Genetic analysis of the human microglial transcriptome across brain regions, aging and disease pathologies

Katia de Paiva Lopes, Gijsje J. L. Snijders, Jack Humphrey, Amanda Allan, Marjolein A. M. Sneeboer, Elisa Navarro, Brian M. Schilder, Ricardo A. Vialle, Madison Parks, Roy Missall, Welmoed van Zuiden, Frederieke A. J. Gigase, Raphael Kübler, Amber Berdenis van Berlekom, Emily M. Hicks, Chotima Böttcher, Josef Priller, René S. Kahn, Lot D. de Witte and Towfique Raj

Microglia have emerged as important players in brain aging and pathology. To understand how genetic risk for neurological and psychiatric disorders is related to microglial function, large transcriptome studies are essential. Here we describe the transcriptome analysis of 255 primary human microglial samples isolated at autopsy from multiple brain regions of 100 individuals. We performed systematic analyses to investigate various aspects of microglial heterogeneities, including brain region and aging. We mapped expression and splicing quantitative trait loci and showed that many neurological disease susceptibility loci are mediated through gene expression or splicing in microglia. Fine-mapping of these loci nominated candidate causal variants that are within microglia-specific enhancers, finding associations with microglial expression of USP6NL for Alzheimer’s disease and P2RY12 for Parkinson’s disease. We have built the most comprehensive catalog to date of genetic effects on the microglial transcriptome and propose candidate functional variants in neurological and psychiatric disorders.

Microglia, the myeloid immune cells of the brain, are a cell type of compelling interest in the pathogenesis of several brain disorders. Microglia play critical roles in inflammatory responses, regulation of brain homeostasis, neurodevelopment and neurogenesis. Microglia are highly dynamic cells that are strongly influenced by different environmental signals, which result in distinct phenotypes and functions across brain regions. In addition, microglial functions vary across different ages, disease pathologies and between the sexes. For decades, changes in microglial density, morphology and transcriptional state have been observed in postmortem brain tissue of patients with neurological and psychiatric disorders. This was initially suggested to reflect a response of the immune system to underlying disease processes. However, recent evidence from genome-wide association studies (GWAS) and other follow-up analyses suggested that a proportion of the genetic risk of neurological and psychiatric diseases acts through microglial cells. As the microglial cells of the nervous system, microglia may therefore play a causal role in disease.

To better understand this potential causal role of microglia in brain pathology and identify microglia-related targets for treatment, there is a critical need to identify which gene(s) are influenced by disease-associated genetic risk variants in microglia. This is a complicated task because most of the common variants that have been identified are located outside protein-coding regions. These variants influence gene expression through complex regulatory mechanisms, such as altering enhancer activity, which often affect a gene beyond the nearest gene. Applying a combination of genetic and transcriptomic analyses on the same samples of a set of different donors, one can elucidate which gene is under the influence of which genetic variant by calling quantitative trait loci (QTLs). Investigations of QTLs in microglia have been limited by the availability of microglial samples from the number of individuals required to perform well-powered genomic analyses. Recently, Young et al. constructed expression QTLs (eQTLs) in primary human microglia (n = 93 individuals per samples) and detected 401 eQTLs, some of which colocalized with Alzheimer’s disease (AD) loci, including BIN1. However, the microglial transcriptome is highly heterogeneous compared to other cell types, so larger sample sizes are needed to further identify statistically significant eQTLs. In addition, it is increasingly clear that genetic risk can also be mediated through messenger RNA splicing. For instance, the CD33 locus in AD influences CD33 splicing, resulting in isoforms with different biological and likely pathological functions.

In the present study, we describe the Microglia Genomic Atlas (MiGA), a genetic and transcriptomic resource consisting of 255 primary human microglia samples isolated ex vivo from 4 different brain regions of 100 individuals with neurodegenerative, neurological or neuropsychiatric disorders, as well as unaffected controls.
(Fig. 1). We performed systematic analyses to investigate sources of microglial heterogeneity, including brain region, age and sex. We further performed eQTL and splicing QTL (sQTL) analyses in each region and performed a meta-analysis across the four regions to increase our discovery power. We then performed colocalization and used fine-mapping and microglia-specific epigenomic data to prioritize genes and variants that influence neurological disease susceptibility through gene expression and splicing in microglia. With this approach, we built the most comprehensive resource to date of cis-genetic effects on the microglial transcriptome and propose the underlying molecular mechanisms of potentially causal functional variants in several brain disorders.

**Results**

**Biological factors driving the microglial transcriptome.** We isolated microglia from different brain regions of 115 donors. Details of the donors and quality controls are described in the Supplementary Note, Supplementary Figs. 1–4 and Supplementary Table 1. We explored the variation of a wide range of biological factors in driving the human microglia transcriptome before and after controlling for technical confounders (Supplementary Figs. 5 and 6). Using principal component analysis (PCA), we observed no clear separation by any factor after regressing out the technical factors, with the exception of age (Supplementary Fig. 7). Sex explained little variance (Fig. 2a) and we observed no differentially expressed genes (DEGs) between males and females (Supplementary Table 2). Using a linear mixed model to estimate the variance explained for a set of factors per gene, we found that donor identity explained the most variance per gene (mean = 13.5%) (Fig. 2a). Brain regions explained comparatively little variance overall (mean = 2.95%) but we identified a subset of genes that were strongly variable between regions. We performed pairwise comparisons of differential gene expression between each pair of regions, accounting for shared donors in a linear mixed model (Fig. 2b and Supplementary Tables 3–8). The largest number of DEGs (false discovery rate (FDR) <0.05; log fold change >1) were between the subventricular zone (SVZ) and the two cortical regions (609 in the medial frontal gyrus (MFG), 909 in the superior temporal gyrus (STG)), whereas comparing the STG to the MFG revealed the fewest (6 genes). We compared our findings to a published dataset of white and gray matter microglia and found small but significant overlaps with our MFG versus STG to the MFG revealed the fewest (6 genes). We compared our findings to a published dataset of white and gray matter microglia and found small but significant overlaps with our MFG versus SVZ comparison (upregulated odds ratio (OR) = 18.4; P < 1 × 10^{-45}; downregulated OR = 4.83; P = 9 × 10^{-4}; Fisher’s exact test; Fig. 2c).

We then performed k-means clustering of the genes found in the pairwise comparisons. We identified k = 4 as the optimal clustering partitioning after minimizing the total within-cluster sum of squares (Fig. 2d). Cluster 1 contained genes that were upregulated in the cortical regions compared to subcortical regions, such as P2RY12, CD36 and MRCl, and was enriched in genes that were downregulated in the AD brain and in response to in vitro culture. Cluster 2 contained genes that were downregulated in the cortex compared to the subcortical brain (for example, FCER1A, IL15, RGS1). Cluster 3 contained genes specifically downregulated in the SVZ (for example, CX3CL1, CCR2, FCGR3B) and cluster 4 contained genes upregulated in SVZ (for example, IL10, CLU, CD83) compared to the other 3 regions (Fig. 2e). We found that genes implicated in inflammatory processes were highly expressed in cluster 2 (Fig. 2f), whereas genes related to the homeostatic functions of microglial cells were mainly present in cluster 1. Cluster 4 included genes that were involved in biological functions related to hormonal signaling and interferon response (Fig. 2f).

Analysis of upstream regulators of the four clusters using ingenuity pathway analysis (IPA) was inconclusive (Supplementary Table 9). We overlapped the region-specific genes with gene sets altered after stimulation with lipopolysaccharide (LPS) or interferon-γ (IFN-γ; generated in house), after in vitro culture, and in microglia derived from the brains of patients with AD compared to controls. Cluster 1 genes were enriched in genes that were downregulated in the AD brain and in response to in vitro culture. Cluster 2 genes were significantly enriched for genes upregulated after in vitro culture and in AD-derived microglia. The genes in clusters 2, 3 and 4 showed enrichment, with LPS-responsive genes in both directions (Fig. 2g and Supplementary Table 9).

We examined changes in splicing between microglial regions using a differential transcript usage (DTU) framework; 176 transcripts in 132 genes had evidence of DTU (log OR >1; empirical FDR <0.1), with most transcripts coming from comparisons with the SVZ (Extended Data Fig. 1a). Thirty-one DTU genes were also differentially expressed between pairs of regions (OR = 5.47, P = 2.9 × 10^{-13}; Fisher’s 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 (Extended Data Fig. 1b). The regional DTU gene set includes genes involved in mitochondrial functions, glucocorticoid receptor signaling pathways and host defense against infections (Extended Data Fig. 1c), pathways also observed in the regional expression analysis.

