Single nucleus sequencing fails to detect microglial activation in human tissue

N. Thrupp[1,2], C. Sala Frigerio[1,2,3], L. Wolfs[1,2], N. G. Skene[4], S. Poovathingal[1,2], Y. Fourne[1,2], P. M. Matthews[4], T. Theys[5], R. Mancuso[1,2], B. de Strooper[1,2,3,6], M. Fiers[1,2,3*]*

[1] Centre for Brain and Disease Research, Flanders Institute for Biotechnology (VIB), Leuven, Belgium

[2] Department of Neurosciences and Leuven Brain Institute, KU Leuven, Leuven, Belgium

[3] UK Dementia Research Institute at UCL, University College London, London, UK

[4] UK Dementia Research Institute at Imperial College London and Department of Brain Science, Imperial College London, London, UK

[5] Department of Neurosciences, Research group experimental neurosurgery and neuroanatomy, KU Leuven, Leuven, Belgium

** corresponding authors*

[6] lead contact

Abstract

Single nucleus RNA-Seq (snRNA-Seq) methods are used as an alternative to single cell RNA-Seq methods, as they allow transcriptomic profiling of frozen tissue. However, it is unclear whether snRNA-Seq is able to detect cellular state in human tissue. Indeed, snRNA-Seq analyses of human brain samples have failed to detect a consistent microglial activation signature in Alzheimer's Disease. A comparison of microglia from single cells and single nuclei of four human subjects reveals that ~1% of genes is depleted in nuclei compared to whole cells. This small population contains 18% of genes previously implicated in microglial activation, including *APOE, CST3, FTL, SPP1,* and *CD74*. We confirm our findings across multiple previous single nucleus and single cell studies. Given the low sensitivity of snRNA-Seq to this population of activation genes, we conclude that snRNA-Seq is not suited to detecting cellular activation in microglia in human disease.

Keywords

Microglia ; activation ; Alzheimer's Disease ; single nucleus RNA-Seq

Introduction

Single cell approaches allow us to study cell-to-cell heterogeneity (Habib et al., 2017), in brain material however, it is difficult to dissociate individual cells (Habib et al., 2017; Lake et al., 2016). This is further complicated if one is interested in studying the human brain, where often only frozen material is available. One alternative to study cellular transcriptional heterogeneity in brain tissue is single nucleus transcriptomics. Single nucleus RNA-Seq (snRNA-Seq) studies have shown concordance between single cell and single nucleus transcriptome profiles in mice (Bakken et al., 2018; Habib et al., 2017; Lake et al., 2017), but have limited the comparison to the identification of major cell types. It is unclear whether a snRNA-Seq approach is equally effective in identifying dynamic cellular substates such as microglial activation in human tissue.

A recent breakthrough in the field of Alzheimer's Disease (AD) using single cell RNA-Seq (scRNA-Seq) demonstrated clearly that microglia become activated in response to amyloid plaques in mouse models (Keren-Shaul et al., 2017). This response comprises a transcriptional switch to a state called Activation Response Microglia (ARM) (Sala Frigerio et al., 2019), or Disease-Associated Microglia (DAM, MGnD) (Keren-Shaul et al., 2017; Krasemann et al., 2017) . Ample evidence suggests that this microglial response is also relevant in human AD: microglia are believed to play a role in amyloid clearance (Efthymiou and Goate, 2017) and complement-mediated synapse loss (Fonseca et al., 2017), and histological studies have demonstrated considerable microgliosis around plaques in humans (McGeer et al., 1987). In addition, there is significant overlap between those genes involved in the microglial response, and genes within loci carrying AD genetic risk, as identified in Genome-Wide Association Studies (GWAS) (Efthymiou and Goate, 2017; Jansen et al., 2019; Kunkle et al., 2019; Lambert et al., 2013; Marioni et al., 2018), for example, *APOE*, *TREM2*, *APOC1*, *CD33* (Sala Frigerio et al., 2019). Most recently, the engrafting of human microglia into AD mouse models, followed by single cell RNA-sequencing, identified 66 DAM genes relevant to human activation[15], and a bulk RNA-Seq study of AD patients identified 64 DAM genes[16]. In stark contrast, a number of high-profile snRNA-Seq studies of microglia in human AD (Del-Aguila et al., 2019; Grubman et al., 2019; Mathys et al., 2019; Zhou et al., 2020) have not recovered a consistent microglial activation signature. A recent

cluster analysis by Mathys *et al.* of 48 AD patients and controls reported only 28 of 257 orthologous activation genes in common with the DAM signature (Mathys et al., 2019). Differential expression analysis between AD and control patients revealed 22 genes upregulated in AD patients (5 overlapping with the DAM signature). Of these AD genes, only 8 were also upregulated in another snRNA-Seq study of human AD (Grubman et al., 2019), and only 4 were also upregulated in another snRNA-Seq study of AD TREM2 variants (Zhou et al., 2020). The AD TREM2 variant study also only identified 11 DAM genes enriched in AD patients compared with controls. Del Aguila *et al.,* analysing single nucleus transcriptomics from 3 AD patients, were unable to recapitulate an activation signature (Del-Aguila et al., 2019). This has led to speculation that there is no such DAM signature in humans.

Here we compared the performance of snRNA-Seq to scRNA-Seq for the analysis of microglia from human cortical biopsies, and demonstrated that technical limitations inherent to snRNA-Seq provide a more likely explanation for this lack of consistency in snRNA-Seq studies of AD. We confirmed our results using publicly-available data.

Results

snRNA-Seq recovers major cell types from human tissue, but not microglial state

scRNA-Seq of FACS-sorted microglia was performed on temporal cortices of four human subjects who had undergone neocortical resection (see Supplementary Table 1 for subject data)(Mancuso et al., 2019). We generated snRNA-Seq libraries from these same subjects. Following quality filtering, PCA analysis and clustering of 37,060 nuclei, we identified 7 major cell types (Supplementary Fig. 1a, b): oligodendrocytes (ODC, 34.0%), excitatory neurons (27.0%), interneurons (11.2%), oligodendrocyte precursors (OPC 9,4%), microglia (11.3%), astrocytes (6.0%), and endothelial cells (1.1%). We focus here on the microglial population, which was extracted from the main dataset.

We first checked whether clustering analysis of single nuclei could recover subpopulations of microglia comparable to the single cell approach. A comparison of single nucleus and single cell clustering suggested that we could only partially recover similar microglial subcluster structure using both methods (see Supplementary Text and Supplementary Fig. 1c-e).

Gene expression profiling of human nuclei and cells

To compare gene abundance in single microglial cells (14,823 cells) and nuclei (3,940 nuclei), we performed a differential abundance analysis between cells and nuclei from the 4 subjects (Fig. 1a). As demonstrated in previous studies (Bakken et al., 2018; Gerrits et al., 2019; Habib et al., 2017; Lake et al., 2017), the majority of genes showed similar normalized abundance levels in cells and nuclei, with 98.6% of genes falling along the diagonal in Fig. 1a (Pearson's correlation coefficient = 0.92 , $p < 2.2e\text{-}16$). However, we identified a group of 246 genes (1.1% of detected genes) that was less abundant in nuclei (fold change < -2, $p_{adj} < 0.05$, blue in Fig. 1a). A second population of 68 genes (0.3%) was found to be more abundant in nuclei (fold change > 2, $p_{adj} < 0.05$, red in Fig. 1a). Additionally, 3,248 genes were exclusively detected in cells, and 5,068 genes exclusively detected in nuclei.

The observed differences in abundance between cells and nuclei were consistent across all four subjects (Fig. 1b, Supplementary Fig. 2a). Downsampling of cellular reads indicated that differences in abundance were not the result of different sequencing depths (Supplementary Fig. 2b,c). The full differential abundance results can be found in Supplementary Table 2.

To assess the robustness of this finding, we used our nuclei-abundant genes and cell-abundant genes to compare enrichment across all pairs of 8 publicly-available single cell or single nucleus datasets (Supplementary Table 3, Fig. 1c). We consistently found our nuclei-depleted (cell-abundant) genes to be depleted in other single nucleus microglia compared to single cell microglia (mean microglial Z-score of cell-abundant genes was 7.95 when comparing cells to nuclei, whereas cell-to-cell comparisons yielded a mean of 0.01, and nuclei-to-nuclei comparisons yielded a mean of 0.81, for Z-scores with $p_{adj} < 0.05$). We also

found our nuclei-abundant genes to be consistently enriched in other microglial nuclei compared with microglial cells (mean microglial Z-score -2.99 compared to -2.33 in nuclei against nuclei, no significant enrichment was found in cell-to-cell comparisons with $p_{adj} < 0.05$).

To assess functional enrichment among genes found to be more abundant in cells or nuclei, we ranked all genes according to log fold change (genes with a low abundance in nuclei had a negative log fold change) and performed a Gene Set Enrichment Analysis (GSEA, Subramanian et al., 2005) against gene markers from previous studies (Fig. 2a). For these analyses, a positive Normalised Enrichment Score (NES) represented nuclear enrichment, and a negative NES represented nuclear depletion. As expected, cytoplasmic RNA (defined by Bahar Halpern et al., 2015) was clearly enriched among genes found to be more abundant in cells (NES = -1.98, p_{adj} = 3.6e-05), as was mitochondrial mRNA (NES = -1.71, p_{adj} = 1.6e-04, gene set extracted from Ensembl's BioMart (Zerbino et al., 2018)). mRNA found to be more abundant in the nucleus by (Bahar Halpern et al., 2015) tended towards enrichment in nuclei but was not significant (NES = 0.87, p_{adj} = 8.2e-01), which is to be expected as scRNA-Seq captures both nuclear and cytoplasmic RNA. RNA of genes coding for ribosomal proteins were also depleted in nuclei (NES = -2.28, p_{adj} = 3.6e-05), as previously described[1]. Genes with shorter coding sequences (CDS) were depleted in nuclei (NES = -1.38, p_{adj} = 2.5e-02), while longer CDS were enriched (NES = 2.07, p_{adj} = 2.1e-05), as already observed in earlier snRNA-Seq studies[3]. Finally, the genes defined by Gerrits (Gerrits et al., 2019) as cellular-enriched in a differential analysis of microglial cells versus (fresh) nuclei in humans were also enriched in cells in our data, showing a NES score of -2.15 (p_{adj} = 3.6e-05).

To further characterise genes with higher or lower abundance in nuclei compared with cells, we performed GSEA, using Gene Ontology (GO) terms extracted from MSigDb (Liberzon et al., 2011) against the ranked log fold change. We selected the 100 terms with the highest NES, and the 100 terms with the lowest NES ($p_{adj} < 0.05$). Given the high overlap in terms, we clustered ontology terms based on the number of shared genes, in order to define "super" GO clusters (Supplementary Fig. 2e,f). We repeated the GSEA analysis using these *super*-GO clusters (Fig. 2b, Supplementary Tables 4 and 5) and observed an enrichment of neuronal and synaptic terms in nuclei-abundant genes (also shown in the red population in Fig.1a). We

suspect a synaptosome contamination during centrifugation. This is supported by the enrichment of synaptosome genes (NES = 1.82, p_{adj} = 3.6e-05, Fig. 2a; Hafner et al., 2019) and ambient RNA – mRNA originating not from cells/nuclei but from free-floating transcripts in the solution (Macosko et al., 2015) – (NES = 1.71, p_{adj} = 9.2e-05, Fig. 2a, Supplementary Fig. 2d) within the nucleus-abundant genes. The two gene sets share a strong overlap (Supplementary Table 6). These genes, although enriched in nuclei compared with cellular levels, still show low abundance (most of these genes show a normalised abundance of no more than 2 – Fig. 1a).

