as suggested by the authors mentioning the use of five parameters (line 686). However, the genuine task at hand should be a 3D to 3D registration. Without addressing this, the validity of the alignment process remains questionable. That said, again, considering the extensive research on cross-modality registration, it is challenging to stake a claim to novelty without introducing substantive new ideas and robust data, which have yet to be identified in this manuscript.

9. The numerous small fragmentations observed in Figure 3 raise concerns; they seem to be the result of over segmentation rather than meaningful Regions of Interest (ROIs) that convey significant anatomical information about the brain. Additionally, considering their small size, the relatively large registration errors reported in the study likely destroy the precise registration of these fragmented ROIs.

10. The connection between registering Human Connectome Project (HCP) data and the NextBrain initiative is not clear. While HCP data is publicly accessible and has been utilized in numerous studies, merely registering different cohorts of images does not inherently justify scientific significance. Such procedures are commonplace in scholarly works presented at the MICCAI conference and similar forums, suggesting that the study's approach to image registration does not constitute a novel scientific contribution on its own.

11. Claiming that NextBrain is beneficial for ultra-resolution segmentation seems overly ambitious due to two significant concerns. Firstly, the data foundation of NextBrain is confined to older brains, which are associated with substantial, undocumented, unquantified, and unexplained variabilities. These factors limit its direct applicability to new datasets.

Secondly, ex vivo brain images can already be segmented using numerous existing methods, making it unclear how NextBrain would offer any advantage over existing templates. If the segmentation process is unsupervised, NextBrain's relevance is questionable. Conversely, if the process is supervised, a detailed quantitative analysis is necessary to demonstrate how NextBrain was utilized, including a comprehensive comparison with other templates and methodologies, which are both missing.

In sum, the manuscript, as it stands, falls short of making a compelling contribution to the field, lacking in both novel methodologies and insights, as well as in rigorous quantitative analysis.

(Remarks on code availability)

Referee #2

(Remarks to the Author)
The work presented by Casamitjana and colleagues represents a potential significant advance regarding how magnetic resonance imaging (MRI) considers brain parcellations in the context of neuroimaging research. Importantly, the manuscript represents a potential significant advancement in the joint analysis of serial histological sections and MRI data. Many MRI scientists are seeking to improve the specificity of MRI findings, and resources such as the ones presented my, ultimately, be a critical way forward when attempting to perform this type of work. The methodology proposed considers histological labelling, AI-driven optimization of segmentations across several contiguous slices, a hierarchal non-linear registration scheme that considers slice-to-blockface registration, slice-to-slice smoothness correction, confounds of tearing, stretching, and missing data, and whole brain segmentation of novel high-resolution post-mortem MRI and ante-mortem MRI. While I think this paper will be impactful and highly relevant to the community, I do feel that it is undersold in many parts and also does not do a meaningful job of separating itself from recent work in the field.

Novel atlas generation: Given the explosion of data, data processing, and AI methods that being made available to the community, there have also been new atlases and reference maps that are also being developed. The authors mention several in their manuscript: Allen Human Brain Atlase, BigBrain, Juelich Brain, and the recent Glasser-Van Essen parcellation. However, the anatomy appears simply to be a representation of the ontology that was provided by Ding et al for the Allen Brain Atlas. How were the anatomical rules decided upon? How was consistency between raters established? Other than the technological innovation what can be discovered using this parcellation? What differentiates it (using examples) from these previous new parcellations? The authors also mention that the atlases can be customized using novel annotations. Given the framework that is described, it appears that this is indeed the case. How can novel annotations make their way back into to toolkit that was developed? Can they provide evidence of this and how this could help provide novel findings in the context of the lifespan or disease?

Visualization: The authors have taken the time to develop reproducible methods to painstakingly reconstruct 5 hemispheres at equisite resolution. Showing these data and their ultimate in 3D with clearly visualizations of the anatomy would be to the benefit of the authors as a means of highlighting the utility of this incredible resource, if only from the visualization perspective.
Segmentation validation: This is a place where I have significant concerns. It appears that there are two levels of validation that are proposed for the registration and subsequent segmentations. The former was validated using a landmark-based methodology. This seems very limited in a modern context and given the plethora of validation techniques that are available. Can something more robust be used. I was also surprised by the limited validation of the final volumetric segmentations. I do not think validating this against FreeSurfer segmentations is sufficient. The goal of this novel pipeline is to provide greater anatomical resolution at the structure-level, thus a more refined validation of the final segmentions using manual annotations is clearly necessary.

Application on a healthy ageing sample : Overall, I believe that his could have been a far more compreshensive evaluation. The visualization makes the findings difficult to understand. A Spearman correlation is likely to be insufficient to capture characteristics related to the ageing processing? Why some form of nonlinear spline or curve fitting not used to match the state-of-the art that is used in the field? Analogous to my previous questions, what can we learn by using this methodology that was previously unknown? Personally, I find this bioliogical validation to be weak and I believe that this really further consideration so that the authors can help highlight the utility of their work.

(Remarks on code availability)
Seems excellent. As with most code that comes from this group it is easy to understand and reasonable to implement.

Referee #3

(Remarks to the Author)
Casamitjana et al., present a next-generation probabilistic human brain atlas with automatic segmentation methods. Combining ex vivo MRI and histolgical staining, they created high-resolution 3D atlas with histology based segmentation. Furthermore, they demonstrate that their Bayesian tool can create automatic segmentations of both ex vivo and in vivo MRI data. Although a similar approach was used previously (e.g., Julich-Brain), this new atlas presents finer segmentations than previous human MRI atlases. Moreover, new segmentation tools can help to rapidly advance findings in both ex vivo and in vivo MRI for basic science and clinical applications. A web visualization was nicely implemented to browse the altas including histological sections easily. Hence, I feel that the manuscript contains important progress with a potentially significant impact on the human brain atlas. However, I also have several major concerns/comments.

Major comments
- Integration and comparison with previous probabilistic atlas and histological atlases.
There were a few noticeable probabilistic atlas (e.g., Julich atlas) as the authors pointed out in the introduction.
It'd be important to compare how the new atlas compares to the existing one, particularly Julich atlas.
The authors mentioned that the proposed one contains finer segmentations.
To illustrate this point further, the previous altas can be registered to the new atlas, and provide side-by-side comparison in the same space.
The comparison should also include potential differences in the ontology of anatomical areas.

- One of the main emphases in the manuscript is to enhance MRI segmentation by including histological sections.
This presents a unique opportunity to integrate existing histological atlases (e.g., Paxinos) with very fine segmentation.
When this true histology-based segmentation is imported into the new atlas, how does it compare with the labels in the new atlas?
Integration of the histology atlas can greatly improve its usability of atlas not just for the MRI community but also broader community using the histological section.

- Can this atlas be extended to align 2D histological data onto the new atlas?
Such capability will support the utility of this atlas bridging both macroscale MRI and cellular resolution histological data (e.g., pathological specimen).

- Authors nicely presented that their segmentation can be applied to many MRI image of normal brain across different ages.
Can this algorithm be also used to detect abnormality of the brain such as cancer or hemorrhage that will cause changes in the MRI contrast?
It'll be good to have MRI data from pathological conditions and how this algorithm can be utilized.

- How is 0.99mm registration error computed?
When the registration error was calculated, is there any difference across different brain areas such as white matter vs gray matter or cortex vs deep brain structures?

- Fig.5 contain many biological interesting finding.
However, I feel that the current figure presentation captures brain regional changes across aging well.
It'd be good to add graph or other quantitative presentation to show how different brain areas change over aging.

Minor comments
- Fig3A, what is the blue brain?

- Line 434-446, data is not presented clearly.
Please use arrows or some areal indication to indicate in the Fig5 to show specific brain areas discussed in the text.

- Fig4 with high ex vivo MRI and dense segmentation is very nice, please make the data publicly available.

(Remarks on code availability)

Version 2:

Reviewer comments:

Referee #1

(Remarks to the Author)

The author has made some effort to debate with the previous comments. However, it should be noted that the order of the text in the revised manuscript has been altered, particularly in sections related to the references I suggested, resulting in the

omission of important comparisons. Moreover, most of the constructive suggestions appear to have been rejected without providing solid new data to justify the significance of the presented results. It remains unclear how this work advances our understanding of human brain anatomy or introduces groundbreaking techniques that are not already known to the medical imaging community.

In response to R1.1, the referenced works were originally associated with R1.3, not this section. However, the authors' response essentially argues that previous studies did not use the same dataset as in this manuscript, rendering all prior work irrelevant and dismissible. This is a clear case of sophistry.

What was expected was a serious, quantitative comparison of the proposed techniques against these prior methods. The absence of such comparisons raises concerns—either there is little technical breakthrough, or there was an intentional effort to make claims without substantiating them through direct comparison.

Also, in response to R1.1, the authors state: "Further, we note that our base modality translation method (SbR, Casamitjana et al., 2021)..." This raises a critical question: does this imply that the fundamental version of the technique proposed in this study was already published? If so, it becomes even more difficult to substantiate the novelty of this work.

In response to R1.2, the listed “comparisons” do not address what was originally requested. If none of the requested experiments—whether involving human brains or non-human datasets—could be conducted, then how can one objectively assess the advantages of the proposed methods?
Furthermore, the authors generated hand-curated results and labeled them as “ground truth.” This is incorrect. These datasets may serve as gold standard references, but they do not represent absolute ground truth.

In response to R1.3, the authors argue that since they are addressing an intra-modality registration problem, validation is unnecessary. However, I strongly disagree. A meaningful comparison should be provided; otherwise, given the extensive body of work on registration in this field, there is no legitimate basis to claim any advancement.

Please note that a list of potential comparison papers for R1.3 was previously provided; however, the authors have moved them to R1.1 (not sure why). I am still keen to see a concrete, direct comparison that demonstrates the proposed registration method's superiority.

In response to R1.4, the authors stated: “The revised version of the manuscript includes a quantitative comparison, showing that the average error decreases from 1.45 mm to 0.99 mm with respect to our prior pipeline.”