We explored the effect of diagnosis on the microglial transcriptome and detected 24 genes, such as MCF2 and ALDH3B1, differentially expressed in the group with dementia compared to controls (FDR <0.05; Supplementary Table 10). No significant gene expression changes were found for Parkinson’s disease (PD), major depressive disorder and bipolar disorder (BD)/schizophrenia (SCZ) (Supplementary Tables 11–13). To assess the effect of aging on the microglial transcriptome, we fitted a linear mixed model accounting for shared donors across all four regions. We observed 1,693 genes (338 upregulated, 1,355 downregulated at FDR <0.05) associated with the chronological age of individuals (Fig. 3a and Supplementary Table 14). Similarly, we found 225 transcripts from 150 genes exhibiting DTU with age (FDR <0.1; Extended Data Fig. 2a), where the balance between a long and short isoform shifted over age (Extended Data Fig. 2b). Thirty-six of these DTU genes also showed an association with age at the gene expression level (OR = 3.47, P = 7 × 10^{-4}; Fisher’s exact test). Genes upregulated in aging were significantly enriched for several Gene Ontology (GO) biological processes including lipid metabolism, immune responses such as natural killer cell and IFN signaling, and phagosome formation (Fig. 3b). The downregulated genes were significantly enriched for cell motility, polarity, interleukin-6 (IL-6) cytokine signaling (Fig. 3b) and for genes also downregulated after in vitro culture and in AD-derived microglia (Fig. 3c). Genes associated with aging DTU were enriched in similar functions (Extended Data Fig. 2c).

We next used gene sets prioritized by transcriptome-wide association study (TWAS) in different diseases (Supplementary Table 15). The upregulated genes in chronological aging showed overrepresentation for genes in AD (for example, MS4A6A, FCER1G and...
Regional heterogeneity

Medial frontal gyrus (n = 77)

Superior temporal gyrus (n = 63)

Subventricular zone (n = 55)

Thalamus (n = 60)

Genotyping

MiGA

Regional heterogeneity

Genetic drivers (90 donors, 216 samples)

Integration with GWAS:
colocalization and functional fine-mapping
(AD, PD, MS, BD, SCZ)

Prioritization of putative:
- Disease genes
- Causal variants

Available datasets for replication

RNA-seq

Genotyping

Microglia isolation

n = 100 donors

Ctrl AD BD PSP MDD SCZ PD

n = 255 samples

Age

Expression

GG GA AA

mRNA isoforms

Sex

Female Male

Available datasets for replication

Integration with GWAS:
colocalization and functional fine-mapping
(AD, PD, MS, BD, SCZ)

Prioritization of putative:
- Disease genes
- Causal variants

Age

Expression

GG GA AA

Genomic position

MFG

STG

THA

SVZ

MTG

Age-related heterogeneity

Regional heterogeneity

Available datasets for replication

Integration with GWAS:
colocalization and functional fine-mapping
(AD, PD, MS, BD, SCZ)

Prioritization of putative:
- Disease genes
- Causal variants
CR1) or PD (for example, BST1, PTPN22 and TNFSF13) GWAS loci but not for genes in SCZ or BD (Fig. 3d). We replicated our findings using an external microglial aging dataset from the parietal cortex and from peripheral blood (Supplementary Fig. 8). The number of genes that overlapped between the datasets was small but significant (upregulated genes OR = 23.4, $P < 1 \times 10^{-16}$; downregulated genes OR = 5.97, $P < 1 \times 10^{-14}$; Fisher’s exact test; Fig. 3e).

It is not known whether the impact of aging on the microglial transcriptome is uniform throughout the human brain. Although most genes showed concordant effect size and direction across regions (Fig. 3f), 91 genes demonstrated age–region relationships after fitting an interaction term model (FDR < 0.05; Supplementary Fig. 9 and Supplementary Table 16). Thirty-five genes (for example, MRC1, CD24) changed specifically in the SVZ and not in the other regions ($R^2_{\text{adj}} > 3$ times the interquartile range (IQR); Fig. 3g and Supplementary Fig. 9). Together, our results indicate that the microglial phenotype ages in a generally uniform manner across brain regions, with a distinct aging trajectory observed in a minority of genes.

Genetic regulatory effects in microglia. We performed cis-eQTL and cis-sQTL analyses in primary human microglia from four different brain regions. After quality control, 216 samples from 90 individuals of European ancestry were used for the analysis (Supplementary Fig. 10). In the region-specific analysis, we observed between 67 and 199 genes with a cis-eQTL (eGenes) and 253–426 genes with a cis-sQTL (sGenes) per region (FDR < 0.05; Supplementary Table 17 and Supplementary Fig. 11). The cis-QTL discovery was highly correlated with the sample size for each region (Spearman’s $\rho = 0.8$ for eQTLs and $\rho = 1$ for sQTLs), contributing to the low number of eQTLs detected in the region-by-region analysis. Therefore, we performed a meta-analysis across all four regions using the multivariate adaptive shrinkage (mashR) v.0.2-11 method to increase power and assess shared QTLs between regions. In total, we identified 3,611 eGenes and 4,614 sGenes at a local false sign rate (LFSR) $\leq 0.05$ in at least 1 region (Fig. 4a and Supplementary Tables 18 and 19).

We observed a high degree of eQTL sharing (effect estimates that are in the same direction and are of similar sizes within a factor of 2) between MFG and STG (72%), as expected, given that these two cortical regions have similar gene expression patterns (Fig. 4b, upper triangle). Microglia from the SVZ exhibited lower pairwise sharing of eQTLs with other regions, with the lowest sharing by magnitude observed between the SVZ and MFG (41%), which is consistent with observed transcriptomic differences between these two regions. For sQTLs, we found overall higher regional sharing effects compared to eQTLs but still following the same trends as for eQTLs (Fig. 4b, lower triangle). In addition, while most of the eQTLs were shared across regions, we identified 1,791 (49.6%) eQTLs with a stronger effect in 1 region than in any other (LFSR $< 0.05$ and greater than twofold effect size in 1 region compared to others). Microglia from the SVZ had the most region-specific effects with 1,045, most likely because the transcriptomic profile of this region is most distinct (Supplementary Table 20). We include examples of shared and region-specific eQTLs (Fig. 4c).

To assess eQTL reproducibility and cell type-specificity, we compared the MiGA eQTLs with four other external eQTL datasets, including microglia, two monocytes and bulk brain dorsolateral prefrontal cortex (DLPFC) using the Storey’s $\pi$ metric. We found that eQTL sharing was both cell type- and region-dependent (Fig. 4d), with the highest sharing between MiGA and Young et al. microglia ($\pi = 0.81–0.86$) but with a lower sharing in the SVZ ($\pi = 0.51$). Sharing with monocyte eQTLs was generally slightly lower than with microglia and sharing with bulk DLPFC eQTLs was lowest. Together, these results highlight shared genetic regulation between microglia and monocytes, which is only partly captured in whole-tissue brain data.

We performed a cross-study eQTL meta-analysis (MiGA, Young et al., MyND and Fairfax et al.) using METASOFT to assess the sharing of effects between distinct cell types. We focused on genes associated with AD, PD, SCZ and multiple sclerosis (MS) (Fig. 4f). Generally, directions of effect between monocytes and microglia were concordant (Supplementary Fig. 12), with the exception of CASS4. eQTLs for CASS4 were significant in both MiGA and monocytes (MyND) but with opposite directions of effect (Fig. 4g), suggesting that the causative variant is located in a complex regulatory element where both enhancing and repressing mechanisms are at play.

Genetic effects in microglia-mediated neurological disease. We next explored whether disease-associated genetic variants may potentially act through microglia eQTLs or sQTLs using the coloc R package v.3.2–1 and publicly available GWAS summary statistics for AD, PD, SCZ, BD and multiple sclerosis (MS). We compared our MiGA QTLs to the same set of published microglia, monocytes and bulk brain tissue QTLs as before. AD and PD had the highest number of colocalizing loci in each QTL dataset, compared to the other diseases (Fig. 5a and Supplementary Table 21), with $\sim 30\%$ of loci containing at least 1 colocalized gene, depending on the stringency of the H4 posterior probability (PP4), with lower proportions observed in BD, SCZ and MS.

