nature portfolio

Peer Review File

Brain-wide cell-type specific transcriptomic signatures of healthy aging in mice

Corresponding Author: Dr Hongkui Zeng

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

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)
The authors use single-cell RNA sequencing to study transcriptomic changes in six major brain regions (isocortex, hippocampal formation, hypothalamus, cerebral nuclei, midbrain, and hindbrain) of young adult and aged mice. Though this could be one of the comprehensive datasets, the study design lacks originality, as this is not the first study comparing young and aged mouse brain regions (PMID 32669714- cerebellum, cortex, hippocampus, and striatum; PMID 36285248- aging female hypothalamus; PMID 36580914 - frontal cortex and striatum). The authors cite these articles in the introduction but do not clearly discuss how this study might be different from the existing literature on the transcriptome of young and aged brains.

After identifying age-associated differentially expressed (DE) genes in each subclass, they describe gene expression changes in major subclasses such as OPCs, oligodendrocytes, microglia, macrophages, endothelial cells, and astroependymal cells. From here, the analysis becomes redundant (Figure 3, 4, 5, and 6) for each subclass mainly with the representation of subclusters, age-DE genes, or GO enrichment analysis. Further, the authors used a commercial Molecular Cartography platform to generate two sets of in situ spatial RNA data for the following regions: cortex, striatum, midbrain, and hindbrain which they called RSTE1 experiment, and RSTE2 for the hypothalamus. However, in figures 3 and 4, the authors chose to represent gene expression changes identified from RSTE1 as violin plots (Figure 3h, Figure 4 c, j, k) while there has been no representation of the spatially resolved image (as for RSTE2 in figures 5 and 6).

The manuscript in its current form is descriptive with the identification of several age-associated DE genes and prediction of GO terms. The study could potentially benefit from the investigation of uncharacterized age-biased clusters and functional validation of predicted mechanisms using molecular and genetic tools.

Referee #2

(Remarks to the Author)
A. Summary of the key results
This study investigates cellular and molecular differences between the young adult and normally aged mouse brains using single-cell RNA-seq and highly multiplex RNA fluorescence in situ hybridization. The researchers profiled genome-wide RNA expression in ~1.2 million cells from young adult and aged male and female mice, representing ~110 brain regions, and clustered these cells by transcriptomic similarity into over 130 molecular subclasses. Comparing cell populations at multiple levels, researchers identified generalized and, more frequently, cell-type specific changes in gene expression. The greatest gene expression change was found to be among the non-neuronal populations, including pro-inflammatory changes in oligodendrocytes and microglia and a decrease in extracellular matrix related gene expression in vascular and smooth muscle cells. Regionally, the cells around the lower third ventricle which express the transcription factor Tbx3, including neurons of the arcuate hypothalamus and tanycytes, were particularly susceptible to age-related changes in gene expression. This is of particular interest given the role of these cells in energy balance and the well-characterized changes in energy balance that occur with aging. Overall, this study provides an immense and rich resource of the molecular architecture of the mouse brain and how it changes with normal aging and so is likely to be of great utility to the neuroscience and aging research fields.

B. Originality and significance: if not novel, please include reference
The results of this study are original and significant.

C. Data & methodology: validity of approach, quality of data, quality of presentation
The approach is generally valid (see concerns raised below), and the data and presentation are of high quality

D. Appropriate use of statistics and treatment of uncertainties
The use of statistics and treatment of uncertainties is generally appropriate, but please see concerns raised below in section F.

E. Conclusions: robustness, validity, reliability
Please see section F below.

F. Suggested improvements: experiments, data for possible revision
There are several key points in the manuscript lacking sufficient evidence or clarity, which need to be addressed in revision. Listed below are the concerns raised during this review, roughly in order of significance:

• To detect genes significantly affected by aging, the authors relied on a single-cell differential expression (DE) statistical method called MAST (Model-based Analysis of Single-cell Transcriptomics). However, recent work has demonstrated that MAST, like many other popular methods for single-cell DE analysis, is highly prone to false-positive detection of DE genes (PMID: 34584091). Pseudo-bulk methods (e.g., edgeR-LRT), in which cell transcriptomes are aggregated by biological replicate, are instead recommended for single-cell DE analysis. While the authors used stringent cutoffs ($p<0.01$, effect size > 1) for significance in the MAST results, the conclusions of this study would be greatly strengthened by a pseudo-bulk DE analysis, especially for the cells focused on figures 3-6.

• The authors state that 96 mice were used to generate the single-cell RNA-seq data and with 1-6 dissected tissue samples from each mouse. However, it is unclear how many mice comprise each cell subclass. In the “Replication” section of the “Reporting Summary,” the authors state that at least two mice were used. However, if one mouse was adult and the other was aged, then this would be n=1 for DE analysis of that cell subclass. In the interest of generalizability, it would help to know how many male mice, female mice, adult mice, and aged mice are represented in each of the cell subclasses. This could be added to the supplemental table of subclasses and their cell and gene content.

• A major conclusion in the study is that cell populations differ greatly in their susceptibility to the effects of aging on gene expression, measured as differences in the number of aging differentially expressed genes. However, the extent to which these differences in aging DE genes reflect differences in cell numbers is unclear, since more populous clusters/subclasses would have more statistical power to detect DE genes. The authors could strengthen their conclusion by plotting the number of DE genes by cluster/subclass size (n cells) and identifying the outliers, or by applying a method for detecting sensitive cell types, such as Augur (PMID: 32690972)

• Line 526-528 states “In our Tbx3+ clusters, cluster 63_TU-ARH Otp Six6 Gaba shows highly specific expression of Npy and Agrp, while cluster 8_ARH-PVp Tbx3 Glut shows specific expression of Pomc”. The data presented suggest that cluster 63 and cluster 8 may correspond to AgRP neurons and POMC neurons. However, the spatial data indicate that TU-ARH Otp Six6 Gaba neurons are located in both the arcuate nucleus and the tuberal nucleus, though AgRP neurons are only found in the arcuate nucleus. Is the TU-ARH Otp Six6 Gaba cluster a heterogeneous cluster containing AgRP neurons and tuberal neurons? If so, this would affect the interpretation of the age-DE genes, since the signal could be driven by either AgRP neurons or tuberal neurons To address this issue, authors could perform AgRP neuron and POMC neuron label transfer from one of the hypothalamic single-cell molecular atlases they cite (Chen et al., 2017 or Campbell et al., 2017) or from the HypoMap (PMID: 36266547). Also, feature plots of Agrp, Npy, Pomc on the HY UMAP may be helpful in illustrating whether clusters 63 and 8 are purely AgRP neurons and POMC neurons, respectively.

• Line 548-551 “Included in the circadian and rhythmic process related genes, we observed Bhlhe40, Bhlhe41, Nr1d2, and Per3 increasing with age only in the Agrp+ cluster (Figure 6k; Supplementary Table 3), suggesting that temporal and rhythmic control of behaviors like feeding, a known function of Agrp+ neurons76, may become altered with age.” Was the adult and aged hypothalamic tissue collected at the same time of day? The circadian timing of tissue collection can influence expression of clock genes in the hypothalamus.

• Line 782: “For mapping to the oligodendrocyte dataset from Marques et. al.30, we used a list of 195 genes.” Additionally, “For both external datasets, gene lists were assembled based on prominent marker genes from each external reference cluster. When mapping confidence score was needed, we sampled 80% genes from the marker list randomly, and performed mapping 100 times. We define the fraction of times a cell is assigned to a given cell type as the mapping probability to that type.” Will the authors provide code to replicate the marker lists? Or were the genes selected manually for the marker list?

• Line 690 states “We collected 1,508,284 cells without performing FACS.” How many cells were collected without FACS, the balance of the 2,777,165 cells?

• Figure 3h plots expression of Hopx3 but this gene is not mentioned in the results.

• For the UMAP in Figure 6b, do the colors in 6c serve as the color key?

• For the dot plot in Figure 6i, many of the dots are larger than the “100%” reference dot. This may be true of other plots. The reference dots should be re-generated to ensure they are correctly scaled to the data.

G. References: appropriate credit to previous work?
Yes.

H. Clarity and context: lucidity of abstract/summary, appropriateness of abstract, introduction and conclusions
Yes, the writing of the manuscript and appropriateness of these sections are excellent.

Referee #3

(Remarks to the Author)
In this study by Jin et al, the authors performed single-cell transcriptomics profiling of approximately 1.2 million high quality brain cells across 16 broadly dissected regions from 52 young (2-month-old) and 44 aged (18-month-old) male and female mice. Cells from the young animals were previously sequenced and annotated in Yao et al. 2023 and served as a reference for annotating cells from both aged animals and Resolve spatial data, which profiled 100 genes of interest in five different regions of interest. The study’s analysis revealed that the greatest expression changes with age occur in non-neuronal cells (OPCs and oligodendrocytes, immune cells, and vascular cells), as well as cells surrounding the third ventricle of the hypothalamus (tanycytes and Tbx3+ neurons), and hypothalamic Agrp+ and Pomc+ neurons.

Overall, the experimental approaches are valid, the data is of good quality. The manuscript is well-written and provides a large and comprehensive dataset with spatial information for the aging research community. However, despite the substantial amount of work, the analysis is relatively shallow, and lacks rigor in some areas. Further, the biological findings and claims are not backed up with experimental validation. Finally, while this is a useful resource, in this reivewer’s view, the impact of the findings does not reach the highest level. Specific concerns and suggestions regarding the experiments to enhance the impact of the work are provided below.

