A map of transcriptional heterogeneity and regulatory variation in human microglia

Adam M. H. Young1,2,3,10, Natsuhiko Kumasaka2,10, Fiona Calvert2, Timothy R. Hammond4,5, Andrew Knights2, Nikolaos Panousis2, Jun Sung Park2, Jeremy Schwatzentruber6, Jimmy Liu2, Kousik Kundu2, Michael Segel1, Natalia A. Murphy1, Christopher E. McMurran1, Harry Bulstrode3, Jason Correa3, Karol P. Budoshoski3, Alexis Joannides3, Mathew R. Guilfoyle3, Rikin Trivedi3, Ramez Kirolos3, Robert Morris3, Matthew R. Garnett3, Ivan Timofeev2, Ibrahim Jalloh3, Katherine Holland3, Richard Mannion3, Richard Mair3, Colin Watts3, Stephen J. Price3, Peter J. Kirkpatrick3, Thomas Santarius3, Edward Mountjoy2,9, Maya Ghossaini2,9, Nicole Soranzo2, Omer A. Bayraktar2, Beth Stevens4,5, Peter J. Hutchinson3, Robin J. M. Franklin3,11 and Daniel J. Gaffney2,11

Microglia, the tissue-resident macrophages of the central nervous system (CNS), play critical roles in immune defense, development and homeostasis. However, isolating microglia from humans in large numbers is challenging. Here, we performed a population-scale study of human microglia isolated from 141 patients undergoing neurosurgery. Using single-cell and bulk RNA sequencing, we identify how age, sex and clinical pathology influence microglia gene expression and which genetic variants have microglia-specific functions using expression quantitative trait loci (eQTL) mapping. We follow up one of our findings using a human induced pluripotent stem cell based macrophage model to fine-map a candidate causal variant for Alzheimer’s disease at the BIN1 locus. Our study provides a population-scale transcriptional map of a critically important cell for human CNS development and disease.

Microglia are tissue-resident macrophages of the CNS and play critical roles in synaptic pruning and neuronal plasticity, as well as maintaining local immune surveillance within the brain1–4. Disease studies have implicated microglial dysfunctions in a number of neurological disorders4–7, but these highly plastic cells have not been studied at a population level. To date, studies of microglial gene expression have been restricted to relatively small samples of either frozen postmortem tissue from existing brain banks or fresh surgical samples from restricted groups of patients, typically those who had temporal lobe resections for epilepsy or peritumoral tissue. Single-cell transcriptomic studies of similar samples have suggested that microglial function may vary across age, sex and brain region8–13. However, these conclusions are often not replicated in studies of equivalent size.

Here, we performed a population-scale study of human microglia to understand how age, sex, pathology, cortical anatomy and common germline genetic variation influence the microglia transcriptome. We used a unique cohort of patients from whom samples had been taken within 8 hours of an acute hemorrhage or traumatic brain injury, to identify two new signatures of acute activation in human microglia. Finally, we examined how our results replicated in a scalable cell model system of microglia, using induced pluripotent stem cell-derived macrophages (IPSDMac) derived from 133 human induced pluripotent stem cell (iPS) lines created by the Human Induced Pluripotent Stem Cell Initiative4–11.

Characterization of microglial cell populations

We undertook analysis of human microglia isolated from 141 patients undergoing a range of neurosurgical procedures (Fig. 1a). These included a ‘control’ group who had cortical microglia sampled at the beginning of a surgical corridor when the distance to the clinical pathology exceeded 4 cm. We also sampled cortical microglia from patients with hydrocephalus, brain tumors and patients with acute brain injury (spontaneous hemorrhage and trauma) who sustained substantial parenchymal injury, enabling us to capture in vivo microglial activation.

For each individual, we isolated CD11b-positive cells and performed both single-cell (Smart-seq2)13 and bulk RNA-seq. After quality control, we retained 112 bulk RNA-seq samples and 9,538 single cells from 129 patients (Fig. 1b). Our bulk RNA-seq samples clustered closely with microglia from two previous studies16–17, and were distinct from both GTEx brain and BLUEPRINT.

1Wellcome-MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK. 2Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK. 3Division of Neurosurgery, Department of Clinical Neurosciences, Cambridge University Hospitals, Cambridge, UK. 4FM Kirby Neurobiology Center, Boston Children’s Hospital Harvard University, Boston, MA, USA. 5Howard Hughes Medical Institute, Broad Institute of Harvard and MIT, Boston, MA, USA. 6EMBL-EBI, Wellcome Genome Campus, Hinxton, UK. 7Biogen, Cambridge, MA, USA. 8Institute of Cancer and Genomic Sciences, College of Medical and Dental Sciences, Birmingham, UK. 9Open Targets, Wellcome Genome Campus, Hinxton, UK. 10These authors contributed equally: Adam M. H. Young, Natsuhiko Kumasaka. 11These authors jointly supervised this work: Robin J. M. Franklin, Daniel J. Gaffney. 12✉ e-mail: rjf1000@cam.ac.uk; dg13@sanger.ac.uk
monocytes (Fig. 1c). Additional clustering only with myeloid cell types revealed that iPSC-derived microglia were transcriptionally more similar to uncultured primary microglia than iPSC-derived macrophages, although substantial between-laboratory effects on both primary and cultured cells were apparent from this analysis (Extended Data Fig. 1a).

We compared our single-cell data to public datasets of 68,000 peripheral blood mononuclear cells (PBMCs) isolated from a healthy donor\(^18\) and 15,000 brain cells from five GTEx donors\(^19\). A total of 8,662 cells in our Smart-seq2 data formed a cluster with the microglia population found in GTEx samples, and distinct from PBMCs (Fig. 1d), and included a range of known microglial marker genes, including P2RY12, CX3CR1 and TMEM119 at high levels (Extended Data Fig. 1b). We defined this population of cells (excluding GTEx samples) as microglia for the remainder of our analysis. We identified three less common, putatively infiltrating populations of cells that closely resembled other blood cell types, including NK T cells, monocytes or B cells that comprised 8.4%, 0.5% and 0.3% of our single-cell dataset, respectively. These cell types may reflect either infiltration of immune cells as a result of blood–brain barrier breakdown or intravascular contamination within the tissue. Abundance of infiltrating cells was strongly correlated with patient pathology.
with trauma patients especially enriched (odds ratio (OR) = 7.6, Fisher exact test $P = 1.2 \times 10^{-10}$) (Fig. 1c). We also found a significant effect of age on the abundance of infiltrating cells (3.4% increase per year, Wald test $P = 0.014$) after adjusting for all known confounding factors, which could reflect increasing blood–brain barrier permeability over the lifespan (Extended Data Fig. 1c). Batch correction using our linear mixed model showed equivalent performance to Seurat V3 (Extended Data Fig. 1d–h), but was readily scalable to the 129 donors as batches used in the subsequent analysis (Methods).