We then compared different QTL datasets to find shared evidence of colocalization at the level of individual genes within a GWAS locus (Fig. 5b–e). The sharing between our microglia and previously published microglia was low (Fig. 5b), with only a
ARTICLES

NATuRE GENETiCS

VOL 54 | JANUARY 2022 | 4–17 | www.nature.com/naturegenetics

Articles

NATuRE GENETiCS

Enrichment in human microglia RNA-seq gene sets

Culture downregulation
AD-derived microglial downregulation
LPS stimulation upregulation
LPS stimulation downregulation
Culture upregulation
AD-derived microglia upregulation
LPS stimulation upregulation
LPS stimulation downregulation

FCGR3B (cluster 3)
RGS1 (cluster 2)
van der Poel 5
SUCNR1, CCL2, MRC1

Cluster 1
Cluster 2
Cluster 3
Cluster 4

Region

MFG
STG
SVZ
THA

IPA

T2/T1 pathway
Drug metabolism
Reelin signaling
T17 activation pathway
Aryl hydrocarbon receptor signaling
PI3K/AKT signaling
Stellate cell activation
Agranulocyte adhesion and diapedesis
Cell signaling
Aldosterone signaling in epithelial cells
Diacylglycerol signaling
IFN induction and antiviral response

-2
0
2
4
6
8
10

-\log_{10} (q)

Brain region

MFG
STG
SVZ
THA

Cluster 1
(n = 333)
Cluster 2
(n = 108)
Cluster 3
(n = 350)
Cluster 4
(n = 296)

Enrichment in human microglia RNA-seq gene sets

Culture downregulation
AD-derived microglial downregulation
LPS stimulation upregulation
LPS stimulation downregulation
Culture upregulation
AD-derived microglia upregulation
LPS stimulation upregulation
LPS stimulation downregulation

-2
0
5
10
15
20

-\log_{10} (q)

Cluster 1
Cluster 2
Cluster 3
Cluster 4
Fig. 3 | Age-related analysis. 

a, Heatmap of 1,693 genes associated with age (FDR < 0.05). Each gene (row) was plotted as a z-score of median expression averaged first by donor (across multiple regions) and then by age quintiles with 20 donors each. Genes are ordered by Ward's hierarchical clustering. 

b, IPA of age-related genes. Only significantly enriched terms are shown (q < 0.05). 

c, Enrichment analysis with curated human microglia RNA-seq gene sets33,34 using a one-sided Fisher's exact test at Bonferroni-adjusted P < 0.05. 

d, Genes associated with age show overrepresentation for TWAS-prioritized genes for AD or PD but not for genes in SCZ or BD. P value based on one-sided Fisher's exact test. 

e, Replication analysis with an independent dataset of human microglial samples, 49 healthy controls with ages between 31 and 102 years old16. The asterisk indicates significant enrichment by a one-sided Fisher's exact test (upregulated genes OR = 23.4, P < 1 × 10^{-16}; downregulated genes OR = 5.97, P < 1 × 10^{-16}). Selected overlapping genes are highlighted. 

f, Scatter plot showing the size correlation of age-related genes by brain region. Only genes that were significant (FDR < 0.05) in at least one region are shown. The x-axis shows the beta values in the MFG and the y-axis shows the betas in other brain regions (STG, SVZ and THA). 

g, Scatter plot showing the z-score-transformed residual expression for selected genes (MRC1 and MS4A6A) by age and brain region.
few known loci in AD and PD (BIN1, PICALM, CHRNβ1), presumably due to lower power in the Young et al.\textsuperscript{26} data compared to our multi-tissue meta-analysis. Overall, 11\% of MiGA eQTL colocalizations could be reproduced in the Young et al.\textsuperscript{26} data and 15\% of the Young et al. colocalizations could be found in the MiGA data, at a relaxed PP4 > 0.5, whereas sharing between the
Neurological disease loci regulate microglial gene expression.

We next examined whether microglial eQTLs that colocated with disease GWAS loci were due to genetic variation within microglia-specific regulatory regions. As outlined further in the Supplementary Note, we found that 10 out of 17 genes that colocaled in AD, 8 out of 18 in PD, 4 out of 9 in SCZ and 3 out of 17 in MS included SNPs that overlapped with microglial enhancers (Fig. 6a, Extended Data Fig. 7 and Supplementary Fig. 20). This approach allowed us to prioritize disease loci that likely act on disease risk by modulating gene expression specifically in microglia. In this article, we discuss two examples.

The ECHDC3 locus has been associated with AD risk in several GWAS\(^{24,46,49}\). The lead SNP rs7920721 sits in an intergenic region that separates two genes, ECHDC3 and USP6NL. Previous analyses prioritized ECHDC3 since it is upregulated in the post-mortem brains of individuals with AD\(^{13,44}\), an eQTL for ECHDC3 was seen in whole blood\(^{32}\), although it did not colocale with the GWAS SNP\(^{32}\).

USP6NL harbors an eQTL observed in all four microglial regions, with the lead QT SNP rs7912495-G increasing USP6NL expression (Fig. 6b). The meta-analyzed eQTL colocalizes with the ECHDC3 locus in all 4 AD GWAS used in this study, with the highest PP (0.95) seen in Marioni et al.\(^46\) (Fig. 6c and Extended Data Figs. 3 and 4). No colocalization was observed in any other QT dataset, although we note that USP6NL was expressed fivefold higher in microglia than in monocytes (MiGA median transcripts per million mapped reads (TPM) = 15.77; MyND\(^10\) median TPM = 3.13). Fine-mapping of the ECHDC3 locus suggested three additional SNPs as well as the lead GWAS SNP (rs7920721) and lead QT SNP (rs7912495). The GWAS lead SNP and the QT lead SNP were in moderate linkage disequilibrium (LD) \((r^2 = 0.65)\), as were 2 of the 3 fine-mapped SNPs (Fig. 6d and Supplementary Table 22). Of the five SNPs of interest, four overlapped a microglia-specific enhancer. Using proximity ligation-assisted chromatin immunoprecipitation followed by sequencing (ChIP–seq; proximity ligation-assisted ChIP–seq (PLAC–seq)) data\(^{48}\), we observed that the overlapping microglial enhancer region had extended long-range connections to regions overlapping the USP6NL promoter and gene body. Notably, there was no colocalization of the upstream ECHDC3 gene in any tested cell type, suggesting that USP6NL is the AD risk gene at this locus. The lead QT SNP rs7912495-G increases AD risk \((\beta = -0.0492; P = 6.8 \times 10^{-19})\) (ref. \(^{46}\)); we suggest that it achieves this through upregulating USP6NL expression in microglia. 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 (Supplementary Table 23).

The MED12L locus was identified in the latest PD GWAS\(^{47}\). The lead SNP rs11707416 sits within a large intron of the MED12L gene, which overlaps with several smaller genes, one of which is 2PRY12. A previous study prioritized 2PRY12 at this locus due to an overlap with eQTLs in the blood and brain\(^{48}\).
Top microglial colocalizations (PP4 > 0.7)

**Gene** | **Microglia** | **Monocytes** | **Brain** | **eQTL** | **sQTL**
--- | --- | --- | --- | --- | ---
BIN1 | 0.99 | 0.05 | 0.99 | 0.93 | 0.79
CASS4 | 0.71 | 0.04 | 0.98 | 0.09 | 0.08
CD33 | 0.51 | 0.84 | 0.38 | 0.87 | 0.54
ACO10160.1 | 0.08 | 0.08 | 0.08 | 0.08 | 0.09
USP9NL | 0.06 | 0.06 | 0.06 | 0.06 | 0.06
AL5261.1 | 0.08 | 0.06 | 0.06 | 0.06 | 0.06
EPHA1-AS1 | 0.08 | 0.06 | 0.06 | 0.06 | 0.06
ZNF661 | 0.98 | 0.05 | 0.99 | 0.93 | 0.79
MS446A | 0.97 | 0.05 | 0.99 | 0.93 | 0.79
NME8 | 0.96 | 0.05 | 0.99 | 0.93 | 0.79
PICLAM | 0.97 | 0.05 | 0.99 | 0.93 | 0.79
ZNF252 | 0.99 | 0.05 | 0.99 | 0.93 | 0.79
NUP98 | 0.99 | 0.05 | 0.99 | 0.93 | 0.79

Colocalized loci per GWAS

**Dataset** | **Microglia** | **Monocytes** | **Brain** | **eQTL** | **sQTL**
--- | --- | --- | --- | --- | ---
AD | 0.99 | 0.05 | 0.99 | 0.93 | 0.79
PD | 0.99 | 0.05 | 0.99 | 0.93 | 0.79
BD | 0.99 | 0.05 | 0.99 | 0.93 | 0.79
SCZ | 0.99 | 0.05 | 0.99 | 0.93 | 0.79
MS | 0.99 | 0.05 | 0.99 | 0.93 | 0.79