1. One of the major claims of the paper is that the greatest gene expression changes are found in non-neuronal cell-types. However, several choices in the analysis of the dataset may affect this finding. Specifically, on Line 798 the authors state “Only subclasses with at least 50 aged and 50 adult cells were evaluated for DE genes. To decrease running time, for large subclasses, we subsampled them to a maximum of 1,000 cells per age.” In subclasses with smaller numbers of cells, the analysis may be underpowered compared to subclasses with 1000 cells analyzed per group, and so it is difficult to call one group more or less changed with age.

2. It is unclear whether the young samples (published in Yao et al. 2023) were prepared side-by-side with the aged samples, or whether these samples were prepared separately. No consideration for batch or batch-effects was made in this paper, which is concerning as batch affects could introduce variability. An analysis of batch-to-batch variation is important.

3. The statistical reporting is unclear throughout the paper.
• For example, in Figure 3C (line 1195), the manuscript states “asterisks denote statistical significance” but no information for about what test was performed, the number of samples, or significance cut offs are listed. Similarly, in 3D, p-values are reported, but no information about how these values were calculated is given. In Extended figure 4E, it is not clear what test was performed or what is being compared.
• Many plots are missing appropriate labels, for example, the violin plots in 3D and 3H have no label indicating what the numbers on the y axis are.
• Many times, a p-value is reported where it is likely that an adjusted p value should be used, such as in 3D, 3H, 5E. Clearer statistical reporting in figure legends would lend clarity. On line 809, the authors indicate p values for the differential expression analysis are adjusted using the Bonferroni method, and this should be indicated throughout the figures and figure legends.

4. The authors did not mention whether variables such as Estrous cycle and Zeitgeber time for tissue collection were controlled. These variables could have significant effects on hypothalamic cells, such as tanycytes, Agrp and NPY neurons (https://doi.org/10.1073/pnas.0904747106), as they behave differently across different stages of the estrous cycle. Additionally, circadian-related genes mentioned in lines 548-549 can fluctuate across different Zeitgeber times. If these variables were not tightly controlled between the young and aged samples, the age-associated differential expression (DE) signatures could be confounded by variations in estrous cycles and Zeitgeber times.

5. Concerns with the FACS population analysis. In the analysis of adult/aged neuronal cell proportions shown in Fig 2 and Fig 6i, the authors should investigate whether the FACS scheme acts as a confounding factor. This could involve assessing whether FACS-enriched neurons exhibit a similar neuronal subclass/cluster composition as the non-FACS neurons. Additionally, it would be important to determine if the proportion of Snap25+ cells significantly changes with age across different brain regions (e.g., from the FACS plot).

6. Statistical analysis for in situ spatial transcriptomics RSTE1 requires further clarification.
a) In lines 887-890, it is mentioned that for each condition, there are 2 replicate brains and 2 technical replicates per brain, resulting in a total of 4 independent replicates per condition. However, it is important to note that the four replicates from two brains may not be considered independent replicates since they are two pairs from two biological replicates. To improve the experimental design, it would be more appropriate to have at least three biological replicates.

b) Since there are 4 replicates per condition, and each point represents a single replicate, four dots per condition are expected in the analysis. Any deviations from this, such as three or five dots per condition in Fig4j-k and ExData Fig7D, require a clearer explanation and documentation to ensure a proper understanding of the data representation.

7. The claim of increased cellular senescence in microglia, oligodendrocytes, and OPCs with age would benefit from stronger evidence. As there is no single marker with absolute specificity for senescent cells, it is essential to investigate multiple markers simultaneously, such as p16, p21, and SASP geens, to provide more robust evidence for cellular senescence. Additionally, confirming the results of senescent-associated markers in the spatial dataset would further strengthen the findings and validate the observed changes in senescence across different brain regions.

8. Interactions between different cell types, such as neuron-astrocyte interactions in ExData Fig 8 and oligodendrocyte-microglia interactions mentioned in lines 577-578, should be validated in the spatial data or by using other appropriate methods. Similarly, potential interactions revealed by shared GO terms across different cell types, such as MHC class I in endothelial cells, VLMC, and ependymal cells, or interferon response in microglia and ependymal cells, could be further investigated by exploring the spatial data.

9. The results reported for the GO analyses in this manuscript are confusing. In 4D, the figure shows negative numbers in blue, and the legend states this is the -log10(p-value). However, the -log10(p-value) transformation should result in positive numbers. Are these the enrichment scores? As this study relies heavily on the results the GO analysis, proper reporting of these results is essential. Further, on line 880 the authors indicate they use a p-value cutoff of 0.05 for the GO analysis, however, as gene set enrichment analyses involve multiple comparisons, it is customary to report the false-discovery rate to correct for this. The authors should report an FDR or other correction for multiple comparisons or justify their methodology.

Additional comments:
1. Coloring the UMAP by sex and FACS population would help to better understand the batch compensation and the effect of the FACS plans on the data.
2. Line 333- Showing the results of the label transfer (perhaps with a UMAP) would help the reader understand the results.
3. Line 721- The authors should fully describe the methods used, especially the quality control metrics used to generate the data.
4. Line 761- The authors should describe how clusters were determined to be low quality.
5. Line 845- It is unclear how top marker genes were defined, authors should explain this.
6. Line 1219- this sentence is not clear.

Code availability- We commend the authors for including access to the BigCat tool in this analysis. However, this does not include the code used to run the analysis of these specific data and to generate these figures. The authors should provide the code to enhance reproducibility and transparency of the study.

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)
Jin and colleagues have comprehensively studied the transcriptomic changes in six major brain regions (isocortex, hippocampal formation, hypothalamus, cerebral nuclei, midbrain, and hindbrain) of young adult and aged mice.

I would like to commend the authors' efforts in revising the paper, which has now brought significant improvements to the manuscript.

My remaining comments are as follows:

Page 11, in the paragraph about tanycytes, I think it would be important also to mention that they are the cells forming the blood-brain barrier at circumventricular organs, which are characterized by the presence of highly permeable vessels with fenestrated endothelium, as summarized in PMID: 34225934.

Page 11, together with reference 65, the authors should also cite PMID: 34225958, which elaborates on the significance of tanycytic neurogenic ability in primates, including humans.

Referee #2

(Remarks to the Author)
The authors have addressed all concerns raised during this review. In its revised form, this manuscript represents a rigorous study and an important contribution to the field.

Referee #3

(Remarks to the Author)
Most points have been addressed, except for the following:

Major Point on Experimental Validation:
1. As suggested by Reviewer #1 as well, experimental validation of some key biological findings and claims is important and would elevate the impact of the study.

2. Batch Effects and Age-Biased Clusters:
Given that samples were prepared independently over five years, the observed differential expression and age-biased clusters could be influenced by batch effects. Assessing the mean gene detection per library and UMAP plots alone may not sufficiently capture this confounding variable. For instance, the age-biased oligodendrocyte-OPC cluster may result from dissection bias across different batches. Again, there is not validation of these findings. Additionally, the validation performed using Resolve spatial transcriptomic (RSTE) datasets appears arbitrary and may suffer from sampling bias, given the low number of cells per tile (fewer than two in Figure 4g).

Minor points:
1. Minor Point: Lack of Color Code in Heatmaps (Figure 3):
The heatmaps in Figure 3 lack a color code, which should be included for clarity.

2. Visualization and Quantification of Gene Expression:
As with other non-neuronal subtypes, visualization and quantification of Cdh19 and Csmd1 expression in astrocytes using in situ RSTE (Figure 3f) would be expected.

Manuscript #: 2023-07-12067

Original title:
Cell-type specific molecular signatures of aging revealed in a brain-wide transcriptomic cell-type atlas

New title:
Brain-wide cell-type specific transcriptomic signatures of healthy aging in mice

Overall Responses:

We thank the referees for their constructive and informative comments on our manuscript. The comments have been extremely helpful in our efforts to improve the manuscript. Though their opinions varied, in general the referees recognized the value of this study as a comprehensive investigation of cell-type specific transcriptomic changes in the normal aging process, providing a useful resource to the community. At the same time, the referees also raised a number of concerns regarding the robustness of our data analysis, the lack of which could compromise the significance and novelty of the work. In the revised manuscript, we have carefully addressed each of the points raised by the referees, substantially improved our computational data analysis approaches, and extensively reorganized the figures and text to demonstrate the high quality of both our data and our analysis as well as the novelty of our findings in comparison with previous studies. Detailed point-by-point responses are shown in the next section. Here we provide a summary of the major revisions we have made to the manuscript:

- Mapped all cells in our aging scRNA-seq dataset to our recently published adult whole mouse brain (WMB) cell type taxonomy and atlas (Yao et al, Nature 2023), to obtain standardized cell type annotation consistent with the reference WMB atlas. Also refined the de novo clustering of the same dataset to identify age-biased transcriptomic clusters.
- Enhanced computational analysis of the aging scRNA-seq dataset by refining the MAST method previously used and employing two additional methods, Augur and pseudo-bulk, as suggested by the referees. Evaluated the consistency of analysis results among the three computational methods.
- Generated two new Resolve spatial transcriptomic datasets from different brain regions using different probe gene sets, resulting in a total of four spatial transcriptomic datasets with a total of 128 samples. Performed extensive cell type mapping and gene expression quantification to validate the results from scRNA-seq data analysis.
- Extensively revised and reorganized all figures and associated texts around eight major sets of scientific findings: majority of age-DE genes are cell-type specific, adult neurogenesis is affected by aging, vascular cell types show diverse age-associated changes, microglia and border-associated macrophages show pro-inflammatory changes, oligodendrocytes show depleted gliogenesis and altered myelination function, especially in the hindbrain, tanycytes and ependymal cells show strong age-associated changes including increased immune responses and decreased neurogenic potential, hypothalamic neuron types involved in energy homeostasis show strong and diverse age-associated changes, specific hindbrain and retrosplenial cortex neuron types show decreased neuronal function.
- Provided a final integrative analysis to identify major themes of age-associated changes in the many neuronal and non-neuronal cell types across the brain and synthesized them under the following four major hallmarks of aging: 1) stem-cell exhaustion, 2) increased immune response and inflammation, 3) deteriorated neuronal network structure and activity, and 4) dysregulated nutrient signaling and energy homeostasis.