Activation genes identified in mouse models of AD are depleted in human nuclei

More interesting was the depletion of immune-related genes in nuclei (Fig. 2b). We therefore tested whether microglial activation genes[7,8,17] were also depleted in nuclei (Fig. 2c, Supplementary Tables 5 and 6). Remarkably, we found a strong depletion of genes associated with mouse microglial activation: 45 of 257 orthologous DAM genes (Keren-Shaul et al., 2017), (NES = -2.16, p_{adj} = 3.6e-05, Fig. 2c,d), and 28 of 200 orthologous ARM genes (Sala Frigerio et al., 2019) (NES = -2.01, p_{adj} = 3.6e-05, Fig. 2c,e), confirming that mouse microglial activation genes wereless abundant in nuclei. Genes upregulated by LPS stimulation in mice (Gerrits et al., 2019) also showed depletion in nuclei (NES = -1.86, p_{adj} = 3.6e-05, Fig. 2c, Supplementary Figure 2g).

Activation genes identified in mouse studies of AD are depleted in human nuclei

We next examined genes that were identified as markers of the human microglial response to AD in the recent snRNA-Seq study by (Mathys et al., 2019) (Fig. 2c, f, g). Markers of this response (referred to by Mathys *et al.* as "Mic1") had a NES score of -2.14 (p_{adj} = 3.6e-05), indicating that they were depleted in nuclei (Fig. 2c). The study identified 28 DAM genes as marker genes of the Mic1 response cluster (shown in orange in Fig. 2f); however the majority of DAM genes were not recovered using their snRNA-Seq protocol (purple in Fig. 2f). Fig. 2g shows in green all the markers of the human activation cluster Mic1. Clearly, DAM genes and other Mic1 markers showed higher abundance in cells relative to nuclei (confirming the

NES score in Fig. 2c). Further, it seems likely that the recovered DAM genes (orange in Fig. 2f) and Mic1 markers in general (green in Fig. 2g), were detected in the original snRNA-Seq experiment owing to their higher nuclear abundance compared with the nuclear abundance of other genes, including those DAM genes that were not recovered (purple in Fig. 2f).

Discussion

In summary, in our comparison of snRNA-Seq and scRNA-Seq of human microglia, we identified a set of genes (1.1% of the gene population) with at least 2-fold lower abundance in nuclei compared to their cellular levels (Fig. 1a-b). This small set is strongly enriched for genes previously associated with microglial activation in mouse models of AD, for example *APOE*, *CST3*, *FTL*, *SPP1*, and *CD74* (Fig. 2b-e). Thus, while our work agrees with previous experiments demonstrating that snRNA-Seq can determine cell type (Supplementary Fig. 1a,b), we argue that there are important limitations when studying cellular state in humans. This limitation is likely responsible for the difficulty in identifying consistent DAM- or ARM-like gene populations in the human brain in snRNA-Seq-based studies. We identified similar patterns of depletion in other single nucleus microglia (Fig. 2c).

Examination of data from the Mathys *et al.* study of human nuclei in AD (Mathys et al., 2019) shows that only genes with higher nuclear abundance levels were detected (Fig. 2f, g). This suggests that the discordance between human and mouse microglial activation is at least in part a consequence of limitations in the technology, rather than biological differences between the species as current snRNA-Seq suggest. Deeper sequencing (or increased sample sizes) may compensate for this lack of sensitivity. However, the sparse nature of snRNA-Seq and the high level of heterogeneity in human samples, combined with the fact that many relevant genes have a more than two-fold lower abundance in nuclei (*e.g. APOE* fold change = 2.57, *CST3* fold change = 3.44, *FTL* fold change = 6.53), strongly suggests that this will remain a problem.

While our data is (at least partially) in agreement (Fig. 2a, c) with Gerrits *et al.* (Gerrits et al., 2019) which also compares nuclei with cells in human microglia, they did not report a

nuclear depletion of activation genes. We suspect the reason for this is (a) the low human sample number (n=2); (b) for the cluster analysis, Gerrits *et al.* scaled cell and nuclei expression to mitochondrial and ribosomal reads, essentially masking differences between nuclei and cells, and (c) for the differential expression analysis, Gerrits *et al.* compared fresh cells to fresh nuclei, as opposed to frozen nuclei.

Alternative approaches may be more suitable for generating a brain atlas of human disease such as AD, particularly where we are limited to frozen material. *In situ* spatial transcriptomics (ST) negates issues related to tissue dissociation and cell or nucleus isolation (Ståhl et al., 2016), while at the same time retaining spatial information. This approach has recently been applied to examine transcriptomic changes and identify genes that are co-expressed across multiple cell types in the amyloid plaque niche of the mouse brain (Chen et al., 2019). In humans, a similar methodology was recently applied to identify pathway dysregulation and regional differences in cellular states of the postmortem spinal tissue of Amyotrophic Lateral Sclerosis (ALS) patients states (Maniatis et al., 2019). Its application to AD patients may shed light on transcriptomic changes occurring in microglia which localize near plaques, and may also provide insights into the crosstalk occurring between neighbouring cells.

In conclusion, while snRNA-Seq offers a viable alternative to scRNA-Seq for identification of cell types in tissue where cell dissociation is problematic, its utility for detecting cellular states in disease is limited.

Data Availability

Sequencing data from single microglial cells is available on GEO (accession number GSE137444). Sequencing data from single nuclei will be made available on GEO.

Code Availability

Analysis of previous datasets was performed using the EWCE package (Skene and Grant, 2016) for R and the MicroglialDepletion package (https://github.com/NathanSkene/MicroglialDepletion).

Acknowledgements

Work in the De Strooper laboratory was supported by the European Union (ERC- 834682 - CELLPHASE_AD), the Fonds voor Wetenschappelijk Onderzoek (FWO), KU Leuven, VIB, UK-DRI (Medical Research council, ARUK and Alzheimer Society), a Methusalem grant from KU Leuven and the Flemish Government, the "Geneeskundige Stichting Koningin Elisabeth", Opening the Future campaign of the Leuven Universitair Fonds (LUF), the Belgian Alzheimer Research Foundation (SAO-FRA) and the Alzheimer's Association USA. Bart De Strooper is holder of the Bax-Vanluffelen Chair for Alzheimer's Disease. Cell sorting was performed at the KU Leuven FACS core facility, and sequencing was carried out by the VIB Nucleomics Core. Renzo Mancuso is recipient of a postdoctoral fellowship from the Alzheimer's Association, USA. Paul M. Matthews acknowledges generous support from the Edmond J. Safra Foundation and Lily Safra, the NIHR Biomedical Research Centre at Imperial College and the UK Dementia Research Institute.

Author contributions

Conceptualization: B.D.S., M.F., C.S.F., R.M. ; surgery and extraction of patient tissue samples: T.T. ; Investigation: R.M., L.W., S.P. ; formal analysis: N.T., Y.F., M.F., N.G.S., P.M.M. ; Writing – original draft: N.T., C.S.F., M.F. ; writing – review & editing: B.D.S., R.M., N.G.S., P.M.M. ; supervision: M.F., B.D.S.

Competing interests

The authors declare that there are no competing interests.

Materials and Methods

Isolation of human primary microglial cells

Human primary microglial cells from the Mancuso *et. al* study (Mancuso et al., 2019) were used. Briefly, microglia were FAC-sorted from brain tissue samples resected from the lateral temporal neocortex of 4 epilepsy patients during neurosurgery. The full protocol is described in the original study. All procedures followed protocols approved by the local Ethical Committee (protocol number S61186). Sequencing was performed as described for the nuclei.

Isolation of nuclei from human subjects

Nuclei from frozen biopsy tissue were isolated as follows: brain tissue was sliced on dry ice, then homogenized manually (15 gentle strokes) with 1mL ice-cold Homogenisation Buffer (HB) with 5µL Rnasin Plus. The homogenate was strained with a 70µm strainer and washed with 1.65mL to a final volume of 2.65mL. 2.65mL of Gradient medium was added (Vf = 5.3mL). To isolate the nuclei, the sample was added to a 4mL 29% cushion using a P1000, and the weight adjusted with HB. The sample was centrifuged in a SW41Ti rotor at 7,700 rpm for 30 minutes at 4°C. The supernatant was removed with a plastic Pasteur pipette, followed by removal of the lower supernatant with P200. Nuclei were resuspended in 200µL of resuspension buffer, transferred to a new tube, washed again with 100-200µL resuspension buffer, and pooled with the previous solution. Clumps were disrupted by pipetting with P200, then filtered through a Falcon tube with 0.35µm strainer. 9µL of sample was mixed with 1µL of propidium iodide (PI) stain, loaded onto a LUNA-FL slide and allowed to settle for 30 seconds. We viewed nuclei with the LUNA-FL Automated cell counter to check numbers and shape.

Single nucleus sequencing

RNA sequencing was performed using the 10X Genomics Single Cell 3` Reagent Kit (v2) according to manufacturer protocols. cDNA libraries from fresh-frozen nuclei were sequenced on an Illumina HiSeq platform 4000. Supplementary Table 1 provides sequencing information per sample (for cells and nuclei).

Single nucleus analysis

Alignment. Cellranger v2.1.1 was used to demultiplex and align sequencing output to a human reference genome (assembly hg38 build 95). We used a "pre-mRNA" database to align single nuclei to exons as well as introns (10x Genomics, n.d.). Following alignment, nuclei from one patient sample (RM101.1) were removed due to poor quality (low read and gene count). See Supplementary Table 1 for sample information. Unfiltered count matrices were used for downstream analysis.

Extraction of microglial nuclei. Data was processed using the Seurat v3.0.2 package (Butler et al., 2018; Stuart et al., 2019) in R v3.6.1. For each patient, the count matrix was filtered to exclude nuclei with fewer than 100 genes. Counts were depth-normalised, scaled by 10,000 and log-transformed. *FindVariableFeatures* was run using a variance-stabilising transformation ("vst") to identify the 2,000 most variable genes in each sample. Data from the 4 patients was then integrated using Seurat's *FindIntegrationAnchors* with default parameters, and *IntegrateData* using 40 principal components (PCs). The dataset consisted of 37,060 nuclei, with a mean read depth of 4,305 counts per nucleus, and 1,791 genes per nucleus. Integrated data was scaled (default Seurat parameters). We ran a Principal Components Analysis (PCA), then calculated Uniform Manifold Approximation and Projection (UMAP) embeddings using 40 PCs. We identified clusters using Seurat's *FindNeighbours* and *FindClusters* functions, again using 40 PCs. Based on abundance of known celltype markers, we assigned each cluster to a cell type. Microglial clusters were identified using known markers including *P2RY12*, *CSF1R*, *CX3CR1*, and extracted for downstream analysis.

