nature research

Peer Review Information

Journal: Nature Genetics
Manuscript Title: Genetic analysis of the human microglia transcriptome across brain regions, aging and disease pathologies
Corresponding author name(s): Dr Towfique Raj

Reviewer Comments & Decisions:

Decision Letter, initial version:

1st Mar 2021

Dear Dr Raj,

First of all, I can't thank you enough for bearing with me. I am sorry this decision has taken so long.

Your Article, "Atlas of genetic effects in human microglia transcriptome across brain regions, aging and disease pathologies" has now been seen by 3 referees. You will see from their comments below that while they find your work of interest, some important points are raised. We are interested in the possibility of publishing your study in Nature Genetics, but would like to consider your response to these concerns in the form of a revised manuscript before we make a final decision on publication.

To guide the scope of the revisions, the editors discuss the referee reports in detail within the team, including with the chief editor, with a view to identifying key priorities that should be addressed in revision and sometimes overruling referee requests that are deemed beyond the scope of the current study. In this case, we considered all the reviewers comments to be reasonable and we would expect you to address all of them in full. Please do not hesitate to get in touch if you would like to discuss these issues further.

We therefore invite you to revise your manuscript taking into account all reviewer and editor comments. Please highlight all changes in the manuscript text file. At this stage we will need you to upload a copy of the manuscript in MS Word .docx or similar editable format.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your manuscript:

*1) Include a "Response to referees" document detailing, point-by-point, how you addressed each referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the referees along with the revised manuscript.

*2) If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions, available
here.
Refer also to any guidelines provided in this letter.

*3) Include a revised version of any required Reporting Summary:
https://www.nature.com/documents/nr-reporting-summary.pdf
It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review.
A revised checklist is essential for re-review of the paper.

Please be aware of our guidelines on digital image standards.

Please use the link below to submit your revised manuscript and related files:

[REDACTED]
Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We can be flexible with deadlines for the revision. But do let us know if you think that you'll need longer than 16 weeks to complete.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

Nature Genetics is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Sincerely,

Safia Danovi

Editor
Nature Genetics

Referee expertise:

Referee #1: sQTLs / brain

Referee #2: microglia

Referee #3: immune cell genetics

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
Lopes et al. present one of the largest human microglia resources to date, featuring 255 primary microglia samples isolated from four different brain regions (medial frontal gyrus, superior temporal gyrus, thalamus, and subventricular zone) of 100 human subjects. The authors perform pairwise comparisons of differential gene expression between brain regions and discover subsets of genes with variable patterns of expression across brain regions. They also find that age-dependent effects on microglial gene expression are largely shared across brain regions.

The authors proceeded to catalog cis-eQTLs and cis-sQTLs in microglia for different brain regions, focusing on a subset of 90 subjects who were of self-reported European ancestry. They found between 67 and 199 eGenes (genes with cis-eQTL) per region and 253 to 426 sGenes (genes with a cis-sQTL) per region. Given the small sample size per region and the possibility of high donor-to-donor variation in microglia, the authors performed a meta-analysis of all four regions to increase their power in discovering eGenes and sGenes. In doing so, the authors found 3,611 eGenes and 4,614 sGenes in at least one brain region. Although a majority of the discovered eQTLs were shared across brain regions, the authors highlight a few region-specific eQTLs.

Finally, the authors present colocalizations between their QTLs and GWAS variants associated with Alzheimer's disease, Parkinson's disease, schizophrenia, bipolar disorder, and multiple sclerosis. They proceeded to only focus on microglia eQTLs that colocalized with disease GWAS loci and assess if the associated SNPs overlapped microglia-specific regulatory regions. Using microglia chromatin-interaction data, the authors find several examples of eQTLs (two of which were described in detail) where the lead SNP was located in a putative microglia-specific enhancer.

The authors acknowledge that their analysis of the microglia transcriptome data had multiple limitations, all of which were described in the Discussion section. For example, their meta-analysis of QTLs from the four brain regions failed to account for shared donors (which may increase the false discovery rate). They also acknowledge that additional downstream experiments would be needed to validate the prioritized enhancers that they discussed in their two eQTL examples. Overall, the manuscript is well-written and easy to follow. Given the size of their microglia dataset, I would expect the paper to have many citations if published. However, the paper needs to be strengthened by

additional computational and experimental work, before it is suitable for publication in Nature Genetics.

Major Comments:

1. The authors demonstrate that microglia from different brain regions exhibit differential gene expression. If the authors investigated both expression and splicing QTLs in the paper, why didn't the authors test for differential splicing in microglia between the brain regions?

2. As related to point #1 above, the authors should also investigate the impact of aging on splicing in the microglial transcriptome.

3. The description of the method used to test for cis-sQTLs was a bit hard to follow. It's unclear if the effect sizes for SNPs from the linear regression model are accurately calculated if the model includes intron ratios from the same intron cluster as observations. This is because linear regression assumes independence of observations, whereas excision ratios for introns found in an intron cluster can be correlated (e.g. introns flanking a cassette exon exhibit correlated excision patterns).

4. As related to point #3 above, it is somewhat surprising and counter-intuitive to see that the authors detected several times more sGenes than eGenes per region, as splicing analysis requires higher RNA-seq depth and is typically underpowered in regular RNA-seq datasets. Could this be due to the method used for sQTL detection inflating the statistical significance?

5. The authors report examples of two genes (USP6NL and P2RY12) carrying an eQTL where the candidate causal variant overlaps a putative microglia-specific enhancer. Do any of these variants disrupt motifs for known transcription factor binding sites? Given that these two genes are the only two concrete examples presented in this manuscript linking microglial eQTLs to neurological diseases, experiments to test the effects of candidate causal variants and putative microglia-specific enhancers are needed.

6. The manuscript seems to largely focus on differential gene expression in microglia and microglia eQTLs (with two highlighted examples). I was a bit disappointed by the lack of discussion related to microglia splicing QTLs, given that the Abstract and Introduction sections seem to give equal importance to both eQTLs and sQTLs. It would be helpful if the authors can expand the splicing analysis (also see points #1 and #2 above), and discuss at least one interesting sQTL example. The authors' comment in the Discussion section that long-read RNA-seq can improve the analysis and interpretation of sQTLs is valid, but this does not mean that they should not do a thorough splicing analysis on this dataset.

7. Can the authors clarify if the p-values shown for eQTLs and sQTLs in figures (e.g. Fig. 4C, 4G; Fig. 6B, 6E) are nominal or FDR-adjusted p-values? As related to this point, I don't quite understand why the authors stated on line 239-240 that "eQTLs for CASS4 are highly significant in both MiGA and monocytes (MyND) but with opposite directions of effect". The p-values in MiGA seem quite marginal according to Fig. 4G.

8. For reproducibility purposes, codes used for the analysis in this paper should be made publicly available (e.g., GitHub).

9. Raw data for this work are being deposited to GEO/dbGAP. Since a major value of this work is the dataset as a resource, the data need to be made publicly available prior to publication.

Minor Comments:

1. I think there should be a comma after "aging" in the title.

2. In affiliation 2, "Ronald M. Loeb Center for Alzheimer's disease" needs to be changed to "Ronald M. Loeb Center for Alzheimer's Disease" (upper case for 'D').

3. On line 66, the sentence from the previous line ends on a comma, not a period.

4. On Figure 1, is there a typo in the light purple box "Interaction with age-, and regional heterogeneity"?

5. On Figure 2A, there is inconsistent spacing in "Ancestry -PC1".

6. On Figure 5F, there is inconsistent spacing in "PP4> 0.7".

7. In general, there is inconsistent italicization of gene names; the main text mostly has them italicized whereas the figures do not have them italicized. Along this note, there are instances where protein names are italicized (see line 323).

8. Line 360 starts with an unintentional period.

9. The sentence starting on line 440 is a run-on sentence.

10. The sentence starting on line 447 is a sentence fragment.

11. Fixed and random effects of the linear mixed model (DREAM) should be explicitly specified in the Methods section. Currently, the text in the Methods section just lists what features go into the model without explicitly specifying which is a fixed effect and which is a random effect.

12. The supplementary figures could be exported with higher resolution (for example, Supplementary Figures 15 and 17 look very grainy upon zooming in).