With these changes and improvements, we hope the editor and referees agree that, in addition to being the most comprehensive, cell-type specific study of the aging mouse brain, our data-driven discovery approach has led to significant novel insights into the molecular and cellular underpinnings of the normal aging process. We should also point out that there is only one suggestion from the referees that we choose not to follow, which is to conduct additional functional validation experiments. As it stands now, our study is already an unusually large-scale one, spanning ~6 years and involving nearly 40 co-authors. Completing genetic manipulation and functional validation in 18 months-old mice would take at least another 2 years. We hope the referees will recognize the great values our current study can already bring to the neuroscience, aging and disease research fields, through the systematically characterized, large number of cell-type specific gene expression changes in the healthy aging

brain that will enable many functional and mechanistic studies of the aging process and the interactions between aging and diseases. We would be happy to address any further comments they may have.

Point-by-point Responses:

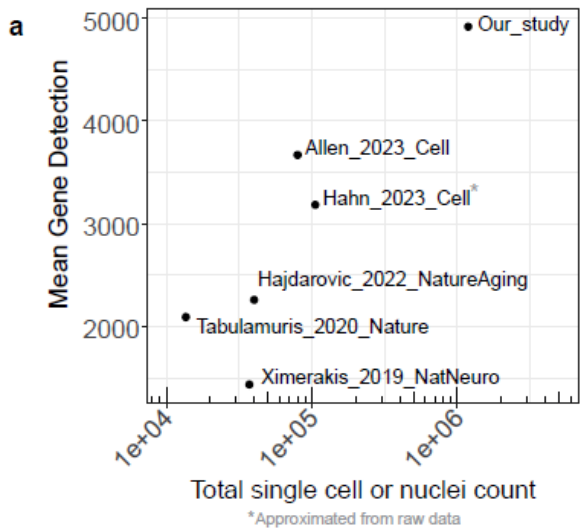
Referee #1 (Remarks to the Author):

The authors use single-cell RNA sequencing to study transcriptomic changes in six major brain regions (isocortex, hippocampal formation, hypothalamus, cerebral nuclei, midbrain, and hindbrain) of young adult and aged mice. Though this could be one of the comprehensive datasets, the study design lacks originality, as this is not the first study comparing young and aged mouse brain regions (PMID 32669714- cerebellum, cortex, hippocampus, and striatum; PMID 36285248- aging female hypothalamus; PMID 36580914 - frontal cortex and striatum). The authors cite these articles in the introduction but do not clearly discuss how this study might be different from the existing literature on the transcriptome of young and aged brains.

We respectfully disagree with the referee's comment here. Our study is not one of the comprehensive studies, but THE most comprehensive study to date on the aging mouse brain. We profiled 16 broad regions under the 6 major brain structures, covering ~110 CCF-defined brain regions. Furthermore, we mapped our scRNA-seq data from the aging brain to our high-resolution whole-mouse-brain (WMB) cell type atlas (now published: Yao et al, Nature 2023. doi: 10.1038/s41586-023-06812-z. PMID: 38092916), and therefore the cell types defined in this study (172 cell subclasses, 434 supertypes and 847 clusters) are also of the highest resolution compared to previous aging studies. Previous studies on the whole brain or very large portions of the brain lacked cell type resolution and sequencing depth to cover diverse cell types, while those targeting smaller brain regions in greater depths were under variable conditions from different labs, making it difficult to compare and integrate into a consistent view. Thus, our study with the greatest number of brain regions covered and the highest cell type resolution enabled an unprecedently comprehensive, detailed and unbiased investigation of how the brain changes with age at molecular and cellular levels. We identified age-associated gene expression changes in cell types at all levels of granularity, including a large number (total 2,449) of age-DE genes in many subclasses, supertypes or clusters, as well as a small number of age-biased clusters. And we confirmed major findings using additional spatial transcriptomic datasets. These findings systematically reveal and categorize the gene specific and cell-type specific underpinnings of four major hallmarks of aging in the brain: 1) stem-cell exhaustion, 2) increased immune response and inflammation, 3) deteriorated neuronal network structure and activity, and 4) dysregulated nutrient signaling and energy homeostasis.

We regret that we didn't make this novelty of study design clear in our original manuscript. We now provide the below tables and a new figure in the revised manuscript (Extended Data Figure 1) and shown below, to compare the scope of our study with all major previous ones in terms of number of cells profiled, regions covered, number of cell types, and quality of data (number of genes detected per cell).

Author; Journal Year	Title	Region(s) profiled	Method(s) used
Ximerakis; Nat Neurosci 2019	Single-cell transcriptomic profiling of the aging mouse brain	whole brian except for hindbrian regions (no regional dissections)	scRNAseq
Tabula Muris Consortium; Nature 2020	A single-cell transcriptomic atlas characterizes ageing tissues in the mouse	cerebellum, cortex, hippocampus, and striatum	scRNAseq
Hajdarovic; Nature Aging 2022	Single-cell analysis of the aging female mouse hypothalamus	hypothalamus	snRNAseq
Allen; Cell 2023	Molecular and spatial signatures of mouse brain aging at single-cell resolution	frontal cortex and striatum	spatial (MERFISH) + snRNAseq
Hahn; Cell 2023	Atlas of the aging mouse brain reveals white matter as vulnerable foci	MOp, Vis, Ent, Hip ant, Hip post, Hy, Th, Caud. put., Pons, Medulla, Cereb, SVZ, Chor. plx., Corp. cal., Olf	bulk RNAseq + spatial (Visium) + some snRNAseq
Jin; scRNA seq 10Xv3 (our study)	Cell-type specific molecular signatures of aging revealed in a brain-wide transcriptomic cell-type atlas	PL+ILA+ORB, AI, ACA, RSP, Hip, PPP, Ent, Hy, STRd, STRv, PAL, sAMY, PAG+RAmb, SNr+SNc+VTA, Pmot/sat-A, Pmot/sat-P	scRNAseq + spatial (Resolve)



b

Broad region	Jin; scRNA seq 10Xv3 (our study)	Hahn; Cell 2023 (bulk)	Tabula Muris Consortium; Nature 2020	Allen; Cell 2023	Hajdarovic; Nature Aging 2022
Isocortex	Yes	Yes*	Yes	Yes	
Olfactory		Yes*			
Hippocampal formation	Yes	Yes	Yes		
Striatum	Yes	Yes	Yes	Yes	
Palidum	Yes				
Thalamus		Yes*			
Hypothalamus	Yes	Yes*			Yes
Midbrain	Yes				
Hindbrain	Yes	Yes*			
Cerebellum		Yes*	Yes		
White matter	Yes	Yes*		Yes	

* bulk only

From Extended Data Figure 1. Comparison of this study to other published mouse aging scRNA-seq datasets. (a) Scatter plot of mean gene detection plotted against total cell or nucleus count from major scRNA-seq datasets published within the last 5 years. (b) Table documenting major brain structures profiled within each study. Grey cells denote regions that were profiled with bulk-seq rather than single-cell methods.supertypes.

Our study has the greatest resolution of cell types detected

Study name	Resolution (number of preliminary cell types)	Description of cell types
Our study	**139 subclasses**	**Subclasses with ncell.aged and ncell.adult >100**
Allen 2023, Cell	43 types	Similar to subclass/supertype level in AIBS taxonomy
Hahn 2023, Cell	15 regions 8 types from snRNA HIP, 8 types from snRNA CP	15 = Number of regions profiled with bulk; scRNA types are similar to subclass

As a particular example, a recent mouse aging study published in *Cell* in September 2023 (Hahn et al Cell 2023 as shown in the above plots) represents the most comprehensive published study of gene expression changes in aged brain to date. However, it only conducted bulk RNA-seq for 15 dissected brain regions, supplemented with a small set of snRNA-seq data for two of the 15 regions. Consequently, the age-related DE genes they identified could not be quantitatively associated with any specific cell types. While coarse assignment to widely distributed non-neuronal types or dominant neuronal types in a region (e.g., D1 vs D2 medium spiny neurons in striatum) was done, it is impossible to meaningfully interpret DE genes in vast majority of brain regions which contain multiple cell types. A major finding from the study was that different brain regions exhibit different aging trajectories; however, this was also confounded by the different proportions of cell types in different regions, in particular the white matter associated non-neuronal populations. Therefore overall, in comparison with the excellent Hahn et all *Cell* 2023 study, we believe that our current study represents a much more substantial and non-incremental advance in the aging field, as elaborated below.

We also used more stringent criteria to define age-DE genes. Furthermore, based on referees' suggestions below, we have now added two more computational approaches, Augur and pseudo-bulk, to further support the robustness of the genes we identify (Figure 1i; Extended Data Figure 5a-c). As discussed in more detail below, the analysis results across the three computational approaches are highly consistent with each other. This demonstrates that our findings are more solid than those from previous studies as well.