Indeed, in the main text (lines 294–296), the authors write: “(known to be a better proxy for the registration error than similarity of label overlap metrics [59]), is 0.99 ± 0.51 mm – a considerable reduction compared to our previous pipeline [6], which yielded 1.45 ± 1.68 mm.”

However, there are several inconsistencies and ambiguities:

1. Statistical Overlap: The sum of 0.99 + 0.51 already reaches 1.5 mm, which falls within the range of the previous result (1.45 ± 1.68 mm), making it unclear how this represents a substantial improvement.

2. Figure Discrepancy: Figure 3’s legend states the same values, yet the figure itself presents a conflicting number: 1.27 ± 0.59 mm.

3. Lack of Clarity: The methodology for computing these numbers is not well explained, raising concerns about their consistency and validity.
Even if 0.99 ± 0.51 mm is the correct value, as mentioned earlier, the improvement over the previous pipeline is not particularly significant. Greater clarity and justification are needed to substantiate this claim.

For R1.5, the authors failed to address the key question of how they validate the registration of subjects with potential brain atrophy. Instead, they merely stated that the registration template (target) was derived from 305 subjects, which is irrelevant to the issue at hand. Unfortunately, this response falls below the acceptable standard for any technical paper discussing registration.

For R1.6, the authors declined to present quantitative results demonstrating the improvement of their registration method over established approaches on their dataset. However, providing such evidence is essential—without it, the work remains a theoretical argument rather than a state-of-the-art contribution. From a technical standpoint, the absence of supporting data suggests that the method does not outperform existing techniques. From a life sciences perspective, the study should either

offer novel discoveries uniquely derived from this dataset or focus on improving the registration results first.

For R1.7, while the authors reordered the figures, the illustrations remain largely unchanged and still lack meaningful quantitative comparisons. I strongly recommend combining and simplifying Figures 1 and 2, or possibly moving Figure 2 to the supplementary materials. More importantly, the required comparison in Figure 3 is inadequate—the so-called 'quantitative comparison' consists of just two lines of text in the lower right corner, presenting values (1.27 ± 0.59 vs. 1.44 ± 1.8) without significant statistical difference. Worse, the figure legend reports entirely different numbers (0.99 vs. 1.45), raising serious concerns about consistency and reliability. Without supporting data, these inconsistencies undermine confidence in the results, putting the study's reproducibility into question.

For other previous comments, the authors frequently chose to argue rather than provide the supporting data necessary to strengthen their work. Since none of the critical questions have been adequately addressed, I believe it is essential to resolve these issues first before I invest further time—out of goodwill—in offering constructive technical suggestions for improvement. If the authors aim to claim a state-of-the-art contribution without solid evidence, the work will struggle to earn respect, even if it eventually gets published, somewhere.

(Remarks on code availability)

Referee #2

(Remarks to the Author)
Excellent manuscript. This will be an important impactful paper in neuroscience.

(Remarks on code availability)

Referee #3

(Remarks to the Author)
Authors successfully address my concerns.
I can clearly see the improvement of the manuscript.
Thank you for incorporating my suggestions.

(Remarks on code availability)

Version 3:

Reviewer comments:

Referee #3

(Remarks to the Author)

(Remarks on code availability)

Letter of response to reviewers' comments

First of all, we would like to thank the reviewers and the editor for their comprehensive feedback. We believe that incorporating their comments has made our manuscript considerably stronger. The new version:

- Better positions our work - both the atlas and the methods that were used to build it.
- Includes new experiments for more comprehensive quantification and comparison with existing atlases:
    - Allen MNI template: quantitative segmentation and classification results.
    - Mai-Paxinos atlas and Allen reference brain: an extensive qualitative comparison is provided in the revised supplementary materials.
- Connects the results with different brain networks and existing biological hypotheses.

The updated data release also includes a 3D manual segmentation of the 100um isotropic scan by *Edlow et al. (2019)*, which we used to evaluate the accuracy of our automated segmentation algorithm in high-resolution *ex vivo* data (please see Supplementary Videos 3 and 4; the data are publicly available at https://openneuro.org/datasets/ds005422/versions/1.0.1). We believe that this segmentation will be an invaluable resource for the neuroimaging community, e.g., to augment the value of this scan as a common coordinate frame, or to develop and validate new segmentation methods.

In this letter, the original reviewers' comments are in *italics*, whereas our responses are in regular font. Changes on the manuscript have been marked in red.

Reviewer #1

R1.1. *Casamitjana et al. conducted a study aimed at aligning histological (H&E and LFB) images with MR images of five human hemispheres, in the creation of a data resource named NextBrain, which was intended to serve as a probabilistic atlas of human brains. Additionally, they explored the application of NextBrain in aiding the segmentation of further ex vivo MR images and in the study of brain aging. This study falls into an established territory as similar methodologies of cross-modality registration and information fusion for atlas building have been extensively documented in prior research (see below for some references and examples).*

[a] Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., & Marchal, G. (1995, June). Automated multi-modality image registration based on information theory. In IPMI Vol. 3, No. 6, pp. 263-274.

[b] Yang, Q., Li, N., Zhao, Z., Fan, X., Chang, E. I. C., & Xu, Y. (2020). MRI cross-modality image-to-image translation. Scientific Reports, 10(1), 3753.

[c] Kieselmann, J. P., Fuller, C. D., Gurney-Champion, O. J., & Oelfke, U. (2021). Cross-modality deep learning: contouring of MRI data from annotated CT data only. Medical Physics, 48(4), 1673-1684.

[d] van Tulder, G., & de Bruijne, M. (2018). Learning cross-modality representations from multi-modal images. IEEE Trans Medical Imaging, 38(2), 638-648.

[e] Xu, Z., Luo, J., Yan, J., Pulya, R., Li, X., Wells, W., & Jagadeesan, J. (2020). Adversarial uni- and multi-modal stream networks for multimodal image registration. In MICCAI 2020: 23rd International Conference, Lima, Peru, Proceedings, Part III 23 (pp. 222-232). Springer International Publishing.

[f] Wang, H., Rivenson, Y., Jin, Y., Wei, Z., Gao, R., Günaydın, H., ... & Ozcan, A. (2019). Deep learning enables cross-modality super-resolution in fluorescence microscopy. Nature Methods, 16(1), 103-110.

[g] Santella, A., Kolotuev, I., Kizilyaprak, C., & Bao, Z. (2022). Cross-modality synthesis of EM time series and live fluorescence imaging. eLife, 11, e77918.

[h] Van de Plas, R., Yang, J., Spraggins, J., & Caprioli, R. M. (2015). Image fusion of mass spectrometry and microscopy: a multimodality paradigm for molecular tissue mapping. Nature Methods, 12(4), 366-372.

The original submission may have not sufficiently clarified how our 3D histology reconstruction problem falls out of the scope of existing inter-modality registration techniques. Crucially, prior work (including references [a-h] above) did not address joint 3D constraints on thousands of nonlinearly deformed sections acquired in non-parallel planes as part of different blocks (leading to ~100,000 parameters) - and in presence of artefacts like folds and tears, which we address automatically (as opposed to, e.g., the costly manual intervention used in *BigBrain*).

Further, we note that our base modality translation method (SbR, *Casamitjana et al., 2021*): *(i)* unlike [b,d] above, does not require aligned pairs of images, which are not available in our setup; *(ii)* outperforms mutual information [a] and unpaired translation based on CycleGAN [c,e]; *(iii)* unlike [g], does not require landmarks or manual intervention. We also note that, while potentially an interesting direction of work, it is not our goal to super-resolve the MRI to the resolution of the histology (as [f,h]).

We have edited the introduction of the manuscript (and, to a small extent, the abstract) to carefully clarify these contributions on the image registration side of our work.

R1.2. *Notably, this study fails to exhibit a marked improvement over existing techniques applied to both human and non-human animal brains—a detailed and careful comparison is notably absent. Furthermore, the manuscript lacks evidence to back the alignment and segmentation of Regions of Interest (ROIs); it merely offers an ad-hoc description of the procedure accompanied by elementary quantification. Consequently, it remains uncertain whether the NextBrain atlas will yield any novel insights or knowledge.*

Following this and other comments below, the revised manuscript better positions our work in the context of existing human atlases (please see revised introduction, with new Fig. 1), and contains new experiments and results, including: *(i)* more detailed quantitative results on the accuracy of the 3D histology reconstruction, including a comparison with *Mancini et al., 2020*; *(ii)* where applicable, quantitative comparisons with the Allen MNI template (the only competing method that enables *in vivo* segmentation with histology-like labels); *(iii)* an exhaustive qualitative comparison

with the Mai&Paxinos atlas and Allen Reference brain; *(iv)* Dice scores on the 100um *ex vivo* scan, using a new ground truth segmentation that we generated specifically for this purpose; *(v)* a proof-of-concept study on Alzheimer's disease classification, where *NextBrain* achieves better discriminative ability than existing atlases; and *(vi)* more detailed ageing curves that showcase the nuanced effects of ageing in the brain that *NextBrain* can detect.

R1.3. *This study's conceptual foundation lies in the cross-modality, cross-scale registration of imaging data, a field that has seen considerable exploration. Past endeavors have spanned a wide range of techniques, such as the co-registration of specimens' MR images with CT images, PET images, gene expression patterns, spectrometry and various forms of microscopy, including light and electron microscopy, among others. In light of these extensive prior efforts, asserting any form of conceptual innovation within this domain would be challenging. This study overlooks a crucial opportunity by failing to benchmark against any existing approaches, not providing any compelling evidence of superior performance.*

While we agree with the reviewer that the inter-modality registration literature is vast, ours is not a "standard" inter-modality registration problem. Our reconstruction requires solving a registration problem with approximately 100,000 parameters, with joint 3D constraints on thousands of nonlinearly deformed 2D sections acquired in non-parallel planes as part of different blocks (and often in presence of strong artefacts like folding and tearing). As explained in R1.1 above, there are not any existing methods that can tackle this problem. *BigBrain*, for example, required extensive (and very costly) manual intervention to cope with artefacts, even with unblocked tissue cut with a custom whole-brain microtome. We, on the other hand, handle them automatically, even in presence of non-parallel blocks.