Reviewer #2:
Remarks to the Author:
In the manuscript "Atlas of genetic effects in human microglia transcriptome across brain regions, aging and disease pathologies" de Paiva Lopes et al describe genetic and transcriptome analysis of 255 human microglia samples from 100 human donors.
The paper addresses the important question to assess how genetic risk is related to microglia heterogeneity, function and CNS disease.
Overall, it is well performed, elegant study, that addresses a pertinent question in the field.

Below a series of comments/points that I have with the paper:

Major comments
- Microglia were isolated using CD11b beads (line 88). This is relatively poor way to isolate microglia. Using this strategy also other immune cells might be obtained, such as (infiltrating) monocytes and the abundance of these may vary across brain regions and diseases. Furthermore, potential cell doublets will be collected, if a CD11B cell is still attached to a second cell. And dead cells will also be obtained. CD11B bead for microglia isolation is substandard.

Additionally, for the microglia isolation enzymes were used at 37C. It is well known that this causes microglia activation, affected gene expression, and as such likely interferes with the results.

- In fig 2E the overlap between grey and white matter microglia is compared with the findings in Van der Poel et al. This overlap is statistically significant, however the numbers indicated in the venn diagrams this overlap seems low. For example, in the "upregulated grey matter" venn only 13 out of 385 genes from MiGA overlap with Van der Poel. I have the same comment on figure 3E.

- The transcriptome data is used to compare different brain region's and aging, but why not to identify disease-signatures? It would be interesting to see whether there are DEGs between the different diagnosis groups.

- Line 574: why no logFC threshold?

Minor comments
- In line 69 the paper from Young et al is cited where 93 donors were studied. In the present manuscript 100 donors are studied. Why do the authors claim that they have a larger sample size? The difference is a not so dramatic 7 donors.

- The figures are not mentioned in the right order in the text. E.g Fig 2E is described before fig 2C.

- The effect of ageing on the microglia transcriptome is investigated and shares some features with AD genes. What was the amyloid-b load in these ageing brains? It is known that amyloid can induce a specific gene signature in microglia. The authors need to show neuropathological data of the analysed tissues.

Reviewer #3:
Remarks to the Author:

A:Summary of the key results

Microglia form a myeloid derived immune cell subset of the central nervous system which has been implicated to play a sentinel role in neuronal health and disease. In particular, there is a large overlap between GWAS risk variants for neurodegenerative conditions, namely Alzheimer's and regulatory variants in non CNS myeloid subsets (namely monocytes). These data highlight the importance of further dissecting determinants of microglia activity - both genetic and otherwise.

Pavia Lopes, Snijders, Humphrey and colleagues present a study exploring age, anatomical and

genetic determinants of microglia gene expression, with particular reference to neurodegenerative disease genetics. They do this by examining gene expression using bulk RNAseq of microglia across a cohort of 115 individuals, which after QC is reduced to 255 samples from 200 donors. Whilst the cohort is relatively small from a genetic perspective, it is to be appreciated that these samples are not easy to obtain - in terms of both the source and also the methods of extraction.

The authors proceed to describe anatomical determinants of microglial expression (examining four brain regions thalamus/ subventricular zone/, medial frontal and superior temporal gyri), and then explore how these interact with age and also genetics.

They define four major clusters of genes with divergent properties in terms of expression based on anatomy – the SVZ looks most dissimilar. They demonstrate that in general age regulates expression in a similar manner across regions, although they show a number of genes that are regulated by age divergently according to anatomical location.

From a genetic perspective they describe both eQTL and splicing QTL, demonstrating enrichment of risk loci for neurodegenerative diseases specially in microglia for eQTL, whereas control datasets for bipolar and schizophrenia do not show such association. The observation of splicing effects was interesting, although perhaps the authors could move to demonstrate the proposed transcripts affected – in fact little was remarked on here.

In general though, I thought the paper was well written and a valuable contribution to the literature. The authors make clear distinctions between the genetic effect on gene expression in monocytes and microglia – showing that whilst there is commonality across most loci, a number show microglia distinct effects and, crucially these are enriched in neurodegenerative disease risk loci.

Minor points of concern/ areas to be expanded:
1. The authors discuss 25 genes that are differentially regulated inMFG/STG but not THA/SVZ – is there commonality in these genes. Are they under shared control of a transcription factor that is in itself differentially regulated according to region?
2. The directional effect of genetics on CASS4 expression in monocytes versus microglia is particularly interesting. Are the authors able to further elucidate the mechanism here at the genomic level?
3. The references seem out of kilter – they refence Young et al - a preprint describing microglia eQTL across 93 samples – this is ref. 27 – but ref 27 is to Stevens et al (!?). I have not assiduously checked references, but suggest this is done.
4. The figures seem somewhat out of order especially for figure 2 where [anels exist not in the order that they are cited in the text – this is confusing and should be corrected (2a, 2b, 2d, 2e, 2c, 2f, 2g).
5. Line 66 page 1, a full stop is missing between (eQTL) and Investigations.
6. Page 2 – the authors say they observed no differentially expressed genes between males and females – yet they use differentially expressed genes in sex checking as far as I can see – this should be modified/ expanded on.

B: Originality and significance: if not novel, please include reference
There is a preprint of a smaller cohort which the authors reference – this dataset to my mind is less thorough than that presented and it does not detract significantly from this work.

C:Data & methodology: validity of approach, quality of data, quality of presentation

The figures were exceptionally clear (albeit slightly out of order). The text was easy to read.

D:Appropriate use of statistics and treatment of uncertainties
The authors have been robust throughout using the latest statistical approaches and pipelines. I have no major concerns here.

E: Conclusions: robustness, validity, reliability
The findings appear robust and the authors do not make any far-fetched statements.

F:Suggested improvements: experiments, data for possible revision
Limited – see minor comments. In general, this appears ready to publish.

G:References: appropriate credit to previous work?
I have not gone through every reference, but it would appear so.

H: Clarity and context: lucidity of abstract/summary, appropriateness of abstract, introduction and conclusions
No issue.

Author Rebuttal to Initial comments

Response letter

NG-A56637: Atlas of genetic effects in human microglia transcriptome across brain regions, aging and disease pathologies

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
Lopes et al. present one of the largest human microglia resources to date, featuring 255 primary microglia samples isolated from four different brain regions (medial frontal gyrus, superior temporal gyrus, thalamus, and subventricular zone) of 100 human subjects. The authors perform pairwise comparisons of differential gene expression between brain regions and discover subsets of genes with variable patterns of expression across brain regions. They also find that age-dependent effects on microglial gene expression are largely shared across brain regions.

The authors proceeded to catalog cis-eQTLs and cis-sQTLs in microglia for different brain regions, focusing on a subset of 90 subjects who were of self-reported European ancestry. They found between 67 and 199 eGenes (genes with cis-eQTL) per region and 253 to 426 sGenes (genes with a cis-sQTL) per region. Given the small sample size per region and the

possibility of high donor-to-donor variation in microglia, the authors performed a meta-analysis of all four regions to increase their power in discovering eGenes and sGenes. In doing so, the authors found 3,611 eGenes and 4,614 sGenes in at least one brain region. Although a majority of the discovered eQTLs were shared across brain regions, the authors highlight a few region-specific eQTLs.

Finally, the authors present colocalizations between their QTLs and GWAS variants associated with Alzheimer's disease, Parkinson's disease, schizophrenia, bipolar disorder, and multiple sclerosis. They proceeded to only focus on microglia eQTLs that colocalized with disease GWAS loci and assess if the associated SNPs overlapped microglia-specific regulatory regions. Using microglia chromatin-interaction data, the authors find several examples of eQTLs (two of which were described in detail) where the lead SNP was located in a putative microglia-specific enhancer.

The authors acknowledge that their analysis of the microglia transcriptome data had multiple limitations, all of which were described in the Discussion section. For example, their meta-analysis of QTLs from the four brain regions failed to account for shared donors (which may increase the false discovery rate). They also acknowledge that additional downstream experiments would be needed to validate the prioritized enhancers that they discussed in their two eQTL examples. Overall, the manuscript is well-written and easy to follow. Given the size of their microglia dataset, I would expect the paper to have many citations if published. However, the paper needs to be strengthened by additional computational and experimental work, before it is suitable for publication in Nature Genetics.

Thanks for your thorough assessments.

Major Comments:

1. The authors demonstrate that microglia from different brain regions exhibit differential gene expression. If the authors investigated both expression and splicing QTLs in the paper, why didn't the authors test for differential splicing in microglia between the brain regions?