In our revised paper, we have refined the mapping of aging cells to our high-resolution WMB cell type atlas from P56 mice and pushed the analyses to supertype and cluster levels as much as the cell numbers allow, and supplemented with more spatial data, to pinpoint the cell type-specific changes with age more accurately. We have now provided more detailed comparisons between our findings, which are the most comprehensive in the largest number of cell types to date, and those from previous studies, with many cited references. In particular,

we believe that our findings of age-related gene expression changes in both neuronal and non-neuronal cell types in the hypothalamus around the third ventricle are truly novel.

After identifying age-associated differentially expressed (DE) genes in each subclass, they describe gene expression changes in major subclasses such as OPCs, oligodendrocytes, microglia, macrophages, endothelial cells, and astroependymal cells. From here, the analysis becomes redundant (Figure 3, 4, 5, and 6) for each subclass mainly with the representation of subclusters, age-DE genes, or GO enrichment analysis. Further, the authors used a commercial Molecular Cartography platform to generate two sets of in situ spatial RNA data for the following regions: cortex, striatum, midbrain, and hindbrain which they called RSTE1 experiment, and RSTE2 for the hypothalamus. However, in figures 3 and 4, the authors chose to represent gene expression changes identified from RSTE1 as violin plots (Figure 3h, Figure 4 c, j, k) while there has been no representation of the spatially resolved image (as for RSTE2 in figures 5 and 6).

We thank the referee for pointing out these issues. In the revised manuscript, we have now extensively reorganized the figures by consolidating repetitive aspects of the original figures and generating new figures each cohesively illustrating a set of key findings. For example, Figure 3 now summarizes major findings across the four major non-neuronal classes which were previously described in separate figures in the original manuscript. Figure 3 also highlights and summarizes age-biased non-neuronal clusters (clusters that are made up of greater or fewer aged cells than expected by chance). The remaining main figures now highlight key neuronal and non-neuronal populations that were of particular interest, including age-enriched mature oligodendrocytes present in the aged hindbrain (Figure 4), tanycytes and ependymal cells (Figure 5), and hypothalamic neurons involved in energy homeostasis (Figure 6).

We have now also added spatial depictions of the Resolve spatial transcriptomic data everywhere Resolve data is described and summarized all quantification of Resolve gene expression results by sample means. In our original manuscript, we were not trying to hide anything by not showing Resolve images; we just thought it was not as informative as the quantitative violin plots. But we agree with the referee that showing the in situ spatial images can be more convincing.

The manuscript in its current form is descriptive with the identification of several age-associated DE genes and prediction of GO terms. The study could potentially benefit from the investigation of uncharacterized age-biased clusters and functional validation of predicted mechanisms using molecular and genetic tools.

We would like to emphasize that this is a large-scale atlasing paper, in which we not only describe the datasets but also perform data-driven discovery in a hypothesis-free manner, like many other such papers published in Nature before. We report a number of novel findings, including identifying thousands of (not “several”) age-DE genes from specific neuronal or non-neuronal cell types in different brain regions, associating them (through GO terms) with major hallmarks of aging (as mentioned above), and identifying the hypothalamic region surrounding the third ventricle that regulates food intake, nutrient sensing and energy homeostasis as a hotspot for aging. We acknowledge that in our original manuscript, we did not do a good job in clearly demonstrating these points. We have now put in great efforts in the revised manuscript to organize the Results sections and figures around each of the key findings, with hopefully a clearer narrative.

We believe that functional validation of our findings is beyond the scope of the current study. Performing such experiments using molecular and genetic tools on aging mice would easily take at least two years, and thus substantially delay the publication of our study, diminishing its impact to accelerate research in the field. In our opinion, the right way to do functional validation in this case should be a similar scope of systematic characterization of cell types across the brain following certain functional manipulation, in order to obtain a comprehensive picture of how different cell types change compared to normal aging, rather than cherry-picking a particular positive result as “validation”. To perform such systematic characterization, it will take substantial amount of time, efforts and resources, which is why we think it should belong to a separate study.

Referee #2 (Remarks to the Author):

A. Summary of the key results

This study investigates cellular and molecular differences between the young adult and normally aged mouse brains using single-cell RNA-seq and highly multiplex RNA fluorescence in situ hybridization. The researchers profiled genome-wide RNA expression in ~1.2 million cells from young adult and aged male and female mice, representing ~110 brain regions, and clustered these cells by transcriptomic similarity into over 130 molecular subclasses. Comparing cell populations at multiple levels, researchers identified generalized and, more frequently, cell-type specific changes in gene expression. The greatest gene expression change was found to be among the non-neuronal populations, including pro-inflammatory changes in oligodendrocytes and microglia and a decrease in extracellular matrix related gene expression in vascular and smooth muscle cells. Regionally, the cells around the lower third ventricle which express the transcription factor Tbx3, including neurons of the arcuate hypothalamus and tanycytes, were particularly susceptible to age-related changes in gene expression. This is of particular interest given the role of these cells in energy balance and the well-characterized changes in energy balance that occur with aging. Overall, this study provides an immense and rich resource of the molecular architecture of the mouse brain and how it changes with normal aging and so is likely to be of great utility to the neuroscience and aging research fields.

We thank the referee for the very positive comments and the recognition of the great utility of our study as an immense and rich resource to the neuroscience and aging research fields.

B. Originality and significance: if not novel, please include reference
The results of this study are original and significant.

C. Data & methodology: validity of approach, quality of data, quality of presentation
The approach is generally valid (see concerns raised below), and the data and presentation are of high quality.

D. Appropriate use of statistics and treatment of uncertainties
The use of statistics and treatment of uncertainties is generally appropriate, but please see concerns raised below in section F.

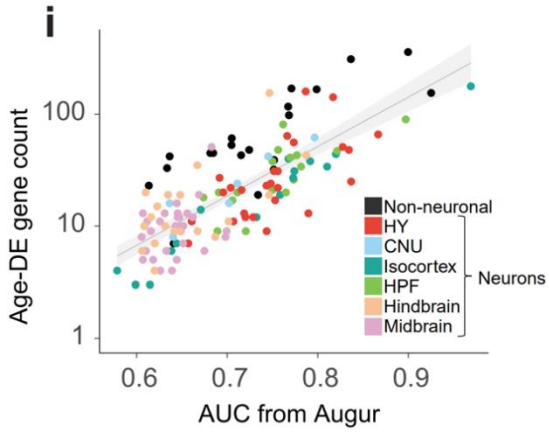
E. Conclusions: robustness, validity, reliability
Please see section F below.

F. Suggested improvements: experiments, data for possible revision
There are several key points in the manuscript lacking sufficient evidence or clarity, which need to be addressed in revision. Listed below are the concerns raised during this review, roughly in order of significance:

• To detect genes significantly affected by aging, the authors relied on a single-cell differential expression (DE) statistical method called MAST (Model-based Analysis of Single-cell Transcriptomics). However, recent work has demonstrated that MAST, like many other popular methods for single-cell DE analysis, is highly prone to false-positive detection of DE genes (PMID: 34584091). Pseudo-bulk methods (e.g., edgeR-LRT), in which cell transcriptomes are aggregated by biological replicate, are instead recommended for single-cell DE analysis. While the authors used stringent cutoffs ($p<0.01$, effect size > 1) for significance in the MAST results, the conclusions of this study would be greatly strengthened by a pseudo-bulk DE analysis, especially for the cells focused on figures 3-6.

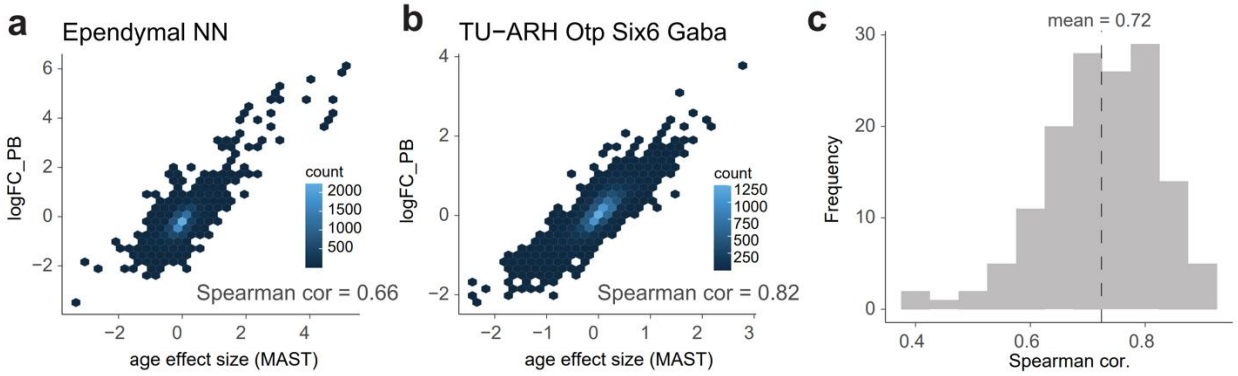
We thank the referee for the great suggestion. We have now performed both pseudo-bulk and Augur DE gene analyses as the referee suggested. We find that our original findings using the stringent cutoffs recognized by the referee here are validated by the new analyses of both pseudo-bulk and Augur, as shown in Figure 1i, Figure 3c, Figure 6a, and Extended Data Figure 5a-c (also see below). This really strengthens the robustness of our findings and leads to more findings that we are now confident to report. We have added these new results into the revised manuscript, and we summarize them below.