Pre-processing of microglial nuclei per patient. Microglia from each patient sample were analysed individually as described for all cell types above, with the following modifications: raw counts were filtered to remove genes and counts that were ± 3 standard deviations away from the median value. After normalization, doublets were identified using DoubletFinder v2.0.2 (McGinnis et al., 2019) using 40 PCs, assuming a 7.5% doublet rate. Following removal of doublets, filtering and Seurat normalization were performed again. Data from patients was then integrated and clusters were identified as above. We discarded small clusters than contained markers for microglia as well as other cell types. After pre-processing, 3,927 nuclei remained, with a mean count depth of 1,295 and a mean gene count of 879 genes per nucleus.

Single cell analysis

Full details of single cell processing are available in Mancuso *et al.*(Mancuso et al., 2019). Only cells from the four patients included in the single nucleus study were used here.

Comparisons of single cells and single nuclei

Cluster analysis (see supplementary text). In order to identify microglial cell states in the nuclei data we calculated gene markers for each cluster using Seurat's *FindMarkers* function, selecting only markers with a positive fold change. Gene markers for cell clusters were extracted from the original Mancuso *et al.* study. Markers for nuclei and cells are available in Supplementary Table 7. For the analysis, we kept the top 40 significant markers ($p_{adj} < 0.05$) based on log fold change for the nuclear clusters and cellular clusters. For each nucleus, we calculated the mean abundance levels of each cell cluster marker set against the aggregated abundance of random control gene sets, using Seurat's *AddModuleScore* function. This gave us the MS40 score for each marker set. We performed two-sided Fisher's Exact tests with Benjamini Hochberg corrections to determine the overlap of cell cluster markers with nuclear cluster markers (selecting the top 40 markers for each set), using the union of all genes in the cell and nuclei datasets as a background ($p_{adj} < 0.05$ was considered significant).

Differential Abundance. We discarded all non-microglial clusters (brain macrophages, neutrophils), leaving 3,721 nuclei and 14,435 cells. Differential abundance analysis was performed with the Seurat package, using a two-sided Wilcoxon rank sum test, with a Bonferroni correction for multiple testing. Genes with $p_{adj} < 0.05$ and fold change > |2| were considered significant. As Seurat applies a pseudocount of +1 to data before calculating log fold changes, a fold change of 2 corresponds to a log fold change of 0.63. Log fold changes calculated by Seurat were used for further analysis in gene set enrichment analysis.

Scatter plots. We calculated the mean of the normalized abundance levels for cells and for nuclei, and log-transformed these values.

Assessment of nuclear-enriched or cell-enriched gene sets in public scRNA-Seq and snRNA-Seq datasets. We followed the methodology described in (Skene et al., 2018): genes that were significantly more abundant in nuclei or more abundant in cells (see "Differential abundance" methodology above) were used, creating two gene sets. 8 public datasets were reduced to contain six major cell types: pyramidal neurons, interneurons, astrocytes, interneurons, microglia and oligodendrocyte precursors. Within each dataset, for each gene in our gene sets,

we calculated a celltype specificity score using the EWCE R package (Github version committed July 29, 2019; Skene and Grant, 2016). For each pair of datasets, *X* and *Y*, we subtracted the mean microglial specificity score of *Y* from *X*. We then calculated the same scores for 10,000 random gene sets: the probability and z-score for the difference in specificity for the dendritic genes is calculated using these. Finally, the depletion z-score for each gene set was equal to: (mean subtracted microglial specificity score – bootstrapped mean) / (bootstrapped standard deviation). A large positive z-score thus indicated that the gene set was depleted in microglia of dataset *Y* relative to dataset *X*. Benjamini-Hochberg multiple testing corrections were applied.

Public datasets. For the Karolinska Institutet (KI) dataset (Skene et al., 2018), we used S1 pyramidal neurons. For the Zeisel 2018 dataset (Zeisel et al., 2018) we used all ACTE* cells as astrocytes, TEGLU* as pyramidal neurons, TEINH* as interneurons, OPC as oligodendrocyte precursors and MGL* as microglia. For the Saunders dataset (Saunders et al., 2018), we used all Neuron.Slc17a7 celltypes from the frontal cortex (FC), hippocampus (HC) or posterior cortex (PC) as pyramidal neurons; all Neuron.Gad1Gad2 cell types from FC, HC or PC as interneurons; Polydendrocye as OPCs; Astrocyte as astrocytes, and Microglia as microglia. The Lake datasets both came from a single publication (Lake et al., 2018) which had data from frontal cortex, visual cortex and cerebellum. The cerebellum data was not used here. Data from frontal and visual cortices were analyzed separately. All other datasets - Dronc Human (Habib et al., 2017), Dronc Mouse (Habib et al., 2017), Allen Institute for Brain Science (AIBS) (Hodge et al., 2019), Tasic (Tasic et al., 2016) and Habib (Habib et al., 2016) – were used as described previously (Skene et al., 2018). Supplementary Table 3 lists all datasets. An R package is available for the analysis at https://github.com/NathanSkene/MicroglialDepletion.

Functional analysis. We performed Gene Set Enrichment Analysis (GSEA) using the R package fgsea v1.8.0 (Sergushichev, 2016), using default parameters. Gene sets were mapped against a list of genes ranked according to fold change between cellular abundance and nuclear abundance. Gene ontology (GO) sets were obtained from MSigDB (Liberzon et al., 2011; Subramanian et al., 2005). Other gene sets were obtained from previous studies (see Supplementary Table 6). $p_{adj} < 0.05$ (Benjamini-Hochberg correction) was considered significant.

Clustering of gene ontology terms. GSEA of GO terms resulted in many functional categories with overlapping genes. In order to reduce this redundancy, the top and bottom 100 GO terms

according to normalized enrichment score (with $p_{adj} < 0.05$) were clustered as follows: a Jaccard index (the size of the intersection of the two datasets, divided by the size of the union of the two datasets, multiplied by 100) of the overlapping genes was calculated between each significant GO set. The resulting similarity matrix was converted to a dissimilarity matrix, and hierarchical clustering was performed on the matrix. We selected a k value of 16 to group the GO terms based on the hierarchical clustering (see Supplementary Table 2). Gene sets were merged, and each new "super" GO was assigned an annotation manually. GSEA analysis was performed on these *super*-GO gene sets as described above.

Gene sets from previous studies. We extracted gene sets from previous studies for this analysis. A full list of gene sets is available in Supplementary Table 6. Where data was selected from mouse datasets, we converted the mouse gene to its human ortholog using R's BioMaRt package v2.40.5 (Durinck et al., 2009), selecting only orthologs that displayed 1-to-1 orthology. For the ARM gene set we selected the top 200 ARM genes based on log fold change (Sala Frigerio et al., 2019). For the Gerrits human gene set, we took the union of all genes that showed significant differential abundance between cells and nuclei (microglia) from donor 1 and donor 2 (Gerrits et al., 2019). For the LPS gene set, we took the union of all genes significantly upregulated in LPS in cells and in nuclei (microglia) from the Gerrits study (Gerrits et al., 2019).

Downsampling of cell counts. To match cell sequencing depth to nucleus sequencing depth (see Supplementary Fig. 2b,c), we sampled without replacement the number of reads in the cells by a proportion of 0.32, using the *downsampleMatrix* function of the DropletUtils R package v1.4.3 (Griffiths et al., 2018; Lun et al., 2019). This resulted in a read depth of 1,304 compared with the original read depth of 3,979 reads per cell.

Definition of ambient RNA profile in nuclei. We extracted nuclei with less than 700 counts from the original unfiltered raw count matrix of all cell types (resulting in 2,414 nuclei with a mean read depth of 590), and summed the gene counts, under the assumption that these were empty drops rather than nuclei. We took the top 150 genes to represent the ambient RNA profile. The mean read depth of these genes in the empty drops was 121 reads per cell.

References

10x Genomics, n.d. Creating a Reference Package with cellranger mkref -Software -Single Cell Gene Expression -Official 10x Genomics Support [WWW Document]. URL https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/advanced/references#premrna (accessed 11.13.19).

Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I., Itzkovitz, S., 2015. Nuclear Retention of mRNA in Mammalian Tissues. Cell Rep. 13, 2653–2662. https://doi.org/10.1016/j.celrep.2015.11.036

Bakken, T.E., Hodge, R.D., Miller, J.A., Yao, Z., Nguyen, T.N., Aevermann, B., Barkan, E., Bertagnolli, D., Casper, T., Dee, N., Garren, E., Goldy, J., Graybuck, L.T., Kroll, M., Lasken, R.S., Lathia, K., Parry, S., Rimorin, C., Scheuermann, R.H., Schork, N.J., Shehata, S.I., Tieu, M., Phillips, J.W., Bernard, A., Smith, K.A., Zeng, H., Lein, E.S., Tasic, B., 2018. Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. PLOS ONE 13, e0209648. https://doi.org/10.1371/journal.pone.0209648

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., Satija, R., 2018. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420. https://doi.org/10.1038/nbt.4096

Chen, W.-T., Lu, A., Craessaerts, K., Pavie, B., Sala Frigerio, C., Mancuso, R., Qian, X., Lalakova, J., Kühnemund, M., Voytyuk, I., Wolfs, L., Snellinx, A., Munck, S., Jurek, A., Fernandez Navarro, J., Saido, T.C., Lundeberg, J., Fiers, M., De Strooper, B., 2019. Spatial and temporal transcriptomics reveal microglia-astroglia crosstalk in the amyloid-β plaque cell niche of Alzheimer's disease (preprint). Neuroscience. https://doi.org/10.1101/719930

Del-Aguila, J.L., Li, Z., Dube, U., Mihindukulasuriya, K.A., Budde, J.P., Fernandez, M.V., Ibanez, L., Bradley, J., Wang, F., Bergmann, K., Davenport, R., Morris, J.C., Holtzman, D.M., Perrin, R.J., Benitez, B.A., Dougherty, J., Cruchaga, C., Harari, O., 2019. A single-nuclei RNA sequencing study of Mendelian and sporadic AD in the human brain. Alzheimers Res. Ther. 11, 71. https://doi.org/10.1186/s13195-019-0524-x

Durinck, S., Spellman, P.T., Birney, E., Huber, W., 2009. Mapping Identifiers for the Integration of Genomic Datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4, 1184–1191. https://doi.org/10.1038/nprot.2009.97

Efthymiou, A.G., Goate, A.M., 2017. Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. Mol. Neurodegener. 12, 43. https://doi.org/10.1186/s13024-017-0184-x

Fonseca, M.I., Chu, S.-H., Hernandez, M.X., Fang, M.J., Modarresi, L., Selvan, P., MacGregor, G.R., Tenner, A.J., 2017. Cell-specific deletion of C1qa identifies microglia as the dominant source of C1q in mouse brain. J. Neuroinflammation 14, 48. https://doi.org/10.1186/s12974-017-0814-9