**Location**

**CCAR2** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**B7T1** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**TUFM** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**SULT1A2** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**SULT1A1** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**SPNS1** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**ECP3** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**ATXN2** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**CTSB** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**AC020614.1** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**FAM49B** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**CNIH** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**NUP98** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**KLH7.6.2** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**WDR6** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**ARH1** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**TIPK2** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**PARK** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
**TMEM163** | 0.95 | 0.08 | 0.06 | 0.06 | 0.06
P2RY12 is an eQTL (lead SNP rs3732765) identified in the METASOFT meta-analysis, with the lead QTL SNP rs3732765-A decreasing P2RY12 expression (Fig. 6e). The eQTL colocalizes with the PD GWAS MED12L locus (PP4 = 0.88; Fig. 6f). Colocalization was also observed with a P2RY12 eQTL in the DLPFC (PP4 = 0.93; Fig. 5d and Extended Data Fig. 5). Fine-mapping revealed that the lead GWAS SNP rs11707416 was a causal SNP by multiple fine-mapping tools (a consensus SNP) and is in perfect LD (r² = 1) with the lead QTL SNP rs3732765 (Fig. 6g). In addition, there were four other SNPs prioritized by fine-mapping, two of which were in perfect or very high LD with the lead QTL SNP. Of the seven SNPs in the set, five overlapped a microglia-specific enhancer region on either side of the P2RY12 promoter. PLAC–seq revealed long-range connections between the enhancer and the P2RY12 promoter but not to MED12L (Fig. 6g). No colocalization was observed with any MED12L QTL. Altogether this suggests that P2RY12 is the causal gene at the locus. The lead QTL SNP rs3732765-A decreases PD risk (β = −0.06; P = 2.4 × 10⁻⁴) (ref. 39) and we suggest that it acts through downregulating P2RY12 expression in microglia. The effects on transcription factor binding were predicted for rs11707416, rs41366744, rs4680405 and rs62285879, again for multiple motifs (Supplementary Table 23).

sQTLs identify additional disease-associated 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 (Supplementary Table 21). We highlight two examples of sQTLs associated with AD and identify key challenges ahead for the interpretation of such events. The CD33 risk locus has been implicated in susceptibility to AD12. Previous analyses in peripheral monocytes found an association between the lead GWAS SNP rs3865444 and the inclusion of CD33 exon 2 (ref. 40). 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 (Fig. 7a,c,e). Another sQTL was identified in MS4A6A. The MS4A gene cluster is a gene-dense region spanning 600 kilobases (kb) and containing 12 genes. We observed colocalization with eQTLs and sQTLs in MS4A6A, as well as eQTLs in MS4A4A and MS4A4E (Fig. 5f). In MiGA, we observed colocalization solely with sQTLs in MS4A6A (Fig. 7b,d,f). We overlaid all 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, two transcripts containing this intron had a premature polyadenylation site. rs2162254-A was associated with an increased usage of this intron, which may result in increased production of the shorter isoforms; this could have a downstream consequence on MS4A6A protein function.

Discussion
In this study, we present the MiGA, a comprehensive genetic and transcriptomic resource consisting of primary human microglial samples across multiple disease pathologies. We demonstrate that transcriptional heterogeneity in human microglia varies between brain regions and across aging. We generated a catalog of eQTLs and sQTLs in microglia and thereby validated and extended the list of disease genes and putative causal variants underlying the risk for neurodegenerative and psychiatric diseases.

Regional and age-related differences in microglial density, morphology, gene and protein expression were previously described for both animals and humans51,52,53,54. Our analyses suggest some pathways that may be involved in regulating regional heterogeneity, such as reelin, IFN and glucocorticoid signaling pathways. In addition, we found an age-related change in the genes involved in a wide range of inflammatory responses, in line with previous results of aging in microglia55,56 and peripheral blood57. Of interest are a downregulation of C2, P2RY12 and P2RY13, key players in microglia–neuron interactions58,59, as well as genes related to age-related disorders: MS4A4A, MS4A6A, BST1, and P2RY12. Our pathway analyses identified immune-related pathways that may be of relevance for the mechanisms of microglial aging, including STAT3 and IL-6 signaling, as well as liver X receptor (LXR)/retinoid X receptor (RXR) activation, which has emerged as a key player in regulating cholesterol homeostasis and inflammation in the brain with a potential role in neurodegenerative disorders60,61,62. Based on previous studies in humans and mice63, we expected to find region-specific patterns of age-related changes in microglia. MS4A6A, a gene related to AD risk39,40, was one of the genes that showed a region-specific effect of age41,64. By mapping both eQTLs and sQTLs in human microglia, we have created a resource that has informed our own genetic studies and will be useful for the genetics and neuroscience community. We have identified specific disease colocalizations that may not be captured in monocytes or bulk brain tissue, like BIN1, USP6NL and PICALM in AD, P2RY12 in PD, PLXNB2 in MS and IFRD1 in SCZ. We also found colocalizations with opposing effects, such as CASS4. The disease-associated eQTL results were partly shared between the MiGA and the microglia eQTL study by Young et al.65. Differences between the studies in age, diagnoses of the included donors, studied brain region(s), recruitment of tissue (surgical versus autopsy), postmortem delay and sample size (93 individuals per samples versus 90 individuals per 216 samples) have likely contributed to a lack of sharing of part of the hits. By mapping sQTLs, we have shown that the known AD risk association with CD33 exon 2 splicing is also present in microglia, and added disease associations that may act through splicing such as MS4A6A in AD, SPP1A1L2 and FAM49B in PD, IRF3 in SCZ, STK4 and GMIP in BD, and CD37 and EFCAB13 in MS. Interpretation of the qQTLs will be improved with the generation of long-read RNA sequencing (RNA-seq) in microglia to identify new transcripts.
We performed comprehensive fine-mapping of GWAS loci in five diseases through an ensemble of four different methods and microglia-specific epigenomic datasets to identify credible sets of putative causal variants. This approach allowed us to identify candidate functional variants in multiple disease susceptibility loci that modulate microglia-specific enhancer activity and regulate causal

**Articles**

NATuRE GENETiCS | VOL 54 | JANUARY 2022 | 4–17 | www.nature.com/naturegenetics
Fig. 7 | sQTLs in CD33 and MS4A6A colocalize with the AD risk loci. a, Intron usage in CD33 is associated with rs3865444 in all four microglial regions. The nominal P value from the linear regression model in the region-by-region sQTL analysis is indicated above the box plots. The beta and P values from the meta-analysis are also indicated above the figure. The box plots show the median, the box spans from the first to the third quartiles and the whiskers extend 1.5 times the IQR from the box. The meta-analyzed CD33 sQTL colocalizes with an AD risk locus (PP4 = 0.92). b, The lead GWAS SNP and lead sQTL SNP share the same causal variant rs3866444, and three other SNPs prioritized by fine-mapping are in high LD. The lead SNP falls near the CD33 sQTL locus colocalizes with an AD risk locus.

gene expression, which in turn likely modify disease risk by altering the function of microglia (or other myeloid cells) in the brain. In AD, we propose USP6NL to be the causal gene in the ECHDC3 locus, due to both a convincing colocalization with AD GWAS and eQTL and the overlap of fine-mapped putative causal SNPs within a defined microglial enhancer that connects with the USP6NL
promoter. USP6NL, a GTPase-activating protein involved in the control of endocytosis, adds to a growing list of genes (BIN1, PICALM, RABEP1, RIN3 and SORL1) that implicate the dysfunction of the myeloid endosomal system in AD.7,76.

In PD, we propose P2RY12 in the MED12L locus through a similar mechanism. P2RY12 is a particularly interesting gene due to the increasing body of literature on its importance for the functioning of microglia7,77 and the proposed link between PD and purinergic signaling.28, P2RY12 is one of the P2Y metabotropic G protein-coupled purinergic receptors, which is highly expressed in microglia compared to other brain and myeloid cell types. P2RY12 expression is lost on microglial activation28 and culture28; in our analyses, we showed that expression is decreased with aging. P2RY12 plays a role in microglial migration, activation and neuronal activity63,74. Further validation work is required to test whether the enhancers we prioritized with fine-mapping regulate these genes specifically in microglia.

We recognize several limitations to the current study. First, our sample size was still small compared to monocyte and brain datasets20–22. We increased power by combining the four regions in a meta-analysis, with the caveat of not adjusting for shared donors, which will have increased our FDR. Another limitation is a variety of known and unknown pre- and postmortem factors that have an impact on the microglial transcriptome, as shown by our variance partition analyses, which we could not control for in our analyses. 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 datasets27,28. We sorted the microglial cells with CD11b+ beads. This marker is not restricted to microglia and may capture small fractions of other myeloid cells. 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 changes in gene expression27,28. Furthermore, our ability to detect additional disease-associated eQTLs may be obscured due to the use of bulk RNA-seq data. Future work with large numbers of single-cell RNA-seq profiles from many individuals creates opportunities for mapping eQTLs across microglial subpopulations79,80, although single-cell data are in general sparse and noisy, which may result in reduced power compared to bulk RNA-seq79. Lastly, many eQTLs are conditional and only revealed after specific stimuli that change the activation state of specific cell types. Thus, mapping response-eQTLs after stimulation with specific stimuli in primary microglia may reveal additional associations that may provide further mechanistic insights into the disease-associated variants.