We have now added analyses of differential transcript usage between microglial regions to the manuscript. We chose a transcript-based approach rather than the event based approach of Leafcutter due to the ease of biological interpretability. We identified 176 transcripts differentially used between regions and provide plots for an example gene *RGS1*, as well as ingenuity pathway

analysis of the gene set. The figures have been included in the supplement (**Fig S8**). We have also added methods and results for our analysis to the **Methods** and **Results** section.

Method section**:**

Differential Transcript Usage

Transcript expression was estimated in each sample using RSEM with the GENCODE v30 transcript reference. Lowly expressed transcripts were removed with the threshold transcript counts per million > 1 in at least 30% of all samples. This retained 49,694 transcripts from 10,818 genes. Differential transcript usage was tested simultaneously between each region and with donor age using satuRn[103], a fast method for computing differential transcript usage. No current differential transcript usage tool can model random effects so we were unable to account for shared donors, but otherwise the same technical covariates were used as in the differential expression modelling. Pairwise comparisons between each pair of regions were extracted using the limma::makeContrasts() function. To correct for test statistic inflation due to correlation across transcripts and donors, we employed a more stringent empirical FDR correction[104], but used a more liberal FDR threshold of 0.1 to increase our discovery set of transcripts. Transcripts were considered differentially used between regions at a |log odds ratio| > 1 and an empirical FDR < 0.1. For aging, transcripts were filtered at an FDR < 0.1 but no effect size cutoffs were used.

and

In addition, we analyzed the canonical pathways associated with splicing in the regional differential transcript usage (DTU) gene set (n = 132) and aging DTU gene set (n = 150) at FDR < 0.01 in IPA.

Results section:

We examined changes in splicing between microglia regions using a differential transcript usage (DTU) framework. 176 transcripts in 132 genes had evidence of DTU (log odds ratio > 1; empirical FDR < 0.1), with the majority of transcripts coming from comparisons with the SVZ (**Figure S8A)**. 31 DTU genes were also differentially expressed between pairs of regions (OR = 5.47, $P = 2.9 \times 10^{-12}$; Fisher exact test). *RGS1* is an example of a gene with a shift in the ratio of the two most abundant isoforms in the SVZ compared to the other regions (**Figure S8B**). The regional DTU geneset includes genes that are involved in mitochondrial functions, glucocorticoid receptor signalling pathways, and host defense mechanism against infection **(Figure S8C),** pathways also observed in the regional expression analysis**.**

2. As related to point #1 above, the authors should also investigate the impact of aging on splicing in the microglial transcriptome.

We have now performed differential transcript usage across aging. We identified 225 transcripts differentially used across aging and highlight *P2RY12* as having a shift between a pair of long and short isoforms as microglia age. The figures have been included in the supplement (**Fig S9**) and we added the following sentences to the **Results** section:

Similarly, we found 225 transcripts from 150 genes exhibiting DTU with age (FDR < 0.1) (**Fig S9A**), including *P2RY12,* where the balance between a long and short isoform shifts over age (**Fig S9B**). 36 of these DTU genes also showed an association with age at the gene expression level (OR = 3.47, $P = 7 \times 10^{-8}$, Fisher Exact test).

and

The genes associated with aging DTU were enriched in similar functions (**Fig S9C**).

3. The description of the method used to test for cis-sQTLs was a bit hard to follow. It's unclear if the effect sizes for SNPs from the linear regression model are accurately calculated if the model includes intron ratios from the same intron cluster as observations. This is because linear regression assumes independence of observations, whereas excision ratios for introns found in an intron cluster can be correlated (e.g. introns flanking a cassette exon exhibit correlated excision patterns).

We apologise for not being clear on this point. Our pipeline for cis-sQTLs is taken directly from the Genotype Tissue Expression (GTEx) consortium **(Auguet et al. 2020)**. Although overlapping introns are initially clustered together, the actual regression treats each intron as an independent phenotype. You are correct that there is redundancy in this approach and we are actively working on methods that account for this, for a future study. We added the following sentence to the **Methods** section:

Although junctions were initially grouped together into clusters, we tested each SNP-junction pair separately, which is the standard approach[113,117].

4. As related to point #3 above, it is somewhat surprising and counter-intuitive to see that the authors detected several times more sGenes than eGenes per region, as splicing analysis requires higher RNA-seq depth and is typically underpowered in regular RNA-seq datasets. Could this be due to the method used for sQTL detection inflating the statistical significance?

The reviewer is correct and we also expected a lower discovery rate for sQTLs. However several other factors are probably also involved. We deliberately set the cis-window size for sQTLs at

100kb around the center of each splicing cluster, whereas the eQTLs had a window of 1MB. This was partly due to biological considerations (majority of splicing QTLs are very close to splice sites) but also due to the practicality of working with smaller summary statistic files. This would have the effect of reducing multiple testing burden, which might account for the increased eGenes. The other potential mitigating factor is variability in our microglia, either biological or technical. We expected to see far more eGenes per region given our sample size. We note that the number of shared sGenes was similar to the number of shared eGenes in our MASHR analysis.

5. The authors report examples of two genes (USP6NL and P2RY12) carrying an eQTL where the candidate causal variant overlaps a putative microglia-specific enhancer. Do any of these variants disrupt motifs for known transcription factor binding sites? Given that these two genes are the only two concrete examples presented in this manuscript linking microglial eQTLs to neurological diseases, experiments to test the effects of candidate causal variants and putative microglia-specific enhancers are needed.

This is an important question. We ran a computational prediction tool, motifbreakR, to assess whether any of the fine-mapped or prioritized SNPs potentially disrupt transcription factor bindings. However, for both *USP6NL* and *P2RY12*, multiple SNPs were associated with multiple motifs (see **Table S23**), and the directionality of the disruption ran in both directions. We therefore felt it would be inappropriate to highlight one particular association and therefore added the following text to the **Results** section:

USP6NL:
Transcription factor binding motif analysis was inconclusive, with three of the tested SNPs rs143807787, rs74347557, and rs7912495 predicted to disrupt multiple motifs in different directions (**Table S23**).

P2RY12:
Effects on transcription factor binding were predicted for rs11707416, rs41366744, rs4680405, and rs62285879, again for multiple motifs (**Table S23**).

And we included the following to the **Method** section:

Predicting transcription factor binding motif disruption

We used motifbreakR[126], a package that computes the similarity of a genomic sequence to a range of transcription factor motifs, and calculates the potential disruption to each motif caused by a SNP with a score and P-value. We used as input 426 human transcription factor motifs from the HOCOMOCO database[127] as provided by motifbreakR. We ran the package on all lead QTL, lead GWAS and fine-mapped SNPs found at the USP6NL locus (using the AD GWAS from

Jansen et al.) and the P2RY12 locus from the Nalls et al PD GWAS. Motifbreakr returns metrics for each tested motif on how much each allele of a SNP disrupts or enhances predicted binding. Each comparison then undergoes a background correction to compute p-values. Full results for both loci are in **Table S23.**

We believe it is highly important to understand the mechanistic consequences of these enhancer variants but we feel that the level of experimental validation required (electrophoretic mobility shift assays or luciferase assays etc) is beyond the scope of the current work.

6. The manuscript seems to largely focus on differential gene expression in microglia and microglia eQTLs (with two highlighted examples). I was a bit disappointed by the lack of discussion related to microglia splicing QTLs, given that the Abstract and Introduction sections seem to give equal importance to both eQTLs and sQTLs. It would be helpful if the authors can expand the splicing analysis (also see points #1 and #2 above), and discuss at least one interesting sQTL example. The authors' comment in the Discussion section that long-read RNA-seq can improve the analysis and interpretation of sQTLs is valid, but this does not mean that they should not do a thorough splicing analysis on this dataset.

This was a shortcoming on our part. We have now added an additional figure (**Figure 7**) showing two splicing QTLs which colocalize with Alzheimer's disease risk loci. CD33 has been previously observed and characterised in myeloid cells **(Raj et al. 2014 HMG)** and we show that the same mechanism of exon 2 splicing is also in microglia. For our second locus, MS4A6A, we show a complex splicing pattern where introns from multiple transcripts are associated with the risk locus. We believe this example in particular is a motivating example for future long read RNA-seq experiments that we bring up in the discussion.