Within microglia, we found a high level of consistency across all patients. We identified four populations of cells (Fig. 2a). By examining their relative abundance in different patient groups, we identified one naïve and three distinct microglial activation states. Population A was most common in control and patients with hydrocephalus, while population B, which we identify as subacute activation, was enriched in patients with tumors (OR = 4.9, $P = 7.6 \times 10^{-10}$). Two acute activation populations, C and D, were common in patients with spontaneous hemorrhage and traumatic brain injury (comprising 25–76% of cells), but were rare in other pathologies (<5% of cells) (Fig. 2b,c).

To characterize these populations further, we performed differential expression analysis between different microglial populations. The x axis shows the $P$ value obtained by gProfiler2 with multiple testing corrections (Methods). Bars are colored according to the combinations of clusters in which genes are upregulated. The upregulated cluster IDs are also shown beside the bars.

![Fig. 2 | Transcriptional heterogeneity in human microglia.](image)

**a.** UMAP of 8,662 microglia cells after removing putative infiltrating cells. Colors show four clusters defined using Louvain clustering (Methods). **b.** Microglial population variation between patient pathologies. The four different colors in **a** illustrate population compositions for each pathology. Points colored gray are all other cells. **c** and **d** show population compositions for each pathology. **e** illustrates population compositions for each pathology. Points colored gray are all other cells. **f** shows population compositions for each pathology. Points colored gray are all other cells. **g** shows population compositions for each pathology. Points colored gray are all other cells. **h** shows population compositions for each pathology. Points colored gray are all other cells.
stress-induced senescence and DNA damage (HIST1H2BG), whereas population D expressed genes associated with cell proliferation (FLT1) and chemotaxis (CCL4, CCL8, CXCL16), the latter also being shared with population B.

Populations C and D are of particular interest as they are unlikely to have been previously observed (Fig. 2b,c). To test this hypothesis, we compared our population-specific markers with differentially expressed genes from Alzheimer's disease-associated microglia and glioma-associated microglia (Extended Data Fig. 2a). This analysis confirmed that populations of microglia identified in previous studies most closely resembled populations A or B in our dataset, and that the activated populations C and D were transcriptionally distinct.

To further validate our findings, we performed differential expression between microglia from different patient pathologies and controls (Extended Data Fig. 2b). We selected four candidate marker genes CD63, BIN1, C3 and CCL4 that were upregulated relative to controls in groups of patients with hemorrhage, trauma, hydrocephalus and tumors, respectively (Extended Data Fig. 2c). Immunohistochemistry for each marker in a fixed tissue section confirmed differences in expression between patient groups at the protein level (Extended Data Fig. 2d). Finally, we confirmed that expression of activation markers in populations C and D was not driven by sample processing, using RNAscope in fresh frozen tissue sections (Extended Data Fig. 2e,f).

**Biological determinants of microglial expression**

Our sampling design enabled us to explore the relative importance of a wide range of biological factors in determining microglial gene expression, while controlling for important technical confounders using variance components analysis (Fig. 3a). Of the biological factors we examined, clinical pathology explained more variation than all other factors combined, although all factors except sex, including age, brain region, dominant hemisphere and ancestry, explained some of the variation that was significantly different from zero (likelihood ratio test, family-wise error rate = 0.05). Individuality of the patient explained the most variability of any single factor in the model. However, although this factor captures the contribution of genetic background, it is also likely to be dominated by unmeasured technical batch effects, such as variability in cell dissociation and surgical sampling. We found 260 genes where age explained a fraction of variation that was significantly different from zero, showing upregulation of genes related to inflammation (CLEC7A, CIITA and TLR2) and downregulation of cell identity (P2RY12, CX3CR1), motility and proliferation genes (CSF1R) with increasing age (Fig. 3b–e and Supplementary Table 2). Although sex accounted for some of the variation that was significantly different from zero, the effect size was small.

![Fig. 3 | Single-cell RNA-seq reveals how microglial transcriptional heterogeneity is driven by clinical factors. a, Bar plot shows the variance explained by each factor. Bars colored gray are technical factors (N.exp.genes: the number of expressed genes in each cell, which is expected to reflect cell health or quality; Plate: cells undergoing library preparation and sequencing together on the same 384-well plate; ERCC%: ERCC spike-in percentage among all mapped fragments for each cell; N.fragments: the number of fragments for each cell mapped on autosomes; 96-well position: the position of a cell on the 96-well plate processed in the Smart-seq2 protocol; MT%: percentage of mitochondrial RNA fragments among all mapped fragments for each cell) and colored pink are clinical factors (pathology: age of patient; brain region; brain hemisphere; sex of patient). Blue bars are partly related to the genetic background of the patient (Patient and Ancestry). b, Heat map showing the strength and direction of the age effect for differentially expressed (DE) genes at L TSR greater than 0.9 (Methods). The effect size is the posterior mean estimate weighted by the empirical prior distribution, where hyperparameters were estimated from the data using a linear mixed model (Methods). c, Pathway enrichment for DE genes by age. Pathways colored red are enriched only for DE genes upregulated by age, and the pathways colored blue are enriched for DE genes downregulated by age. d, Example genes upregulated (e) or downregulated (f) by age. f, Heat map showing the average normalized expression levels for DE genes by sex at L TSR greater than 0.9 (Methods). The effect size is the posterior mean estimate weighted by the prior distribution calibrated in the linear mixed model (Methods). g, Pathway enrichment by sex. h, Pathway enrichment for combinations of brain regions. The blue bars show pathways upregulated in cerebellum and occipital lobe. The red bars show pathways upregulated in frontal, parietal and temporal lobes.](image-url)
for little of the overall variation, we found 97 genes that were differentially expressed between males and females (Fig. 3f). These included multiple genes in the complement pathway and synaptic pruning mechanisms (C1QA, C1QC and C3) that were more highly expressed in females than in males (Fig. 3g and Supplementary Table 3). Anatomical region of sampling also had a subtle effect on transcriptional variation, with cerebellar microglia, which are known to exhibit a distinct, less ramified morphology, having increased expression of several recruitment chemokines (CCL4, CCL3, CCL4L2, CCL3L3) (Fig. 3h and Supplementary Table 4). We repeated our differential expression analysis using bulk RNA-seq and found a reasonable correlation in effect sizes between bulk and single-cell datasets for both the age and sex comparisons (r = 0.61 and 0.39 with P = 4.1 \times 10^{-20} and 4.5 \times 10^{-8} for age and sex, respectively) (Extended Data Fig. 3).

**eQTL mapping in human microglia**

Next, we constructed a map of eQTLs. After excluding samples with low genotyping quality or substantial non-European ancestry, we retained bulk RNA-seq data from 93 individuals, and detected 585 eQTLs at a false discovery rate (FDR) of 5% using a simple linear regression model. For comparison, we mapped eQTLs using the same pipeline and significance thresholds in both the BLUEPRINT monocyte (n = 193) and our own IPSDMac data (n = 133). This analysis suggested that the number of eQTLs that we detected in primary microglia was unexpectedly low. In part, this is likely to...
be because of the higher between-sample variability in primary microglia, compared with other cell types (Fig. 4a). To confirm the eQTLs that we detected with linear regression, we performed eQTL mapping only using allele-specific expression. For eQTLs detected at FDR = 5%, the effect sizes estimated from linear regression and from allele-specific analysis were highly correlated \((r = 0.71)\), suggesting that the majority are real (Extended Data Fig. 4a).