In summary, we performed a comprehensive assessment of the transcriptomic landscape of human microglia from multiple brain regions. We generated an atlas of genetic effects on the human microglial transcriptome, which allowed us to identify potential causal genes and variants underlying the risk for neurodegenerative and psychiatric diseases. Our findings represent mechanistic hypotheses that can now be tested with further experimental work at both the level of individual variants and candidate genes.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00976-y.

Received: 29 December 2020; Accepted: 19 October 2021; Published online: 6 January 2022

References
1. Priller, J. & Prinz, M. Targeting microglia in brain disorders. Science 365, 32–33 (2019).
2. Ransohoff, R. M. & El Khoury, J. Microglia in health and disease. Cold Spring Harb. Perspect. Biol. 8, a025660 (2015).
3. Prinz, M., Jung, S. & Priller, J. Microglia biology: one century of evolving concepts. Cell 179, 292–311 (2019).
4. Tan, Y.-L., Yuan, Y. & Tian, L. Microglial regional heterogeneity and its role in the brain. Mol. Psychiatry 25, 351–367 (2020).
5. van der Poel, M. et al. Transcriptional profiling of human microglia reveals grey–white matter heterogeneity and multiple sclerosis-associated changes. Nat. Commun. 10, 1139 (2019).
6. Grabert, K. et al. Microglial brain region-dependent diversity and selective regional sensitivities to aging. Nat. Neurosci. 19, 504–516 (2016).
7. De Biase, L. M. et al. Local cues establish and maintain region-specific phenotypes of basal ganglia microglia. Neuroeut 95, 341–356.e6 (2017).
8. Soreq, E. L. et al. Major shifts in glial regional identity are a transcriptional hallmark of human brain aging. Cell Rep. 18, 557–570 (2017).
9. Masuda, T. et al. Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution. Nature 566, 388–392 (2019).
10. Olah, M. et al. A transcriptomic atlas of aged human microglia. Nat. Commun. 9, 539 (2018).
11. Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer’s disease. Nature 570, 332–337 (2019).
12. Mrdjen, D. et al. High-dimensional single-cell mapping of central nervous system immune cells reveals distinct myeloid subsets in health, aging, and disease. Immunity 48, 599 (2018).
13. Hammond, T. R. et al. Single-cell RNA sequencing of microglia throughout the mouse lifespan and in the injured brain reveals complex cell-state changes. Immunity 50, 253–271.e6 (2019).
14. Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer’s disease. Cell 169, 1276–1290.e17 (2017).
15. Masuda, T. et al. Author correction: spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution. Nature 568, E4 (2019).
16. Galatro, T. F. et al. Transcriptomic analysis of purified human cortical microglia reveals age-associated changes. Nat. Neurosci. 20, 1162–1171 (2017).
17. McGeer, P. L. et al. Microglia in degenerative neurological disease. Glia 7, 84–92 (1993).
18. Kreutzberg, G. W. Microglia: a sensor for pathological events in the CNS. Trends Neurosci. 19, 312–318 (1996).
19. Trapanier, M. O., Hopperton, K. E., Mizrahi, R., Mechawar, N. & Bazinet, R. P. Postmortem evidence of cerebral inflammation in schizophrenia: a systematic review. Mol. Psychiatry 21, 1009–1026 (2016).
20. Hopperton, K. E., Mohammad, D., Trapanier, M. O., Giuliani, V. & Bazinet, R. P. Markers of microglia in post-mortem brain samples from patients with Alzheimer’s disease: a systematic review. Mol. Psychiatry 23, 177–198 (2018).
21. Parikshak, N. N. et al. Genome-wide changes in IncRNA, splicing, and regional gene expression patterns in autism. Nature 540, 423–427 (2016).
22. Raj, T. et al. Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. Science 344, 519–523 (2014).
23. Kuniske, B. W. et al. Genetic meta-analysis of diagnosed Alzheimer’s disease identifies new risk loci and implicates Aα, tau, immunity and lipido–lipid processing. Nat. Genet. 51, 414–430 (2019).
24. McCauley, M. E. & Baleh, R. H. Inflammation in ALS/FTD pathogenesis. Acta Neuropathol. 137, 715–730 (2019).
25. Gandal, M. J. et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science 362, eaat8127 (2019).
26. Young, A. M. H. et al. A map of transcriptional heterogeneity and regulatory variation in human microglia. Nat. Genet. 53, 861–868 (2021).
27. Masuda, T., Sankowski, R., Staszewski, O. & Prinz, M. Microglia heterogeneity in the single-cell era. Cell Rep. 30, 1271–1281 (2020).
28. Li, Y. et al. RNA splicing is a primary link between genetic variation and disease. Science 352, 600–604 (2016).
29. Raj, T. et al. CD33: increased inclusion of exon 2 implicates the Ig V-set domain in Alzheimer's disease susceptibility. Hum. Mol. Genet. 23, 2729–2736 (2014).
30. Hoffmann, G. E. & Schadi, E. E. variancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinformatics 17, 483 (2016).
31. Hoffmann, G. E. & Roussou, P. Dream: powerful differential expression analysis for repeated measures designs. Bioinformatics 37, 192–201 (2021).
32. Hartigan, J. A. & Wong, M. A. A K-means clustering algorithm. J. R. Stat. Soc. Ser. C Appl. Stat. 28, 100–108 (1979).
33. Srinivasan, K. et al. Alzheimer's patient microglia exhibit enhanced aging and unique transcriptional activation. Cell Rep. 31, 107843 (2020).
34. Gosselin, D. et al. An environment-dependent transcriptional network specifies human microglia identity. Science 356, eaaf3222 (2017).
35. Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat. Genet. 51, 793–803 (2019).

36. Gusev, A. et al. Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. Nat. Genet. 50, 538–548 (2018).

37. Raj, T. et al. Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer’s disease susceptibility. Nat. Genet. 50, 1584–1592 (2018).

38. Li, Y. L., Wong, G., Humphrey, J. & Raj, T. Prioritizing Parkinson’s disease genes using population-scale transcriptomic data. Nat. Commun. 10, 994 (2019).

39. Peters, M. J. et al. The transcriptional landscape of age in human peripheral blood. Nat. Commun. 6, 8570 (2015).

40. Storey, J. D. The positive false discovery rate: a Bayesian interpretation and the q-value. Ann. Stat. 31, 2013–2035 (2003).

41. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am. J. Hum. Genet. 88, 586–598 (2011).

42. Marioni, R. E. et al. GWAS on family history of Alzheimer’s disease. Transl. Psychiatry 8, 99 (2018).

43. Ghiambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 10, e1004383 (2014).

44. Lambert, J. C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease. Nat. Genet. 45, 1452–1458 (2013).

45. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. Nat. Genet. 51, 404–413 (2019).

46. Nalls, M. A. et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson’s disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 12, 1610 (2013).

47. Ripke, S. et al. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–427 (2014).

48. Patapolpipat, N. A. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science 365, eea17888 (2019).

49. Zhang, B. et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer’s disease. Cell 153, 707–720 (2013).

50. Desikan, R. S. et al. Polygenic overlap between C-reactive protein, plasma lipids, and Alzheimer disease. Circulation 131, 2061–2069 (2015).

51. Zhermakova, D. V. et al. Identification of context-dependent expression quantitative trait loci in whole blood. Nat. Genet. 49, 139–145 (2017).

52. Nott, A. et al. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. Science 366, 1134–1139 (2019).

53. Grem, P. P. et al. The Parkinson’s disease genome-wide association study locus browser. Mov. Disord. 35, 2056–2067 (2020).

54. Hollingworth, P. et al. Common variants at ARCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD32AP are associated with Alzheimer’s disease. Nat. Genet. 43, 429–435 (2011).

55. Böttcher, C. et al. Human microglia regional heterogeneity and phenotypes determined by multiplexed single-cell mass cytometry. Nat. Neurosci. 22, 78–90 (2019).

56. Mittelbronn, M., Dietz, K., Schluesener, H. J. & Meyermann, R. Local distribution of microglia in the normal adult human central nervous system differs by up to one order of magnitude. Acta Neuropathol. 101, 249–255 (2001).