Furthermore, we would like to emphasise that the 3D reconstruction is just one of the contributions of our work, along with the atlas, segmentation method, data sharing, and software release. We have edited the introduction of the revised manuscript to better position our work and state our contributions.

R1.4. *The cornerstone of this study lies in the alignment, or registration, of images. However, as acknowledged in the manuscript (lines 202-211), the essential methodologies had been previously published. Furthermore, the discussion scarcely addresses how one might achieve a biologically accurate digital reconstruction of human brains, which are renowned for their complexity. From a life sciences viewpoint, a critical omission is the lack of validation for the alignment process to ensure that the resultant NextBrain accurately represents a human brain, especially in comparison to prior efforts. In essence, the procedural description provided in this manuscript fails to instill confidence that the presented data are unequivocally accurate and also offer new knowledge missed in prior work.*

Compared with our prior work *(Mancini et al.)*, the 3D reconstruction pipeline in this article has two main improvements: *(i)* Joint (rather than sequential) optimization of affine and nonlinear transforms; and *(ii)* Integration of our contrastive AI method SbR. The revised version of the manuscript includes a quantitative comparison, showing that the average error decreases from 1.45mm to 0.99mm with respect to our prior pipeline (Wilcoxon $p < 10^{-21}$). It also includes an analysis of the spatial distribution of the error (as suggested by Reviewer 3, see R3.6 below), which is relatively uniform (Extended Data Fig. 5).

We would also like to emphasise that we quantitatively evaluate our 3D reconstruction accuracy with 250 pairs of manually placed landmarks, as opposed to cheaper proxies frequently used in registration papers, like Dice scores, which often correlate poorly with accuracy (see, e.g., *"Image Similarity and Tissue Overlaps as Surrogates for Image Registration Accuracy: Widely Used but Unreliable", Rohlfing, 2013*). Finally, we would also like to clarify that, as a probabilistic template, the atlas is *not* expected to resemble a single human brain, but rather capture the anatomy of a new "arbitrary" collection of brains.

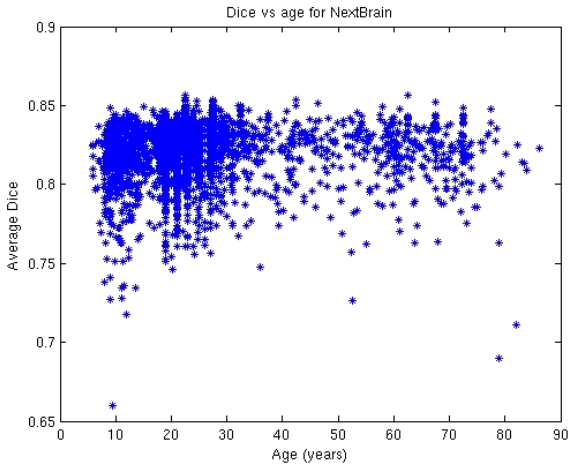
R1.5. *This study utilized five donated brains from elderly individuals aged between 78 and 94 years. Notably, there is an absence of documented information or discussion concerning potential brain atrophy in these tissues compared to normal brains—an issue that is prevalent within this specific age demographic and could significantly affect both registration and atlas construction. Such considerations warrant cautious interpretation of any claims regarding the resulting atlas. Given the uncertainties and the limited sample size (5) over the wide range of ages, and two genders (2 males and 2 females and another female with tumor), it's unclear about the validity of the study's findings on aging as presented in this paper.*

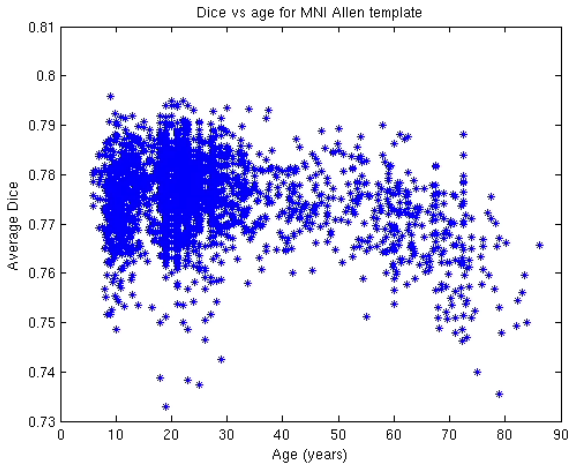
Working with *ex vivo* data at high resolution indeed has the disadvantages (not exclusive to this study) that sample sizes are lower and that the age of controls is skewed towards older subjects. These disadvantages were mitigated with two strategies:

- We "anchor" the atlas with the MNI template (by initialisation), rather than one of the subjects. This atlas is a well-established common coordinate frame built from 305 subjects. Using it as initialisation (rather than one of the five cases) prevents biases towards the selected case.
- Generative modelling is known to provide excellent generalisation ability, not only with *ex vivo* brain atlases (e.g., our thalamic atlas was built with 6 cases) but also in general (see, e.g., "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes.", Ng & Jordan, 2001).

The results on OpenBHB (which spans a wide range of ages), where NextBrain agrees very well with FreeSurfer (better than the MNI Allen template, the only histology-inspired competitor), support the accuracy of NextBrain segmentation in subjects younger than those in the training dataset. In fact, the MNI Allen template works much better for younger rather than older subjects, while the performance of *NextBrain* is more uniform across ages. Quantitatively: the correlation between average Dice and age is *much* stronger for the MNI Allen template ($r=-0.274$, $p<10^{-55}$) than for NextBrain ($r=0.046$, $p=0.009$); see scatter plots below.

These issues have been clarified in the methods and results section of the revision.





R1.6. *It is somewhat surprising that the authors frequently mention their use of state-of-the-art algorithms without introducing any novel aspects to the algorithms employed, nor do they provide comparisons with other methods. This approach resembles a preliminary exercise characterized by the production of colorful but superficial plots without a thorough explanation of the biological significance underlying the data.*

As mentioned in previous comments, we have edited the manuscript to highlight the novel aspects of our work and include new comparisons and figures, while trying to keep the article to a reasonable length. We believe that the new results reinforce the significance of NextBrain: as a dataset, set of methods (and open-source implementations), and ready-to-use tool.

R1.7. *I must emphasize that Figures 1 and 3 could be combined, yet they require augmentation with rigorously conducted significance tests. Furthermore, Figure 2 demands substantial improvement through quantitative analysis, even if the work were to be submitted to a standard medical image analysis journal. For consideration in a leading Results or Discovery journal, the presentation of genuine, significant biological discoveries is essential to justify publication.*

We believe that the overview in Fig. 1 (Fig. 2 in the revised version) is crucial in an article with so many moving parts, but we are of course willing to simplify it or combine it with another figure if the review and editor insist. Likewise, we believe that Fig. 3 (now Fig. 4), which we have augmented with the Allen MNI template, is also important, to understand the difference between NextBrain and competing atlases; we have however decreased its size.

Following the reviewer's advice, we have added to Fig. 2 (now Fig. 3) the quantitative results of the previous 3D reconstruction method, for comparison. We have also added these results to Extended Data Figs. 1-4. Further, as mentioned in Response R1.4 above, we have included a Wilcoxon test to assess the reduction in error provided by the pipeline presented here (from 1.45mm to 0.99mm, $p < 10^{-21}$). As mentioned in R1.4, we have also added an analysis of the spatial distribution of the 3D registration error (as suggested by Reviewer 3 in R3.6). Again, we

believe that the additional quantitative results (on 3D registration, atlas construction, image segmentation, and group analysis) and benchmarking to existing atlases (including the new comparison with Mai-Paxinos atlas and Allen reference brain) has considerably increased the strength of our manuscript.

R1.8. *Figure 4 ought to be relocated to the supplementary materials.*

Figure 4 (which is Fig. 5 in the revised version) now includes a comparison with the ground truth segmentation of the ultra-high resolution scan that we manually labelled and made publicly available, so we believe that it should stay in the main manuscript. Again, we are of course willing to relocate it if the editor and reviewer insist.

R1.9. *Figure 5, which currently seems to lack relevance (see below as well), requires significant enhancement to both elucidate the application of NextBrain and affirm the validity of the underlying data (e.g. age-related brain deformation etc.).*

Figure 5 (now Fig. 6) is, to the best of our knowledge, the most comprehensive map of regional ageing of the human brain to date. At the whole region level, the map is compatible with known effects studied with prior, coarser atlases (e.g., “Evolution of deep grey matter volume across the human lifespan”, *Narvacan et al., 2017*). At the subregion level: *(i)* the results are consistent with different brain networks and existing biological hypotheses, that suggest different factors causing grey matter volumetric reduction with ageing, particularly damaged processes of synaptic renewal and neuronal repairs (see, e.g., “Normal brain development and ageing: quantitative analysis at in vivo MR imaging in healthy volunteers”, *Courchesne et al., 2000*); and *(ii)* in regions previously studied with existing “focal” atlases (e.g., our hippocampal or thalamic atlases), the results are consistent with those from such atlases. In other regions, future work (possibly with higher resolution data) will be needed to further elucidate our findings. We have edited the manuscript to extend the discussion of these results.

A related issue is how the difference in resolution between the atlas and *in vivo* scans influences accuracy. The revised manuscript includes an experiment where we downsample the ultra-high resolution *ex vivo* scan from *Edlow et al. (2019)* to *in vivo* like resolution, estimate a high-resolution segmentation with *NextBrain*, and quantitatively compare it with the high-resolution ground truth. Albeit lower than when processing the high-resolution scan directly, the Dice scores are still fair for such small ROIs (median Dice 0.590 vs 0.667), which further supports the validity of our ageing study.

R1.10. *It is notable that the manuscript lacks in-depth discussion on aligning histological data with MR images, despite the inevitable presence of deformations, along with missing and damaged*

sections. Additionally, the alignment challenge seems to be framed as a 2D to 3D problem, as suggested by the authors mentioning the use of five parameters (line 686). However, the genuine task at hand should be a 3D to 3D registration. Without addressing this, the validity of the alignment process remains questionable. That said, again, considering the extensive research on cross-modality registration, it is challenging to stake a claim to novelty without introducing substantive new ideas and robust data, which have yet to be identified in this manuscript.