Friedman, B.A., Srinivasan, K., Ayalon, G., Meilandt, W.J., Lin, H., Huntley, M.A., Cao, Y., Lee, S.-H., Haddick, P.C.G., Ngu, H., Modrusan, Z., Larson, J.L., Kaminker, J.S., Brug, M.P. van der, Hansen, D.V., 2018. Diverse Brain Myeloid Expression Profiles Reveal Distinct Microglial Activation States and Aspects of Alzheimer's Disease Not Evident in Mouse Models. Cell Rep. 22, 832–847. https://doi.org/10.1016/j.celrep.2017.12.066

Gerrits, E., Heng, Y., Boddeke, E.W.G.M., Eggen, B.J.L., 2019. Transcriptional profiling of microglia; current state of the art and future perspectives. Glia n/a. https://doi.org/10.1002/glia.23767

Griffiths, J.A., Richard, A.C., Bach, K., Lun, A.T.L., Marioni, J.C., 2018. Detection and removal of barcode swapping in single-cell RNA-seq data. Nat. Commun. 9, 1–6. https://doi.org/10.1038/s41467-018-05083-x

Grubman, A., Chew, G., Ouyang, J.F., Sun, G., Choo, X.Y., McLean, C., Simmons, R.K., Buckberry, S., Vargas-Landin, D.B., Poppe, D., Pflueger, J., Lister, R., Rackham, O.J.L., Petretto, E., Polo, J.M., 2019. A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. Nat. Neurosci. 1–11. https://doi.org/10.1038/s41593-019-0539-4

Habib, N., Avraham-Davidi, I., Basu, A., Burks, T., Shekhar, K., Hofree, M., Choudhury, S.R., Aguet, F., Gelfand, E., Ardlie, K., Weitz, D.A., Rozenblatt-Rosen, O., Zhang, F., Regev, A., 2017. Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat. Methods 14, 955–958. https://doi.org/10.1038/nmeth.4407

Habib, N., Li, Y., Heidenreich, M., Swiech, L., Avraham-Davidi, I., Trombetta, J.J., Hession, C., Zhang, F., Regev, A., 2016. Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. Science 353, 925–928. https://doi.org/10.1126/science.aad7038

Hafner, A.-S., Donlin-Asp, P.G., Leitch, B., Herzog, E., Schuman, E.M., 2019. Local protein synthesis is a ubiquitous feature of neuronal pre- and postsynaptic compartments. Science 364. https://doi.org/10.1126/science.aau3644

Hasselmann, J., Coburn, M.A., England, W., Figueroa Velez, D.X., Kiani Shabestari, S., Tu, C.H., McQuade, A., Kolahdouzan, M., Echeverria, K., Claes, C., Nakayama, T., Azevedo, R., Coufal, N.G., Han, C.Z., Cummings, B.J., Davtyan, H., Glass, C.K., Healy, L.M., Gandhi, S.P., Spitale, R.C., Blurton-Jones, M., 2019. Development of a Chimeric Model to Study and Manipulate Human Microglia In Vivo. Neuron. https://doi.org/10.1016/j.neuron.2019.07.002

Hodge, R.D., Bakken, T.E., Miller, J.A., Smith, K.A., Barkan, E.R., Graybuck, L.T., Close, J.L., Long, B., Johansen, N., Penn, O., Yao, Z., Eggermont, J., Höllt, T., Levi, B.P., Shehata, S.I., Aevermann, B., Beller, A., Bertagnolli, D., Brouner, K., Casper, T., Cobbs, C., Dalley, R., Dee, N., Ding, S.-L., Ellenbogen, R.G., Fong, O., Garren, E., Goldy, J., Gwinn, R.P., Hirschstein, D., Keene, C.D., Keshk, M., Ko, A.L., Lathia, K., Mahfouz, A., Maltzer, Z., McGraw, M., Nguyen, T.N., Nyhus, J., Ojemann, J.G., Oldre, A., Parry, S., Reynolds, S., Rimorin, C., Shapovalova, N.V., Somasundaram, S., Szafer, A., Thomsen, E.R., Tieu, M., Quon, G., Scheuermann, R.H., Yuste, R., Sunkin, S.M., Lelieveldt, B., Feng, D., Ng, L., Bernard, A., Hawrylycz, M., Phillips, J.W., Tasic, B., Zeng, H., Jones, A.R., Koch, C., Lein, E.S., 2019. Conserved cell types with divergent features in human versus mouse cortex. Nature 1–8. https://doi.org/10.1038/s41586-019-1506-7

Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., Sealock, J., Karlsson, I.K., Hägg, S., Athanasiu, L., Voyle, N., Proitsi, P., Witoelar, A., Stringer, S., Aarsland, D., Almdahl, I.S., Andersen, F., Bergh, S., Bettella, F., Bjornsson, S., Brækhus, A., Bråthen, G., de Leeuw, C., Desikan, R.S., Djurovic, S., Dumitrescu, L., Fladby, T., Hohman, T.J., Jonsson, P.V., Kiddle, S.J., Rongve, A., Saltvedt, I., Sando, S.B., Selbæk, G., Shoai, M., Skene, N.G., Snaedal, J., Stordal, E., Ulstein, I.D., Wang, Y., White, L.R., Hardy, J., Hjerling-Leffler, J., Sullivan, P.F., van der Flier, W.M., Dobson, R., Davis, L.K., Stefansson, H., Stefansson, K., Pedersen, N.L., Ripke, S., Andreassen, O.A., Posthuma, D., 2019. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat. Genet. https://doi.org/10.1038/s41588-018-0311-9

Keren-Shaul, H., Spinrad, A., Weiner, A., Matcovitch-Natan, O., Dvir-Szternfeld, R., Ulland, T.K., David, E., Baruch, K., Lara-Astaiso, D., Toth, B., Itzkovitz, S., Colonna, M., Schwartz, M., Amit, I., 2017. A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. Cell 169, 1276-1290.e17. https://doi.org/10.1016/j.cell.2017.05.018

Krasemann, S., Madore, C., Cialic, R., Baufeld, C., Calcagno, N., Fatimy, R.E., Beckers, L., O'Loughlin, E., Xu, Y., Fanek, Z., Greco, D.J., Smith, S.T., Tweet, G., Humulock, Z., Zrzavy, T., Conde-Sanroman, P., Gacias, M., Weng, Z., Chen, H., Tjon, E., Mazaheri, F., Hartmann, K., Madi, A., Ulrich, J.D., Glatzel, M., Worthmann, A., Heeren, J., Budnik, B., Lemere, C., Ikezu, T., Heppner, F.L., Litvak, V., Holtzman, D.M., Lassmann, H., Weiner, H.L., Ochando, J., Haass, C., Butovsky, O., 2017. The TREM2-APOE Pathway Drives the Transcriptional Phenotype of Dysfunctional Microglia in Neurodegenerative Diseases. Immunity 47, 566-581.e9. https://doi.org/10.1016/j.immuni.2017.08.008

Kunkle, B.W., Grenier-Boley, B., Sims, R., Bis, J.C., Damotte, V., Naj, A.C., Boland, A., Vronskaya, M., Lee, S.J. van der, Amlie-Wolf, A., Bellenguez, C., Frizatti, A., Chouraki, V., Martin, E.R., Sleegers, K., Badarinarayan, N., Jakobsdottir, J., Hamilton-Nelson, K.L., Moreno-Grau, S., Olaso, R., Raybould, R., Chen, Y., Kuzma, A.B., Hiltunen, M., Morgan, T., Ahmad, S., Vardarajan, B.N., Epelbaum, J., Hoffmann, P., Boada, M., Beecham, G.W., Garnier, J.-G., Harold, D., Fitzpatrick, A.L., Valladares, O., Moutet, M.-L., Gerrish, A., Smith, A.V., Qu, L., Bacq, D., Denning, N., Jian, X., Zhao, Y., Zompo, M.D., Fox, N.C., Choi, S.-H., Mateo, I., Hughes, J.T., Adams, H.H., Malamon, J., Sanchez-Garcia, F., Patel, Y., Brody, J.A., Dombroski, B.A., Naranjo, M.C.D., Daniilidou, M., Eiriksdottir, G., Mukherjee, S., Wallon, D., Uphill, J., Aspelund, T., Cantwell, L.B., Garzia, F., Galimberti, D., Hofer, E., Butkiewicz, M., Fin, B., Scarpini, E., Sarnowski, C., Bush, W.S., Meslage, S., Kornhuber, J., White, C.C., Song, Y., Barber, R.C., Engelborghs, S., Sordon, S., Voijnovic, D., Adams, P.M., Vandenberghe, R.,