57. Olah, M. et al. Identification of a microglia phenotype supportive of remyelination. Glia 60, 306–312 (2012).

58. Stevens, B. et al. The classical complement cascade mediates CNS synapse elimination. Cell 131, 1164–1178 (2007).

59. Radimon, A. et al. Negative feedback control of neuronal activity by microglia. Nature 586, 417–423 (2020).

60. Savage, J. C. et al. Nuclear receptors license phagocytosis by TREM2* myeloid cells in mouse models of Alzheimer’s disease. J. Neurosci. 35, 6532–6543 (2015).

61. Courtine, R. & Landreth, G. E. LXR regulation of brain cholesterol: from development to disease. Trends Endocrinol. Metab. 27, 404–414 (2016).

62. Kao, Y.-C., Ho, P.-C., Tu, Y.-K., Jou, I.-M. & Tsai, K.-J. Lipids and Alzheimer’s disease. Int. J. Mol. Sci. 21, 1903 (2020).

63. Pritts, P. et al. Alzheimer’s disease susceptibility variants in the MS4A6A gene are associated with altered levels of MS4A6A expression in blood. Neurobiol. Aging 35, 279–290 (2014).

64. Huang, K.-L. et al. A common haplotype lowers P1.1 expression in myeloid cells and delays onset of Alzheimer’s disease. Nat. Neurosci. 20, 1052–1061 (2017).

65. Deming, Y. et al. The MS4A4A gene cluster is a key modulator of soluble TREM2 and Alzheimer’s disease risk. Sci. Transl. Med. 11, eaau2291 (2019).

66. Novikova, G. et al. Integration of Alzheimer’s disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. Nat. Commun. 12, 1610 (2021).

67. Mildner, A., Huang, H., Radke, J., Stenzel, W. & Priller, J. P2Y12 receptor is expressed on human microglia under physiological conditions throughout development and is sensitive to neuroinflammatory diseases. Glia 65, 375–387 (2017).

68. Toth, A., Antal, Z., Bereczki, D. & Sperlágh, B. Purinergic signalling in Parkinson’s disease: a multi-target system to combat neurodegeneration. Neurochem. Res. 44, 2413–2422 (2019).

69. van Wageningen, T. A. et al. Regulation of microglial TMEM119 and P2RY12 immunoreactivity in multiple sclerosis white and grey matter lesions is dependent on their inflammatory environment. Acta Neuropathol. Commun. 7, 206 (2019).

70. Haynes, S. E. et al. The P2Y12 receptor regulates microglial activation by extracellular nucleotides. Nat. Neurosci. 9, 1512–1519 (2006).

71. Marsh, S. E. et al. Single cell sequencing reveals glial specific responses to tissue processing & enzymatic dissociation in mice and humans. Preprint at bioRxiv https://doi.org/10.1101/2020.12.03.408542 (2020).

72. Mattei, D. et al. Enzymatic dissociation induces transcriptional and proteotypic bias in brain cell populations. Int. J. Mol. Sci. 21, 7944 (2020).

73. Lee, M. N. et al. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. Science 343, 1246980 (2014).

74. Ramdhani, S. et al. Tensor decomposition of stimulated monocyte and macrophage gene expression profiles identifies neurodegenerative disease-specific trans-eQTLs. PLoS Genet. 16, e1008549 (2020).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. © The Author(s), under exclusive licence to Springer Nature America, Inc. 2022
were sequenced as 150 base pairs (bp) on fragments with an average read depth 100 nanograms of RNA. Libraries for PD; 104 for SCZ; 29 for BD; and 137 for MS.

the four AD GW AS were given consensus names using the most recent GW AS as a protein tau H1/H2 haplotype region (hg19 chr17:43,628,944–44,571,603) were antigen region (hg19 chr6:28,477,797–33,448,354) or the microtubule-associated loci overlapping the human major histocompatibility complex/human leukocyte with the lowest P value. 

Failed final filtering and QC' column. To avoid double-counting in colocalization, none of the donor can have different covariates, we modeled the individual as a random effect and added selected covariates to adjust for possible technical and biological confounders. The final model accounted for sex, donor ID, age, region, cause of death, the first four ancestry MDS values (C1–C4), percentage of mRNA bases, median insert size and percentage of ribosomal bases (Picard), percentage read alignment (Picard) and sequencing lane. Biological covariates were donor ID, age, sex, brain region, cause of death, sample pH, main diagnosis, PMI in minutes and the first four genotyping ancestry multidimensional scaling (MDS) values (C1–C4).

Differential expression analysis. Differential expression analysis was performed between the brain regions using the R package Differential expression for repeated measures (dream) from VariancePartition2. Dream uses a linear model to increase power and decrease false positives for RNA-seq datasets with repeated measurements. For the analysis, inputs included the count matrix and the covariate fist data were normalized using the function fitdistrplus (v.2.9.4) which also performs voom transformation. Since one donor can have different covariates, we modeled the individual as a random effect and added selected covariates to adjust for possible technical and biological confounders. The final model accounted for sex, donor ID, age, region, cause of death, the first four ancestry MDS values (C1–C4), percentage of mRNA bases, median insert size and percentage of ribosomal bases. P values were then adjusted for multiple testing correction using the Benjamini–Hochberg FDR correction. For all the differential expression analysis, the donor ID and cause of death covariates were modeled as random effects and the other covariates were modeled as fixed effects (https://github.com/RajLabMSSM/MiGA_public_release). Details about the differential age-by-region, sex-related and diagnosis analysis are described in the Supplementary Note.

Pathway and gene set enrichment analyses. Pathway analysis. We performed canonical pathway analyses in the IPA software v.6.8752261 independently using the following input gene sets: upregulated DEGs aging (n = 338 genes); downregulated DEGs aging (n = 1,355 genes); and clusters of gene sets for specific brain regions, that is, cluster 1 (n = 333 genes upregulated in the MFG and STG), cluster 2 (n = 108 genes upregulated in the SVZ and THA), cluster 3 (n = 350 genes downregulated in the SVZ) and cluster 4 (n = 296 genes upregulated in the SVZ) at FDR < 0.05. In addition, we analyzed the canonical pathways associated with splicing in the regional DTU gene set (n = 132) and aging DTU gene set (n = 150). 

Gene set enrichment analysis. To test specific pathways, we used curated gene sets and tested statistical enrichment using a Fisher’s exact test at FDR < 0.05 for the following curated gene lists: (1) human AD curated lists (53 upregulated and 22 downregulated genes from Srinivasan et al.33); (2) culture microglia curated lists (raw counts were extracted from Gosselin et al.26); (3) IFN-γ-stimulated gene list (159 genes); (4) LPS-stimulated microglia curated gene list; (5) microglia-specific curated list (249 genes from Pol et al.). The software Enrichr (v.1.32.0) was employed to determine differential gene expression between ex vivo and microglial samples cultured for 7 (3,674 upregulated and 4,121 downregulated genes were detected and used in further analysis) and 10 days (14 IFN-γ-stimulated genes; 74 upregulated and 6 downregulated genes were detected using the methods described below in IFN-γ- and LPS-stimulated microglia; (4) LPS-stimulated microglia curated gene list (472 upregulated and 316 downregulated genes were detected using the methods described below in IFN-γ- and LPS-stimulated microglia); (5) aging in human peripheral blood curated gene list (600 upregulated and 897 downregulated genes from Peters et al.33); and (6) microglia-specific curated list (249 genes from Patir et al.26). Additionally, we included specific disease-related lists based on the latest TWAS results: (1) PD curated gene list (36 genes from Raj et al.13); (2) PD curated gene list (77 genes from Li et al.82); (3) SCZ curated gene list (43 genes from Gusev et al.;(4) BD curated gene list (16 genes from the TWAS results from latest BD TWAS).
applied an initial genotyping quality control using bcftools v.1.9 and VCFtools v.0.115, keeping SNPs with a call rate > 95%, MAF > 5%, Hardy–Weinberg equilibrium P > 1 x 10^−8 and a sample call rate > 95%.

Duplicated and up to third-degree-related samples were removed based on pairwise kinship coefficients estimated using KING v.2.2.5 (Supplementary Fig. 10a). DNA samples were matched to the RNA-seq data to confirm the same donor origin using the MBV tool from QTLOoQ v.1.3 (Supplementary Fig. 10b) and sex mismatching samples were removed by comparing DNA-inferred sex from PLINK v.1.9b6.10 to RNA gene expression of the UTY and KIT genes (Supplementary Fig. 10c). This resulted in 593,748 genotyped variants passing all quality control steps in 98 donors, of which 90 donors had European ancestry. The genetic ancestry of samples was confirmed by PCA using the PLINK program4. The MDS values of the study participants were compared to those of the 1000 Genome Project samples (phase 3; Supplementary Fig. 10d).