As mentioned in R1.1-R1.4 above, we have included new results and edited the manuscript to better position our registration method in the context of the inter-modality registration literature.

The 5-parameter transform mentioned by the reviewer is only used to robustly initialise the registration (which we call “first level of complexity”). As explained in Section “Refined alignment with preliminary nonlinear model”, the final level of complexity (“level 3 – stage 4”) solves a registration problem with complex joint constraints and approximately 100,000 parameters (about 100 parameters per section). These nonlinear parameters model complex transformations to the 3D space of the MRI. We have edited the methods section to clarify these levels of increasing complexity.

R1.11. *The numerous small fragmentations observed in Figure 3 raise concerns; they seem to be the result of over segmentation rather than meaningful Regions of Interest (ROIs) that convey significant anatomical information about the brain. Additionally, considering their small size, the relatively large registration errors reported in the study likely destroy the precise registration of these fragmented ROIs.*

This “fragmentation” is not visible in the probabilistic visualisation (panel A of the figure), as it is just a common discretization artefact when taking the most likely label at each voxel in areas where two or more labels mix continuously (e.g., the two layers of the cerebellar cortex, or the grey and white matter of the pons). That is why this fragmentation is not present in most brain regions. We have added a line to the caption of this figure explaining this artefact.

We would also like to clarify that: *(i)* the registration process in atlas building is independent of the 3D histology reconstruction; *(ii)* this atlas building registration can actually *compensate* for 3D reconstruction errors, since it seeks to co-register 3D labels, irrespective of whether they were slightly shifted by the imperfect 3D reconstruction in first place; and *(iii)* the atlas building registration is expected ***not*** to perfectly register ROIs, particularly in areas where ROIs can have very different shapes or even topology: the imperfect registration yields blurriness that enables the atlas to adapt to new cases (see, e.g., “Unified Segmentation”, Ashburner & Friston, 2005).

R1.12. *The connection between registering Human Connectome Project (HCP) data and the NextBrain initiative is not clear. While HCP data is publicly accessible and has been utilized in numerous studies, merely registering different cohorts of images does not inherently justify scientific significance. Such procedures are commonplace in scholarly works presented at the*

MICCAI conference and similar forums, suggesting that the study's approach to image registration does not constitute a novel scientific contribution on its own.

NextBrain does indeed enable co-registration of cohorts into a common coordinate frame (a common procedure in neuroimaging studies). However, our contribution is not a mere co-registration of HCP subjects, but rather: the new atlas, the methods that were used to build it, the segmentation method, the shared data, and the software release. As we said above, we have edited the introduction of the revised manuscript to better position our work and state our contributions.

R1.13. *Claiming that NextBrain is beneficial for ultra-resolution segmentation seems overly ambitious due to two significant concerns. Firstly, the data foundation of NextBrain is confined to older brains, which are associated with substantial, undocumented, unquantified, and unexplained variabilities. These factors limit its direct applicability to new datasets. Secondly, ex vivo brain images can already be segmented using numerous existing methods, making it unclear how NextBrain would offer any advantage over existing templates. If the segmentation process is unsupervised, NextBrain's relevance is questionable. Conversely, if the process is supervised, a detailed quantitative analysis is necessary to demonstrate how NextBrain was utilized, including a comprehensive comparison with other templates and methodologies, which are both missing.*

The concerns with the training cohort have been addressed in the response to R1.5 above. Regarding the second point: *ex vivo* scans, particularly at ultra-high resolution, have until recently been uncommon and restricted to few select sites. Even though the reviewer claims that there are numerous methods to segment such scans, we are only aware of two approaches: *(i)* using modified versions of FreeSurfer; and *(ii)* registration-based segmentation. Both approaches segment with a limited number of ROIs and limited resolution, and are unable to capitalise on histology. Only recently, researchers have started to manually label *a handful* of select regions in ultra-high resolution ex vivo scans to train neural networks (e.g., *Khandelwal et al., 2023*).

NextBrain is the first tool that can readily segment *ex vivo* MRI at ultra-high resolution, into hundreds of ROIs, and using information derived from histology. Having said that, we agree with the reviewer evaluating segmentation accuracy with a gold standard is important. For this purpose, we have created a ground truth segmentation of the ultra-high resolution *ex vivo* scan from *Edlow et al. (2019)*, which we have used to report Dice scores in the revised version of the manuscript. We have made this ground truth segmentation publicly available, since we believe that it may be highly valuable to the research community (e.g., for ROI analysis in the coordinate frame of the scan, or for development and validation of segmentations methods).

Reviewer #2

R2.1. *The work presented by Casamitjana and colleagues represents a potential significant advance regarding how magnetic resonance imaging (MRI) considers brain parcellations in the context of neuroimaging research. Importantly, the manuscript represents a potential significant advancement in the joint analysis of serial histological sections and MRI data. Many MRI scientists are seeking to improve the specificity of MRI findings, and resources such as the ones presented may, ultimately, be a critical way forward when attempting to perform this type of work. The methodology proposed considers histological labelling, AI-driven optimization of segmentations across several contiguous slices, a hierarchical non-linear registration scheme that considers slice-to-blockface registration, slice-to-slice smoothness correction, confounds of tearing, stretching, and missing data, and whole brain segmentation of novel high-resolution post-mortem MRI and ante-mortem MRI. While I think this paper will be impactful and highly relevant to the community, I do feel that it is undersold in many parts and also does not do a meaningful job of separating itself from recent work in the field.*

We greatly appreciate the kind words from the reviewer. As explained in the response to Rev 1 (particularly R1.1-R1.4), we have reshaped the introduction of the article (including the new Fig. 1) to better position our work in the context of the existing literature.

R2.2. *Novel atlas generation: Given the explosion of data, data processing, and AI methods that are being made available to the community, there have also been new atlases and reference maps that are also being developed. The authors mention several in their manuscript: Allen Human Brain Atlases, BigBrain, Juelich Brain, and the recent Glasser-Van Essen parcellation. However, the anatomy appears simply to be a representation of the ontology that was provided by Ding et al for the Allen Brain Atlas. How were the anatomical rules decided upon?*

Leaving aside the sparsity of the annotated sections and the low sample size (N=1), the Allen reference brain *(Ding et al., 2016)* leads the pack in terms of number and detail of labels; this is why we used their ontology as a basis. The set of atlases selected to label each region (listed in the Supplementary Methods) was decided by our senior neuroanatomists (JCA, MB, DK) based on the visibility and clarity of ROIs in our LFB sections. We have edited the Supplementary Methods section to clarify these issues.

R2.3. *How was consistency between raters established?*

To maximise the consistency in the annotations, a single senior neuroanatomist was in charge of defining the protocol of each brain region. This way, the junior neuroanatomists in charge of delineating a given region were all trained by the same person. This is clarified in the Supplementary Methods of the revised manuscript.

Unfortunately, no formal inter-/intra-rater analysis was performed: given the huge size of our labelling endeavour, we prioritised labelling more cases, rather than having our neuroanatomists label the same sections multiple times. We note that this is a common situation in projects requiring large-scale annotation of neuroanatomy - such as the Allen reference brain *(Ding et a., 2016).*

R2.4. *Other than the technological innovation what can be discovered using this parcellation? What differentiates it (using examples) from these previous new parcellations?*

As mentioned above, we have reshaped the introduction (including the new Fig. 1) to better position our atlas in the context of previous atlases. Moreover, we have included new experiments and results comparing *NextBrain* with competing algorithms, showing that it is more accurate than the Allen MNI template when segmenting brain ROIs; that it is better than the Allen MNI template and our previous single-structure *ex vivo* atlases at discriminating Alzheimer's vs control subjects in a group study; and that it can segment ultra-high resolution MRI with quite high accuracy (which we can now quantify with our new ground truth segmentation, which we have publicly released). We have also extended our discussion of the insights provided by the ageing experiment and included ageing plots with B-spline fits for individual subregions (see also R2.8). Finally, we have also included a comprehensive comparison with the available digital sections of the Mai-Paxinos atlas (in the supplement) and corresponding sections of the Allen reference brain.

R2.5. *The authors also mention that the atlases can be customized using novel annotations. Given the framework that is described, it appears that this is indeed the case. How can novel annotations make their way back into to toolkit that was developed? Can they provide evidence of this and how this could help provide novel findings in the context of the lifespan or disease?*

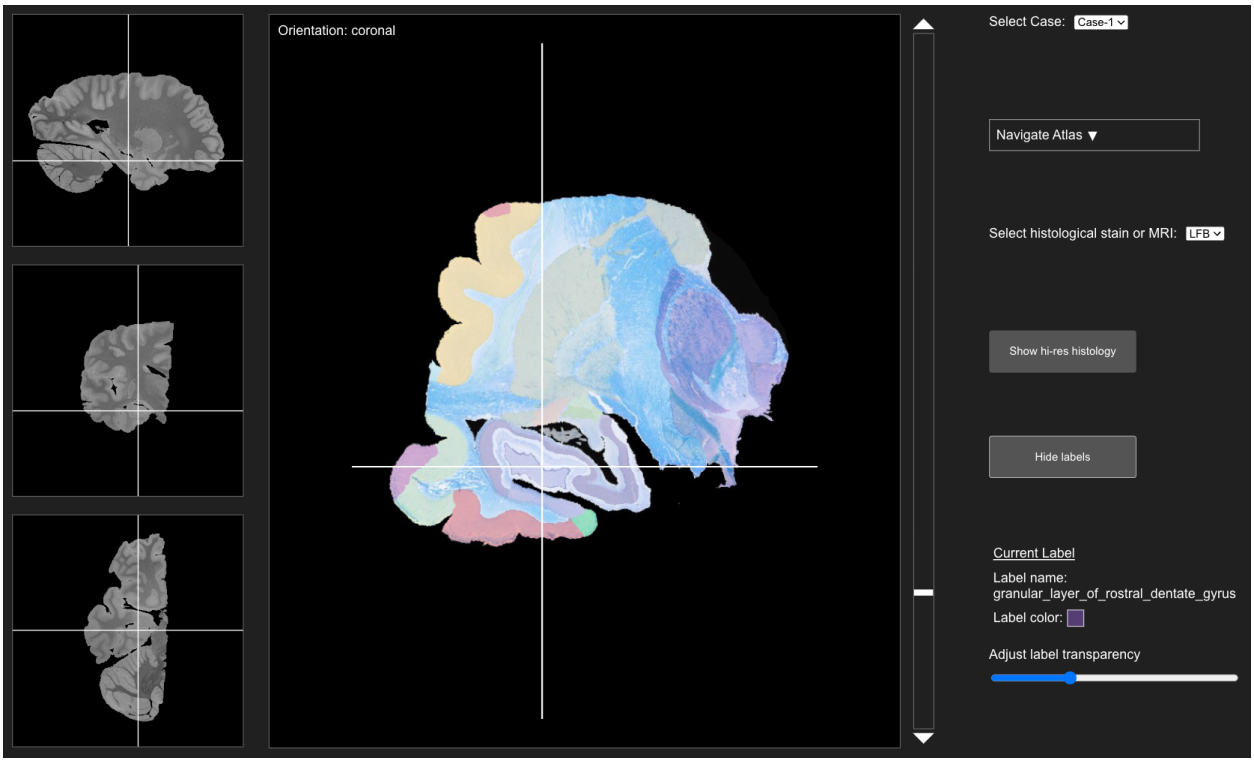
To modify or extend the atlas, one would have to download the data, modify (or extend) the manual annotations, and then rerun all the scripts to rebuild the atlas. This is possible as we have shared all our code and data, but does require domain expertise and compute power. Automatising this process to make it more accessible is definitely desirable, but it also quite a large engineering endeavour and remains as future work. This is now discussed in the Discussion section of the revised manuscript.