Mayhaus, M., Cupples, L.A., Albert, M.S., Deyn, P.P.D., Gu, W., Himali, J.J., Beekly, D., Squassina, A., Hartmann, A.M., Orellana, A., Blacker, D., Rodriguez-Rodriguez, E., Lovestone, S., Garcia, M.E., Doody, R.S., Munoz-Fernadez, C., Sussams, R., Lin, H., Fairchild, T.J., Benito, Y.A., Holmes, C., Karamujić-Čomić, H., Frosch, M.P., Thonberg, H., Maier, W., Roschupkin, G., Ghetti, B., Giedraitis, V., Kawalia, A., Li, S., Huebinger, R.M., Kilander, L., Moebus, S., Hernández, I., Kamboh, M.I., Brundin, R., Turton, J., Yang, Q., Katz, M.J., Concari, L., Lord, J., Beiser, A.S., Keene, C.D., Helisalmi, S., Kloszewska, I., Kukull, W.A., Koivisto, A.M., Lynch, A., Tarraga, L., Larson, E.B., Haapasalo, A., Lawlor, B., Mosley, T.H., Lipton, R.B., Solfrizzi, V., Gill, M., Longstreth, W.T., Montine, T.J., Frisardi, V., Diez-Fairen, M., Rivadeneira, F., Petersen, R.C., Deramecourt, V., Alvarez, I., Salani, F., Ciaramella, A., Boerwinkle, E., Reiman, E.M., Fievet, N., Rotter, J.I., Reisch, J.S., Hanon, O., Cupidi, C., Uitterlinden, A.G.A., Royall, D.R., Dufouil, C., Maletta, R.G., Rojas, I. de, Sano, M., Brice, A., Cecchetti, R., George-Hyslop, P.S., Ritchie, K., Tsolaki, M., Tsuang, D.W., Dubois, B., Craig, D., Wu, C.-K., Soininen, H., Avramidou, D., Albin, R.L., Fratiglioni, L., Germanou, A., Apostolova, L.G., Keller, L., Koutroumani, M., Arnold, S.E., Panza, F., Gkatzima, O., Asthana, S., Hannequin, D., Whitehead, P., Atwood, C.S., Caffarra, P., Hampel, H., Quintela, I., Carracedo, Á., Lannfelt, L., Rubinsztein, D.C., Barnes, L.L., Pasquier, F., Frölich, L., Barral, S., McGuinness, B., Beach, T.G., Johnston, J.A., Becker, J.T., Passmore, P., Bigio, E.H., Schott, J.M., Bird, T.D., Warren, J.D., Boeve, B.F., Lupton, M.K., Bowen, J.D., Proitsi, P., Boxer, A., Powell, J.F., Burke, J.R., Kauwe, J.S.K., Burns, J.M., Mancuso, M., Buxbaum, J.D., Bonuccelli, U., Cairns, N.J., McQuillin, A., Cao, C., Livingston, G., Carlson, C.S., Bass, N.J., Carlsson, C.M., Hardy, J., Carney, R.M., Bras, J., Carrasquillo, M.M., Guerreiro, R., Allen, M., Chui, H.C., Fisher, E., Masullo, C., Crocco, E.A., DeCarli, C., Bisceglio, G., Dick, M., Ma, L., Duara, R., Graff-Radford, N.R., Evans, D.A., Hodges, A., Faber, K.M., Scherer, M., Fallon, K.B., Riemenschneider, M., Fardo, D.W., Heun, R., Farlow, M.R., Kölsch, H., Ferris, S., Leber, M., Foroud, T.M., Heuser, I., Galasko, D.R., Giegling, I., Gearing, M., Hüll, M., Geschwind, D.H., Gilbert, J.R., Morris, J., Green, R.C., Mayo, K., Growdon, J.H., Feulner, T., Hamilton, R.L., Harrell, L.E., Drichel, D., Honig, L.S., Cushion, T.D., Huentelman, M.J., Hollingworth, P., Hulette, C.M., Hyman, B.T., Marshall, R., Jarvik, G.P., Meggy, A., Abner, E., Menzies, G.E., Jin, L.-W., Leonenko, G., Real, L.M., Jun, G.R., Baldwin, C.T., Grozeva, D., Karydas, A., Russo, G., Kaye, J.A., Kim, R., Jessen, F., Kowall, N.W., Vellas, B., Kramer, J.H., Vardy, E., LaFerla, F.M., Jöckel, K.-H., Lah, J.J., Dichgans, M., Leverenz, J.B., Mann, D., Levey, A.I., Pickering-Brown, S., Lieberman, A.P., Klopp, N., Lunetta, K.L., Wichmann, H.-E., Lyketsos, C.G., Morgan, K., Marson, D.C., Brown, K., Martiniuk, F., Medway, C., Mash, D.C., Nöthen, M.M., Masliah, E., Hooper, N.M., McCormick, W.C., Daniele, A., McCurry, S.M., Bayer, A., McDavid, A.N., Gallacher, J., McKee, A.C., Bussche, H. van den, Mesulam, M., Brayne, C., Miller, B.L., Riedel-Heller, S., Miller, C.A., Miller, J.W., Al-Chalabi, A., Morris, J.C., Shaw, C.E., Myers, A.J., Wiltfang, J., O'Bryant, S., Olichney, J.M., Alvarez, V., Parisi, J.E., Singleton, A.B., Paulson, H.L., Collinge, J., Perry, W.R., Mead, S., Peskind, E., Cribbs, D.H., Rossor, M., Pierce, A., Ryan, N.S., Poon, W.W., Nacmias, B., Potter, H., Sorbi, S., Quinn, J.F., Sacchinelli, E., Raj, A., Spalletta, G., Raskind, M., Caltagirone, C., Bossù, P., Orfei, M.D., Reisberg, B., Clarke, R., Reitz, C., Smith, A.D., Ringman, J.M., Warden, D., Roberson, E.D., Wilcock, G., Rogaeva, E., Bruni, A.C., Rosen, H.J., Gallo, M., Rosenberg, R.N., Ben-Shlomo, Y., Sager, M.A., Mecocci, P., Saykin, A.J., Pastor, P., Cuccaro, M.L., Vance, J.M., Schneider, J.A., Schneider, L.S., Slifer, S., Seeley, W.W., Smith, A.G., Sonnen, J.A., Spina, S., Stern, R.A., Swerdlow, R.H., Tang, M., Tanzi, R.E., Trojanowski, J.Q., Troncoso, J.C., Deerlin, V.M.V., Eldik, L.J.V., Vinters, H.V., Vonsattel, J.P., Weintraub, S., Welsh-Bohmer, K.A., Wilhelmsen, K.C., Williamson, J., Wingo, T.S., Woltjer, R.L., Wright, C.B., Yu, C.-E., Yu, L., Saba, Y., Pilotto, A., Bullido, M.J., Peters, O., Crane, P.K., Bennett, D., Bosco, P., Coto, E., Boccardi, V., Jager, P.L.D., Lleo, A., Warner, N., Lopez, O.L., Ingelsson, M., Deloukas, P., Cruchaga, C., Graff, C., Gwilliam, R., Fornage, M., Goate, A.M., Sanchez-Juan, P., Kehoe, P.G., Amin, N., Ertekin-Taner, N., Berr, C., Debette, S., Love, S., Launer, L.J., Younkin, S.G., Dartigues, J.-F., Corcoran, C., Ikram, M.A., Dickson, D.W., Nicolas, G., Campion, D., Tschanz, J., Schmidt, H., Hakonarson, H., Clarimon, J., Munger, R., Schmidt, R., Farrer, L.A.,

572 Broeckhoven, C.V., O'Donovan, M.C., DeStefano, A.L., Jones, L., Haines, J.L., Deleuze, J.-F., Owen, M.J.,
573 Gudnason, V., Mayeux, R., Escott-Price, V., Psaty, B.M., Ramirez, A., Wang, L.-S., Ruiz, A., Duijn, C.M. van,
574 Holmans, P.A., Seshadri, S., Williams, J., Amouyel, P., Schellenberg, G.D., Lambert, J.-C., Pericak-Vance, M.A.,
575 2019. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau,
576 immunity and lipid processing. Nat. Genet. 51, 414–430. https://doi.org/10.1038/s41588-019-0358-2
577 Lake, B.B., Ai, R., Kaeser, G.E., Salathia, N.S., Yung, Y.C., Liu, R., Wildberg, A., Gao, D., Fung, H.-L., Chen, S.,
578 Vijayaraghavan, R., Wong, J., Chen, A., Sheng, X., Kaper, F., Shen, R., Ronaghi, M., Fan, J.-B., Wang, W., Chun,
579 J., Zhang, K., 2016. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human
580 brain. Science 352, 1586–1590. https://doi.org/10.1126/science.aaf1204
581 Lake, B.B., Chen, S., Sos, B.C., Fan, J., Kaeser, G.E., Yung, Y.C., Duong, T.E., Gao, D., Chun, J., Kharchenko, P.V.,
582 Zhang, K., 2018. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain.
583 Nat. Biotechnol. 36, 70–80. https://doi.org/10.1038/nbt.4038
584 Lake, B.B., Codeluppi, S., Yung, Y.C., Gao, D., Chun, J., Kharchenko, P.V., Linnarsson, S., Zhang, K., 2017. A
585 comparative strategy for single-nucleus and single-cell transcriptomes confirms accuracy in predicted cell-type
586 expression from nuclear RNA. Sci. Rep. 7, 6031. https://doi.org/10.1038/s41598-017-04426-w
587 Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham,
588 G.W., Grenier-Boley, B., Russo, G., Thorton-Wells, T.A., Jones, N., Smith, A.V., Chouraki, V., Thomas, C.,
589 Ikram, M.A., Zelenika, D., Vardarajan, B.N., Kamatani, Y., Lin, C.F., Gerrish, A., Schmidt, H., Kunkle, B.,
590 Dunstan, M.L., Ruiz, A., Bihoreau, M.T., Choi, S.H., Reitz, C., Pasquier, F., Cruchaga, C., Craig, D., Amin, N.,
591 Berr, C., Lopez, O.L., De Jager, P.L., Deramecourt, V., Johnston, J.A., Evans, D., Lovestone, S., Letenneur, L.,
592 Moron, F.J., Rubinsztein, D.C., Eiriksdottir, G., Sleegers, K., Goate, A.M., Fievet, N., Huentelman, M.W., Gill,
593 M., Brown, K., Kamboh, M.I., Keller, L., Barberger-Gateau, P., McGuiness, B., Larson, E.B., Green, R., Myers,
594 A.J., Dufouil, C., Todd, S., Wallon, D., Love, S., Rogaeva, E., Gallacher, J., St George-Hyslop, P., Clarimon, J.,
595 Lleo, A., Bayer, A., Tsuang, D.W., Yu, L., Tsolaki, M., Bossu, P., Spalletta, G., Proitsi, P., Collinge, J., Sorbi, S.,
596 Sanchez-Garcia, F., Fox, N.C., Hardy, J., Deniz Naranjo, M.C., Bosco, P., Clarke, R., Brayne, C., Galimberti, D.,
597 Mancuso, M., Matthews, F., Moebus, S., Mecocci, P., Del Zompo, M., Maier, W., Hampel, H., Pilotto, A.,
598 Bullido, M., Panza, F., Caffarra, P., Nacmias, B., Gilbert, J.R., Mayhaus, M., Lannefelt, L., Hakonarson, H.,
599 Pichler, S., Carrasquillo, M.M., Ingelsson, M., Beekly, D., Alvarez, V., Zou, F., Valladares, O., Younkin, S.G.,
600 Coto, E., Hamilton-Nelson, K.L., Gu, W., Razquin, C., Pastor, P., Mateo, I., Owen, M.J., Faber, K.M., Jonsson,
601 P.V., Combarros, O., O'Donovan, M.C., Cantwell, L.B., Soininen, H., Blacker, D., Mead, S., Mosley, T.H.J.,
602 Bennett, D.A., Harris, T.B., Fratiglioni, L., Holmes, C., de Bruijn, R.F., Passmore, P., Montine, T.J., Bettens, K.,
603 Rotter, J.I., Brice, A., Morgan, K., Foroud, T.M., Kukull, W.A., Hannequin, D., Powell, J.F., Nalls, M.A., Ritchie,
604 K., Lunetta, K.L., Kauwe, J.S., Boerwinkle, E., Riemenschneider, M., Boada, M., Hiltuenen, M., Martin, E.R.,
605 Schmidt, R., Rujescu, D., Wang, L.S., Dartigues, J.F., Mayeux, R., Tzourio, C., Hofman, A., Nothen, M.M., Graff,
606 C., Psaty, B.M., Jones, L., Haines, J.L., Holmans, P.A., Lathrop, M., Pericak-Vance, M.A., Launer, L.J., Farrer,
607 L.A., van Duijn, C.M., Van Broeckhoven, C., Moskvina, V., Seshadri, S., Williams, J., Schellenberg, G.D.,
608 Amouyel, P., 2013. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's
609 disease. Nat. Genet. 45, 1452–1458. https://doi.org/10.1038/ng.2802
610 Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., Mesirov, J.P., 2011. Molecular signatures
611 database (MSigDB) 3.0. Bioinformatics 27, 1739–1740. https://doi.org/10.1093/bioinformatics/btr260
612 Lun, A.T.L., Riesenfeld, S., Andrews, T., Dao, T.P., Gomes, T., Marioni, J.C., participants in the 1st Human Cell Atlas
613 Jamboree, 2019. EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA
614 sequencing data. Genome Biol. 20, 63. https://doi.org/10.1186/s13059-019-1662-y
615 Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N.,
616 Martersteck, E.M., Trombetta, J.J., Weitz, D.A., Sanes, J.R., Shalek, A.K., Regev, A., McCarroll, S.A., 2015.

Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 161, 1202–1214. https://doi.org/10.1016/j.cell.2015.05.002

Mancuso, R., Daele, J.V.D., Fattorelli, N., Wolfs, L., Balusu, S., Burton, O., Liston, A., Sierksma, A., Fourne, Y., Poovathingal, S., Arranz-Mendiguren, A., Frigerio, C.S., Claes, C., Serneels, L., Theys, T., Perry, V.H., Verfaillie, C., Fiers, M., Strooper, B.D., 2019. Stem-cell-derived human microglia transplanted in mouse brain to study human disease. Nat. Neurosci. 1–6. https://doi.org/10.1038/s41593-019-0525-x

Maniatis, S., Äijö, T., Vickovic, S., Braine, C., Kang, K., Mollbrink, A., Fagegaltier, D., Saiz-Castro, G., Cuevas, M., Watters, A., Lundeberg, J., Bonneau, R., Phatnani, H., 2019. Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis 6.

Marioni, R.E., Harris, S.E., Zhang, Q., McRae, A.F., Hagenaars, S.P., Hill, W.D., Davies, G., Ritchie, C.W., Gale, C.R., Starr, J.M., Goate, A.M., Porteous, D.J., Yang, J., Evans, K.L., Deary, I.J., Wray, N.R., Visscher, P.M., 2018. GWAS on family history of Alzheimer's disease. Transl. Psychiatry 8, 99. https://doi.org/10.1038/s41398-018-0150-6

Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J.Z., Menon, M., He, L., Abdurrob, F., Jiang, X., Martorell, A.J., Ransohoff, R.M., Hafler, B.P., Bennett, D.A., Kellis, M., Tsai, L.-H., 2019. Single-cell transcriptomic analysis of Alzheimer's disease. Nature 1. https://doi.org/10.1038/s41586-019-1195-2

McGeer, P.L., Itagaki, S., Tago, H., McGeer, E.G., 1987. Reactive microglia in patients with senile dementia of the Alzheimer type are positive for the histocompatibility glycoprotein HLA-DR. Neurosci. Lett. 79, 195–200. https://doi.org/10.1016/0304-3940(87)90696-3

McGinnis, C.S., Murrow, L.M., Gartner, Z.J., 2019. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. 8, 329-337.e4. https://doi.org/10.1016/j.cels.2019.03.003

Sala Frigerio, C., Wolfs, L., Fattorelli, N., Thrupp, N., Voytyuk, I., Schmidt, I., Mancuso, R., Chen, W.-T., Woodbury, M.E., Srivastava, G., Möller, T., Hudry, E., Das, S., Saido, T., Karran, E., Hyman, B., Perry, V.H., Fiers, M., Strooper, B.D., 2019. The Major Risk Factors for Alzheimer's Disease: Age, Sex, and Genes Modulate the Microglia Response to Aβ Plaques. Cell Rep. 27, 1293-1306.e6. https://doi.org/10.1016/j.celrep.2019.03.099

Saunders, A., Macosko, E.Z., Wysoker, A., Goldman, M., Krienen, F.M., de Rivera, H., Bien, E., Baum, M., Bortolin, L., Wang, S., Goeva, A., Nemesh, J., Kamitaki, N., Brumbaugh, S., Kulp, D., McCarroll, S.A., 2018. Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. Cell 174, 1015-1030.e16. https://doi.org/10.1016/j.cell.2018.07.028

Sergushichev, A.A., 2016. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. bioRxiv 060012. https://doi.org/10.1101/060012

Skene, N.G., Bryois, J., Bakken, T.E., Breen, G., Crowley, J.J., Gaspar, H.A., Giusti-Rodriguez, P., Hodge, R.D., Miller, J.A., Muñoz-Manchado, A.B., O'Donovan, M.C., Owen, M.J., Pardiñas, A.F., Ryge, J., Walters, J.T.R., Linnarsson, S., Lein, E.S., Sullivan, P.F., Hjerling-Leffler, J., 2018. Genetic identification of brain cell types underlying schizophrenia. Nat. Genet. 50, 825–833. https://doi.org/10.1038/s41588-018-0129-5

Skene, N.G., Grant, S.G.N., 2016. Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. Front. Neurosci. 10. https://doi.org/10.3389/fnins.2016.00016

Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., Mollbrink, A., Linnarsson, S., Codeluppi, S., Borg, Å., Pontén, F., Costea, P.I., Sahlén, P., Mulder, J., Bergmann, O., Lundeberg, J., Frisén, J., 2016. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science 353, 78–82. https://doi.org/10.1126/science.aaf2403

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., Satija, R., 2019. Comprehensive Integration of Single-Cell Data. Cell 177, 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031

662 Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L.,
663 Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: A knowledge-based approach for
664 interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. 102, 15545–15550.
665 https://doi.org/10.1073/pnas.0506580102
666 Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T.,
667 Bertagnolli, D., Goldy, J., Shapovalova, N., Parry, S., Lee, C., Smith, K., Bernard, A., Madisen, L., Sunkin, S.M.,
668 Hawrylycz, M., Koch, C., Zeng, H., 2016. Adult mouse cortical cell taxonomy revealed by single cell
669 transcriptomics. Nat. Neurosci. 19, 335–346. https://doi.org/10.1038/nn.4216
670 Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L.E.,
671 La Manno, G., Codeluppi, S., Furlan, A., Lee, K., Skene, N., Harris, K.D., Hjerling-Leffler, J., Arenas, E., Ernfors,
672 P., Marklund, U., Linnarsson, S., 2018. Molecular Architecture of the Mouse Nervous System. Cell 174, 999-
673 1014.e22. https://doi.org/10.1016/j.cell.2018.06.021
674 Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G.,
675 Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., To, J.K.,
676 Laird, M.R., Lavidas, I., Liu, Z., Loveland, J.E., Maurel, T., McLaren, W., Moore, B., Mudge, J., Murphy, D.N.,
677 Newman, V., Nuhn, M., Ogeh, D., Ong, C.K., Parker, A., Patricio, M., Riat, H.S., Schuilenburg, H., Sheppard, D.,
678 Sparrow, H., Taylor, K., Thormann, A., Vullo, A., Walts, B., Zadissa, A., Frankish, A., Hunt, S.E., Kostadima, M.,
679 Langridge, N., Martin, F.J., Muffato, M., Perry, E., Ruffier, M., Staines, D.M., Trevanion, S.J., Aken, B.L.,
680 Cunningham, F., Yates, A., Flicek, P., 2018. Ensembl 2018. Nucleic Acids Res. 46, D754–D761.
681 https://doi.org/10.1093/nar/gkx1098
682 Zhou, Y., Song, W.M., Andhey, P.S., Swain, A., Levy, T., Miller, K.R., Poliani, P.L., Cominelli, M., Grover, S., Gilfillan,
683 S., Cella, M., Ulland, T.K., Zaitsev, K., Miyashita, A., Ikeuchi, T., Sainouchi, M., Kakita, A., Bennett, D.A.,
684 Schneider, J.A., Nichols, M.R., Beausoleil, S.A., Ulrich, J.D., Holtzman, D.M., Artyomov, M.N., Colonna, M.,
685 2020. Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent
686 cellular responses in Alzheimer's disease. Nat. Med. 26, 131–142. https://doi.org/10.1038/s41591-019-0695-9

687

688

689

690 **Figures**

691

692 **Fig. 1: Gene abundance in single microglial cells versus single microglial nuclei of**
693 **human cortical tissue. a.** Mean normalised gene abundance in cells (x axis) and nuclei. (y
694 axis). Red: genes with significantly higher abundance in nuclei ($p_{adj} < 0.05$, fold change > 2).
695 Blue: genes that are significantly less abundance in nuclei ($p_{adj} < 0.05$, fold change < -2).
696 Genes were normalized to read depth (per cell), scaled by 10,000 and log-transformed.
697 *MALAT1* (which had normalized abundance levels of 6.0 and 6.9 respectively in cells and
698 nuclei) has been removed for visualisation purposes. FC = fold change. Full results are
699 available in Supplementary Table 2. **b.** Scatterplot as in a), per patient (with the same genes
700 highlighted). Supplementary Table 1 contains patient data. **c.** Each bar represents a
701 comparison between two datasets (*X* versus *Y*), with the bootstrapped *z*-scores representing
702 the extent to which cell-enriched genes (top panel) and nuclear-enriched genes (bottom panel)
703 have lower specificity for microglia in dataset *Y* relative to that in dataset *X*. Larger *z*-scores
704 indicate greater depletion of genes, and red bars indicate a statistically significant depletion
705 ($p._{adj} < 0.05$, by bootstrapping). KI = Karolinska Institutet ; AIBS = Allen Institute for Brain
706 Science.