New results included in the **Results** section:

Splicing QTLs implicate complex isoform changes at particular GWAS loci

We repeated our colocalization and fine-mapping analyses with sQTLs across the different diseases. Overall we found 81 splicing junctions in 31 genes with a colocalized sQTL at PP4 > 0.7 with 26 GWAS loci (**Table S20**). We highlight two examples of sQTLs associated with Alzheimer's Disease and identify key challenges ahead for the interpretation of such events. The CD33 risk locus has been implicated in AD susceptibility[67]. Previous analyses in peripheral monocytes found association between lead GWAS SNP rs3865444 and the inclusion of CD33 exon 2[67]. In MiGA, we also found a strong colocalization with an sQTL associating the same SNP rs3865444-A with reduced intron usage of intron 1, corresponding to reduced inclusion of exon 2 (**Figure 7A-E).** Another example of sQTL is *MS4A6A*. The MS4A gene cluster is a gene-dense

region spanning 600kb, containing 12 genes. We observed colocalization with eQTLs and sQTLs in *MS4A6A*, as well as eQTLs in *MS4A4A* and *MS4A4E* (**Figure 5F)**. In MiGA, we observed colocalization solely with sQTLs in *MS4A6A* (**Figure 7F-J)**. We overlaid all associated sQTL junctions that colocalized with the AD risk locus and found that the strongest colocalization signals highlighted a cluster of introns in the middle of the gene, with the 5' intron in the cluster having the strongest colocalization. Notably, 2 transcripts containing this intron have a premature polyadenylation site. rs2162254-A is associated with an increased usage of this intron, which may result in increased production of the shorter isoforms, which could have a downstream consequence of MS4A6A protein function.

7. Can the authors clarify if the p-values shown for eQTLs and sQTLs in figures (e.g. Fig. 4C, 4G; Fig. 6B, 6E) are nominal or FDR-adjusted p-values? As related to this point, I don't quite understand why the authors stated on line 239-240 that "eQTLs for CASS4 are highly significant in both MiGA and monocytes (MyND) but with opposite directions of effect". The p-values in MiGA seem quite marginal according to Fig. 4G.

The p-values are nominal from the region by region QTL analysis. The figures 6B and 6E also shows the p-value (RE2 model) from the meta-analysis using METASOFT. We have clarified it in the legends of the respective figures, as follow:

Fig. 4C: Examples of shared (*CTSB* gene, rs12338) and region-specific effect (*RNF40*, rs56039835). eQTL boxplots with residual gene expression (PEER adjusted) per individual stratified by genotype. The eQTL nominal P-value and effect size are listed on top.

Fig. 4G: Example of discordant eQTL effects for *CASS4* (rs6069736) between microglia and monocytes. The P-values are nominal from the region by region eQTL analysis.

Fig. 6B: *USP6NL* expression is associated with the rs7912495 genotype in all four microglia regions. The nominal P-value from eQTL analysis is indicated on top of the boxplots by region. The beta and P-value from the meta-analysis are also indicated on top of the Figure.

Fig. 6E: *P2RY12* expression is associated with the rs3732765 genotype. The nominal P-value from eQTL analysis is indicated on top of the boxplots by region. The beta and P-value from the meta-analysis are also indicated on top of the Figure.

About the eQTLs for *CASS4*, we removed the highly from the phrase as follow:

eQTLs for *CASS4* are significant in both MiGA and monocytes (MyND) but with opposite directions of effect.

8. For reproducibility purposes, codes used for the analysis in this paper should be made publicly available (e.g., GitHub).

The code is publicly available at https://github.com/RajLabMSSM/MiGA_public_release. We also included the link in the text under the URLs subtitle.

9. Raw data for this work are being deposited to GEO/dbGAP. Since a major value of this work is the dataset as a resource, the data need to be made publicly available prior to publication.

Raw and processed RNA-seq and genotype data sets are deposited in the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS; Accession ID: ng00105). All raw data will be available for download via NIAGADS https://dss.niagads.org/.

Minor Comments:

10. I think there should be a comma after "aging" in the title.

We added a comma after aging in the title.

11. In affiliation 2, "Ronald M. Loeb Center for Alzheimer's disease" needs to be changed to "Ronald M. Loeb Center for Alzheimer's Disease" (upper case for 'D').

We changed the typographical error.

12. On line 66, the sentence from the previous line ends on a comma, not a period.

We corrected this in the revision.

13. On Figure 1, is there a typo in the light purple box "Interaction with age-, and regional heterogeneity"?

We corrected the typo in Figure 1.

14. On Figure 2A, there is inconsistent spacing in "Ancestry -PC1".

We adjusted the inconsistent spacing in Figure 2A.

15. On Figure 5F, there is inconsistent spacing in "PP4> 0.7".

We adjusted the inconsistent spacing on Figure 5F.

16. In general, there is inconsistent italicization of gene names; the main text mostly has them italicized whereas the figures do not have them italicized. Along this note, there are instances where protein names are italicized (see line 323).

We corrected this throughout the revised manuscript.

17. Line 360 starts with an unintentional period.

We removed the unintentional period in line 360.

18. The sentence starting on line 440 is a run-on sentence.

We rewrote the sentence as following:

There are several methodological differences (recruitment of tissue, studied brain region, postmortem delay, pH, age, diagnosis, medication use) that could interfere with the interpretation of comparisons between MiGA and other microglial datasets[45,15,5].

19. The sentence starting on line 447 is a sentence fragment.

We rewrote the sentence as following:

although single-cell data is in general sparse and noisy, which may result in reduced power compared to bulk RNA-seq[85].

20. Fixed and random effects of the linear mixed model (DREAM) should be explicitly specified in the Methods section. Currently, the text in the Methods section just lists what features go into the model without explicitly specifying which is a fixed effect and which is a random effect.

For all the Differential Expression Analysis, donor ID and cause of death covariates were modeled as random effects and the others covariates modeled as fixed effects. As described in the Variance Partition (R package with DREAM embedded) vignette, "the difference between modeling a categorical variable as a fixed versus random effect is minimal when the sample size is large compared to the number of categories **(Hoffman, 2020)**." So, variables like diagnosis or

sex will not be sensitive to modeling as a fixed or random effect. However, Donor ID must be modeled as a random effect in order to obtain statistically valid results, as we have considered in our model. The design for each Differential Expressed Analysis is described along with the code on the Github page.

We have included the following text in the **Methods** section:

For all the Differential Expression Analysis, donor ID and cause of death covariates were modeled as random effects and the others covariates modeled as fixed effects (see URL session for GitHub repository page with code).

21. The supplementary figures could be exported with higher resolution (for example, Supplementary Figures 15 and 17 look very grainy upon zooming in).

We have now updated the supplementary colocalization figures (**Figures S16-25**) with higher resolution versions.

Reviewer #2:
Remarks to the Author:
In the manuscript "Atlas of genetic effects in human microglia transcriptome across brain regions, aging and disease pathologies" de Paiva Lopes et al describe genetic and transcriptome analysis of 255 human microglia samples from 100 human donors.
The paper addresses the important question to assess how genetic risk is related to microglia heterogeneity, function and CNS disease.
Overall, it is well performed, elegant study, that addresses a pertinent question in the field.

Thank you for the remarks.

Below a series of comments/points that I have with the paper:

Major comments
1. Microglia were isolated using CD11b beads (line 88). This is relatively poor way to isolate microglia. Using this strategy also other immune cells might be obtained, such as (infiltrating) monocytes and the abundance of these may vary across brain regions and diseases.

We agree with the reviewer that different isolation methods challenge different strengths and weaknesses. We used the more gentle approach of CD11b+ MACS sorting to reduce phenotypical microglial changes that may occur after FACS sorting. However, the purity of the

cells is less with MACS sorting compared to FACS. To check for the purity of the microglial cells and the amount of contamination with other (immune) cell types we applied several methods:

1. In the manuscript, we checked the expression of several cell-type specific markers for astrocytes, neurons, oligodendrocytes in our RNA-seq data (see **Figure S2C**) and these markers were either not or lowly expressed. In addition, we checked the expression levels of two well-known monocyte markers (*CCR, FCER1A*) in the RNA seq data and added these findings to the heatmap in **Figure S2C**. We changed the **Results** section accordingly:

   All microglial samples expressed known microglia-specific genes at high levels, with marker genes for neurons, astrocytes, oligodendrocytes, and monocytes being lowly or not expressed (**Figure S2C**).

2. We did additional single-cell mass cytometry by time of flight (CyTOF) in a subsample of our cohort and single-cell RNA seq to comprehensively characterize the phenotype of our CD11b+ isolated cells and to check the level of contamination. We analyzed the level of contamination across brain regions.