Genome imputation was performed for the 90 donors using the Michigan Imputation Server v.1.4.1 (Minimac4 (ref. 15)) using the 1000 Genomes Project (phase 3) v.5 (GRCh37) European panel and Eagle v.2.4 phasing8 in quality control and imputation mode with the rsq filter set to 0.3. After imputation, variants were lifted over to the GRCh38 reference to match the RNA-seq data using Picard’s liftover tools8 and liftoverMGS.v2.3 (Supplementary Fig. 10e). Finally, we applied another round of variant quality controls, removing indels and multiallelic SNPs and keeping only variants with MAF > 5% and Hardy–Weinberg P > 1 x 10^−8. After imputation, liftover and quality control, a total of 5,803,004 variants were included in the downstream analyses. These variants were additionally annotated using the Single Nucleotide Polymorphism (dbSNP) v.20180418 Database (All_20180418.vcf.gz) and snpEff v.4.3i (ref. 16).

**QTL mapping.** To perform eQTL mapping, we followed the latest pipeline created by the Genotype-Tissue Expression (GTEx) consortium17. We completed a separate normalization and filtering method compared to previous analyses. Gene expression levels were created from RNA-seq output using the BT2 output format40. The expression of individual genes was filtered by removing genes with an expression level of less than 0.1 TPM in at least 6 counts in 20% of samples. An initial filtering removed 18,430 genes. Then, PEER factors were calculated to estimate hidden confounders within our expression data. We created a combined covariance matrix that included PEER factors and the first four genotyping ancestry MDSs as input to the analysis. We tested the numbers of PEER factors from 0 to 20 and found that between 5 and 10 factors produced the largest number of eGenes in each region (Supplementary Fig. 11).

To test for cis-eQTLs, linear regression was performed using the tensorQTL18 v.1.0.2 cis nominal mode for each SNP–gene pair using a 1-Mb window within the transcription start site of a gene. To test for the association between gene expression and the top variant in cis, we used the tensorQTL cis permutation pass per gene with 1,000 permutations. To identify eGenes, we performed a threshold of 0.05.

We performed the sQTL analysis using the splice junction read counts generated by regtools19 v.0.5.1. Junctions were clustered using LeafCutter20 v.0.2.8, specifying for each junction in a cluster a maximum length of 100 kb. Following the GTEx pipeline, introns without read counts in at least 50% of samples or with fewer than 100 read counts in at least 10% of samples were removed. Counts with insufficient variability across the samples were removed. Filtered counts were then quantile-normalized using prepare_phenotype_table.py from LeafCutter, merged and converted to BED format, using the coordinates from the middle of the intron cluster. We created a combined covariance matrix that included the PEER factors and the first four genotyping ancestry MDS values as input to the analysis. We mapped sQTLs with between 0 and 20 PEER factors as covariates in our QTL model and determined 5 to be optimal in the MFG, STG and THA; 0 PEER factors were used for the SVZ (Supplementary Fig. 11).

To test for cis-sQTLs, linear regression was performed using the tensorQTL nominal pass for each SNP junction pair using a 100-kb window from the center of each intron cluster. Although junctions were independently grouped together into clusters, we tested each SNP junction pair separately, which is the standard approach21,22. To test for the association between intron ratio and the top variant in cis, we used the tensorQTL permutation pass, grouping junctions by their cluster using the grp option. To identify significant clusters, we performed a correction using a threshold of 0.05.

We estimated the pairwise replication (χ^2) of cis-sQTLs with the external eQTL datasets using the R package23. Briefly, this involved taking the SNP–gene pairs that were significant in our microglia data at q < 0.1 and extracting the unadjusted P values for the matched SNP–gene pairs in the external dataset.

**Meta-analysis of microglial QTLs.** METASOFT: Meta-analysis of the four microglial brain regions (MFG, STG, THA and SVZ), along with monocytes (MyND and Fairfax et al.41) and DLPFC (ROSMAP) was performed using METASOFT24 v.2.0.1. Effect sizes and the s.e.m. of each SNP–gene pair were used as input. We carried out a random effects meta-analysis using the RE model, optimized to detect associations under heterogeneity.

**mashR.** To estimate and compare the genetic effects in gene expression and splicing proportions across different brain regions, we performed a mash through the R package mashR. Mash employs an empirical Bayesian method to estimate patterns of similarity among conditions and improve the accuracy of effect estimates.

Following the pipeline applied by the GTEx Consortium, we used as input the nominal associations (P, beta and s.e.m.) from eQTLs and sQTLs (SNP–gene pairs for eQTLs or SNP junction pairs for sQTLs) for each region. Then, we selected the strongest associations after computing a sparse factorization matrix of the z-scores using the Sparse Factor Analysis software v.1.0 (http://stephenslab.uchicago.edu/software.html#sfa) with k = 5 factors. Second, we computed data-driven covariance matrices priors by applying the extreme deconvolution method and computed the canonical covariance matrices, including the identity matrix, and matrices representing condition-specific effects. Next, using the entire dataset, we computed the maximum likelihood estimates of the weights for each combination and learned how each pattern effect size combination occurred in the data. Finally, we computed the posterior statistics using the fitted mash model from the previous step. This step created the tables with posterior means and LFSR, a measure analogous to the PDR, which accounts for the effect size and s.e.m. rather than only P-values. This approach improved the effect size estimates and allowed for more quantitative assessments of effect size heterogeneity compared to simple region-specific assessments.

**Statistics and reproducibility.** We generated genotyping and RNA-seq, including both whole-blood and single-cell data from brain regions. We included more quantitative assessments of effect size heterogeneity compared to simple region-specific assessments.

**Data availability**

Raw and processed RNA-seq and genotype datasets have been deposited in the National Institute on Aging Genetics of Alzheimer’s Disease Data Storage Site (NIAGADS) at https://dos.niagads.org/datasets/ng00105/ under accession no. NG00105.v1. The user will need to log on the NIAGADS Data Access Request to start an application. Instructions to download the dataset can be found at https://www.niagads.org/data/request/data-request-instructions. All differential expression, gene lists and fine-mapping results are presented as supplementary tables, while the GWAS fine-mapping results are available at the ecolohistor/Shiny application at 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 the eQTL and sQTL meta-analyses (mashR and METASOFT) and colocalization (COLOC) are available from Zenodo at https://doi.org/10.5281/zenodo.4118676.

**Code availability**

All the code used to perform the analysis is available at https://github.com/RajLabMSSM/MiGA_public_release. To perform eQTL mapping, we followed the latest pipeline created by the GTEx consortium (https://github.com/broadinstitute/gtex-pipeline). To estimate and compare the genetic effects in gene expression and splicing proportions across different brain regions, we used the mashR pipeline (https://stephenslab.igh.ucl.ac.uk/gtexresults/gtex.html). The tools used for genotyping quality control or specific R packages are described in the Methods and Supplementary Note.

**References**

79. de Lange, G. M., Rademaker, M., Boks, M. P. & Palmen, S. J. M. C. Brain donation in psychiatry: results of a Dutch prospective donor program. BMC Psychiatry 17, 347 (2017).
80. Mellef, J. et al. Characterizing primary human microglia: a comparative study with myeloid subsets and culture models. Glia 64, 1857–1868 (2016).
81. Snoeberger, M. A. M. et al. Microglia in post-mortem brain tissue of patients with bipolar disorder are not immune activated. Transl. Psychiatry 9, 153 (2019).
82. Shah, H. PgmNr 1856. Rapid—An Agile and Dependable RNA-Seq Framework (American Society of Human Genetics, 2015).

NATURE GENETICS | www.nature.com/naturegenetics
104. Cotto, K. C. et al. RegTools: integrated analysis of genomic and transcriptomic data for the discovery of splicing variants in cancer. Preprint at bioRxiv https://doi.org/10.1101/436634 (2021).

105. Li, Y. I. et al. Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* 50, 151–158 (2018).

106. Stephens, M. False discovery rates: a new deal. *Biostatistics* 18, 275–294 (2017).

Acknowledgements

We thank members of the Raj and de Witte labs for their feedback on the manuscript. We thank the teams of the NBB and Mount Sinai Neuropathology Brain Bank and Research CoRE for their services. We thank the study participants for their generous gifts of brain donation. Microglia were isolated through the efforts of a large team, and we thank M. Litjens, R. D. van Dijk, A. Fernández-Andreu, P. R. Ormel, H. C. van Mierlo, Y. He, S. Gumbs, M. E van Strien, S. Burm, V. Donega and E. M. Hol for all their contributions to this effort. We thank M. Chao for his assistance with genotyping quality control. This work was supported in part through the computational and data resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai. The research reported in this paper was supported by the Office of Research Infrastructure of the National Institutes of Health (NIH) under award no. S10OD026880. T.R. is supported by grants from the NIH (nos. R01-AG054005, NIA U01-AG086880, NIA R01-AG063130, NIA R01-AG054005, NIA U01-AG086880, NIA R01-AG063130, NIA U01-AG086880, NIA R01-AG063130, NIA U01-AG086880, NIA R01-AG063130, NIA R01-AG063130, NIA R01-AG063130, NIA R01-AG063130, NIA R01-AG063130).