Next, we explored the level of cell-type specificity of the eQTLs we detected by comparing microglia, monocytes and IPSDMac using a three-way empirical Bayesian hierarchical model (Methods). Here, numbers of eQTLs were computed by summing over model posterior probabilities and are therefore not expected to be identical to those from our linear regression analysis. We discovered 855 microglia eQTLs, of which 108 were microglia specific, 449 were shared across all three cell types, 192 were shared with IPSDMacs, of which 108 were microglia specific, 449 were across all three cell types, 192 were shared with IPSDMacs, of which 108 were microglia specific, 449 were across all three cell types, 192 were shared with IPSDMacs, of which 108 were microglia specific, 449 were IPSDMac and 7% of eQTLs shared between all three cell types (Fig. 4c).

We then tested for colocalization of microglia eQTLs with risk loci from 146 genome-wide association studies (GWAS), of which 25 were broadly neurological, including cognitive developmental measures, such as intelligence, neuropsychiatric disorders with adolescent/young adult onset and neurodegenerative diseases (Methods). We found 245 unique gene–trait combinations with the posterior probability of a single shared causal variant between a microglia eQTL and a GWAS locus (PP4) greater than 0.5 for 84 different traits and 129 unique genes (excluding HLA genes). The number of colocalized genes for each trait most likely reflects the statistical power of the study. For example, we detected 13 colocalizations with neutrophil percentage, which also has a sample size of 349,861. We did, however, observe an excess of colocalized microglial eQTLs for certain traits, including Alzheimer’s disease (AD), Parkinson’s disease (PD) and inflammatory bowel disease, probably reflecting the known involvement of microglia or macrophages in the pathology of each of these diseases (Fig. 4d). We also discovered eQTLs that were absent from other tissues, which colocalized with a wide range of GWAS traits. For example, we discovered an eQTL for DAGI, which produces a protein that is involved in the dystrophin–glycoprotein complex, with associations with fed-up feelings, intelligence-related traits and autoimmune diseases (Fig. 4e and Extended Data Fig. 4b). Interestingly, this eQTL is detected in both

**Fig. 5 | Fine-mapping of the BIN1 eQTL/Alzheimer’s disease association.** a, Posterior probability of colocalization between Alzheimer’s disease and the three myeloid cells and GTEx eQTLs for the BIN1 gene. The y axis is based on the AD GWAS primary signal of the BIN1 locus and the x axis is based on the secondary signal at BIN1 found by the conditional analysis. b, Sequencing coverage depth of ATAC-seq and RNA-seq stratified by individuals (top ATAC-seq panel) or the three genotype groups at BIN1 lead eQTL SNP (rs6733839C>T) (bottom three panels). The top two panels show data from the primary microglia (Methods) and the bottom two panels were obtained from IPS cell-derived macrophage (Methods). The MEF2CA motif overlaps with the lead SNP and the alternative allele (T) increases predicted binding affinity. c, Regional Manhattan plot around the BIN1 gene. The y axis shows the statistical significance of AD GWAS in log10 BF. Regional plot shows the statistical significance of IPSDMac and BIN1 gene in log10 BF. d, Regional plot shows the statistical significance of microglia eQTL for BIN1 gene in log10 BF. e, Regional plot shows the statistical significance of IPSDMac eQTL for BIN1 gene in log10 BF. f, Regional plot shows the statistical significance of IPSDMac chromatin accessibility QTL (log10 BF) at the chromatin accessibility peak involving the putative causal variant rs6733839C>T. Tissue type annotation: artery tibial (AT), esophagus gastroesophageal junction (EGJ), colon sigmoid (CS), skin sun exposed lower leg (SSELL), heart left ventricle (HLV), colon transverse (CT), esophagus mucosa (EM), pituitary (PI).
Fine-mapping primary microglia eQTLs using an in vitro model

Given the involvement of microglia in neurodegenerative disease, we next selected Alzheimer's disease to undertake a detailed analysis of colocalization of microglia eQTLs with GWAS loci. Using different AD GWAS eQTLs, we found between 2 and 11 AD risk loci with a PP4 greater than 0.5 with an eQTL in primary microglia (Fig. 4e). These included well-known AD loci, such as BIN1, and less well-studied AD associations, for example, EP301. We repeated our analysis using microglia eQTLs mapped by RASQUAL, a method that boosts power to detect eQTLs using allele-specific expression (Supplementary Table 5). This analysis detected additional colocalizations at other well-known AD GWAS loci, such as CD33 (Extended Data Fig. 5a). Here, analysis of splicing patterns revealed a splice QTL at exon 2 (Extended Data Fig. 5b), consistent with previous studies. One explanation for this result is that the allele-specific signal captured by RASQUAL is more sensitive to the changes in splice pattern. However, we discovered that the test statistics produced by RASQUAL may be inflated by additional overdispersion in our microglia dataset (Fig. 4a).

The challenges of studying primary microglia make the use of IPSDMac an attractive alternative. We therefore next asked whether any of the 11 primary microglia eQTLs that colocalized with an AD risk association could also be detected as an IPSDMac eQTL. We identified three AD association signals (BIN1, the EP301/EP301-AS locus and PTK2B) that colocalized with an eQTL both in primary microglia and in IPSDMac eQTLs (Extended Data Fig. 4c). At the EP301/EP301-AS locus, we found an eQTL for the EP301-AS1 noncoding RNA that colocalized with the AD risk association, but no equivalent signal for the EP301 protein-coding gene in most tissues (Extended Data Fig. 4b,c). We have previously reported that an eQTL for the gene PTK2B colocalized with an AD risk association on chromosome 8 (ref. 29). When we compared this with primary microglia we found a difference in the direction of effect between primary microglia and IPSDMac (Extended Data Fig. 5e).

Finally, our analysis revealed that the AD association signal at BIN1 was highly cell-type specific, being found in primary microglia and IPSDMac, but no other cell types or tissues (Fig. 5a). To fine-map causal variants at BIN1, we generated ATAC-seq data from 5 primary microglia and 89 IPSDMac. We found that the lead SNP of this association signal, rs6733839C>T, was located in a region of open chromatin in both microglia and IPSDMac. rs6733839C>T was also associated with a significant change in chromatin accessibility in IPSDMac (a chromatin accessibility QTL, caQTL) (Fig. 5b, P < 6.1 x 10^-10). This caQTL also colocalized (PP4 = 0.996) with the AD association signal (Fig. 5c–f), strongly suggesting that the causal variant driving the AD risk association directly or indirectly alters chromatin openness in this region. Analysis of the sequence context of this variant revealed that the AD risk allele at rs6733839C>T created a predicted high-affinity binding site for the MEF2C, a transcription factor with established roles in hippocampal learning and memory (ref. 28). A recent study has examined chromatin interactions between the BIN1 promoter and nearby AD risk variants. Our results suggest that rs6733839C>T increases AD risk by increasing the binding of MEF2C, in turn increasing the expression of BIN1. Although BIN1 and MEF2C are broadly expressed in many tissues, coexpression of both genes was found only in primary microglia and IPSDMac (Extended Data Fig. 5g).