707

Figure 1

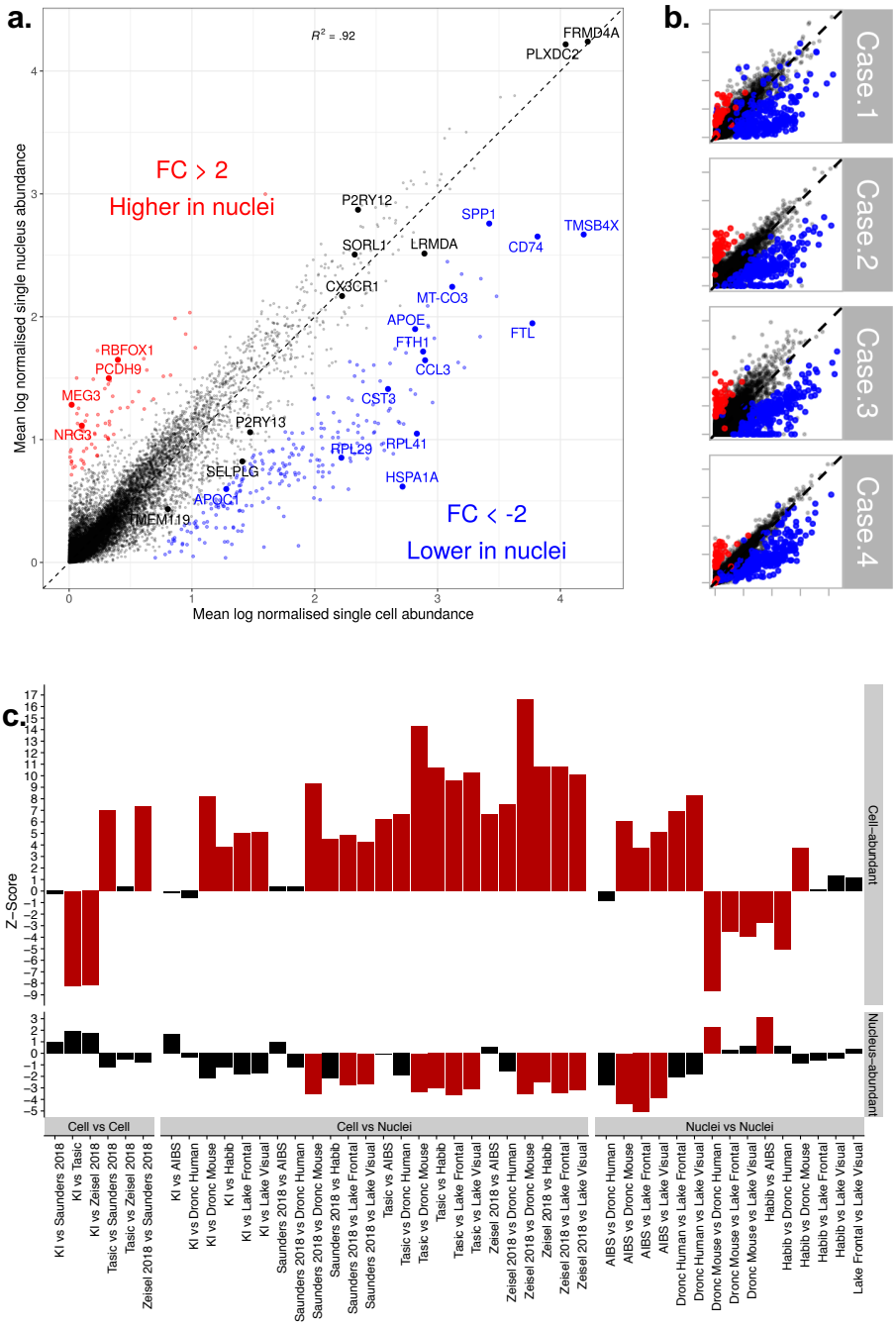
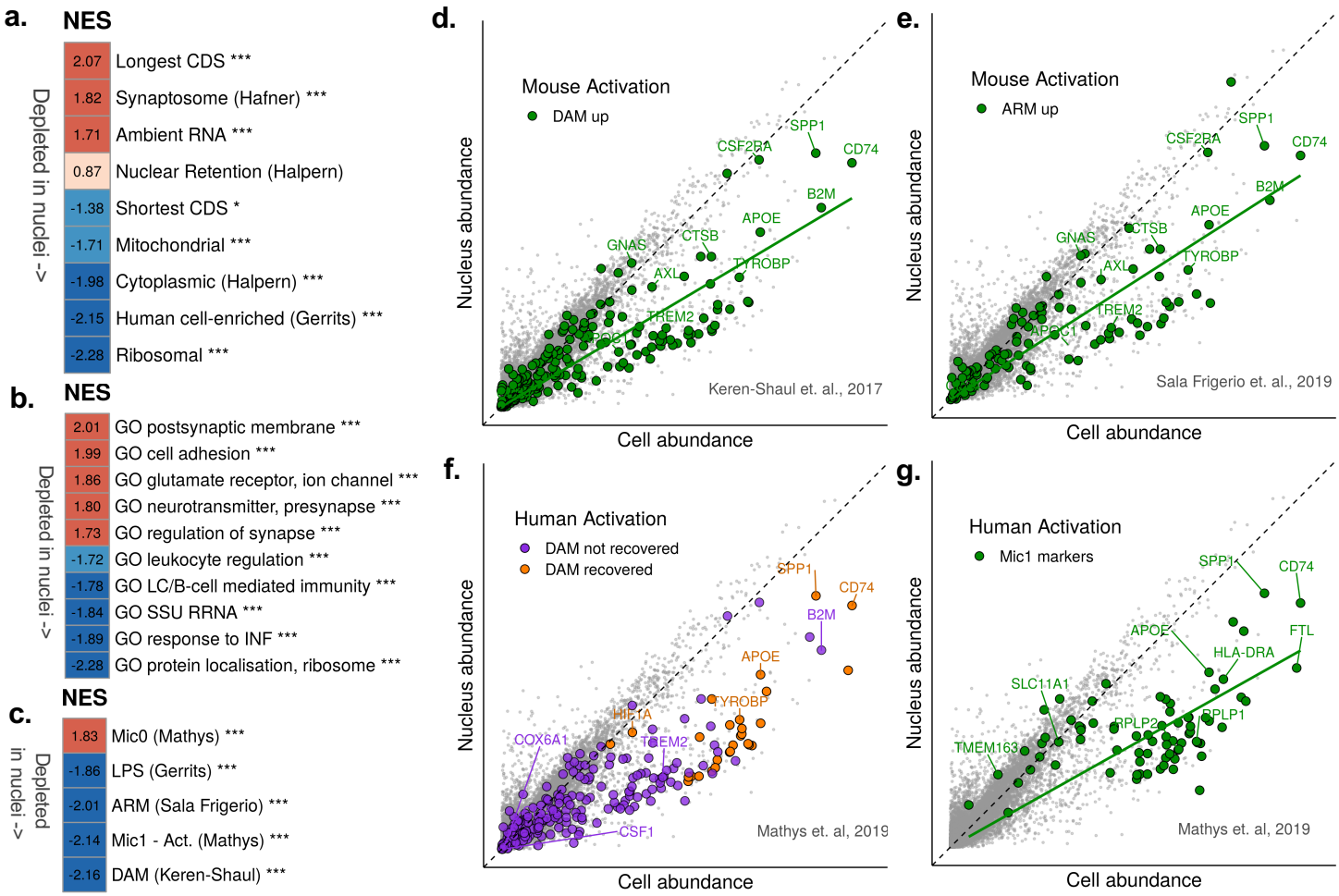


Fig. 2: Functional analysis of genes that are enriched or depleted in nuclei. a. Gene Set Enrichment Analysis (GSEA) of gene sets related to cellular location and gene coding sequence length (CDS). Background genes were ranked according to log fold change of nuclei versus cells. Red: higher Normalised Enrichment Score (NES), i.e. more genes associated with nuclear enrichment ; blue: negative NES scores (depletion in nuclei). *** represents significance ($p_{adj} < 0.0005$). CDS = coding sequence. **b.** GSEA of *super*-Gene Ontology gene sets against ranked nucleus-cell log fold changes. Only top and bottom categories (according to NES) are shown. Colours as in a). GO = Gene Ontology ; SSU RRNA = small subunit ribosomal RNA ; INF = Interferon ; LC = leukocyte. **c.** GSEA of selected gene sets from previous studies of microglial activation, against fold change as in a). *** represents significance ($p_{adj} < 0.0005$). Mic0 = markers of microglial cluster 0 in human brain tissue ; Mic1 = markers of microglial cluster 1 (response to plaques) defined by (Mathys et al., 2019) in human brain tissue. ARM = Activation Response Microglia (Sala Frigerio *et al.*[8]). DAM = Disease-Associated Microglia (Keren-Shaul *et al.*[7]) **d.** Scatterplot as in Fig. 1a), highlighting in green the DAM genes. A regression line for the highlighted genes is shown in green (slope = 0.60). **e.** Scatterplot as in d), highlighting in green the ARM genes. A regression line for the highlighted genes is shown in green (slope = 0.64). **f.** Scatterplot as in d), highlighting the DAM genes recovered in the study of human activation in AD (Mathys et al., 2019). Purple: DAM genes not recovered in their study ; orange: DAM genes recovered in their study. **g.** Scatterplot as in d), Green: human activation marker genes defined by (Mathys et al., 2019). A regression line for the highlighted genes is shown in green (slope = 0.56). Gene sets are available in Supplementary Table 6.

Figure 2



Supplementary Data

Figures

Fig. S1: Clustering of single nuclei from human cortical tissue.

Fig. S2: Gene abundance in single microglial cells versus single microglial nuclei of human cortical tissue.

Text

Supplementary Text: Clustering of microglial cells and nuclei in human cortical tissue.

Tables

Supplementary Table 1: Patient Metadata

Supplementary Table 2: Differential abundance nuclei versus cells

Supplementary Table 3: Public single cell / single nucleus datasets

Supplementary Table 4: Clustering of Gene Ontology (GO) terms

Supplementary Table 5: Results of GSEA analysis

Supplementary Table 6: Lists of all gene sets

Supplementary Table 7: Cluster markers for single nucleus cluster analysis

Fig. S1: Clustering of single nuclei from human tissue. a. UMAP plot of 37,060 nuclei from cortical tissue of 4 neurosurgical patients, coloured according to cell type. Mg = Microglia ; OPC = oligodendrocyte precursor cells ; ODC = oligodendrocytes ; Astro = Astrocytes ; Endoth = Endothelial cells ; Exc. neurons = excitatory neurons. **b.** Violin plots show selected markers of the different cell types (data is normalised for count depth and log-transformed). **c.** UMAP plot of 3,721 microglial nuclei from cortical tissue of 4 neurosurgical patients, coloured according to cluster number, after *in silico* extraction of microglia (based on markers such as *P2RY12*) and reclustering. **d.** Module scores for gene sets extracted from the original Mancuso *et al.* single cell microglia study (Mancuso et al., 2019). The top 40 genes according to log fold change were selected for each gene set. **e.** Overlap of top 40 marker genes from cellular clusters on the horizontal axis (Mancuso *et al.*) and nuclear clusters on the vertical axis. The blue scale represents the number of genes in common, numbers represent p_{adj} values. Vertical coloured bars correspond to the clusters shown in c). N.S. = not significant ($p_{adj} > 0.05$). MS40 = Module Score of top 40 gene markers ; CAM = macrophages ; CRM = cytokine response ; in vitro 1 = activation-like module (similar to *in vitro* macrophages) ; in vitro 2 = activation-like module (similar to *in vitro* monocytes) ; in vivo HM = homeostatic. Nuc = Nuclear clusters. Cluster markers are provided in Supplementary Table 7.