   CyTOF:

   We show in a subsample of our cohort (n = 92 samples derived from 30 donors) that the mean percentage of P2Y12+CD64+ expressing cells is 96%, ranging between 56.6% to 99.9%. The P2Y12 receptor is a purinergic receptor that is exclusively expressed by microglial cells in the central nervous system. CD64 is a well-known marker for myeloid cells. The mean percentage of P2Y12-CD64+ cells was 3%, indicating low levels of infiltrating myeloid cells. This is in line with previous results by mass cytometry, showing that the percentage of (perivascular) macrophages ($CD206^{high}$) was < 5% in P2Y12+ expressing microglial cells (**Böttcher et al. 2019**). In addition, we checked the percentage of (infiltrating) monocytes by measuring migration inhibitory factor related protein 14 (MRP14+) cells in a subsample of the cohort (n = 57 samples derived from 17 donors). The percentage of MRP14+ was 0.37 % on average, suggesting very little infiltration of monocytes. We did not see any differences between brain regions as could be seen in **Figure S2A and SB**. The disease groups were underpowered to detect differences.

   We show these findings in **Figure S2A and S2B** and added these findings to the **Methods** section as follows:

   However, we have previously shown by mass cytometry that the percentage of macrophages ($CD206^{high}$) was low[31]. In a subsample of our cohort we checked the

percentage of (infiltrating) monocytes by measuring migration inhibitory factor related protein 14 (MRP14+) cells (n = 56 samples derived from 19 donors).The percentage of (infiltrating) monocytes was 0.04 % on average. Moreover, we showed in a subsample (*n* = 91 samples derived from 30 donors) of our cohort that the mean percentage of microglia ($P2Y12^{+}CD64^{+}$) cells is 95%, ranging from 56.6% to 99.9%. The percentage of P2Y12-CD64+ is 3.2%, suggesting low levels of infiltration of macrophages/monocytes (see **Figure S2**).

Single-cell RNA sequencing

We generated *new* single-cell data to reveal the level of contamination of other cell types. We isolated CD11b+ microglial cells of the medial frontal gyrus and hippocampus of one donor (diagnosis: Parkinson disease) and constructed scRNAseq barcoded libraries with 10x genomics Chromium Single-Cell 3'. Around 94% of the CD11b+ isolated microglial cells express microglial specific markers, such as *P2Y12* and *TMEM119*, 5% represent myeloid cells, the remaining part contains non-myeloid cells. We added these findings **Figure S3**.

We explained the single-cell analyses in the **Method** section as follows:

Single-cell sequencing

To generate single-cell data we isolated CD11b+ microglial cells from the medial frontal gyrus and hippocampus of one additional donor (MG-22). We used the instructions of the Single-Cell 3' Reagent Kits v2 (10 x Genomics) to construct scRNAseq barcoded libraries. In short, we loaded ~10.000 microglial cells from both regions separately into a slot of Chromium Chip. GEMs were incubated in a thermal cycler for the generation of barcoded cDNA. The cDNA was fragmented after amplification and processed for sequencing by ligating adapters and sample indices. The libraries were sequenced on an Illumina Hi-Seq system. The average sequencing depth was 2197.959 raw reads per cell, and the average was 1108.241 UMIs per cell. We analyzed 8589 cells for the two brain regions combined. Low-quality cells with >10% mitochondrial gene content were removed. Duplicate cells were filtered by removing cells with greater than 10000 transcripts. Genes not detected in at least 3 cells were removed. After rigorous quality control, we analyzed 5990 single cells in total. We used the standard Seurat workflow (Seurat v. 3.2.3) to normalize (NormalizeData; default parameters), scale (ScaleData; on all genes regressing out total UMI count and mitochondrial percentage per cell), and performed dimensionality reduction using principal component analysis with the top 2000 highly

variable genes (RunPCA). The first 13 principal components were used for PCA-Louvain clustering (FindNeighbors and FindClusters). To gain sufficient detail to detect small subpopulations within one donor, the cluster resolution was set to 0.1. We used UMAP, a non-linear dimensionality reduction, to visualize the data (RunUMAP with the first 13 PCs) and produce feature plots in low dimensional space (DimPlot and FeaturePlot). Cluster-enriched genes were identified using logistic regression (FindAllMarkers function with default thresholds and only.pos = TRUE) and were used to manually annotate clusters by cell type.

We added both the CyTOF and scRNAseq findings to the **Result** section:

Mass cytometry (CyTOF) proteomic and scRNA-seq analyses were used to analyze the cell composition and showed an average of 95% of cells being positive for the microglia specific markers P2Y12 and TMEM119 (Figure S2A and S3). The other cells are mainly from myeloid origin, based on the expression of monocyte markers (CD64+/P2Y12- and MRP+ in CyTOF; MRC1/LYV1 in scRNA-seq).

Additionally, we added the following limitation to the **Discussion** section:

We sorted the microglial cells with CD11b+ beads. This marker is not restricted to microglia and may capture small fractions of other myeloid cells.

Furthermore, potential cell doublets will be collected, if a CD11B cell is still attached to a second cell. And dead cells will also be obtained. CD11B bead for microglia isolation are substandard.

Cell doublets are associated with higher contamination. Additional mass spectrometry and sRNAseq analyses showed that the majority of our cells are microglial cells. Also, an assessment of cell viability is critical. We used a microglia protocol that has been validated and extensively described by (J. Melief et al. 2016), which includes an assessment of the cell viability using the DNA-intercalator 7-amino-actinomycin D (7AAD) to detect non-viable cells with flow cytometry. After each isolation, we performed a trypan blue staining of the microglial cells to check the amount of dead cells with microscopy. The cell viability was between 70 and 98%. We now state this more clearly in the **Methods** section:

We performed this study using a validated protocol for post-mortem microglia isolation[30–32,89], which includes the assessment of DNA-intercalator 7-amino-actinomycin D (7-AAD) staining to detect non-viable cells with flow cytometry. After each isolation, we performed trypan blue staining

of the microglial cells to check the amount of death cells with microscopy. The resulting cell viability was between 70% and 98%.

Additionally, for the microglia isolation enzymes were used at 37C. It is well known that this causes microglia activation, affected gene expression, and as such likely interferes with the results.

Recent mouse studies (Marsh et al. 2020; Mattei et al. 2020) compared transcriptional and proteome profiles of isolated microglial cells using different cell isolation techniques. Overall, these studies have shown that several technical factors, such as processing the brain tissue on 37 degrees (compared to 4 degrees) and enzymatic digestion/dissociation cause microglial activation. These comparative studies are based on animal studies, where conditions such as antemortem factors can be controlled and there is no or very limited postmortem delay. It is not yet clear how the isolation procedures influence postmortem-derived human derived microglia, but we agree that this may have interfered with the results.

We added the following sentence as a limitation in the **Discussion** section:

Besides neuroinflammation, hypoxia, and long postmortem intervals, technical artifacts (enzymatic digestion, temperature changes, sorting) may cause microglial activation. We could not control for all these potential confounders, even though these factors could contribute to gene expression changes[83,84].

3. In fig 2E the overlap between grey and white matter microglia is compared with the findings in Van der Poel et al. This overlap is statistically significant, however the numbers indicated in the venn diagrams this overlap seems low. For example, in the "upregulated grey matter" venn only 13 out of 385 genes from MiGA overlap with Van der Poel. I have the same comment on figure 3E.

We agree with the reviewer that the overlap of the number of genes between the two datasets in **Fig. 2E** and **3E** are low. However, we used a null background of 20,000 genes in our enrichment analyses so the chance of any overlap between two random datasets is low, hence the low p-value. The overlap between our dataset and van der Poel et al. or Galatro et al. is small, but still much higher than expected by random chance. We stated this more clearly in the **Results** section:

We compared our findings in a published dataset of white and grey matter microglia[5] and found small, but significant overlaps with our MFG vs SVZ comparison (upregulated OR = 18.4; $P < 1 \times 10^{-16}$, downregulated OR = 4.83; $P = 9 \times 10^{-6}$; Fisher's exact test; **Figure 2C**).

and

We replicated our findings using an external microglia aging dataset from the parietal cortex[15], and from peripheral blood[43] (**Figure S10**). The number of genes that overlap between the datasets was small, but significant (upregulated genes OR = 23.4; $P < 1 \times 10^{-16}$, downregulated genes OR = 5.97; $P < 1 \times 10^{-16}$; Fisher's Exact test; **Figure 3E)**.