Taken together, our results show that one of the largest common variant associations with AD outside that with apolipoprotein E can be studied using a scalable and relatively straightforward iPS-based macrophage model.

Discussion

Here we present a population-level study of human primary microglia. By sampling cells from living donors, we defined transcriptional signatures of in vivo microglial activation, avoiding artifacts from postmortem index and in vitro cell culture. We identified multiple microglial populations and showed how these populations are shaped by pathology and other life history factors. In particular, we identify two populations of microglia that reflect different in vivo acute activation states. We also created an eQTL map in primary human microglia, identified high confidence causal genes and variants underlying risk loci for a range of neurological traits and identified a subset that replicated in a scalable iPS model system. Among other findings, our study revealed that the well-known AD risk locus near the BIN1 gene on chromosome 2 is likely to be driven by a microglia-specific eQTL and suggested that antagonism of BIN1 in microglia would be therapeutically beneficial in AD.

Our results underscore the variability between microglia from different individuals and clinical pathologies. One implication of the variation we observed between different patient pathologies is that the full spectrum of microglial function, in particular following trauma, is not well captured by small studies of a single patient population. The most obvious example of this is the populations of activated microglia we identified that account for less than 5% of cells in nontrauma patients.

Our analysis also provides a picture of the function of microglia following severe injury, producing cell populations that exhibit a mixture of proinflammatory and chemotactic phenotypes. Notably, although animal models of acute brain injury suggest rapid expansion of microglia following trauma, we only observed one population that we identified as having a proliferative phenotype, and which showed downregulation of CSF1R.

In contrast to previous reports, we found relatively subtle effects of age on the expression of individual genes in microglia, with the modest changes we did detect consistent with increased inflammatory senescence in microglia over lifespan. However, our single-cell data also revealed an increase in the influx of putatively infiltrating cells into the brain with increasing age. One explanation is that this phenomenon reflects decreasing blood–brain barrier integrity with age. Differences in microglia expression between males and females were relatively small, although we did observe increased complement activity in females, perhaps suggesting a role for complement pathways in the higher incidence of AD in women.

Our eQTL analysis revealed a number of candidate risk genes for a range of traits, with function in microglia. This was most obvious for neurodegenerative diseases, such as Alzheimer's and Parkinson's disease, and included well-known risk genes, such as BIN1. At BIN1, we demonstrated how our microglia eQTL map can be used as a reference to establish the validity of different model systems to study the subtle effects of common disease risk variants. At this locus, our study provides evidence that IPSDMac are a suitable model system to explore the role of a putative causal variant, rs6733839C>T, its effects on BIN1 expression and its role in AD risk. More generally, although more complicated protocols for IPS microglia differentiation exist, our results highlight that IPSDMac may be sufficient in specific cases. Equally importantly, our results highlight where IPSDMac may not be suitable, for example at the PTK2B locus. Finally, we note that we observed some variability in colocalization results between different AD GWAS studies. This is likely to reflect differences in power, but also some variation in methodology, for example, the use of GWAS by proxy approaches versus direct phenotyping.
An obvious extension of our approach will be to map microglia population-specific genetic effects, for example, to detect eQTLs that manifest only in activated microglia. This analysis was not possible here, due to the low number of individuals with one or more cells in different populations. In particular, the activated populations C and D are composed of 1,209 and 210 cells from 62 and 23 patients, respectively. We anticipate that future studies with larger sample sizes will be sufficiently powered to detect such effects.

In summary, we have generated a population-scale map of gene expression in primary human microglia across a diverse set of clinical pathologies. We demonstrate the human microglial response to an acute insult of the brain parenchyma. Our study provides a systematic exploration of microglia diversity, defines a reference dataset of microglial expression and provides a foundation for robust future functional studies of neurodegenerative disease mechanisms using iPSC-based models.

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/s41588-021-00875-2.

Received: 15 May 2020; Accepted: 23 April 2021; Published online: 3 June 2021

References
1. Schafer, D. P. & Stevens, B. Microglia function in central nervous system development and plasticity. Cold Spring Harb. Perspect. Biol. 7, a020545 (2015).
2. Li, Q. & Barres, B. A. Microglia and macrophages in brain homeostasis and disease. Nat. Rev. Immunol. 18, 225–242 (2018).
3. Saltie, M. W. & Stevens, B. Microglia emerge as central players in brain disease. Nat. Med. 23, 1018–1027 (2017).
4. Guerreiro, R. et al. TREM2 variants in Alzheimer’s disease. N. Engl. J. Med. 368, 117–127 (2013).
5. Jonsson, T. et al. Variant of TREM2 associated with the risk of Alzheimer’s disease. N. Engl. J. Med. 368, 107–116 (2013).
6. Tansey, K. E., Cameron, D. & Hill, M. J. Genetic risk for Alzheimer’s disease is concentrated in specific macrophage and microglial transcriptional networks. Genome Med. 10, 14 (2018).
7. Gjoneska, E. et al. Conserved epigenetic signals in mice and humans reveal immune basis of Alzheimer’s disease. Nature 518, 365–369 (2015).
8. Olah, M. et al. A transcriptional atlas of aged human microglia. Nat. Commun. 9, 539 (2018).
9. Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer’s disease. Cell 169, 1276–1290.e17 (2017).
10. 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).
11. Masuda, T. et al. Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution. Nature 566, 388–392 (2019).
12. Mrđen, 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, 380–395.e8 (2018).
13. Mathys, H. et al. Temporal tracking of microglia activation in neurodegeneration at single-cell resolution. Cell Rep. 21, 366–380 (2017).
14. Kilpinen, H. et al. Common genetic variation drives molecular heterogeneity in human iPSCs. Nature 546, 370–375 (2017).
15. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. 9, 171–181 (2014).
16. Zhang, Y. et al. Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. Neuron 89, 37–53 (2016).
17. Gosselin, D. et al. An environment-dependent transcriptional network specifies human microglia identity. Science 356, eaal3222 (2017).
18. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8, 14049 (2017).
19. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat. Methods 14, 955–958 (2017).
20. Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer’s disease. Nature 570, 332–337 (2019).
21. Sankowski, R. et al. Mapping microglia states in the human brain through the integration of high-dimensional techniques. Nat. Neurosci. 22, 2098–2110 (2019).
22. Kumasaka, N., Knights, A. J. & Gaffney, D. J. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. Nat. Genet. 48, 206–213 (2016).
23. Jansen, I. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. Nat. Genet. 51, 404–413 (2019).
24. Kunkle, B. W. et al. Genetic meta-analysis of diagnosed Alzheimer’s disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nat. Genet. 51, 414–430 (2019).
25. 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).
26. Marioni, R. E. et al. GWAS on family history of Alzheimer’s disease. Transl. Psychiatry 8, 99 (2018).
27. Schwartzentruber, J. et al. Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer’s disease risk genes. Nat. Genet. 53, 392–402 (2021).
28. 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).
29. Alasoo, K. et al. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. Nat. Genet. 50, 424–431 (2018).
30. Barbosa, A. C. et al. MEF2C, a transcription factor that facilitates learning and memory by negative regulation of synapse numbers and function. Proc. Natl Acad. Sci. USA 105, 9391–9396 (2008).
31. Nott, A. et al. Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. Science 366, 1134–1139 (2019).
32. Vela, J. M., Yáñez, A., González, B. & Castellano, B. Time course of proliferation and elimination of microglia/macrophages in different neurodegenerative conditions. J. Neurotrauma 19, 1503–1520 (2002).
33. Abd, E. M. et al. iPSC-derived human microglia-like cells to study neurological diseases. Neuron 94, 278–293.e9 (2017).
34. Okbay, A. et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat. Genet. 48, 624–633 (2016).
35. Luciano, M. et al. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. Nat. Genet. 50, 6–11 (2018).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Methods