R2.6. *Visualization: The authors have taken the time to develop reproducible methods to painstakingly reconstruct 5 hemispheres at exquisite resolution. Showing these data and their ultimate in 3D with clear visualizations of the anatomy would be to the benefit of the authors as a means of highlighting the utility of this incredible resource, if only from the visualization perspective.*

NextBrain includes a website dedicated exclusively to visualising the data:

https://github-pages.ucl.ac.uk/NextBrain/#/atlas

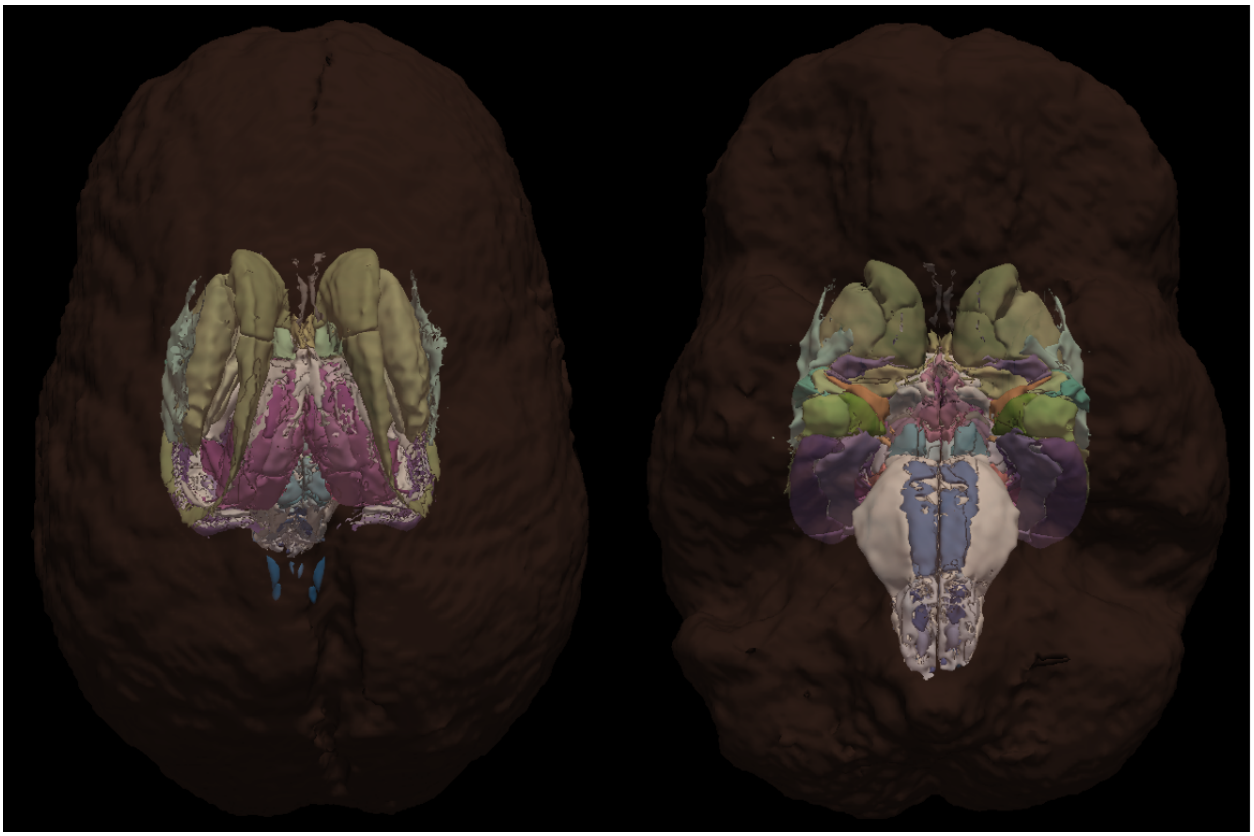
In this website, one can explore the co-registered 3D MRI and histology, e.g., by clicking on the MRI, and navigating to the corresponding location in the histology, while displaying the labels (see sample screenshot below).



A tutorial for this interface can be found in the main website:

https://github-pages.ucl.ac.uk/NextBrain

For 3D rendering: one can easily generate 3D renderings of the atlas or segmentation using a viewer like Freeview (which we distribute with FreeSurfer) or 3D Slicer. We believe that there are already plenty of tutorials online explaining how to do this, but we are glad to add a dedicated tutorial to the NextBrain page if the reviewer considers it appropriate. An example of 3D visualisation of brainstem and subcortical regions is shown below (left: superior view; right: inferior view).



R2.7. *Segmentation validation: This is a place where I have significant concerns. It appears that there are two levels of validation that are proposed for the registration and subsequent segmentations. The former was validated using a landmark-based methodology. This seems very limited in a modern context and given the plethora of validation techniques that are available. Can something more robust be used. I was also surprised by the limited validation of the final volumetric segmentations. I do not think validating this against FreeSurfer segmentations is sufficient. The goal of this novel pipeline is to provide greater anatomical resolution at the structure-level, thus a more refined validation of the final segmentions using manual annotations is clearly necessary.*

For the registration: as already mentioned in the response to R1.4, we would like to emphasise that our set or 250 manually placed pairs of landmarks enable us to obtain a *direct* estimate of the registration error (in mm). This is in contrast with surrogates like Dice scores or similarity measures, which are used more frequently in the registration literature just because they are cheap to obtain, but do not correlate well with the actual registration error in mm (again: please see, e.g., “Image Similarity and Tissue Overlaps as Surrogates for Image Registration Accuracy: Widely Used but Unreliable”, Rohlfing, 2013). Nevertheless, we have enhanced our evaluation with: *(i)* a comparison against our previous framework; *(ii)* statistical testing; and *(iii)* an analysis of the spatial distribution of the error (as suggested by Reviewer 3, please see response to R3.6 below), which turned out to be relatively uniform (Extended Data Fig. 5).

For the segmentation: we have considerably extended the validation in the revised manuscript. We have kept the FreeSurfer comparison because it enables us to evaluate the robustness of our method in a big sample with a wide range of ages, while comparing our method to the Allen MNI template. In addition, we have included new results based on a fine-grained manual segmentation of the 100um ex vivo scan from *Edlow et al. (2019)* with 98 ROIs. First, we report Dice scores of *NextBrain* at the resolution of the labels (200um isotropic). This setup enabled us to assess the accuracy of *NextBrain* when segmenting ultra-high resolution data. And second, we downsampled the scan to a typical *in vivo* resolution (1mm isotropic), segmented it at the resolution of the labels (200um), and computed Dice scores in this high-resolution space. This setup assesses the accuracy of the subregion segmentation when processing standard-resolution scans. Once more, we highlight that we are sharing this 200um segmentation (which was a huge annotation effort; see Supplementary Videos 3 and 4) on OpenNeuro, so that other researchers can use it for ROI analysis in the space of the *ex vivo* scan, or for development and validation of their own segmentations tools.

R2.8. *Application on a healthy ageing sample: Overall, I believe that his could have been a far more comprehensive evaluation. The visualization makes the findings difficult to understand. A Spearman correlation is likely to be insufficient to capture characteristics related to the ageing processing? Why some form of nonlinear spline or curve fitting not used to match the state-of-the art that is used in the field? Analogous to my previous questions, what can we learn by using this methodology that was previously unknown? Personally, I find this biological validation to be weak and I believe that this really further consideration so that the authors can help highlight the utility of their work.*

In the revised manuscript, we have enhanced the ageing experiments with:

- A deeper discussion of the results in Fig. 6 and Extended Data Fig. 7 (see R1.9), which connects our results with different brain networks and existing biological hypotheses.
- A new figure (Extended Data Fig. 8) with the B-spline fits suggested by the reviewer. We note that, as clearly seen in this new figure, many of the trajectories are nonlinear, which makes the Spearman correlation (which is rank based) more appropriate than the Pearson correlation (which assumes a linear relationship). While the correlation is a single number that does not tell the full story (compared with, e.g., the B-spline fit), it enables joint visualisation in a single image (i.e., as in Fig. 6 and Extended Data Fig. 7).
- A proof-of-concept group study of Alzheimer's disease, where *NextBrain* achieves better discriminative ability than existing atlases.

R2.9. *(Remarks on code availability): Seems excellent. As with most code that comes from this group it is easy to understand and reasonable to implement.*

We thank the reviewer for these kind words.