4. The transcriptome data is used to compare different brain region's and aging, but why not to identify disease-signatures? It would be interesting to see whether there are DEGs between the different diagnosis groups.

Thank you for this suggestion. We included the disease subgroup analyses (controls, dementia, Parkinson's disease, major depressive disorder, or bipolar disorder/schizophrenia) in the revised manuscript, with the caveat that these analyses are underpowered to detect differences between disease groups. We only found 24 differentially expressed genes (FDR < 5%) in the dementia group compared to controls. No significant gene expression changes were found for the other disease groups. The findings in Parkinson's disease are discussed in more detail in the preprint by **(Navarro et al. 2020)**.

We added the following sentences to the **Method** section:

The effect of diagnosis was analyzed with DREAM in microglial samples from subjects with dementia (n = 15 samples derived from 9 donors), Parkinson's disease (PD; n = 18 samples derived from 12 donors, **Table S10)**, major depressive disorder (MDD; n = 74 samples derived from 21 donors, **Table S11**; bipolar disorder or schizophrenia (BD/SCZ; n = 37 samples derived from 14 donors, and controls (n = 96 samples derived from 38 donors, **Table S12**). Disease groups were compared to controls seperatly while accounting for the effect of age, sex, brain region, technical covariates, donor variation, ancestry, and cause of death.

We changed the **Results** section as follows:

We explored the effect of diagnosis on the microglia transcriptome and detected 24 genes, such as *MCF2* and *AIDH3B1*, differentially expressed in the dementia group compared to controls (FDR < 0.05; **Table S10**). No significant gene expression changes were found for PD, MDD and BD/SCZ (**Table S11-S13)**.

List of differentially expressed genes are displayed in **Table S10-S13.**

5. Line 574: why no logFC threshold?

Line 574 refers to the threshold used in the Differential Expressed Analysis using an interaction model between age and region. For the quantitative covariate of age, only an FDR adjusted p-value was used as a threshold because the log(FC) corresponds to the estimated increase of age as a unit per year. It is different from a case-control analysis, for example, since age is a continuous variable, not categorical.

Minor comments

6. In line 69 the paper from Young et al is cited where 93 donors were studied. In the present manuscript 100 donors are studied. Why do the authors claim that they have a larger sample size? The difference is a not so dramatic 7 donors.

We included in total 216 samples (from 90 individuals) compared to 93 samples (from 93 individuals) in the study of Young et al. By combining more samples (from shared individuals) we increased power for statistical testing, even though the number of included donors did not differ. We changed the following sentence in the **Introduction** section for clarification:

Recently, Young *et al.* constructed expression QTLs (eQTLs) in primary human microglia (*n* = 93 individuals/samples), and detected 401 eQTLs, some of which colocalized with AD loci, including *BIN1* **(Young et al. 2019).**

7. The figures are not mentioned in the right order in the text. E.g Fig 2E is described before fig 2C.

As suggested, we changed the figures in chronological order following the main text.

8. The effect of ageing on the microglia transcriptome is investigated and shares some features with AD genes. What was the amyloid-b load in these ageing brains? It is known that amyloid can induce a specific gene signature in microglia. The authors need to show neuropathological data of the analysed tissues.

We have added the available neuropathological data for part of the samples. This was not available for some of the most recent autopsies. The NBB determines AD pathology according to Braak's staging for neurofibrillary tangles (NFT) **(H. Braak and Braak 1991)** and CERAD criteria **(Mirra et al. 1991)** adjusted for age, respectively, and were assessed in sections stained with Gallyas and Bodian silver staining. An 'ABC' score was attributed to each subject for amyloid pathology, which includes an assessment of braak staging for Aβ plaques depending on the affected brain regions (none (O), low (A), intermediate (B), high (C)) (Hyman et al. 2012). For

some samples the presence and topographical distribution of α-synuclein was rated according to Braak's staging scheme **(Heiko Braak et al. 2003)**. We included all available information about braak NFT and a-synuclein staging, amyloid deposition (ABC scores) in the sample overview file (metadata). The sample sizes per group were too small to perform meaningful additional analyses.

We added the following sentences to the **Method** section:
Neuropathological assessments have been performed by the NBB (see URLs). Detailed information per donor, including tissue type, age, sex, postmortem interval, pH of cerebrospinal fluid, cause of death, diagnosis, use of medication and neuropathological information is provided in **Table S1**.

Reviewer #3:
Remarks to the Author:

A:Summary of the key results

Microglia form a myeloid derived immune cell subset of the central nervous system which has been implicated to play a sentinel role in neuronal health and disease. In particular, there is a large overlap between GWAS risk variants for neurodegenerative conditions, namely Alzheimer's and regulatory variants in non CNS myeloid subsets (namely monocytes). These data highlight the importance of further dissecting determinants of microglia activity - both genetic and otherwise.

Pavia Lopes, Snijders, Humphrey and colleagues present a study exploring age, anatomical and genetic determinants of microglia gene expression, with particular reference to neurodegenerative disease genetics. They do this by examining gene expression using bulk RNAseq of microglia across a cohort of 115 individuals, which after QC is reduced to 255 samples from 200 donors. Whilst the cohort is relatively small from a genetic perspective, it is to be appreciated that these samples are not easy to obtain - in terms of both the source and also the methods of extraction.

The authors proceed to describe anatomical determinants of microglial expression (examining four brain regions thalamus/ subventricular zone/, medial frontal and superior temporal gyri), and then explore how these interact with age and also genetics.

They define four major clusters of genes with divergent properties in terms of expression based on anatomy – the SVZ looks most dissimilar. They demonstrate that in general age

regulates expression in a similar manner across regions, although they show a number of genes that are regulated by age divergently according to anatomical location.

From a genetic perspective they describe both eQTL and splicing QTL, demonstrating enrichment of risk loci for neurodegenerative diseases specially in microglia for eQTL, whereas control datasets for bipolar and schizophrenia do not show such association. The observation of splicing effects was interesting, although perhaps the authors could move to demonstrate the proposed transcripts affected – in fact little was remarked on here.

In general though, I thought the paper was well written and a valuable contribution to the literature. The authors make clear distinctions between the genetic effect on gene expression in monocytes and microglia – showing that whilst there is commonality across most loci, a number show microglia distinct effects and, crucially these are enriched in neurodegenerative disease risk loci.

Thank you for the kind remarks. In response to your suggestion on increasing the analysis of the genetic effects of splicing, we have now included a new figure (**Fig 7**) where we go into detail for two sQTL loci colocalizing with AD risk variants and demonstrate their complexity.

Minor points of concern/ areas to be expanded:
1. The authors discuss 25 genes that are differentially regulated inMFG/STG but not THA/SVZ – is there commonality in these genes. Are they under shared control of a transcription factor that is in itself differentially regulated according to region?

We performed a k-means clustering analysis on a subset of genes that were differentially expressed between each brain region (see **Fig 2B)** and we identified four distinct groups of genes that may be differentially regulated. More in detail, cluster 1 contained genes that were upregulated in MFG/STG compared to THA/SVZ. Cluster 2 contained genes that are downregulated in MFG/STG compared to THA/SVZ (see **Fig 2E**).
As the reviewer suggested, we used the upstream regulator analysis tool in IPA to predict upstream molecules which may be causing the specific observed gene expression changes in cluster 1 or 2. A few interesting upstream molecules were identified in cluster 1, most of them are related to homeostatic processes. These include FKBP5, SOST, and CNOT3. The upstream molecules in cluster 2 appeared to be related to cytokine signaling (TNF, IL4, IFNG, CD40lg, CSF2). But overall, results lack the specificity to draw meaningful conclusions. We therefore felt it would be inappropriate to highlight them in the manuscript, so we decided to only add a table in supplements with the upstream molecules and add the following sentence to the **Results** section:

Analysis of upstream regulators of the four clusters using IPA was inconclusive (**Table S9).**

and added the following sentences to the **Methods** section:

Additionally, to identify upstream transcriptional regulators that may explain the observed gene expression changes across the different regional clusters we used the IPA upstream regulator analysis. We show the top 20 upstream molecules in **Table S9**.

2. The directional effect of genetics on CASS4 expression in monocytes versus microglia is particularly interesting. Are the authors able to further elucidate the mechanism here at the genomic level?