Tissue sampling. Human brain tissue was obtained with informed consent under protocol REC 16/LO/2168 approved by the NHS Health Research Authority. Adult brain tissue biopsies were taken from the site of neurosurgery resection for the original clinical diagnosis. Serial sections were collected into five main categories: a 'control' group in which cortical microglia were sampled at the beginning of a surgical corridor when the distance to the clinical pathology exceeded 4 cm. This group was utilized to identify any iatrogenic factors influencing the transcriptional dynamics of human microglia. Additionally, samples were obtained from the cortex of patients with hydrocephalus, brain tumors, patients who had sustained a spontaneous hemorrhage and a traumatic brain injury. Paired venous blood was sampled. Tissue was transferred to Hibernate A low fluorescence (HALF) supplemented with 1x SOS (Cell Guidance Systems), 2% Glutamax (Life Technologies), 1% penicillin–streptomycin (P/S) (Sigma), 0.1% BSA (Sigma), insulin (4 g ml−1 Sigma), pyruvate (220 g ml−1, Gibco) and DNase I type IV (40 g ml−1 Sigma) on ice and transported to a containment level 2 laboratory.

Dissociation of brain tissue. Brain tissue was mechanically digested in fresh ice-cold HALF supplemented with 1x SOS (Cell Guidance Systems), 2% Glutamax (Life Technologies), 1% P/S (Sigma), 0.1% BSA (Sigma), insulin (4 g ml−1 Sigma), pyruvate (220 g ml−1, Gibco) and DNase I type IV (40 g ml−1 Sigma). The prepared mix was spun in HBSS® (Life Technologies) at 300g for 5min and the supernatant discarded. The digested tissue was rigorously triturated at 4°C and filtered through a 70-μm nylon cell strainer (Falcon) to remove large cell debris and undigested tissue. Filtrate was spun in a 22-g Percoll (Sigma) gradient with DMEM F12 (Sigma) and spun at 800g for 20min. Supernatant was discarded and the pellet was re-suspended in ice-cold supplemented HALF.

Fluorescence-activated cell sorting. For single-cell smart sequencing, human microglia were using fluorescence-activated cell sorting (FACS). The isolated cell suspension was incubated with conjugated PE anti-human CD11b (BioLegend) for 20 min at 4°C. Cells were washed twice in ice-cold supplemented HALF and stained with Helix NP viability marker. Cell sorting was performed on BD Aria III cell sorter (Becton Dickinson) at the University of Cambridge Cell Phenotyping Hub at Cambridge University Hospital, UK.

Magnetic-activated cell sorting. To avoid sustained stress on microglia as a result of prolonged sorting times for bulk sequencing, magnetic-activated cell sorting (MACS) was performed on these cells. An isolated suspension of cells was incubated with anti-CD11b conjugated magnetic beads (1:50, Miltenyi, catalog no. 130-049-601) for 15 min at 4°C. Cells were washed twice with supplemented HALF and stained with Helix NP viability marker. Cell sorting was performed on BD Aria III cell sorter (Becton Dickinson) at the University of Cambridge Cell Phenotyping Hub at Cambridge University Hospital, UK.

Immunohistochemistry. Tissue was fixed with 4% PFA at 4°C for 18h overnight and subsequently submerged in 30% sucrose w/v in PBS for cryoprotection at 4°C until it settled down to the bottom (~48hours). Cryoprotected brain was then embedded in cryomold filled with OCT. Brain tissue was then frozen in isopentane submerged in preheated citrate buffer pH 6.0 (Sigma) in a water bath at 95 °C for 5 min and the supernatant discarded. The digested tissue was rigorously triturated at 4°C and dried through a 70-μm nylon cell strainer (Falcon) to remove large cell debris and undigested tissue. Filtrate was spun in a 22-g Percoll (Sigma) gradient with DMEM F12 (Sigma) and spun at 800g for 20min. Supernatant was discarded and the pellet was re-suspended in ice-cold supplemented HALF.

iPS cell culture and macrophage differentiation. We cultured 133 iPS cell lines from normal control donors1, 2. iPS cell culture and macrophage differentiation was carried out as previously described3 with some minor modifications (see Supplementary Note for details).

Single-cell RNA-seq of primary microglia. Single primary microglia cells were processed as previously described2, but with some minor modifications to the Nextera library making process: 0.5 ng of complementary DNA was used as input for the Nextera Illumina library. A reagent volume was scaled down 100-fold. Tagmentation was quenched with 0.2% SDS. Libraries were amplified with KAPA HiFi (Kapa Biosystems, catalog no. KK2601) with indexing primers ordered from Integrated DNA Technologies.

Low-input bulk RNA-seq and ATAC-seq library preparation for primary microglia and iPS-derived macrophages. For RNA-seq samples, between 0.3 ng and 10 ng of bulk total RNA from primary microglia cells or iPS-derived macrophage cells was used as input for a modified Smart-seq2 library preparation (see Supplementary Note for detailed protocol). ATAC-seq library preparation was performed as previously described4. Pools of 96 libraries were sequenced over 8 lanes or 24 lanes of a HiSeq SBS v4 for RNA-seq and ATAC-seq preparations, respectively, collecting 75 base pair (bp) paired-end reads.