The below plot shows high correspondence between our original MAST results (nDEgenes; log scale y axis) and Augur scores (x axis) for all cell subclasses analyzed. The subclasses with the highest values in both methods (in the top-right portion of the panel) are the ones we report in detail in our manuscript. Additional plots in the revised manuscript show more detailed correspondences between MAST and Augur for non-neuronal and immature neuronal subclasses (Figure 3c) and neuronal subclasses (Figure 6a), respectively.



From Figure 1i. Correspondence between the age-DE gene count from MAST (y axis, log scale) versus the AUC score calculated from Augur (x axis) for all neuronal subclasses and non-neuronal supertypes.

The below plots show high correspondence between effect sizes calculated from MAST and pseudo-bulk methods for 2 example subclasses (Extended Data Figure 5a,b), and a summary of spearman correlations from all subclasses in the study (Extended Data Figure 5c).



From Extended Data Figure 5a-c. **(a-b)** Select two-dimensional density plots comparing the log(fold change) per gene calculated from a pseudo-bulk (PB) method (y axis) versus the age effect size (comparable to log2(FC); see Methods) estimated by MAST (x axis) for two subclasses: ependymal cells (a) and hypothalamic neuron subclass TU-ARH Otp Six6 Gaba (b). **(c)** Distribution of spearman correlation coefficients from all pseudo-bulk to MAST comparisons (such as those visualized in a and b) for all subclasses tested (n = 146).

• The authors state that 96 mice were used to generate the single-cell RNA-seq data and with 1-6 dissected tissue samples from each mouse. However, it is unclear how many mice comprise each cell subclass. In the “Replication” section of the “Reporting Summary,” the authors state that at least two mice were used. However, if one mouse was adult and the other was aged, then this would be n=1 for DE analysis of that cell subclass. In the interest of generalizability, it would help to know how many male mice, female mice, adult mice, and aged mice are represented in each of the cell subclasses. This could be added to the supplemental table of subclasses and their cell and gene content.

We have added the numbers of donors and number of libraries to each subclass in Supplementary Table 2. We have also split the numbers by sex and age. The mean number of donors per age for all subclasses is 14 ± 10. There is only one subclass, i.e., DG-PIR Ex IMN (standing for dentate gyrus and piriform cortex immature neurons), with only 1 aged donor; we believe this indicates the scarcity of the DG-PIR IMN cells in the aged brain, likely due to age-associated decline in adult neurogenesis. There is one other subclass (L2/3 IT ENT Glut) that has 2 young adult donors; for all other cases, there are at least 4 young adult or aged donors in each subclass. We have clarified this point in the first section of Results of the revised manuscript (lines 165-175).

• A major conclusion in the study is that cell populations differ greatly in their susceptibility to the effects of aging on gene expression, measured as differences in the number of aging differentially expressed genes. However, the extent to which these differences in aging DE genes reflect differences in cell numbers is unclear, since more

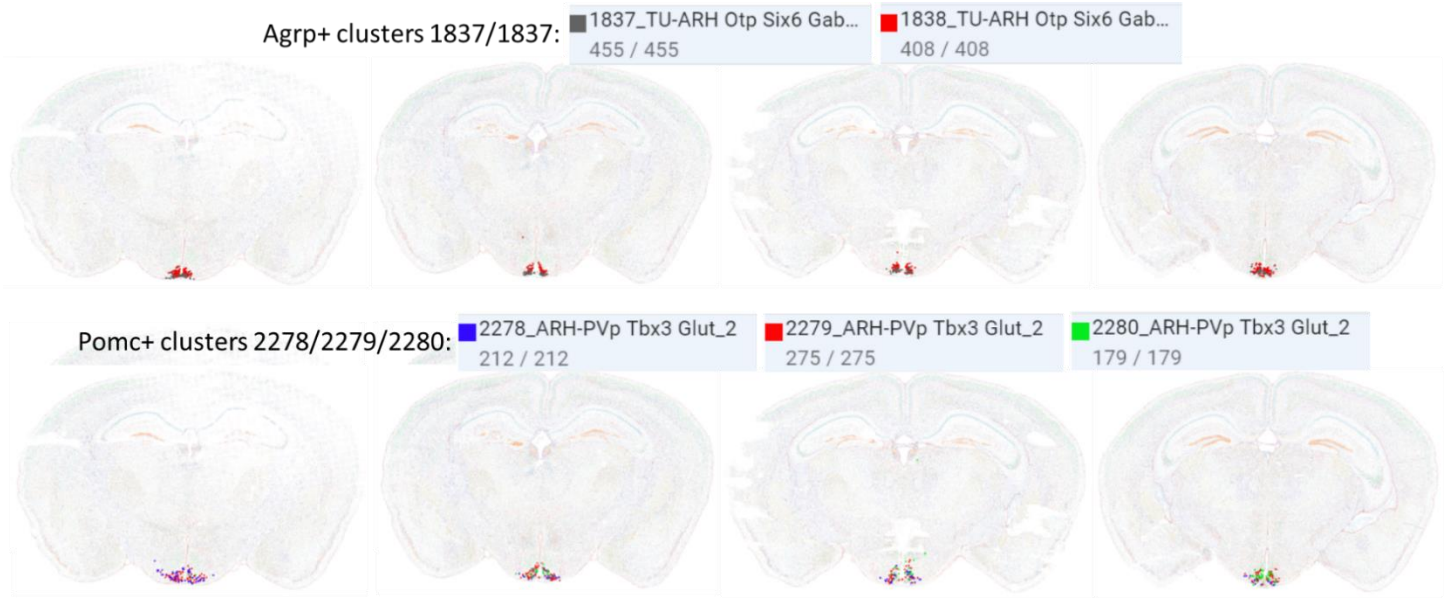
populous clusters/subclasses would have more statistical power to detect DE genes. The authors could strengthen their conclusion by plotting the number of DE genes by cluster/subclass size (n cells) and identifying the outliers, or by applying a method for detecting sensitive cell types, such as Augur (PMID: 32690972)

We thank the referee for the great suggestion. As mentioned above, we have now used Augur to analyze our entire dataset and the results are consistent with those using the MAST method (Figures 1i, 3c, 6a). We have now also shown the plot of number of DE genes by number of cells in each subclass, which shows no correlation with each other (R^2 = 0.007), demonstrating that the ability to detect DE genes is not due to cell numbers (Extended Data Figure 5d). For example, the numbers of tanycytes and hypothalamic ARH neurons are relatively small and yet we have detected many robust age-DE genes at subclass level (Figures 3c, 6a), even further down to cluster level for both tanycytes (Figure 5, Extended Data Figure 9) and ARH neurons (Figure 6d-m, Extended Data Figure 11c).

• Line 526-528 states “In our Tbx3+ clusters, cluster 63_TU-ARH Otp Six6 Gaba shows highly specific expression of Npy and Agrp, while cluster 8_ARH-PVp Tbx3 Glut shows specific expression of Pomc”. The data presented suggest that cluster 63 and cluster 8 may correspond to AgRP neurons and POMC neurons. However, the spatial data indicate that TU-ARH Otp Six6 Gaba neurons are located in both the arcuate nucleus and the tuberal nucleus, though AgRP neurons are only found in the arcuate nucleus. Is the TU-ARH Otp Six6 Gaba cluster a heterogeneous cluster containing AgRP neurons and tuberal neurons? If so, this would affect the interpretation of the age-DE genes, since the signal could be driven by either AgRP neurons or tuberal neurons To address this issue, authors could perform AgRP neuron and POMC neuron label transfer from one of the hypothalamic single-cell molecular atlases they cite (Chen et al., 2017 or Campbell et al., 2017) or from the HypoMap (PMID: 36266547). Also, feature plots of Agrp, Npy, Pomc on the HY UMAP may be helpful in illustrating whether clusters 63 and 8 are purely AgRP neurons and POMC neurons, respectively.

We thank the referee for pointing out this issue which we neglected to clarify in our original manuscript as we only labeled cells in the Resolve images at subclass level. First, we would like to note that the cluster numbers between the previous and current versions of the manuscript have changed, and what was previously labeled as 8_ARH-PVp Tbx3 Glut is now 379_ARH-PVp Tbx3 Glut_2 and 63_TU-ARH Otp Six6 Gaba is now 336_TU-ARH Otp Six6 Gaba_1 in our revised manuscript.