We agree with the reviewer that the divergent genetic impact on *CASS4* in monocytes and microglia is very interesting from both an Alzheimer's disease as well as a more fundamental genomics perspective. The opposite result in monocytes suggests that the functional variant is located in a complex regulatory region, as suggested before **(Schwartzentruber et al. 2021; Young et al. 2019; Alasoo et al. 2018; Peters et al. 2016; Raj et al. 2014)**, where both enhancing and repressing mechanisms are at play. As the baseline expression of *CASS4* is different between microglia and monocytes, we hypothesize that the expression of co-regulating factors may be different and determine the effect of the variant on gene expression. Functional studies to characterize the transcription and co-factors and to better understand this intergenic region and its effect on *CASS4* expression are needed to understand the complex genetic impact on CASS4 expression. Further functional fine-mapping of cis-eQTLs in a larger monocytes and microglia datasets are also necessary to identify the functional variants in each cell types. We added the following sentence in the **Result** Section in the manuscript:

Generally, directions of effect between monocytes and microglia were concordant **(Figure S14)**, with the exception of *CASS4*. eQTLs for *CASS4* are significant in both MiGA and monocytes (MyND) but with opposite directions of effect **(Figure 4G)**, suggesting that the causal variant is located in a complex regulatory element with where both enhancing and repressing mechanisms are at play.

3. The references seem out of kilter – they refence Young et al - a preprint describing microglia eQTL across 93 samples – this is ref. 27 – but ref 27 is to Stevens et al (!?). I have not assiduously checked references, but suggest this is done.

Thank you for this point, we checked the references throughout the manuscript and changed accordingly.

4. The figures seem somewhat out of order especially for figure 2 where [anels exist not in the order that they are cited in the text – this is confusing and should be corrected (2a, 2b, 2d, 2e, 2c, 2f, 2g).

Please see response to #Reviewer 2 point 7.

5. Line 66 page 1, a full stop is missing between (eQTL) and Investigations.

Please see response to #Reviewer 1 point 12.

6. Page 2 – the authors say they observed no differentially expressed genes between males and females – yet they use differentially expressed genes in sex checking as far as I can see – this should be modified/ expanded on.

First, we have noticed that sex explained little variance in our dataset (accounting for only 0.26%), as indicated by the Variance Partition analysis in Figure 2A. Then, to perform Differential Expressed analysis by sex, we removed the genes from the X and Y chromosomes but, we did not identify any gene at FDR adjusted p-value of 5%. For clarification, we have added the following sentence in the **Methods** section:

Differential sex-related analysis was performed with DREAM **(Hoffman and Schadt 2016)** as described above, with the difference that we have excluded the genes from the chromosomes X and Y.

B: Originality and significance: if not novel, please include reference
There is a preprint of a smaller cohort which the authors reference – this dataset to my mind is less thorough than that presented and it does not detract significantly from this work.

C:Data & methodology: validity of approach, quality of data, quality of presentation

The figures were exceptionally clear (albeit slightly out of order). The text was easy to read.

D:Appropriate use of statistics and treatment of uncertainties
The authors have been robust throughout using the latest statistical approaches and pipelines. I have no major concerns here.

E: Conclusions: robustness, validity, reliability
The findings appear robust and the authors do not make any far-fetched statements.

F:Suggested improvements: experiments, data for possible revision
Limited – see minor comments. In general, this appears ready to publish.

G:References: appropriate credit to previous work?
I have not gone through every reference, but it would appear so.

H: Clarity and context: lucidity of abstract/summary, appropriateness of abstract, introduction and conclusions
No issue.

References

Aguet, François, Alvaro N. Barbeira, Rodrigo Bonazzola, Andrew Brown, Stephane E. Castel, Brian Jo, Silva Kasela, et al. 2019. “The GTEx Consortium Atlas of Genetic Regulatory Effects across Human Tissues.” *BioRxiv*. https://doi.org/10.1101/787903.

Alasoo, Kaur, Julia Rodrigues, Subhankar Mukhopadhyay, Andrew J. Knights, Alice L. Mann, Kousik Kundu, HIPSCI Consortium, Christine Hale, Gordon Dougan, and Daniel J. Gaffney. 2018. “Shared Genetic Effects on Chromatin and Gene Expression Indicate a Role for Enhancer Priming in Immune Response.” *Nature Genetics* 50 (3): 424–31.

Böttcher, Chotima, Stephan Schlickeiser, Marjolein A. M. Sneeboer, Desiree Kunkel, Anniki Knop, Evdokia Paza, Pawel Fidzinski, et al. 2019. “Human Microglia Regional Heterogeneity and Phenotypes Determined by Multiplexed Single-Cell Mass Cytometry.” *Nature Neuroscience* 22 (1): 78–90.

Braak, H., and E. Braak. 1991. “Neuropathological Stageing of Alzheimer-Related Changes.” *Acta Neuropathologica* 82 (4): 239–59.

Braak, Heiko, Kelly Del Tredici, Udo Rüb, Rob A. I. de Vos, Ernst N. H. Jansen Steur, and Eva Braak. 2003. “Staging of Brain Pathology Related to Sporadic Parkinson’s Disease.” *Neurobiology of Aging* 24 (2): 197–211.

Coetzee, Simon G., Gerhard A. Coetzee, and Dennis J. Hazelett. 2015. “MotifbreakR: An R/Bioconductor Package for Predicting Variant Effects at Transcription Factor Binding Sites.” *Bioinformatics (Oxford, England)* 31 (23): 3847–49.

Efron, Bradley. 2004. “Large-Scale Simultaneous Hypothesis Testing.” *Journal of the American Statistical Association* 99 (465): 96–104.

Gilis, Jeroen, Kristoffer Vitting-Seerup, Koen Van den Berge, and Lieven Clement. 2021. “SatuRn: Scalable Analysis of Differential Transcript Usage for Bulk and Single-Cell RNA-Sequencing Applications.” *BioRxiv*. bioRxiv. https://doi.org/10.1101/2021.01.14.426636.

Hoffman, Gabriel E., and Eric E. Schadt. 2016. “VariancePartition: Interpreting Drivers of Variation in Complex Gene Expression Studies.” *BMC Bioinformatics* 17 (1): 483.

Hollingworth, Paul, the Alzheimer’s Disease Neuroimaging Initiative, Denise Harold, Rebecca Sims, Amy Gerrish, Jean-Charles Lambert, Minerva M. Carrasquillo, et al. 2011. “Common Variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP Are Associated with Alzheimer’s Disease.” *Nature Genetics* 43 (5): 429–35.

Hyman, Bradley T., Creighton H. Phelps, Thomas G. Beach, Eileen H. Bigio, Nigel J. Cairns, Maria C. Carrillo, Dennis W. Dickson, et al. 2012. “National Institute on Aging-Alzheimer’s Association Guidelines for the Neuropathologic Assessment of Alzheimer’s Disease.” *Alzheimer’s & Dementia: The Journal of the Alzheimer’s Association* 8 (1): 1–13.

Kulakovskiy, Ivan V., Ilya E. Vorontsov, Ivan S. Yevshin, Ruslan N. Sharipov, Alla D. Fedorova, Eugene I. Rumynskiy, Yulia A. Medvedeva, et al. 2018. “HOCOMOCO: Towards a Complete Collection of Transcription Factor Binding Models for Human and Mouse via Large-Scale ChIP-Seq Analysis.” *Nucleic Acids Research* 46 (D1): D252–59.

Li, Yang I., David A. Knowles, Jack Humphrey, Alvaro N. Barbeira, Scott P. Dickinson, Hae Kyung Im, and Jonathan K. Pritchard. 2018. “Annotation-Free Quantification of RNA Splicing Using LeafCutter.” *Nature Genetics* 50 (1): 151–58.

Marsh, Samuel E., Tushar Kamath, Alec J. Walker, Lasse Dissing-Olesen, Timothy R. Hammond, Adam M. H. Young, Abdulraouf Abdulraouf, et al. 2020. “Single Cell Sequencing Reveals Glial Specific Responses to Tissue Processing & Enzymatic Dissociation in Mice and Humans.” *BioRxiv*. bioRxiv. https://doi.org/10.1101/2020.12.03.408542.

Mattei, Daniele, Andranik Ivanov, Marc van Oostrum, Stanislav Pantelyushin, Juliet Richetto, Flavia Mueller, Michal Beffinger, et al. 2020. “Enzymatic Dissociation Induces Transcriptional and Proteotype Bias in Brain Cell Populations.” *International Journal of Molecular Sciences* 21 (21): 7944.