Bulk RNA-seq data of other myeloid cells and brain tissues. We downloaded fastq files of bulk RNA-seq of 6 primary microglia (pMICs) and 9 iPS-cell-derived microglia (iMICs)5, 10 monocyte-derived macrophage (MDMs) and iPS-cell-derived macrophages (IPSDMacs)6, 10 IMMCS, 8 MDMs and 4 pMICs7, 8 IMMCS, 9 iMICs and 3 pMICs8, 18 IPSDMac and 9 MDMs9, and 3 pMICs. See Supplementary Table 7 for details of cell types and sample sizes. For brain tissues, we downloaded the count table of RNA-seq data for all tissues from GTEx (v.7; Data availability) and extracted 1,671 brain samples. We also downloaded single-cell gene expression files of the BLUEPRINT monocye RNA-seq data from EGA (Data availability) and processed these the same as our sample.

Sequencing data preprocessing. All sequence datasets were aligned to human genome using Bowtie 2 with default settings; then they were mapped to human genome assembly GRCh38 using STAR (v.2.5.3a; https://github.com/alexbor/stars/). We converted the genotypes of single nucleotide variants (SNVs) and copy number variations (CNVs) to the genotype coordinates from GRCh37 to GRCh38 using CrossMap (v.0.5.2; http://crossmap.sourceforge.net/).

Bulk RNA-seq data of other myeloid cells and brain tissues. We downloaded fastq files of bulk RNA-seq of 6 primary microglia (pMICs) and 9 iPS-cell-derived microglia (iMICs)5, 10 monocyte-derived macrophage (MDMs) and iPS-cell-derived macrophages (IPSDMacs)6, 10 IMMCS, 8 MDMs and 4 pMICs7, 8 IMMCS, 9 iMICs and 3 pMICs8, 18 IPSDMac and 9 MDMs9, and 3 pMICs. See Supplementary Table 7 for details of cell types and sample sizes. For brain tissues, we downloaded the count table of RNA-seq data for all tissues from GTEx (v.7; Data availability) and extracted 1,671 brain samples. We also downloaded single-cell gene expression files of the BLUEPRINT monocye RNA-seq data from EGA (Data availability) and processed these the same as our sample.

Sequencing data preprocessing. All sequence datasets were aligned to human genome using Bowtie 2 with default settings; then they were mapped to human genome assembly GRCh38 using STAR (v.2.5.3a; https://github.com/alexbor/stars/). We converted the genotypes of single nucleotide variants (SNVs) and copy number variations (CNVs) to the genotype coordinates from GRCh37 to GRCh38 using CrossMap (v.0.5.2; http://crossmap.sourceforge.net/).
across genome-wide genes (where CPM effect to investigate the statistical significance of age effect. all mapped fragments) as random effects and the age of the patients as the fixed of mapped fragments, 96-well plate position where each cell was sorted, ERCC% (patient, pathology, brain region, brain hemisphere, ancestry and sex) and nonmicroglia) as a binary outcome. We used all possible clinical confounders R to fit the generalized linear mixed model for infiltrating cell status (microglia/our method when fitting donors and plates as batch effects, because these existing random effects with independent variance parameters RNA percentage, brain region, brain hemisphere, ancestry and sex) were fitted as correction approach was valid, we also compared with the three established batch correction methods (Harshley17), Seurat V3 (ref. 19) and Benji correct17. Our model returned a reasonable clustering of cells that was comparable to that from Seurat V3 (Extended Data Fig. 1d–h). It was not possible to compare the performance of our method when setting donors and plates as batch effects, because these existing methods do not scale to large numbers of batches (129 donors and 67 plates).

Characterization of infiltrating cells. We used the lme4 package implemented in R to fit the generalized linear mixed model for infiltrating cell status (microglia/ nonmicroglia) as a binary outcome. We used all possible clinical confounders (patient, pathology, brain region, brain hemisphere, ancestry and sex) and technical confounders (the number of expressed genes for each of the 20000+ cells). We performed a permutation test once for each gene and constructed the empirical value for each gene to perform the multiple testing correction genome wide. We also mapped eQTLs using RASQUAL (v.0.1; https://github.com/natsuhiko/ rasqual) with the raw count data and the same 25 PCs used in linear regression as for the covariates. We used the −no-posterior-update option to keep the posterior genotype dosage identical to the prior genotype dosage, which allowed us to stabilize the convergence of model fitting. We picked up the minimum BH threshold for each gene to perform the multiple testing correction genome wide. We performed a permutation test once for each gene and constructed the empirical null distribution to which the real Q values were compared to calibrate the FDR threshold. Colocalization analysis with GWAS traits was performed using COLOC13 implemented in R.

Bayesian hierarchical model. We extended a standard Bayesian hierarchical model17 to jointly map eQTLs in three different cell types. We employed the association Bayes factor at each variant for each gene to compute the regional Bayes factors (RBFS) in a cis region of 1 megabase (Mb) centered at the transcription start site under 15 different hypotheses. The RBFS were used in a hierarchical model to estimate prior probabilities that eQTLs were colocalized between any two of the three cell types, as well as shared among three cell types. This can provide posterior probability that a gene is an eQTL for each cell type. See Supplementary Note for more details.

Data availability
Patients were consented to share both expression and raw genotype data under managed access, and all data are available under managed access from theEGA, upon approval by the Wellcome Sanger Institute Data Access Committee. More details on how to access these data can be found at https://ega-archive.org/datasets/ EGAD000010005736. Raw data (fastq files and CRAM files) of Smart-seq2 and bulk RNA-seq for the primary microglia samples, as well as the raw genotype data (Illumina Omni 2.5) and imputed genotype data by Beagle software, are available from the European Phenome-Genome Archive (EGA) (accession ID: EGAD000010005736). Summary statistics of eQTLs mapped by linear regression and RASQUAL for primary microglia are also available from EGA (accession ID: EGAD000010005736). The 1000 Genomes Phase III integrated variant set can be obtained from the project website (http://www.internationalgenome.org/data/). GTEx v.7 summary statistics and brain Dnmt-seq data with cell type annotation data can be obtained from the GTEx project website (https://www.gtexportal .org/home/datasets). PMCM 68K single-cell transcriptome data can be obtained from the European Phenome-Genome Archive (EGA) (accession ID: EGAD000010002674). For details on how to access these data, please visit https://ega-archive.org/datasets/EGAD000010002674.

References
36. Alasoo, K. et al. Transcriptional profiling of macrophages derived from monocytes and iPSCs identifies a conserved response to LPS and novel alternative transcription. Sci. Rep. 5, 12524 (2015).
37. Douvaras, P. et al. Directed differentiation of human pluripotent stem cells to microglia. Stem Cell Rep. 8, 1516–1524 (2017).
38. Muffat, J. et al. Efficient derivation of microglia-like cells from human pluripotent stem cells. Nat. Med. 22, 1358–1367 (2016).
39. Zhang, H. et al. Functional analysis and transcriptomic profiling of iPSC-derived macrophages and their application in modeling Mendelian disease. *Circ. Res.* **117**, 17–28 (2015).