Reviewer #3

R3.1. *Casamitjana et al., present a next-generation probabilistic human brain atlas with automatic segmentation methods. Combining ex vivo MRI and histological staining, they created high-resolution 3D atlas with histology based segmentation. Furthermore, they demonstrate that their Bayesian tool can create automatic segmentations of both ex vivo and in vivo MRI data. Although a similar approach was used previously (e.g., Julich-Brain), this new atlas presents finer segmentations than previous human MRI atlases. Moreover, new segmentation tools can help to rapidly advance findings in both ex vivo and in vivo MRI for basic science and clinical applications. A web visualization was nicely implemented to browse the altas including histological sections easily. Hence, I feel that the manuscript contains important progress with a potentially significant impact on the human brain atlas.*

We appreciate the positive comments by the reviewer.

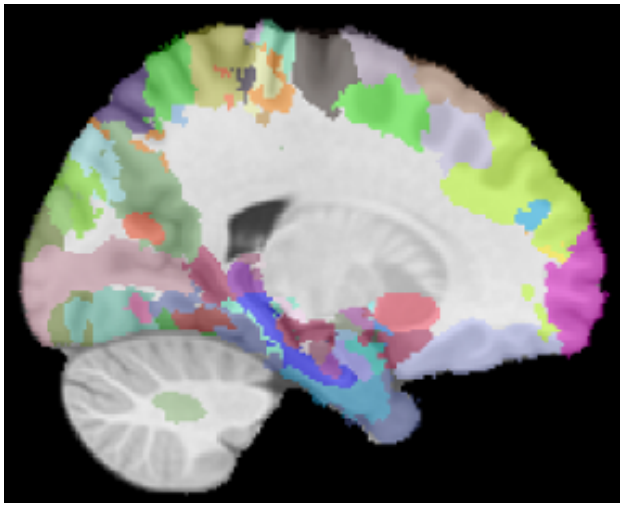
R3.2. *Integration and comparison with previous probabilistic atlas and histological atlases. There were a few noticeable probabilistic atlas (e.g., Julich atlas) as the authors pointed out in the introduction. It'd be important to compare how the new atlas compares to the existing one, particularly Julich atlas. The authors mentioned that the proposed one contains finer segmentations. To illustrate this point further, the previous atlas can be registered to the new atlas, and provide side-by-side comparison in the same space. The comparison should also include potential differences in the ontology of anatomical areas.*

The introduction of the revised version of the manuscript better positions NextBrain in the context of the current literature, highlighting the differences with existing atlases. We have included a new Fig. 1 that summarises these differences.

Rather than Julich, we have included the suggested side-by-side comparison with the Mai-Paxinos atlas and Allen reference brain, both of which contain delineations with great detail (see supplement of the revised version). An example is shown in the figure below. **[REDACTED]**

[FIGURE REDACTED]

For Julich-Brain: while it comprises several specimens and has dense 3D histology, it relies on labels that are made by the community, are not comprehensive, and align poorly with the template intensities (see below) - which is problematic since it only supports segmentation by co-registration with the template.



R3.3. *One of the main emphases in the manuscript is to enhance MRI segmentation by including histological sections. This presents a unique opportunity to integrate existing histological atlases (e.g., Paxinos) with very fine segmentation. When this true histology-based segmentation is imported into the new atlas, how does it compare with the labels in the new atlas? Integration of the histology atlas can greatly improve its usability of atlas not just for the MRI community but also broader community using the histological section.*

Co-registration of the digitised plates of the Mai&Paxinos atlas and NextBrain is very difficult, particularly given the sparsity of the former. In the supplement of the revised manuscript, we compare each Mai&Paxinos plate with our atlas after rough manual alignment, as well as with the closest labelled section of the Allen reference brain. However, we cannot (to the best of our knowledge) create a dense Mai&Paxinos labelling in the space of our atlas for direct voxel-wise comparison.

R3.4. *Can this atlas be extended to align 2D histological data onto the new atlas? Such capability will support the utility of this atlas bridging both macroscale MRI and cellular resolution histological data (e.g., pathological specimen).*

While technically possible, this is an incredibly challenging problem, for several reasons, including:

- Slice-to-volume registration is itself a difficult, ill-posed problem (see, e.g., “Slice-to-volume medical image registration: A survey”, Ferrante & Paragios, 2017).
- 3D histology reconstructions are highly anisotropic, leading to strong interpolation artefacts when slicing in oblique direction in 3D.
- Our final probabilistic only comprises label information; extension to LFB or H&E intensities is straightforward but suffers from the aforementioned interpolation artefacts.

R3.5. *Authors nicely presented that their segmentation can be applied to many MRI image of normal brain across different ages. Can this algorithm be also used to detect abnormality of the brain such as cancer or hemorrhage that will cause changes in the MRI contrast? It’ll be good to have MRI data from pathological conditions and how this algorithm can be utilized.*

Probabilistic atlases like NextBrain have been combined with models of tumours and white matter lesions in order to obtain joint segmentations of brain regions and pathology. Earlier models were simple outlier detectors; more recent works use dedicated shape priors (see examples in list of references below). It is out of the scope of this work to evaluate *NextBrain* in combination with such models for pathology segmentation, which remain as future work. We have, nevertheless, added a discussion item on this matter.

- *“Automated segmentation of multiple sclerosis lesions by model outlier detection”, Van Leemput et al., 2001.*
- *“Automatic Brain and Tumor Segmentation”, Moon et al., 2002.*
- *“Automatic Brain Tumor Segmentation by Subject Specific Modification of Atlas Priors”, Prastawa et al., 2008.*
- *“Multi-modal Brain Tumor Segmentation via Latent Atlases”, Riklin et al., 2012.*
- *“Segmenting glioma in multi-modal images using a generative model for brain lesion segmentation”, Menze et al., 2012.*
- *“Brain Tumor Segmentation Using a Generative Model with an RBM Prior on Tumor Shape”, Agn et al., 2015.*
- *“A contrast-adaptive method for simultaneous whole-brain and lesion segmentation in multiple sclerosis”, Cerri et al., 2021.*

R3.6. *How is 0.99mm registration error computed? When the registration error was calculated, is there any difference across different brain areas such as white matter vs gray matter or cortex vs deep brain structures?*

To compute the registration error, we manually placed pairs of landmarks in corresponding locations of the two modalities, and computed the distance between them after registration.

In terms of differences across brain regions: we computed the spatial distribution of the error in each of the five cases using kernel regression. The results are shown in a new figure (Extended

Data Fig. 5), which reveals that the distribution is approximately uniform, i.e., there is are not any regions that consistently have smaller or larger errors across the five cases.

R3.7. *Fig.5 contain many biological interesting finding. However, I feel that the current figure presentation captures brain regional changes across aging well. It'd be good to add graph or other quantitative presentation to show how different brain areas change over aging*.

The revised manuscript includes a deeper discussion of the results, a new figure with B-spline fits, and a proof-of-concept group study of Alzheimer's disease; please see further details in the response to R2.8.

== Minor comments ==

R3.8. *Fig3A, what is the blue brain?*

BlueBrain was the original name of our atlas, before we realised that there was already another project with that name. We have corrected this typo in the revised version.

R3.9. *Line 434-446, data is not presented clearly. Please use arrows or some areal indication to indicate in the Fig5 to show specific brain areas discussed in the text.*

This is an excellent suggestion; we have added markers to the figure (now Fig.6) and referenced them in the text.

R3.10. *Fig4 with high ex vivo MRI and dense segmentation is very nice, please make the data publicly available.*

As mentioned above, we have made the segmentations publicly available on OpenNeuro. We have also included two videos in the supplement flying over the slices (Supplementary Videos 3 and 4).

Letter of response to reviewers' comments

First of all, we would like to thank the reviewers and the editor for their feedback. We are encouraged by the highly positive feedback of Reviewers 2 and 3 ("this will be an important impactful paper in neuroscience") and are very motivated to write the best possible manuscript. Below is our detailed response to the outstanding issues raised by Reviewer 1 (R1). In this letter, the reviewers' original comments are in *italics*, whereas our responses are in regular font. Changes in the manuscript have been marked in red. For convenience, citations **[a-h]** referenced by R1 in their first round of comments are listed here:

[a] *Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., & Marchal, G. (1995, June). Automated multi-modality image registration based on information theory. In IPMI, Vol. 3, No. 6, 263-274*

[b] *Yang, Q., Li, N., Zhao, Z., Fan, X., Chang, E. I. C., & Xu, Y. (2020). MRI cross-modality image-to-image translation. Scientific Reports, 10(1), 3753*

[c] *Kieselmann, J. P., Fuller, C. D., Gurney-Champion, O. J., & Oelfke, U. (2021). Cross-modality deep learning: contouring of MRI data from annotated CT data only. Medical Physics, 48(4), 1673-1684*

[d] *van Tulder, G., & de Bruijne, M. (2018). Learning cross-modality representations from multi-modal images. IEEE Trans Medical Imaging, 38(2), 638-648*

[e] *Xu, Z., Luo, J., Yan, J., Pulya, R., Li, X., Wells, W., & Jagadeesan, J. (2020). Adversarial uni-and multi-modal stream networks for multimodal image registration. In MICCAI 2020, part 3, 222-232*

[f] *Wang, H., Rivenson, Y., Jin, Y., Wei, Z., Gao, R., Günaydın, H., ... & Ozcan, A. (2019). Deep learning enables cross-modality super-resolution in fluorescence microscopy. Nature Methods, 16(1), 103-110*

[g] *Santella, A., Kolotuev, I., Kizilyaprak, C., & Bao, Z. (2022). Cross-modality synthesis of EM time series and live fluorescence imaging. eLife, 11, e77918*

[h] *Van de Plas, R., Yang, J., Spraggins, J., & Caprioli, R. M. (2015). Image fusion of mass spectrometry and microscopy: a multimodality paradigm for molecular tissue mapping. Nature Methods, 12(4), 366-372*

Reviewer #1

R1.1. *The author has made some effort to debate with the previous comments. However, it should be noted that the order of the text in the revised manuscript has been altered, particularly in sections related to the references I suggested, resulting in the omission of important comparisons.*

A slight reordering was needed to accommodate overlapping suggestions from three different reviewers. Regarding the suggested references **[a-h]**: our introduction does not thoroughly survey the pairwise registration literature because this is not the main topic of the paper (but rather human brain atlasing) and because we already have 104 citations – which is twice the suggested limit in the author guidelines. However, we did cite two relevant surveys on nonlinear medical image registration [56] and 3D histology reconstruction [52]. In Section "Nonlinear AI registration" of the revised version (lines 932-934), we now reference mutual information and CycleGAN – which are the two unsupervised methods that are most relevant to our inter-modality registration problem.