Melief, J., M. A. M. Sneeboer, M. Litjens, P. R. Ormel, S. J. M. C. Palmen, I. Huitinga, R. S. Kahn, E. M. Hol, and L. D. de Witte. 2016. “Characterizing Primary Human Microglia: A Comparative Study with Myeloid Subsets and Culture Models.” *Glia* 64 (11): 1857–68.

Melief, Jeroen, Nathalie Koning, Karianne G. Schuurman, Martijn D. B. Van De Garde, Joost Smolders, Robert M. Hoek, Marco Van Eijk, Jörg Hamann, and Inge Huitinga. 2012. “Phenotyping Primary Human Microglia: Tight Regulation of LPS Responsiveness.” *Glia* 60 (10): 1506–17.

Mirra, S. S., A. Heyman, D. McKeel, S. M. Sumi, B. J. Crain, L. M. Brownlee, F. S. Vogel, J. P. Hughes, G. van Belle, and L. Berg. 1991. “The Consortium to Establish a Registry for Alzheimer’s Disease (CERAD). Part II. Standardization of the Neuropathologic Assessment of Alzheimer’s Disease.” *Neurology* 41 (4): 479–86.

Navarro, Elisa, Evan Udine, Katia de Paiva Lopes, Madison Parks, Giulietta Riboldi, Brian M. Schilder, Jack Humphrey, et al. 2020. “Discordant Transcriptional Signatures of Mitochondrial Genes in Parkinson’s Disease Human Myeloid Cells.” https://doi.org/10.1101/2020.07.20.212407.

Peters, James E., Paul A. Lyons, James C. Lee, Arianne C. Richard, Mary D. Fortune, Paul J. Newcombe, Sylvia Richardson, and Kenneth G. C. Smith. 2016. “Insight into Genotype-Phenotype Associations through EQTL Mapping in Multiple Cell Types in Health and Immune-Mediated Disease.” *PLoS Genetics* 12 (3): e1005908.

Raj, Towfique, Katie Rothamel, Sara Mostafavi, Chun Ye, Mark N. Lee, Joseph M. Replogle, Ting Feng, et al. 2014. “Polarization of the Effects of Autoimmune and Neurodegenerative Risk Alleles in Leukocytes.” *Science (New York, N.Y.)* 344 (6183): 519–23.

Sarkar, Abhishek K., Po-Yuan Tung, John D. Blischak, Jonathan E. Burnett, Yang I. Li, Matthew Stephens, and Yoav Gilad. 2019. “Discovery and Characterization of Variance QTLs in Human Induced Pluripotent Stem Cells.” *PLoS Genetics* 15 (4): e1008045.

Schwartzentruber, Jeremy, Sarah Cooper, Jimmy Z. Liu, Inigo Barrio-Hernandez, Erica Bello, Natsuhiko Kumasaka, Adam M. H. Young, et al. 2021. “Genome-Wide Meta-Analysis, Fine-Mapping and Integrative Prioritization Implicate New Alzheimer’s Disease Risk Genes.” *Nature Genetics* 53 (3): 392–402.

Sneeboer, Marjolein A. M., Gijsje J. L. J. Snijders, Woutje M. Berdowski, Alba Fernández-Andreu, Psychiatric Donor Program of the Netherlands Brain Bank (NBB-Psy), Hans C. van Mierlo,

Amber Berdenis van Berlekom, et al. 2019. “Microglia in Post-Mortem Brain Tissue of Patients with Bipolar Disorder Are Not Immune Activated.” *Translational Psychiatry* 9 (1): 153.

Young, Adam M. H., Natsuhiko Kumasaka, Fiona Calvert, Timothy R. Hammond, Andrew Knights, Nikolaos Panousis, Jeremy Schwartzentruber, et al. 2019. “A Map of Transcriptional Heterogeneity and Regulatory Variation in Human Microglia.” *Genomics*. bioRxiv.

Decision Letter, first revision:

Our ref: NG-A56637R

23rd Jun 2021

Dear Dr. Raj,

Thank you for submitting your revised manuscript "Atlas of genetic effects in human microglia transcriptome across brain regions, aging, and disease pathologies" (NG-A56637R). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Genetics, pending minor revisions to comply with our editorial and formatting guidelines.

The current version of your manuscript is in a PDF format, please email us a copy of the file in an editable format (Microsoft Word or LaTex)-- we can not proceed with PDFs at this stage.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements in about a week. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Thank you again for your interest in Nature Genetics Please do not hesitate to contact me if you have any questions.

Sincerely,

Safia Danovi
Editor
Nature Genetics

Reviewer #1 (Remarks to the Author):

The authors have tried their best to address our prior concerns. I have no further comment.

Reviewer #2 (Remarks to the Author):

de Paiva Lopes and colleagues use a large set of CNS samples to map cis-eQTLs and cis-sQTLs in microglia that drive region-specific ene expresion and some CNS-pathology-associated gene expression.

The authors addressed allthe points I raised and included pertinentst additional data (CyTOF/scRNAseq) to address my concerns. I.e. sample purity, gene induction by isolation protocols, etc.

This is an impressive study, #samples/potential impact wise and will be a highly cited manuscript, deserving of publication in Nature Genetics. I have no further reservations, and support acceptance.

Reviewer #3 (Remarks to the Author):

The authors have responded to each of my concerns and made appropriate changes where possible - I think the review process has improved the manuscript, but I now think it is ready to publish without further amendments. From my perspective I have no other comments - it is a great piece of work.

Final Decision Letter:

In reply please quote: NG-A56637R1 Raj

19th Oct 2021

Dear Dr. Raj,

I am delighted to say that your manuscript "Genetic analysis of the human microglia transcriptome across brain regions, aging and disease pathologies" has been accepted for publication in an upcoming issue of Nature Genetics.

Prior to setting your manuscript, we may make minor changes to enhance the lucidity of the text and with reference to our house style. We therefore ask that you examine the proofs most carefully to ensure that we have not inadvertently altered the sense of your text in any way.

Once your manuscript is typeset and you have completed the appropriate grant of rights, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately.

Your paper will be published online after we receive your corrections and will appear in print in the next available issue. You can find out your date of online publication by contacting the Nature Press Office (press@nature.com) after sending your e-proof corrections. Now is the time to inform your Public Relations or Press Office about your paper, as they might be interested in promoting its publication. This will allow them time to prepare an accurate and satisfactory press release. Include your manuscript tracking number (NG-A56637R1) and the name of the journal, which they will need

when they contact our Press Office.

Before your paper is published online, we shall be distributing a press release to news organizations worldwide, which may very well include details of your work. We are happy for your institution or funding agency to prepare its own press release, but it must mention the embargo date and Nature Genetics. Our Press Office may contact you closer to the time of publication, but if you or your Press Office have any enquiries in the meantime, please contact press@nature.com.

Acceptance is conditional on the data in the manuscript not being published elsewhere, or announced in the print or electronic media, until the embargo/publication date. These restrictions are not intended to deter you from presenting your data at academic meetings and conferences, but any enquiries from the media about papers not yet scheduled for publication should be referred to us.

Please note that Nature Genetics is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. Find out more about Transformative Journals

<B>Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. For submissions from January 2021, if your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route our standard licensing terms will need to be accepted, including our self-archiving policies. Those standard licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

Please note that Nature Research offers an immediate open access option only for papers that were first submitted after 1 January, 2021.

You will not receive your proofs until the publishing agreement has been received through our system.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. Please let your coauthors and your institutions' public affairs office know that they are also welcome to order reprints by this method.

If you have not already done so, we invite you to upload the step-by-step protocols used in this manuscript to the Protocols Exchange, part of our on-line web resource, natureprotocols.com. If you complete the upload by the time you receive your manuscript proofs, we can insert links in your article that lead directly to the protocol details. Your protocol will be made freely available upon publication of your paper. By participating in natureprotocols.com, you are enabling researchers to more readily reproduce or adapt the methodology you use. Natureprotocols.com is fully searchable, providing your protocols and paper with increased utility and visibility. Please submit your protocol to https://protocolexchange.researchsquare.com/. After entering your nature.com username and password you will need to enter your manuscript number (NG-A56637R1). Further information can be found at https://www.nature.com/nprot/.

Sincerely,

Safia Danovi
Editor
Nature Genetics