40. Chen, L. et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* **167**, 1398–1414.e24 (2016).

41. Jiang, H., Lei, R., Ding, S.-W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinform.* **15**, 182 (2014).

42. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

43. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

44. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

45. Kumasaka, N., Knights, A. J. & Gaffney, D. J. High-resolution genetic mapping of putative causal interactions between regions of open chromatin. *Nat. Genet.* **51**, 128–137 (2019).

46. Kang, H. M. et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018).

47. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).

48. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

49. Haghyverdi, L., Iam, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* https://doi.org/10.1038/nbt.4091 (2018)

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

51. Veyrieras, J.-B. et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* **4**, e1000214 (2008).

**Acknowledgements**

R.F. was supported by funding from the UK Multiple Sclerosis Society (MS50), the Adelson Medical Research Foundation and a core support grant from the Wellcome Trust and MRC to the Wellcome Trust-Medical Research Council Cambridge Stem Cell Institute (203151/Z/16/Z). A.Y. is supported by a Wellcome Trust Clinicians PhD Fellowship (203151/Z/16/Z). All data for this study were generated under Open Targets project OTAR039. N.K. and D.J.G. were funded by the Wellcome Trust grant WT206194. We thank the staff in the Cellular Genetics and Phenotyping and Sequencing core facilities at the Wellcome Sanger Institute.

**Author contributions**

The brain samples were obtained under neurosurgery by A.M.H.Y., H.B., J.C., K.P.B., A.J., M.R. Guilfoyle, R.T., R.K., R. Morris, R. Mair, C.W., S.J.P., F.J.K., T.S. and P.J.H. Cell isolation protocols were performed by A.M.H.Y., M.S., N.A.M. and C.E.M. Microglia isolation strategies were designed by A.M.H.Y., T.R.H., B.S. and R.J.M.F. Single-cell and bulk RNA-seq were performed by A.M.H.Y., A.K. and F.C. Cell culture experiments iPSDMacs were performed by A.K. and N.P. The main analyses and data preparations were performed by N.K. E.M. and M.G. processed GWAS summary statistics for colocalization analysis. J.S. and J.L. provided the summary statistics of Alzheimer’s disease. K.K. and N.S. provided the imputed BLUEPRINT genotype data. N.P. preprocessed the iPSeMac RNA-seq data. J.S.P. and O.A.B. performed the RNAscope and immunohistochemistry assay. A.M.H.Y., N.K., R.J.M.F. and D.J.G. wrote the manuscript. T.R.H. and B.S. assisted in editing the manuscript.

**Competing interests**

D.J.G. and E.M. were employees of Genomics PLC at the time the manuscript was submitted.

**Additional information**

Extended data is available for this paper at https://doi.org/10.1038/s41588-021-00875-2.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00875-2.

Correspondence and requests for materials should be addressed to R.J.M.F. or D.J.G.

Peer review information Nature Genetics thanks Julie Williams and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | See next page for caption.
Extended Data Fig. 1 | Overview of bulk and single-cell RNA-seq data. a, UMAP of bulk RNA-seq for myeloid cells. The ‘Primary microglia’ cluster contains samples collected in this study (pink dots) and previous studies (purple dots) (information on the source of previous study data can be found in Supplementary Table 7). ‘Cultured primary and IPS-derived cells’, includes IPS-derived macrophages and microglia (blue dots), cultured primary microglia and monocyte derived macrophages (orange dots). ‘Monocytes’ (green dots) denotes primary monocytes obtained from the BLUEPRINT project. b, Feature plots of three microglia marker genes (P2RY12, CX3CR1 and TMEM119) using the same UMAP coordinates as Fig. 1d. c, Age versus percentage of infiltrating cells. Red line shows the logistic regression line, the red transparent band shows the 95% confidence interval estimated using a generalised linear mixed model for the binary outcome (Materials and Methods). d, UMAP plot identical to Fig. 1d. e, UMAP plot from the first 12 principal components computed from the same input data for the linear mixed model without any batch correction. f, UMAP of the same 12 PCs where the batch effect was corrected by using Harmony47. g, UMAP of batch corrected data using the canonical correlation analysis method implemented in Seurat V348 with a default setting. We computed the 12 PCs from the integrated data for UMAP plot. h, UMAP of batch corrected data using MNN correct49. Note that points were coloured according to the cell types (same as Fig. 1d): glutamatergic neurons from the PFC (exPFC); pyramidal neurons from the hip CA region (exCA); GABAergic interneurons (GABA); granule neurons from the hip dentate gyrus region (exDG); astrocytes (ASC); oligodendrocytes (ODC); oligodendrocyte precursor cells (OPC); neuronal stem cells (NSC); endothelial cells (END); dendritic cell (DC); B cell (B); hematopoietic progenitor cell (CD34+); NK T cell (NK).
Extended Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | Microglia marker gene comparisons and validations. a, Marker gene enrichment analysis with Alzheimer’s disease associated microglia and glioma associated microglia. There are three different comparisons for Alzheimer’s disease associated microglia and 14 different populations for glioma associated microglia. Heatmap shows odds ratios and Benjamini-Hochberg (BH) Q-values of the Fisher exact tests between our marker genes and differentially expressed genes in other studies. b, Differentially expressed genes between microglia from different patient pathologies using single cell RNA-seq data. Heatmap shows averaged, normalised expression level (defined as the posterior mean of pathology random effect term, see Materials and Methods) of differentially expressed genes at local true sign rate (ltsr) greater than 0.9 (Urbut et al. 2019; see Materials and Methods for details). Heatmap is divided into groups based on all possible pairwise groupings of the four cell populations, ordered by most transcriptionally distinct, such that the most different grouping, trauma versus all non-trauma, appears at the top. c, Differential expression of candidate marker genes for immunohistochemistry in fresh frozen patient tissue samples. d, Immunohistochemistry panel of each pathology to validate expression of a differentially expressed gene at the protein level: hydrocephalus (C3), tumour (CCL4), haemorrhage (CD63) and trauma (BIN-1) compared to control. Iba-1 (red) and protein of interest (green). e, RNAscope image of differentially expressed gene panel for cluster C; HAMP (yellow) and RAC2 (purple) with C1QC (green) used to identify microglia. f, RNAscope image of differentially expressed gene panel for cluster D; KLF (yellow) and CCL20 (purple). Scale bar 10μM.
Extended Data Fig. 3 | Differential expression analysis with bulk RNA-seq data. a, Variance components analysis of log CPM values for the bulk RNA-seq data (N=102) with biological and technical factors using the linear mixed model (Online methods). b, Heatmap shows the effect size of age for each gene (each row) estimated by the linear mixed model (Online methods). The genes with LTSR > 0.9 in single-cell data are shown. c, PADI2 normalised expression in bulk RNA-seq data against patients' age. d, P2RY12 expression in bulk RNA-seq data against patients' age. e, Heatmap shows the average expression of males and females for each gene (each row) estimated by the linear mixed model (Online methods). The genes with LTSR > 0.9 in single-cell data are shown. f, C1QA normalised expression in bulk RNA-seq data for males (M) and females (F). g, HLA-DQB1 normalised expression in bulk RNA-seq data for males (M) and females (F). h, Heatmap shows the average expression for 5 different brain regions estimated by the linear mixed model (Online methods). The genes differentially expressed between a combination of Occipital and Cerebellum and the 3 other regions (LTSR > 0.9) in single-cell data are shown.
Extended Data Fig. 4 | See next page for caption.
Extended Data Fig. 4 | Colocalisation of eQTLs with various GWAS traits. a, eQTL effect size comparison for 502 eQTL genes at FDR 5% (linear regression) whose gene body contains at least one feature SNP with sufficient coverage (greater than 5% of average coverage across coding regions). The x-axis shows the eQTL effect size (beta) estimated from linear regression and the y-axis shows the eQTL effect size (pi value) from RASQUAL using only allele-specific count data. The red line shows the least square line crossing (0, 0.5). Note that, x = 0 is the null hypothesis for linear regression and y = 0.5 is the null hypothesis for RASQUAL. b, Examples of colocalised eQTLs in microglia. Colocalisation with Parkinson’s disease at KLHL7-AS1 eQTL (left column), colocalisation with Fed-up feelings at DAG1 eQTL (middle column) and colocalisation with Crohn’s disease at ERAP2 eQTL (right column). The y-axis of each panel shows log10 association Bayes factor for the eQTL or the GWAS trait. The colour of each point indicates LD index (r2 value) to the lead eQTL variant shown by the purple diamond. c, Heatmap of the posterior probability for colocalisation (PP4) between various GWAS traits and cell types/tissues. Each row corresponds to a specific combination of gene and a GWAS trait. Each column corresponds to eQTLs discovered in different cell types and tissues. The first column of the heatmap corresponds to microglia eQTLs, the second column corresponds to eQTLs in IPS cell derived macrophage (IPSDMac) from this study (Materials and Methods), the third column shows eQTLs in primary monocytes from the BLUEPRINT project (Materials and Methods) and the remaining 48 tissues are eQTLs from GTEx V7 (Materials and Methods). The colour of each grid shows the strength of PP4 (white: PP4 = 0.0 and red: PP4 = 1.0). Gray indicates that the gene was very weakly or not expressed, and therefore no eQTL summary statistics were available.
Extended Data Fig. 5 | See next page for caption.
Extended Data Fig. 5 | Fine mapping of microglia eQTLs. a, Regional association plots at the CD33 locus. b, Coverage plot shows the normalised expression level around the CD33 gene stratified by genotype at the putative splice variant (rs12459419C>T). The zoom-in panel shows a coverage plot of expression level around the second exon (ENST0000262262.4). The coverage shows the first intron expression is negatively correlated with the second exon expression, suggesting the expression of non-coding isoform (ENST00000601785.5) is increased by the alternative allele (T) of the splicing QTL. c, Colocalisation between an association with risk for Alzheimer’s disease on chromosome 2 and an eQTL for the noncoding RNA gene EPHA1-AS1 in microglia, GTEx tissues and myeloid cell types. The x-axis shows the posterior probability of colocalisation (PP4) and y-axis shows the average expression level (log10 TPM) for each tissue or cell type. d, Colocalisation between AD risk and expression of the protein-coding EPHA1 gene. The x-axis shows the posterior probability of colocalisation (PP4) and y-axis shows the average expression level (log10 TPM) for each tissue or cell type. e, Boxplots show the relationship between expression at the PTK2B gene and genotype at the lead eQTL variant (rs28834970C>T) three myeloid cell types. The y-axis shows normalised expression levels (log TPM value). Each dot on the box shows the expression level of a single sample. f, Coverage plot shows chromatin accessibility in IPS cell derived macrophages stratified by three genotype groups of the lead AD GWAS/BIN1 eQTL variant g, Scatter plot of MEF2C (x-axis) and BIN1 (y-axis) expression in GTEx brain tissues and myeloid cell type.
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.