Please see R1.6 below for further details on mutual information, CycleGAN, and the lack of comparison with **[a-h]**.

R1.2. *Moreover, most of the constructive suggestions appear to have been rejected without providing solid new data to justify the significance of the presented results.*

Some suggestions were not taken into consideration, for reasons clarified in the rest of this letter of response.

R1.3. *It remains unclear how this work advances our understanding of human brain anatomy or introduces groundbreaking techniques that are not already known to the medical imaging community.*

We respectfully disagree with R1. NextBrain contributes, in an open-source fashion:

- The first probabilistic atlas of the whole human brain built from histology.
- The most comprehensive dataset of registered MRI, histology, and anatomical labels of the whole human brain.
- A Bayesian segmentation tool that automatically extracts the NextBrain labels from any MRI scan (333 distinct anatomical regions).
- A registration method that *jointly* aligns thousands of non-parallel histological sections into a reference MRI scan in 3D, while preventing overlap and gaps in between.
- A web app to explore the data (also useful for educational purposes).
- A reference labeling of a publicly available 100 μm ex vivo brain (from **[3])**.

R1.4. *In response to R1.1, the referenced works were originally associated with R1.3, not this section.*

In their previous review, R.1.1 specifically said *“see below for some references and examples”*, so we moved the reference list **[a-h]** up to R1.1 for convenience – so that the editor and reviewers did not have to go back and forth between R.1.1 and R1.3 to look at these citations.

R1.5. *However, the authors' response essentially argues that previous studies did not use the same dataset as in this manuscript, rendering all prior work irrelevant and dismissible. This is a clear case of sophistry.*

We would like to clarify that:

- The arguments in R1.3 of our previous letter of response (reproduced below for convenience, with main technical points emphasized) were not related to *not using the same dataset* – but

rather *not tackling the problem* of jointly aligning thousands of non-parallel sections in 3D with overlap / gap constraints.

- As explained in R.1.1 above, prior work on nonlinear image registration is only briefly summarized in the introduction because it is not the main focus of this work - but, again, references to mutual information and CycleGAN have been added to this newly revised version of the manuscript.

Response to R1.3 in previous letter of response to reviewers' comments:

While we agree with the reviewer that the inter-modality registration literature is vast, ours is not a 'standard' inter-modality registration problem. Our reconstruction requires solving a registration problem with approximately 100,000 parameters, with **joint 3D constraints** on thousands of nonlinearly deformed 2D sections acquired in **non-parallel planes** as part of different blocks (and often in presence of strong artefacts like **folding and tearing**). As explained in R1.1 above, there are not any existing methods that can tackle this problem. *BigBrain*, for example, required extensive (and very costly) manual intervention to cope with artefacts, even with unblocked tissue cut with a custom whole-brain microtome. We, on the other hand, **handle them automatically**, even in the presence of non-parallel blocks.

Furthermore, we would like to emphasise that the 3D reconstruction is just one of the contributions of our work, along with the atlas, segmentation method, data sharing, and software release. We have edited the introduction of the revised manuscript to better position our work and state our contributions.

R1.6. *What was expected was a serious, quantitative comparison of the proposed techniques against these prior methods. The absence of such comparisons raises concerns—either there is little technical breakthrough, or there was an intentional effort to make claims without substantiating them through direct comparison.*

Since there are no methods that can jointly register non-parallel sections with overlap and gap constraints, and since the prior methods referenced by R1 **[a-h]** do not address this problem, we believe that the reviewer is referring to the SbR [10] component of our novel 2D–3D joint registration framework. SbR is a small component of this framework, and is essentially a 2D modality translation algorithm that we use to measure the similarity between a histology section and a (resampled) slice of MRI. Here, one could potentially insert a registration method indeed, including methods **[a-h]**.

However, most these methods require costly manual intervention or almost perfectly aligned data for pixel-level supervision that are not available in histology-MRI registration (specific details below). The only methods from **[a-h]** that could be used are those based on mutual information (being a generic metric) **[a]** or CycleGAN (being unpaired / unsupervised) **[c,e]**. We have extensively explored mutual information (MI) and CycleGAN in previous work (see, e.g., **[10,92]** or this abstract), and developed SbR precisely to overcome their weaknesses in the context of histology-MRI registration. Specifically, our workshop paper presenting SbR **[10]** demonstrates, on the same problem and with different datasets, that this technique outperforms MI and

CycleGAN. Therefore, MI and CycleGAN will not yield meaningful nor more accurate results in our atlas building application.

Below are the specific details about each of the references suggested by R1 **[a-h]**:

[a] describes mutual information, which is a versatile metric for inter-modality registration and could potentially be used instead of SbR. However, MI is not very accurate in nonlinear inter-modality scenarios like our histology-MRI problem and is clearly outperformed by modern machine learning approaches in such scenarios (see, e.g. “SynthMorph: learning contrast-invariant registration without acquired images”, Hoffmann et al., 2022). In the context of histology-MRI registration, our workshop paper **[10]** shows that SbR outperforms MI on two different public datasets (BigBrain and Allen atlas).

[b] belongs to a family of methods that make one modality resemble the other (“modality translation”) and then use standard intra-modality metrics based on intensity differences or correlations; note that SbR also belongs to this family. Nevertheless, **[b]** relies on *supervised* modality translation, which requires almost perfectly aligned images of the two modalities with direct pixel correspondence. Such aligned data are readily available in their application – brain MRI, where scans of different MRI contrasts can be rigidly aligned to yield training data. However, obtaining perfectly aligned histology and MRI is not feasible due to spatial distortion and tissue damage (folding, tearing).

[c] also relies on modality translation, but uses CycleGAN, a machine learning method that does *not* require paired images, and could potentially replace SbR in our application. While convenient from the data standpoint, the lack of supervision at the pixel level curbs the performance of this algorithm in histology-MRI registration; we show in **[10]** that SbR outperforms CycleGAN approaches on BigBrain and the Allen atlas datasets.

[d] is similar to **[b]**, but seeking to learn a common representation from multi-modal data using autoencoders (rather than predicting one modality from the other). Like **[b]**, it also requires almost perfectly aligned data (to exploit direct pixel correspondence), which are not available in our application.

[e] also builds in CycleGAN and has similar problems to **[c]**.

[f] is a super-resolution (SR) article, where a neural network is trained with almost perfectly aligned images, i.e., similar to **[b,d]**, but with SR instead of modality translation. To generate the training data, the authors use a cross-correlation metric, which only works when registering two images whose intensities follow a linear relationship (which, normally, is only the case when they are from the same modality). This precludes application to our histology-MRI problem.

[g] requires *costly manual annotation*, in the form of either landmarks or segmentations (from which landmarks can be derived); it is not a fully automated method.

[h] registers images by aligning pairs of *manually* placed landmarks using affine (rather than nonlinear) transforms. Therefore, it also requires comprehensive human intervention.

R1.7. *Also, in response to R1.1, the authors state: "Further, we note that our base modality translation method (SbR, Casamitjana et al., 2021)..." This raises a critical question: does this*

imply that the fundamental version of the technique proposed in this study was already published? If so, it becomes even more difficult to substantiate the novelty of this work.

As explained in R1.6 above, SbR is just a small component of the joint 2D/3D registration method. The main contributions from our work are a probabilistic histological atlas of the human brain and the associated collection of methods and materials that we have made publicly available. By focusing solely on SbR, the reviewer places does not fully consider the broader scope of the work, which includes several additional contributions that we believe are also of relevance: the actual atlas, segmentation method, the rest of the 2D/3D registration method (e.g., joint registration aspect, joint constraints, etc.), registered histology-MRI data, densely labeled histology, web app, and densely labeled 100 µm ex vivo brain from **[3]**.

R1.8. *In response to R1.2, the listed "comparisons" do not address what was originally requested. If none of the requested experiments—whether involving human brains or non-human datasets—could be conducted, then how can one objectively assess the advantages of the proposed methods?*

Comment R1.2 in the previous round said that the study "*fails to exhibit a marked improvement over existing techniques"*, that *"a detailed and careful comparison is notably absent",* and that the study *"lacks evidence to back the alignment and segmentation of Regions of Interest"*. We believe that these issues were thoroughly addressed in our revision – which added, with respect to the original submission:

- A summary figure (Fig. 1) that highlights the developments of NextBrain with respect to previous atlases.
- A supplement comparing NextBrain with the Mai-Paxinos and Allen atlases.
- On the joint 2D/3D alignment with overlap and gap constraints:
    - A comparison with our previous pipeline **[6]** – again, the only available method for jointly registering non-parallel sections in 3D with overlap / gap constraints.
- On the segmentation front:
    - A comparison with the Allen MNI template, the only competing methods that enables in vivo segmentation with histology-like labels.
    - An analysis on high-resolution MRI, using a new labelling of the 100 µm *ex vivo* brain from **[3]**, which we made for the purpose of the revision, and which we have made publicly available.
    - A proof-of-concept study on Alzheimer's disease classification using *in vivo* MRI scans, where *NextBrain* achieves better discriminative ability than existing atlases.
    - More detailed ageing curves that showcase the nuanced effects of ageing in the brain that *NextBrain* can detect.

Please see R1.6 on why an experiment replacing SbR by mutual information or CycleGAN was not considered.

R1.9. *Furthermore, the authors generated hand-curated results and labeled them as "ground truth." This is incorrect. These datasets may serve as gold standard references, but they do not represent absolute ground truth.*

Even though labelled histology is commonly regarded as "ground truth" in the medical image computing literature, the reviewer is right that this is not strictly correct. We have therefore replaced the terminology in the text as suggested by R1.