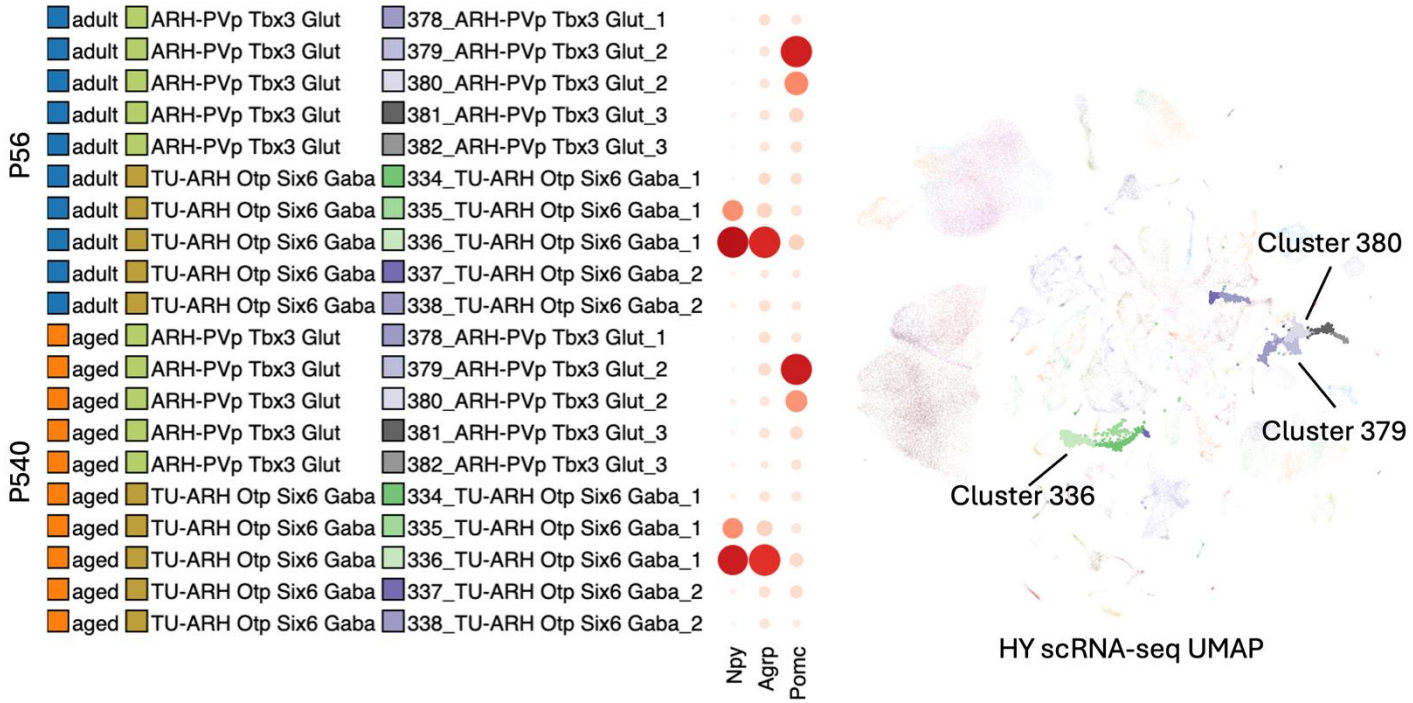
Indeed, in our WMB cell type atlas (Yao et al, Nature 2023; note this has a different set of cluster labels that do not correspond to those presented in this aging manuscript), Agrp/Npy is found in subclass 104 TU-ARH Otp Six6 Gaba (clusters 1837, 1838 with the strongest expression, out of a total of 24 clusters in the subclass), and Pomc is found in subclass 126 ARH-PVp Tbx3 Glut (clusters 2278, 2279, 2280, out of a total of 11 clusters in the subclass). These Agrp+ and Pomc+ clusters have very tight expression in ARH only, as shown in our whole-brain MERFISH data plotted here.



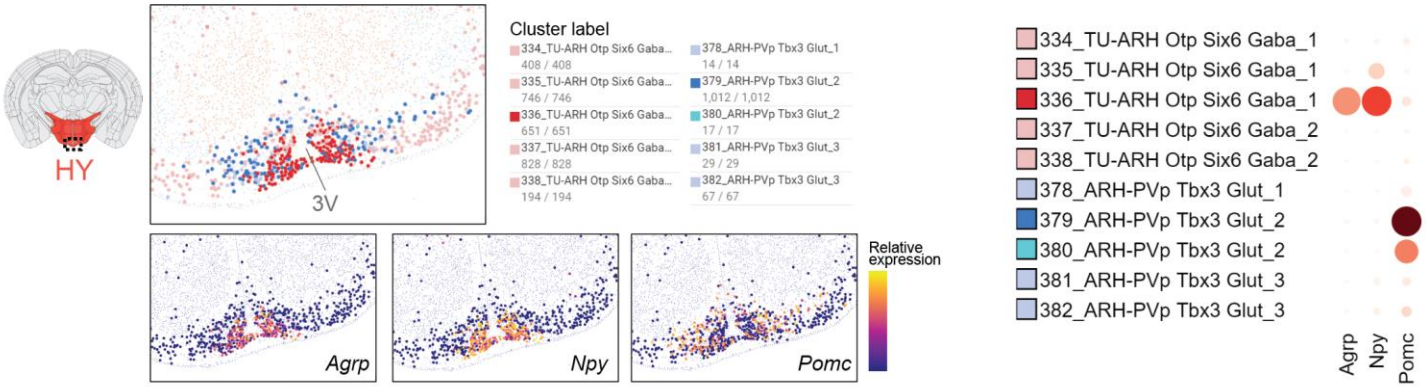
(cluster labels here are unique to ABC-WMB atlas)

We have now mapped all clusters in our aging study to the WMB atlas. We also find that the new clusters in our aging study dataset have dominant expression of Agrp/Npy in cluster 336 (corresponding to WMB clusters 1837/1838) and Pomc in cluster 379 and 380 (corresponding to WMB clusters 2278/2279 and 2280, respectively), in both scRNA-seq data shown in the UMAPs and Resolve sections below. Consistent with the MERFISH data above, the Resolve data below shows that cluster 336 colocalizes with Agrp/Npy in the medial part of ARH (also see Figure 6f), and cluster 379 colocalizes with Pomc in the lateral part of ARH.

scRNAseq gene expression



Resolve *in situ* gene expression



• Line 548-551 “Included in the circadian and rhythmic process related genes, we observed Bhlhe40, Bhlhe41, Nr1d2, and Per3 increasing with age only in the Agrp+ cluster (Figure 6k; Supplementary Table 3), suggesting that temporal and rhythmic control of behaviors like feeding, a known function of Agrp+ neurons76, may become altered with age.” Was the adult and aged hypothalamic tissue collected at the same time of day? The circadian timing of tissue collection can influence expression of clock genes in the hypothalamus.

Yes, all brain tissues were collected at the same time of day, i.e., the morning of each day. Also, all mice used in this study were housed under the same light/dark cycle. Therefore, there is no circadian timing difference between young adult and aged mice for the majority of cells. The only exception to this is the tanycytes, for which we included additional cells collected from the reverse light/dark cycle P56 libraries to boost our cell numbers.

Light/dark cycle did not appear to show significant effect on transcriptome signal for the tanycytes (Extended Data Figure 9b).

• Line 782: “For mapping to the oligodendrocyte dataset from Marques et. al.30, we used a list of 195 genes.” Additionally, “For both external datasets, gene lists were assembled based on prominent marker genes from each external reference cluster. When mapping confidence score was needed, we sampled 80% genes from the marker list randomly, and performed mapping 100 times. We define the fraction of times a cell is assigned to a given cell type as the mapping probability to that type.” Will the authors provide code to replicate the marker lists? Or were the genes selected manually for the marker list?

We have created https://github.com/AllenInstitute/mouse_aging_scRNAseq to make available all the codes used for analysis, including code used to select marker lists.

• Line 690 states “We collected 1,508,284 cells without performing FACS.” How many cells were collected without FACS, the balance of the 2,777,165 cells?

About ½ of the final dataset are made up of libraries collected with “unbiased methods” (i.e., Not sorted for neurons), which include libraries stained and sorted for Hoechst+, Hoechst+/Calcein+, or unsorted (No FACs). About ¼ of the final dataset are made up of “No FACs” cells. This has been clarified in the Methods section. The proportion of each subclass that is made up of unbiased cells has also been added to Supplementary Table 2.

• Figure 3h plots expression of Hopx3 but this gene is not mentioned in the results.

This has been revised in the text. Oligodendrocytes are now discussed in Figure 4, and the Hopx gene is shown in Figure 4e,h.

• For the UMAP in Figure 6b, do the colors in 6c serve as the color key?

The original Figure 6b and 6c are now new Figure 6b and Extended Data Figure 11c. We have clarified the color legends for the UMAPs shown in Figure 6b.

• For the dot plot in Figure 6i, many of the dots are larger than the “100%” reference dot. This may be true of other plots. The reference dots should be re-generated to ensure they are correctly scaled to the data.

We have fixed this and other similar figures.

G. References: appropriate credit to previous work?
Yes.

H. Clarity and context: lucidity of abstract/summary, appropriateness of abstract, introduction and conclusions
Yes, the writing of the manuscript and appropriateness of these sections are excellent.

Referee #3 (Remarks to the Author):

In this study by Jin et al, the authors performed single-cell transcriptomics profiling of approximately 1.2 million high quality brain cells across 16 broadly dissected regions from 52 young (2-month-old) and 44 aged (18-month-old) male and female mice. Cells from the young animals were previously sequenced and annotated in Yao et al. 2023 and served as a reference for annotating cells from both aged animals and Resolve spatial data, which profiled 100 genes of interest in five different regions of interest. The study’s analysis revealed that the greatest expression changes with age occur in non-neuronal cells (OPCs and oligodendrocytes, immune cells, and vascular cells), as well as cells surrounding the third ventricle of the hypothalamus (tanycytes and Tbx3+ neurons), and hypothalamic Agrp+ and Pomc+ neurons.

Overall, the experimental approaches are valid, the data is of good quality. The manuscript is well-written and provides a large and comprehensive dataset with spatial information for the aging research community. However,

despite the substantial amount of work, the analysis is relatively shallow, and lacks rigor in some areas. Further, the biological findings and claims are not backed up with experimental validation. Finally, while this is a useful resource, in this reivewer's view, the impact of the findings does not reach the highest level. Specific concerns and suggestions regarding the experiments to enhance the impact of the work are provided below.

We thank the referee for the positive comments on the large scale and comprehensiveness of our study which results in a useful resource to the aging community. We greatly appreciate their suggestions below in helping to strengthen the depth and rigor of our analysis and the impact of the findings. We are glad to say that we have now performed all these additional analyses, and thus the rigor and impact of our study is demonstrated (detailed below).