The authors declare no competing interests.

Author contributions

L.D.d.W. and T.R. conceived and supervised the study. G.J.L.S., M.A.M.S. and A.B.v.B. isolated the microglia at University Medical Center Utrecht. G.J.L.S., E.N., A.A., M.P., E.A.I.G. and R.K. isolated the microglia at Mount Sinai School of Medicine. E.N., A.A. and M.P. performed genotyping and RNA-seq. W.v.Z. performed RNA-seq on stimulated microglia samples with input from G.J.L.S. and L.D.d.W. K.P.L. performed the data preprocessing and quality control. K.P.L. led the analyses of the region, aging, QTL analyses and meta-analysis, with input from J.H. and G.J.L.S. G.J.L.S. led data interpretation, functional overlaps and replication work. J.H. led the genetic, fine-mapping and epigenomic analyses. B.M.S. assisted with the fine-mapping analyses. R.A.V. assisted with QTL mapping and performed genotyping quality control. E.M.H. performed the single-cell analysis. R.K. provided funding and was involved in establishing the NBB for Psychiatry, providing tissue for this project. R.M. performed the validation work. J.P. and C.B. provided data for validation. J.H., G.J.L.S., K.P.L., L.D.d.W. and T.R. wrote the manuscript with input from all coauthors. All authors read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-021-00976-y.

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00976-y.

Correspondence and requests for materials should be addressed to Lot D. de Witte or TewIQUE Raj.

Peer review information *Nature Genetics* thanks Yi Xing and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | Regional heterogeneity analysis for transcript usage. A) Heatmap of relative transcript usage between regions using all 176 transcripts from pairwise comparisons of differential transcript usage (DTU; empirical FDR < 0.1), plotted as row-scaled z-scores of mean transcript usage per region; red and blue indicates high and low relative transcript usage, respectively. Transcripts form 2 k-means clusters, n refers to the number of transcripts in each cluster. Core microglia genes from Patir et al. highlighted. B) Transcript usage plots for the gene RGS1. The two most abundant transcripts are bolded. The DTU signal is driven by a reduction of the intron retention transcript ENST00000498352.1 and a corresponding increase in the protein-coding transcript ENST00000367459.8 in the SVZ compared to the other regions. Boxplots show the median with the first and third quartiles of the distribution. C) Functional Enrichment Analysis of all 132 genes with regional DTU using Ingenuity Pathway Analysis (IPA). Significantly enriched terms shown (q-value < 0.05).
Extended Data Fig. 2 | Age-related analysis for transcript usage. A) Heatmap of the 225 transcripts associated with age (empirical FDR < 0.1). Each row plotted as Z-score of median expression averaged first by donor (across multiple regions) and then by age quintiles with 20 donors each. Transcripts are ordered by Ward’s hierarchical clustering. Core microglia genes from Patir et al. highlighted. B) Example transcript usage for P2RY12. The association is caused by an increase in the long protein-coding transcript ENST00000302632.3 and a corresponding decrease in the short intron retention transcript ENST00000468596.1 during aging. C) Functional Enrichment Analysis of all 150 genes with DTU in aging using Ingenuity Pathway Analysis (IPA). Only significantly enriched terms shown (q-value < 0.05).
### Extended Data Fig. 3 | Full colocalization results in Alzheimer's Disease

Colocalization PP4 displayed for each GWAS locus (right text) and gene (left text) for each QTL dataset. An empty value means no QTL was present for testing for that gene in that dataset.
Extended Data Fig. 4 | Colocalization results for each regional microglia dataset in Alzheimer’s Disease. Colocalization PP4 displayed for each GWAS locus (right text) and gene (left text) for each QTL dataset. An empty value means no QTL was present for testing for that gene in that dataset.
### Extended Data Fig. 5 | Full colocalization results in Parkinson's Disease

Colocalization PP4 displayed for each GWAS locus (right text) and gene (left text) for each QTL dataset. An empty value means no QTL was present for testing for that gene in that dataset.
**Parkinson’s Disease**

GWAS: Nalls et al 2019; PP4 >= 0.5

| Gene        | dataset | Microglia | Type | eQTL | sQTL |
|-------------|---------|-----------|------|------|------|
| CCAR2       | 0.09    | 0.3       | 0.7  | 0.09 | 0.12 |
| AC037459.3  |         |           |      |      |      |
| SETDB2      | 0.07    | 0.06      |      | 0.09 | 0.14 |
| SULT1A2     | 0.51    | 0.51      | 0.25 | 0.78 | 0.82 |
| SULT1A1     | 0.12    | 0.28      | 0.17 | 0.78 | 0.82 |
| EIF3C       | 0.09    | 0.09      | 0.09 | 0.07 | 0.12 |
| ATXN2L      | 0.08    | 0.07      | 0.08 | 0.07 | 0.07 |
| CTSB        | 0.79    | 0.91      | 0.86 | 0.07 | 0.05 |
| FAM49B      |         | 0.06      |      | 0.71 | 0.89 |
| FGF20       | 0.62    | 0.07      | 0.1  | 0.69 | 0.03 |
| GCH1        |         | 0.06      |      | 0.13 | 0.19 |
| NUPL2       | 0.13    | 0.91      | 0.27 | 0.11 | 0.39 |
| GPNMB       | 0.27    | 0.64      |      | 0.31 | 0.31 |
| ARIL6IP4    | 0.07    | 0.06      |      | 0.26 | 0.13 |
| WDR6        | 0.58    | 0.24      |      | 0.75 | 0.88 |
| ARIH2       |         | 0.63      |      | 0.08 | 0.48 |
| PYCR2       |         |           |      |      |      |
| KPNA1       | 0.06    | 0.06      | 0.06 | 0.56 | 0.1  |
| P2RY12      | 0.22    | 0.06      |      | 0.96 | 0.95 |
| RNF40       | 0.07    | 0.1       |      | 0.71 | 0.09 |
| ITGAX       | 0.07    | 0.08      | 0.25 | 0.53 |      |
| CTF1        | 0.07    | 0.58      |      |      |      |
| SIPA1L2     | 0.07    | 0.1       | 0.06 | 0.2  | 0.98 |

**Extended Data Fig. 6** Colocalization results for each regional microglia dataset in Parkinson’s Disease. Colocalization PP4 displayed for each GWAS locus (right text) and gene (left text) for each QTL dataset. An empty value means no QTL was present for testing for that gene in that dataset.
Extended Data Fig. 7 | Overlap of colocalized microglia eQTLs with epigenomic features in AD and PD. Cell-type specific promoters and enhancers were overlapped with SNP sets for each colocalizing microglia QTL - GWAS locus. SNP sets consisted of the lead GWAS SNP, the lead QTL SNP and any fine-mapped consensus or credible SNPs. Results are summarized here by the number of SNPs in the set that overlap with a particular feature type.
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 (v.19.09.1). RAPID aligns samples to the hg38 genome build using STAR (v.2.7.2a) software with GENCODE v30 transcriptome reference and calculates quality control metrics using Picard. Estimated transcript abundance was obtained using RSEM (v.1.3.1) 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 vcfutils (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 (v0.2.0.1) and mashR (v0.2.2-11)

Colocalization: R package COIQC (v3.2.7-1)

Fine-mapping: echolocatorR https://rajlab.shinyapps.io/Fine_Mapping_Shiny

Analysis of transcription factor motifs: motifbreakR with HOMOCOCO database

All plots were created using ggplot in R (version 3.6.0), with ggrepel, ggfortify, patchwork, and ggbiom 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/ng00105v; Accession number: NG00105v1). 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-list.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. (2017) and Peters et al. (2015). Galatro et al. 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 dorsalateral prefrontal cortex (DPFC - Ng et al. 2017) using the q-value n1 metric. We found the highest sharing between MiGA and the Young et al. microglia, and sharing with bulk DPFC 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 colocated 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 19-22.

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 interval (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; 56 of them were female. All analyses were adjusted for age, gender, and other covariates. The details are described in Figure 1 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 hospitals, 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.