- 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 any 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 used

Data analysis

- R 3.6.1 (https://www.r-project.org/)
- RASQUAL (https://github.com/natsuhiko/rasqual)
- Beagle 4.0 (https://faculty.washington.edu/browning/beagle/b4_0.html)
- bwa 0.7.4 (https://sourceforge.net/projects/bio-bwa/files/)
- skewer 0.1.27 (https://github.com/repliomics/skewer)
- STAR 2.5.3a (https://github.com/alexdobin/STAR/releases)
- featureCounts 2.0.0 (http://subread.sourceforge.net/)

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

Patients were consented to share both expression and raw genotype data under managed access and all data are available under managed access from the EGA, upon approval by the Wellcome Sanger Institute Data Access Committee. More details on how to access these data can be found at https://ega-archive.org/
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**: The total sample size is 141 patients undergoing neurosurgery. The number was determined so that the probability to observe at least one heterozygous/minor homozygous patient at a genetic variant with minor allele frequency 0.05 is greater than 0.9. With N=141, the probability was 0.9999995.

- **Data exclusions**: We haven’t excluded any sample.

- **Replication**: eQTLs were mapped in a population with the 141 unrelated patients.

- **Randomization**: We adjusted confounding effects on eQTLs by introducing principal components calculated both from the SNP genotypes and RNA-seq data.

- **Blinding**: Patients were enrolled randomly in the UK without knowing genetic data.

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 |

**Antibodies**

**Antibodies used**

- Primary antibodies; Iba-1 [1:1000, Wako, O19-19741], Iba-1 [1:300, abcam, ab5076], C3 (1:200 abcam, ab97462), CCL4 (1:50, r&d systems, MAB271), CD63 (1:300, abcam, ab59479) and BIN1 (1:500, abcam, ab182562)

**Validation**

- Iba-1 Wako: [https://www.fujifilmci.com/wp/wp-content/uploads/2020/07/Gel-Microglia-ICC-Application-Protocol_190508.pdf]
- Iba-1 Abcam [https://www.abcam.com/iba1-antibody-ab5076.html]
- C3 Abcam [https://www.abcam.com/c3-antibody-ab97462.html]
- CCL4 R&D [https://www.rndsystems.com/products/human-ccl4-mip-1beta-antibody-24006_mab271]
- CD63 Abcam [https://www.abcam.com/cd63-antibody-t63-bsa-and-azide-free-ab59479.html]
- BIN1 [https://www.abcam.com/bin1-antibody-epr13463-ab182562.html]
### Human research participants

Policy information about [studies involving human research participants](#).

| Population characteristics | Adult brain tissue biopsies were taken from the site of neurosurgery resection for the original clinical indication. |
|-----------------------------|---------------------------------------------------------------------------------------------------------------|
| Recruitment                 | Patients were offered the opportunity to donate waste tissue when undergoing a neurosurgical procedure for the clinical indication. |
| Ethics oversight            | Human brain tissue was obtained with informed consent under protocol REC 16/LO/2168 approved by the NHS Health Research Authority. |

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