R1.10. *In response to R1.3, the authors argue that since they are addressing an intra-modality registration problem, validation is unnecessary. However, I strongly disagree. A meaningful comparison should be provided; otherwise, given the extensive body of work on registration in this field, there is no legitimate basis to claim any advancement.*

Our original response to this comment (reproduced under R1.5 above) did not argue that validation was unnecessary. We did validate the registration with 250 pairs of manually placed landmarks, and compared against **[6]** – which is the only existing method for joint registration of non-parallel sections in 3D with overlap and gap constraints. For the possibility of replacing SbR by mutual information or CycleGAN as the component to measure 2D similarity in our framework, please see R1.6 above.

R1.11. *Please note that a list of potential comparison papers for R1.3 was previously provided; however, the authors have moved them to R1.1 (not sure why).*

Please see R1.4.

R1.12. *I am still keen to see a concrete, direct comparison that demonstrates the proposed registration method's superiority.*

Please see R1.1, R1.6, R1.8, and R1.10.

R1.13. *In response to R1.4, the authors stated: "The revised version of the manuscript includes a quantitative comparison, showing that the average error decreases from 1.45 mm to 0.99 mm with respect to our prior pipeline."*
Indeed, in the main text (lines 294–296), the authors write: "(known to be a better proxy for the registration error than similarity of label overlap metrics [59]), is 0.99 ± 0.51 mm – a considerable reduction compared to our previous pipeline [6], which yielded 1.45 ± 1.68 mm."

However, there are several inconsistencies and ambiguities:
1. Statistical Overlap: The sum of 0.99 + 0.51 already reaches 1.5 mm, which falls within the range of the previous result (1.45 ± 1.68 mm), making it unclear how this represents a substantial improvement.
2. Figure Discrepancy: Figure 3's legend states the same values, yet the figure itself presents a conflicting number: 1.27 ± 0.59 mm.
3. Lack of Clarity: The methodology for computing these numbers is not well explained, raising concerns about their consistency and validity. Even if 0.99 ± 0.51 mm is the correct value, as mentioned earlier, the improvement over the previous pipeline is not particularly significant. Greater clarity and justification are needed to substantiate this claim.

We would like to clarify the 1.27 mm error reported in Fig. 3 is for the specimen in that figure, whereas 0.99 mm is the average across all specimens - as stated in the main text (line 294 in the previous version, now line 300) and in the caption of Fig. 3. The errors for the other four cases are shown in Extended Data Figures 1-4 and average to 0.99 mm, as stated in the main text. However, we acknowledge that it is more convenient to group the results in a single place, so we have created a new table (Table 1) with the reconstruction errors for all cases and methods. We have also made edit to the figures (Fig. 3 and Extended Data Figures 1-4) and their captions to further clarify this issue.

We would also like to clarify that the improvement with respect to **[6]** is highly significant (paired Wilcoxon $p<10^{-21}$, as was specified in line 297 – now line 308), and that 0.51 mm is the standard deviation, rather than the standard error, which is sqrt(N) = sqrt(250) = 15.8 times smaller (i.e., 0.03 mm) and explains the small p-value. At $p<10^{-21}$, there is virtually no overlap between the distribution of the medians of the errors of the two approaches (medians rather than means, since we use non-parametric testing). In the revised version, we provide both the standard deviation and standard error to avoid misunderstandings. We have also included specimen-wise p-values in the new Table 1.

We would like to particularly thank R1 for their close scrutiny of these results, as double-checking the landmarks for the new table made us realize that: *(i)* there were 5 landmarks in the two first cases (those from **[6]**) that had inaccurate annotations; and *(ii)* that we had not multiplied the standard deviations of the baseline method by the voxel size (0.4 mm) in the previous submission. These points have now been corrected and, importantly, do not change any of the presented results or conclusions.

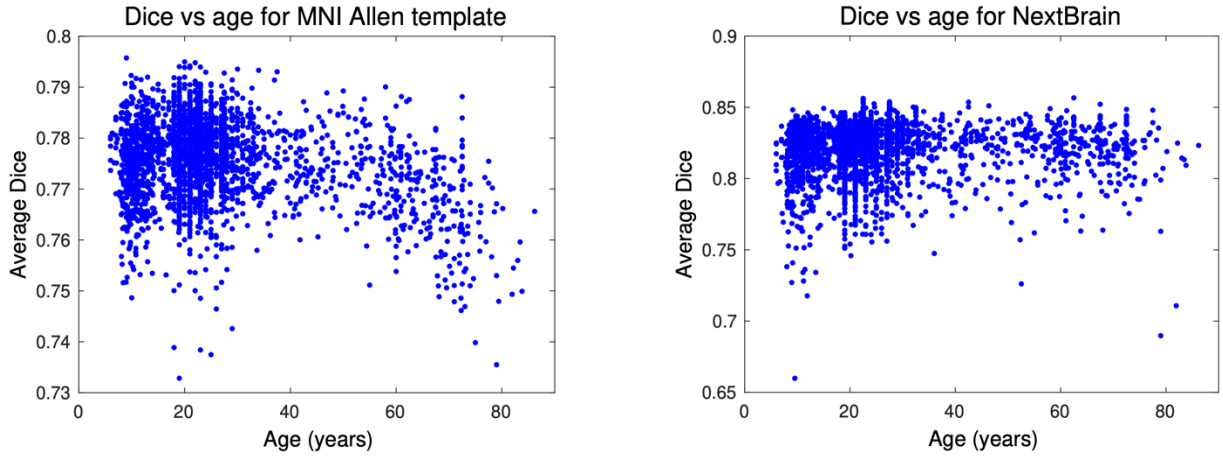
R1.14. *For R1.5, the authors failed to address the key question of how they validate the registration of subjects with potential brain atrophy. Instead, they merely stated that the registration template (target) was derived from 305 subjects, which is irrelevant to the issue at hand. Unfortunately, this response falls below the acceptable standard for any technical paper discussing registration.*

We respectfully disagree with this statement. Anchoring the atlas in MNI305 space (which is an average of 305 subjects of diverse ages) and segmenting with generative models (Bayesian

inference) are the two strategies that we use to *mitigate* the age bias of our atlas. The *effectiveness* of the approach was validated in an experiment that assessed the segmentation accuracy of NextBrain and the MNI Allen template as a function of age, using a dataset with a very wide age range (OpenBHB, with ages from 6 to 86 years). We included this experiment in the response to R1.5 (reproduced below for convenience), and we have also added it to the revised version of the manuscript (lines 561-569 and Extended Data Figure 6B-C).

Response to R1.5 in previous letter:

The results on OpenBHB (which spans a wide range of ages), where NextBrain agrees very well with FreeSurfer (better than the MNI Allen template, the only histology-inspired competitor), support the accuracy of NextBrain segmentation in subjects younger than those in the training dataset. In fact, the MNI Allen template works much better for younger rather than older subjects, while the performance of *NextBrain* is more uniform across ages. Quantitatively: the correlation between average Dice and age is *much* stronger for the MNI Allen template ($r=-0.274$, $p<10^{-55}$) than for NextBrain ($r=0.046$, $p=0.009$); see scatter plots below.



R1.15. *For R1.6, the authors declined to present quantitative results demonstrating the improvement of their registration method over established approaches on their dataset. However, providing such evidence is essential—without it, the work remains a theoretical argument rather than a state-of-the-art contribution. From a technical standpoint, the absence of supporting data suggests that the method does not outperform existing techniques. From a life sciences perspective, the study should either offer novel discoveries uniquely derived from this dataset or focus on improving the registration results first.*

Please see R1.1, R1.4, R1.6, R1.8, and R1.10.

R1.16. *For R1.7, while the authors reordered the figures, the illustrations remain largely unchanged and still lack meaningful quantitative comparisons. I strongly recommend combining*

and simplifying Figures 1 and 2, or possibly moving Figure 2 to the supplementary materials. More importantly, the required comparison in Figure 3 is inadequate—the so-called 'quantitative comparison' consists of just two lines of text in the lower right corner, presenting values (1.27 ± 0.59 vs. 1.44 ± 1.8) without significant statistical difference. Worse, the figure legend reports entirely different numbers (0.99 vs. 1.45), raising serious concerns about consistency and reliability. Without supporting data, these inconsistencies undermine confidence in the results, putting the study's reproducibility into question.

We believe that the potential move of figures to the supplement is a minor issue; we believe that both Fig. 1 and Fig. 2 are important enough to be in the main text, but are of course willing to move them to the supplement if the editor requests so.

For the quantitative results in the figures: as explained R1.13 above, we have realized the inconvenience of having the results of each case (mean, standard deviation, histogram) in a separate figure, and have consolidated them in Table 1. As we also explained in R1.13:

- The inconsistencies are not such (but again, we acknowledge the inconvenience of having the results spread across five figures).

- Results of a Wilcoxon test were provided for the full set of landmarks (the results now include p-values per case as well).

R1.17. *For other previous comments, the authors frequently chose to argue rather than provide the supporting data necessary to strengthen their work. Since none of the critical questions have been adequately addressed, I believe it is essential to resolve these issues first before I invest further time—out of goodwill—in offering constructive technical suggestions for improvement. If the authors aim to claim a state-of-the-art contribution without solid evidence, the work will struggle to earn respect, even if it eventually gets published, somewhere.*

Despite our disagreements on several points, we greatly appreciate the time that R1 is dedicating to reviewing this submission. We are however convinced that:

- Existing approaches (other than our previous version of the pipeline **[6]**) cannot jointly align multiple oblique 2D images into a 3D space while satisfying joint geometric constraints in 3D.

- Existing inter-modality nonlinear registration approaches could potentially replace the SbR component in the more complex 2D/3D registration framework, but they either *(i)* require costly manual intervention of perfectly aligned data that cannot be obtained with large histological sections, or *(ii)* rely on generic mutual information or unpaired machine learning approaches (like CycleGAN) that underperform compared with SbR [10].

- While there is obviously a registration component in our submission, its main contributions are a probabilistic histological atlas of the human brain and the associated collection of methods and materials that we have made publicly available, including the actual atlas, segmentation method, the 2D/3D registration framework with joint constraints, registered histology-MRI data, densely labeled histology, web app, and densely labeled 100 μm ex vivo brain from **[3]**.