nature research

Corresponding author(s): Towfique Raj

Last updated by author(s): Aug 3, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	All custom code is available in the GitHub repository: https://github.com/RajLabMSSM/MiGA_public_release RNASeq was processed using the RAPiD pipeline (v19.09.1). RAPiD aligns samples to the hg38 genome build using STAR (v2.7.2a) software with GENCODE v30 transcriptome reference and calculates quality control metrics using Picard. Estimated transcript abundance was obtained using RSEM (v1.3.1) and transcripts were summed to the gene level with tximport. # Softwares used for downstream analysis Sources of variation from RNA-seq data: variancePartition (v1.17.7) Differential Expression Analysis: R package "Differential expression for repeated measures" (DREAM) from the variancePartition (v1.17.7) package. Pathway analysis: Ingenuity Pathway Analysis (IPA) Genotype, genetic association and QC: bcftools (v1.9) and vcftools (v0.1.15) Imputing data: Michigan Imputation Server v1.4.1 (Minimac 4) https://imputationserver.sph.umich.edu/index.html Mapping QTL: tensorQTL (v1.0.2) Calling local splicing events or intronic excision: Regtools (v0.5.1) and Leafcutter (v0.2.8) Meta-analysis: METASOFT (v2.0.1) and mashR (v0.2-11) Colocalization: R package COLOC (v3.2-1) Fine-mapping: echolocatoR https://rajlab.shinyapps.io/Fine_Mapping_Shiny Analysis of transcription factor motifs: motifbreakR with HOCOMOCO database All plots were created using ggplot in R (version 3.6.0), with ggrepel, ggfortify, patchwork, and ggbio for additional layers of visualization.

All software is freely available.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

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 at https://dss.niagads.org/datasets/ng00105/; Accession number: NG00105.v1). The user will need to log into NIAGADS Data Access Request (DAR) to start an application. All differential expression, gene lists, and fine-mapping results are present as supplementary tables. The GWAS fine-mapping results are available from the echoLocatoR Shiny application at https://rajlab.shinyapps.io/Fine_Mapping_Shiny. Full nominal and permuted eQTL and sQTL summary statistics per brain region are available from Zenodo at https://doi.org/10.5281/zenodo.4118605 (eQTL) and https://doi.org/10.5281/zenodo.4118403 (sQTL). Results for eQTL and sQTL meta-analysis (mashR and METASOFT) and colocalization (COLOC) are available from Zenodo at https://doi.org/10.5281/zenodo.4118676.

All code to perform the analysis is available at https://github.com/RajLabMSSM/MiGA_public_release.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed for this study. The number of samples (n = 255 after quality control), was determined by the availability of high quality brain tissues to isolate microglia.
Data exclusions	In total, 59 out of 314 samples were excluded due to insufficient RNASeq quality or insufficient sample size by brain region. Supplementary Figure 1 shows a flowchart of quality control, and all measures applied are available Online and in the Methods section.
Replication	Our results were successfully replicated in external datasets. Regional heterogeneity analysis: van der Poel et al. (2019) dataset of white and grey matter microglia was used to validate the results of differentially expressed genes. We performed Fisher's exact test, and the results are highlighted in Figure 2E. Age-related analysis: The results were replicated in two different datasets, Galatro et al. (2017) and Peters et al. (2015). Galatro et al. is a dataset of human microglia samples, 49 healthy controls with ages of donors between 31 and 102 years old. Peters et al. performed a whole-blood gene expression meta-analysis in 14,983 individuals. Both comparisons showed significant overlap (p-value <0.05) from Fisher's exact test (Figure 3E and Supplementary Figure 7, respectively). Genetic regulatory effects in microglia: We compared our eQTL results with four other external eQTL datasets, including microglia (Young et al. 2019), monocytes (Navarro et al. 2020, Fairfax et al. 2014), and bulk brain dorsolateral prefrontal cortex (DLPFC - Ng et al. 2017) using the q-value π1 metric. We found the highest sharing between MiGA and the Young et al. microglia, and sharing with bulk DLPFC eQTLs was the lowest (Figure 4D). Colocalization and Fine-mapping: We downloaded public GWAS summary statistics for Alzheimer's Disease (Kunkle et al. 2019, Marioni et al. 2018, Lambert et al. 2013, Jansen et al. 2019), Parkinson's Disease (Nalls et al. 2019), Schizophrenia (Psychiatric Genomics Consortium 2014), and Multiple Sclerosis (Multiple Sclerosis Genetics Consortium 2019). We performed colocalization with these GWAS and our own QTLs and the publicly available QTL datasets described above. Several colocalized genes in our microglia eQTLs were replicated in the different published QTLs. We present full plots for all colocalizations in Figures 5 and Supplementary Figures 13-22. Single-cell: Considering the sparsity of this type of material and the consistency when comparing single-cell experiments with triplicates, we ran single experiments.
Randomization	No allocation into groups was performed. After death, pathological assessment was performed to measure post mortem inverval (PMI) and pH. Biological covariates also include diagnosis, region of the brain from where microglia were isolated, sex, age and cause of death from each donor.
Blinding	The investigators were not blinded for group allocation (diagnosis, sex, age etc.) during data analysis, since adjustment for these factors was necessary in the data analyses.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Primary human CD11b+ microglia were isolated at autopsy from 100 donors with neurological and psychiatric diseases, as well as unaffected subjects (controls), generating a total of 255 samples from four brain regions: medial frontal gyrus (MFG), superior temporal gyrus (STG), subventricular zone (SVZ) and thalamus (THA). The age range of the 100 donors is between 21 and 103 years old; 58 of them were female. All analyses were adjusted for age, gender, and other covariates. The details are described in Figure1 and Supplementary Table 1.
Recruitment	The brain banks are non-profit organizations that collect human brain tissue of donors with neurological, psychiatric disorders, but also non-diseased donors. The brain banks collaborate with several hospital, in order to approach the participants of clinical cohorts with the requests to consider brain donation. The included donors may not be a population-based study group, but a selected subgroup of neurological, psychiatric and non-diseased controls that are more likely to visit the hospital. Therefore, the results of this study should be carefully generalized to the total population because the design might be prone to specific selection biases.
Ethics oversight	Post-mortem brain samples were obtained from the Netherlands Brain Bank (NBB) and the Neuropathology Brain Bank and Research CoRE at Mount Sinai Hospital. The permission to collect human brain material was obtained from the Ethical Committee of the VU University Medical Center, Amsterdam, The Netherlands, and the Mount Sinai Institutional Review Board. Ethics approval and consent to participate (from Mount Sinai): All autopsies were performed with written consent from the legal next-of-kin. All tissue samples were obtained de-identified under approved Institutional Review Board (IRB) protocols at the Mount Sinai Hospital. The study was performed under IRB-approved guidance and regulations to keep all patient information strictly de-identified. All research conformed to the principles of the Helsinki Declaration.

Note that full information on the approval of the study protocol must also be provided in the manuscript.