As mentioned above in our response to referee 1, we think that functional validation will involve extensive experimental work and should belong to a separate study in the future.

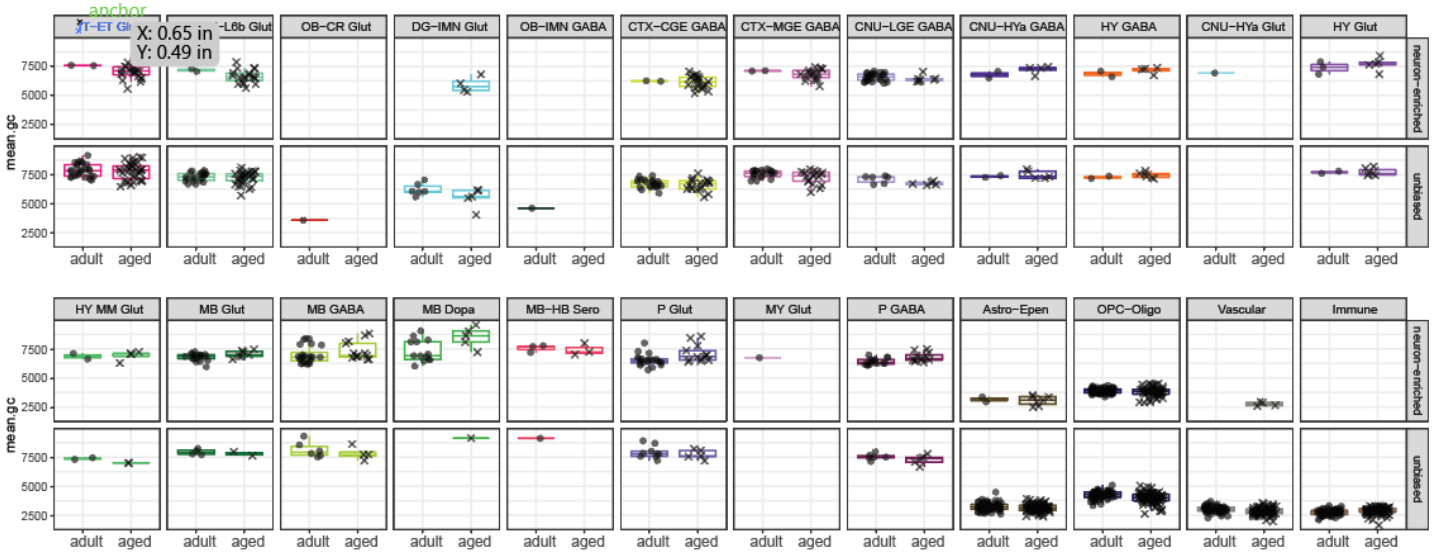
1. One of the major claims of the paper is that the greatest gene expression changes are found in non-neuronal cell-types. However, several choices in the analysis of the dataset may affect this finding. Specifically, on Line 798 the authors state "Only subclasses with at least 50 aged and 50 adult cells were evaluated for DE genes. To decrease running time, for large subclasses, we subsampled them to a maximum of 1,000 cells per age." In subclasses with smaller numbers of cells, the analysis may be underpowered compared to subclasses with 1000 cells analyzed per group, and so it is difficult to call one group more or less changed with age.

In our study, we identified 172 neuronal and non-neuronal subclasses and conducted analysis for 132 neuronal subclasses and 20 non-neuronal supertypes (contained within 14 non-neuronal subclasses), obtaining a total of 2,449 highly significant age-DE genes that are distributed widely across cell types (lines 194-199, Figures 1h-l, 3a-b, Supplementary Tables 2 and 3). As mentioned above in our response to referee 2, we have now use two additional computational methods, Augur and pseudo-bulk, to analyze our data, and obtained consistent results across all three methods (Figures 1i, 3c, 6a, Extended Data Figure 5a-c). In particular, the Augur and pseudo-bulk methods are not biased by sample size. We have also added a plot with the number of DE genes versus the number of cells per subclass, to demonstrate that the number of age DE genes detected in our study does not correlate with the number of cells in each group (Extended Data Figure 5d).

2. It is unclear whether the young samples (published in Yao et al. 2023) were prepared side-by-side with the aged samples, or whether these samples were prepared separately. No consideration for batch or batch-effects was made in this paper, which is concerning as batch affects could introduce variability. An analysis of batch-to-batch variation is important.

Samples from each animal (regardless of age) were prepared independently across a period of 5 years with collection of libraries from the two ages intermingled over time using Allen Institute's standard pipeline with standardized QC metrics to minimize batch-to-batch variations using the same sampling conditions as described in ABC-WMB atlas (Yao et al, 2023; Nature). When we examine the mean gene detection per library split by major library prep method (rows in figure below), there is not a significant difference between aged and adult values across different cell classes (columns in figure below). These values are also plotted across all cells as density distributions in Extended Data Figure 2.

We have now also included UMAPs colored by additional meta data information including sex, library, external donor, and QC score in Extended Data Figure 5f to provide additional insight into the effect that these variables have on this dataset.



Mean gene detection per library and cell class

3. The statistical reporting is unclear throughout the paper.
• For example, in Figure 3C (line 1195), the manuscript states “asterisks denote statistical significance” but no information for about what test was performed, the number of samples, or significance cut offs are listed. Similarly, in 3D, p-values are reported, but no information about how these values were calculated is given. In Extended figure 4E, it is not clear what test was performed or what is being compared.
• Many plots are missing appropriate labels, for example, the violin plots in 3D and 3H have no label indicating what the numbers on the y axis are.
• Many times, a p-value is reported where it is likely that an adjusted p value should be used, such as in 3D, 3H, 5E. Clearer statistical reporting in figure legends would lend clarity. On line 809, the authors indicate p values for the differential expression analysis are adjusted using the Bonferroni method, and this should be indicated throughout the figures and figure legends.

We in fact used the same statistical method and criterion for all the analyses and described the statistical method in the Methods section in our original manuscript. We apologize for the inconsistently simplified wording in other parts of the text and figure legends, and we have now corrected it in all the places.

4. The authors did not mention whether variables such as Estrous cycle and Zeitgeber time for tissue collection were controlled. These variables could have significant effects on hypothalamic cells, such as tanycytes, Agrp and NPY neurons (https://doi.org/10.1073/pnas.0904747106), as they behave differently across different stages of the estrous cycle. Additionally, circadian-related genes mentioned in lines 548-549 can fluctuate across different Zeitgeber times. If these variables were not tightly controlled between the young and aged samples, the age-associated differential expression (DE) signatures could be confounded by variations in estrous cycles and Zeitgeber times.

The Zeitgeber time of the light/dark cycle was similar (within a 3-hour window) for all the tissue collections. But we did not keep track of the estrous cycle for female mice. We have stated this explicitly in the Methods and acknowledge that estrous cycle is a potential source of variation. However, it is worth noting that in our study, male and female cell types show the same age DE genes in nearly all the cases (except for a trend of sex difference in hypothalamus which was described in our original manuscript). Therefore, it is unlikely that our findings are due to variability in estrous cycle. We have also now included sex in the MAST statistical model that should help account for variability due to sex.

5. Concerns with the FACS population analysis. In the analysis of adult/aged neuronal cell proportions shown in Fig 2 and Fig 6i, the authors should investigate whether the FACS scheme acts as a confounding factor. This could involve assessing whether FACS-enriched neurons exhibit a similar neuronal subclass/cluster composition as the non-FACS neurons. Additionally, it would be important to determine if the proportion of Snap25+ cells significantly changes with age across different brain regions (e.g., from the FACS plot).

We have added proportions of unbiased sampling cells (this includes No FACS, Calcein+, Calcein+/Hoechst+ cells), as well as proportion of Snap25+ cells to each subclass in Supplementary Table 2. As discussed in the Methods section of our original manuscript, we indeed found that FACS condition showed an effect (but minor) on library signal, as indicated by a few differences in mean QC scores (Extended Data Figure 5e). To account for this, we included QC score as a covariate in our MAST model for identifying DE genes. We would like to emphasize again that the findings reported in our revised manuscript are now demonstrated using three different computational approaches.

6. Statistical analysis for in situ spatial transcriptomics RSTE1 requires further clarification.
a) In lines 887-890, it is mentioned that for each condition, there are 2 replicate brains and 2 technical replicates per brain, resulting in a total of 4 independent replicates per condition. However, it is important to note that the four replicates from two brains may not be considered independent replicates since they are two pairs from two biological replicates. To improve the experimental design, it would be more appropriate to have at least three biological replicates.
b) Since there are 4 replicates per condition, and each point represents a single replicate, four dots per condition are expected in the analysis. Any deviations from this, such as three or five dots per condition in Fig4j-k and ExData Fig7D, require a clearer explanation and documentation to ensure a proper understanding of the data representation.

Please note that we actually had two replicate brains per sex, and thus a total of four independent biological replicates. We pooled male and female brains together because we were not claiming any sex-specific findings. We have clarified this in the revised manuscript, with the exact sampling plan shown in Extended Data Figure 4 and in Methods.

If there are <4 dots, that is because of low-quality images, or not enough cells of a type detected in that particular image. As mentioned above, we have now provided Resolve images along with the quantification plots summarized by sample in all relevant figures.

7. The claim of increased cellular senescence in microglia, oligodendrocytes, and OPCs with age would benefit from stronger evidence. As there is no single marker with absolute specificity for senescent cells, it is essential to investigate multiple markers simultaneously, such as p16, p21, and SASP geens, to provide more robust evidence for cellular senescence. Additionally, confirming the results of senescent-associated markers in the spatial dataset would further strengthen the findings and validate the observed changes in senescence across different brain regions.

We agree with the referee on this point. We were not able to confidently identify more than one senescence related DE genes in a given cell type and we did not profile enough genes in our spatial dataset to make any strong claims regarding coordinated cellular senescence. Therefore, we have now removed these claims from the paper.

8. Interactions between different cell types, such as neuron-astrocyte interactions in ExData Fig 8 and oligodendrocyte-microglia interactions mentioned in lines 577-578, should be validated in the spatial data or by using other appropriate methods. Similarly, potential interactions revealed by shared GO terms across different cell types, such as MHC class I in endothelial cells, VLMC, and ependymal cells, or interferon response in microglia and ependymal cells, could be further investigated by exploring the spatial data.

We thank the referee for the great suggestion. We have now provided extensive spatial quantification of cell types and genes throughout figures. We have found that there appears to be a significant age-dependent interaction with BAM and ventral tanycytes in the third ventricle and have included this finding in the revised manuscript (Figure 5l,m).

9. The results reported for the GO analyses in this manuscript are confusing. In 4D, the figure shows negative numbers in blue, and the legend states this is the -log10(p-value). However, the -log10(p-value) transformation should result in positive numbers. Are these the enrichment scores? As this study relies heavily on the results the GO analysis, proper reporting of these results is essential. Further, on line 880 the authors indicate they use a p-value cutoff of 0.05 for the GO analysis, however, as gene set enrichment analyses involve multiple

comparisons, it is customary to report the false-discovery rate to correct for this. The authors should report an FDR or other correction for multiple comparisons or justify their methodology.

We apologize for the confusions. All of the figures demonstrating GO analysis results have now been replotted and most figures showing negative numbers have been replaced, except for the summary plot in Figure 7a. These numbers represent a GO term significance score, for which the absolute value is equivalent to the -log10(p-value) of the term (i.e., higher is more significant), and the sign of the value (positive or negative) demonstrates the direction of the change (positive == increasing with age; negative == decreasing with age). This has now been described more clearly in the figure legends and Methods text.

GO analysis results were corrected for multiple testing using the g:SCS method that is the default method implemented by g:Profiler2 (the R package that was used for enrichment, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6602461/), and these are the p-values that are reported in all the relevant figures of the paper. Briefly, this method accounts for the dependency of multiple tests in the context of enrichment analysis by taking into account the overlap of functional terms. It is more conservative than FDR, but less strict than Bonferroni correction. We have added this explanation to the Methods section.

Additional comments:
1. Coloring the UMAP by sex and FACS population would help to better understand the batch compensation and the effect of the FACS plans on the data.
2. Line 333- Showing the results of the label transfer (perhaps with a UMAP) would help the reader understand the results.
3. Line 721- The authors should fully describe the methods used, especially the quality control metrics used to generate the data.
4. Line 761- The authors should describe how clusters were determined to be low quality.
5. Line 845- It is unclear how top marker genes were defined, authors should explain this.
6. Line 1219- this sentence is not clear.

We have addressed all these minor points. We have now added additional UMAPs colored by meta data to Extended Data Figure 5f. We have added illustrative panels to Figure 1c-d that are meant to clarify our label transfer and analysis approaches. We have expanded on the details of our QC metrics in the Methods section.

Code availability- We commend the authors for including access to the BigCat tool in this analysis. However, this does not include the code used to run the analysis of these specific data and to generate these figures. The authors should provide the code to enhance reproducibility and transparency of the study.

All analysis code has now been deposited to https://github.com/AllenInstitute/mouse_aging_scRNAseq.

Manuscript (2023-07-12067A-Z) title:
Brain-wide cell-type specific transcriptomic signatures of healthy aging in mice

Response to referees

Referee #1 (Remarks to the Author):

Jin and colleagues have comprehensively studied the transcriptomic changes in six major brain regions (isocortex, hippocampal formation, hypothalamus, cerebral nuclei, midbrain, and hindbrain) of young adult and aged mice.

I would like to commend the authors' efforts in revising the paper, which has now brought significant improvements to the manuscript.

My remaining comments are as follows:

Page 11, in the paragraph about tanycytes, I think it would be important also to mention that they are the cells forming the blood-brain barrier at circumventricular organs, which are characterized by the presence of highly permeable vessels with fenestrated endothelium, as summarized in PMID: 34225934.

Page 11, together with reference 65, the authors should also cite PMID: 34225958, which elaborates on the significance of tanycytic neurogenic ability in primates, including humans.

We appreciate this referee's positive comments about our revised manuscript which has shown significant improvements. We have incorporated the referee's suggestion here, by adding a sentence about CVOs (page 10, lines 375-376) and the two references PMIDs 34225934 and 34225958 (new ref numbers 60 and 62).

Referee #2 (Remarks to the Author):

The authors have addressed all concerns raised during this review. In its revised form, this manuscript represents a rigorous study and an important contribution to the field.

We thank this referee's very positive comments about the important contribution our study could make to the field.

Referee #3 (Remarks to the Author):

Most points have been addressed, except for the following:

Major Point on Experimental Validation:
1. As suggested by Reviewer #1 as well, experimental validation of some key biological findings and claims is important and would elevate the impact of the study.

We acknowledge that the key biological findings from our study will need experimental validations. Given that there are many findings from many neuronal and non-neuronal cell types in the brain (as summarized in the new Figure 6), it will take many efforts and substantial amounts of time and resources to conduct these validation experiments. We should also be cautious about the interpretation of the results from any single validation experiment due to the many cell types and genes involved and their interactions in the aging process. Thus, we believe that extensive experimental validations should be carried out in the future by us and others in the field, and we have stated this in the Discussion section (page 17, lines 650-651).

2. Batch Effects and Age-Biased Clusters:
Given that samples were prepared independently over five years, the observed differential expression and age-biased clusters could be influenced by batch effects. Assessing the mean gene detection per library and UMAP plots alone may not sufficiently capture this confounding variable. For instance, the age-biased oligodendrocyte-OPC cluster may result from dissection bias across different batches. Again, there is not validation of these findings. Additionally, the validation performed using Resolve spatial transcriptomic (RSTE) datasets appears arbitrary and may suffer from sampling bias, given the low number of cells per tile (fewer than two in Figure 4g).

We agree that batch effects are a major issue in the single-cell genomics studies that need to be addressed to ensure the robustness of the outcomes and conclusions of the studies. In this manuscript, we have put in great efforts, both experimentally and computationally, to overcome batch effects and ensure the validity and robustness of the results and findings we report. As described in the manuscript, we generated the scRNA-seq dataset from 284 libraries from 96 mice. Each of the 16 regions of interest (ROIs) was profiled with 10-31 libraries, sex and age balanced. We used the most stringent QC criteria in the field (compared to other published studies) and applied QC both before and after clustering to filter out low-quality single-cell transcriptomes from a total of 2,777,165 cells profiled. After clustering, we also excluded small populations of out-of-ROI clusters, as well as any clusters that were composed of >80% of cells from a single library so donor/library bias was minimized. These QC and exclusion processes led to a final of 1,162,568 high-quality single-cell transcriptomes for analyses, and they were classified into 172 cell subclasses and 847 cell clusters. The average number of donor mice per age for all subclasses is 14 ± 10 (mean ± SD). We used three independent computational methods (MAST, Augur and pseudo-bulk) to analyze age-associated gene expression changes and showed that the results from the three methods are highly consistent with each other. In MAST, we used donor and QC metrics as covariates in modeling, and we used a stringent cutoff equivalent to log2 fold change > 1 (i.e., more than 2-fold change), plus a significant adjusted p value, to call age-DE genes. With all these measures, we believe we have demonstrated that our results and findings are robust and convincing.

Specifically, regarding the age-biased oligodendrocyte clusters, we have now added a bar to panel e of the new Fig 3 (corresponding to the old Fig 4) to show the number of donor mice for each cluster. For the two aging-enriched MOL clusters, 816 and 819, there are 33 and 18

donors contributing to each cluster, respectively. And for the 8 aging-depleted transitional oligodendrocyte clusters (806-813), there are 41-96 donors for each. As such, we think it is unlikely the age-biased nature of these clusters is due to donor or dissection bias.

Each of the RSTE experiments also has multiple biological replicates, and each imaged area is large. Note that each tile imaged by Resolve and used as a unit for our quantification of cell densities (e.g., those shown in Fig 3d and 3g) is 294 µm x 294 µm. We have now added the range of number of tiles for each imaged area to Extended Data Figure 5, which shows that a single image sample captured by Resolve ranged between 15-22 tiles for RSTE1, 14-37 tiles for RSTE2, 9-66 tiles for RSTE3 (varied between the two ROIs), and 53-67 tiles for RSTE4. So while Fig 3g indeed shows a relatively low number of cells per tile, each sample captures 15-22 tiles, so we think the sampling is quite robust.

Minor points:

1. Minor Point: Lack of Color Code in Heatmaps (Figure 3):
The heatmaps in Figure 3 lack a color code, which should be included for clarity.

We note that the heatmaps in original Fig 3 did have color legends. Fig 3 has now been changed to Fig 2, and we have made sure color legend for each heatmap is shown.

2. Visualization and Quantification of Gene Expression:
As with other non-neuronal subtypes, visualization and quantification of Cdh19 and Csmd1 expression in astrocytes using in situ RSTE (Figure 3f) would be expected.

These two genes were not included in any Resolve experiment. And thus we can only show their quantification in the scRNA-seq data (as shown in the new Fig 2f) but not Resolve data.