Figure S1

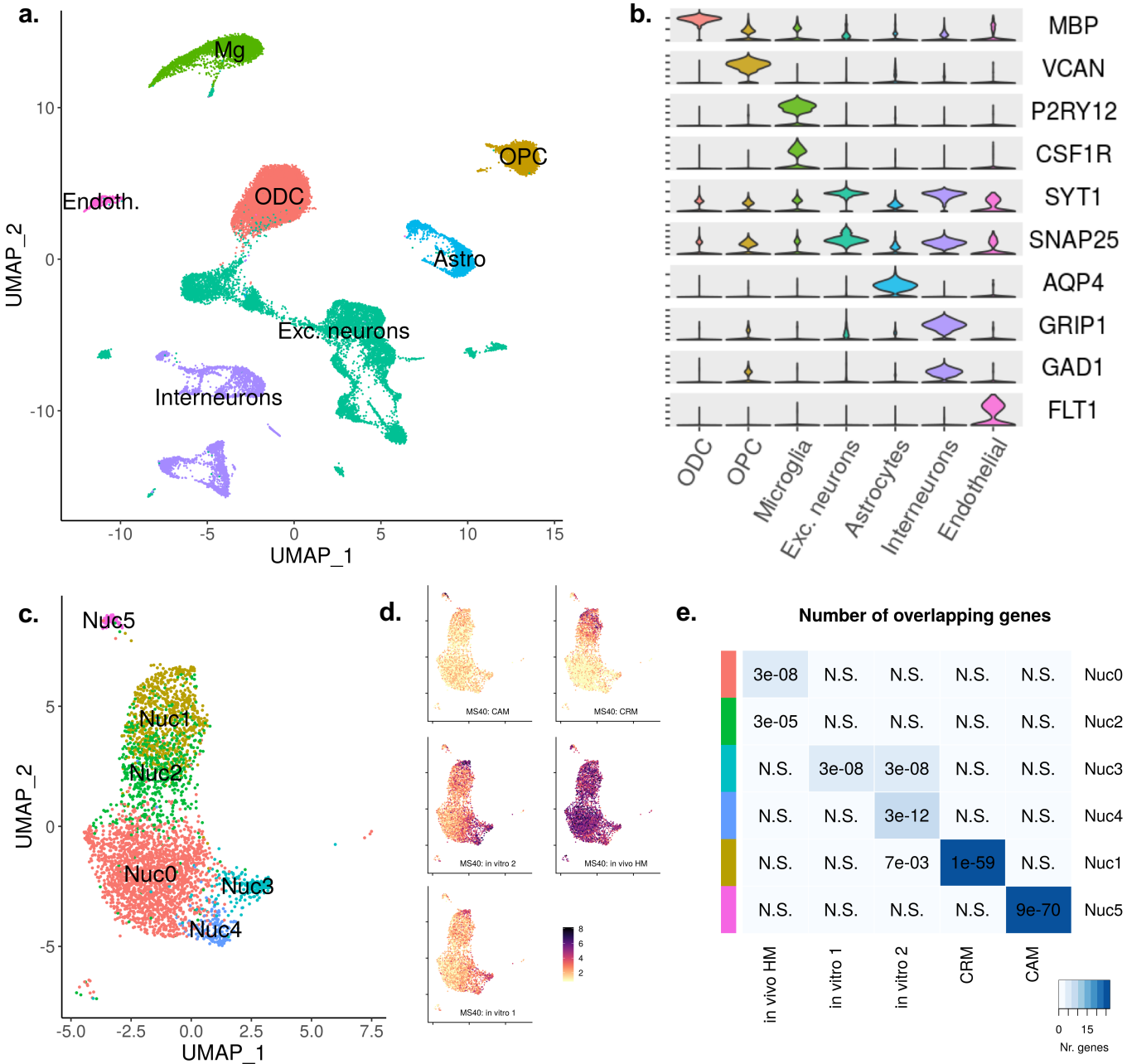
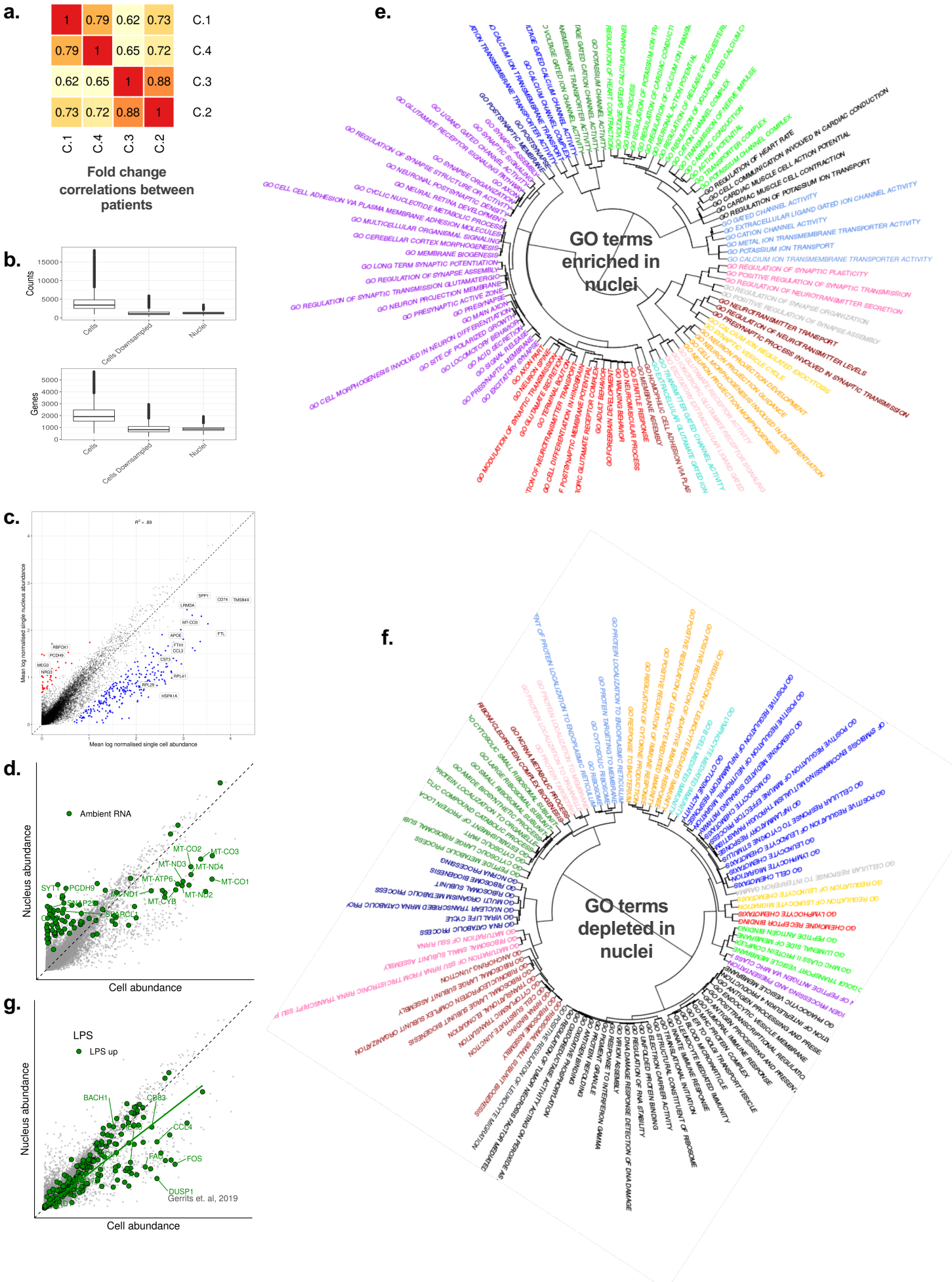


Fig. S2: Gene abundance in single microglial cells versus single microglial nuclei of human cortical tissue. a. Correlation matrix of gene abundance fold changes (single cell vs single nucleus abundance) between patients. **b.** Downsampling of reads: boxplots for numbers of reads (top) and numbers of genes (bottom) for single cells before downsampling, single cells after downsampling, and single nuclei. Boxplots show median, with 25% and 75% quantiles. **c.** Scatterplot of mean gene abundance in cells against mean gene abundance in nuclei (as in Fig. 1a) after downsampling of reads in cells. Data is normalised to count depth and log-transformed. Points in red represent genes with significantly higher abundance in nuclei, while those in blue are significantly less abundant in nuclei ($p_{adj} < 0.05$, fold change > |2|). **d.** Scatter plot, as in Fig. 1a) showing the ambient mRNA in green (the same dataset was used in Fig. 2d). Ambient RNA is defined as the 150 most abundant genes in the 700 nuclei with the lowest total read counts. **e, f.** Dendrograms of **e.** top 100 Gene Ontology (GO) terms enriched in nuclei, and **f.** top 100 GO terms depleted in nuclei. GO terms were clustered based on overlap between their gene sets. The colours show how GO terms were clustered. These clusters are described in Supplementary Table 4. **g.** Scatterplot as in Fig. 1a), highlighting in green genes that are upregulated during LPS stimulation in mice (Gerrits et al., 2019). A regression line for the highlighted genes is shown in green (slope = 0.78).

Figure S2



Supplementary Text: Clustering of microglial cells and nuclei in human cortical tissue

We sequenced nuclei from cortical tissue of 4 neurosurgical patients. Single cell sequencing of FAC-sorted microglia was performed on cortical tissue of the same patients in a previous study(Mancuso et al., 2019). Subject data is available in Supplementary Table 1. Following quality filtering, data integration, PCA analysis and clustering of nuclei, we identified 7 main cell types in 37,060 nuclei: oligodendrocytes (ODC, 34.0%), excitatory neurons (27.0%), interneurons (11.3%), oligodendrocyte precursors (OPC 9,4%), microglia (11.3%), astrocytes (6.0%), and endothelial cells (1.1%). Supplementary Fig. 1a and Supplementary Fig. 1b show UMAP embeddings for all nuclei, coloured by cell type, and selected markers for each cell type, respectively.

Microglial nuclei were isolated and reclustered. We identified 3,721 microglia (expressing *MEF2A*, *P2RY12*, *CX3CR1*, *CSF1R*), a macrophage cluster (enriched for *CD163* and *MRC1*, 67 nuclei), a neutrophil cluster (72 nuclei), and a cluster containing microglial as well as astrocytic markers (marked by *GFAP*, 68 nuclei). The neutrophil and ambiguous clusters were discarded, leaving only microglia and brain macrophages for downstream analysis (Supplementary Fig. 1c). Cluster markers are provided in Supplementary Table 6.

In order to determine if nuclei could recover microglial clusters identified in cells, we selected the top 40 markers defined by Mancuso *et al.* (Mancuso et al., 2019) for each of the clusters they identified in the original analysis of microglial cells. For each nucleus, we scored each set of markers based on the abundance of those markers in the nucleus, using Seurat's *AddModuleScore* function. These scores, referred to as MS40 scores, are highlighted in Supplementary Fig. 1d. Our nuclei were able to recover a cytokine response cluster (CRM), marked by *CCL3*, *CCL4*, and an activation-like cluster, equivalent to the "*in vitro* microglia" identified in the original study (original markers included *APOC1*, *GPNMB*, *SPP1*, *APOE*). Homeostatic markers appeared ubiquitously through-out the nuclei dataset, and we were not able to distinguish a reduction of these markers in the activation-like response cluster, as we would expect from transcriptomic profiling of microglia in mice (Keren-Shaul et al., 2017; Sala Frigerio et al., 2019). Finally, the CAM (macrophage) cluster

(*CD163*, *MRC1*), separated out from the bulk of the microglia, and was easily-recognisable based on its MS40 score. Cluster markers are provided in Supplementary Table 6.

In order to quantify the differences between cells and nuclei in more detail, we examined the overlap of the top 40 markers between nuclei clusters and cell clusters (Supplementary Fig. 1e). The cell macrophage (CAM) and cell cytokine (CRM) clusters showed the largest overlaps with Nuc1 and Nuc7 (27 and 24 of 40 markers, respectively). Other clusters only showed overlaps of between 1 and 5 genes. Cluster Nuc3 showed similar overlaps between "*in vitro* 1" and "*in vitro* 2" (5 genes). Cluster Nuc0 showed an overlap of 5 genes with "*in vivo* HM", and cluster Nuc2 showed an overlap of 2 genes with "*in vivo* HM". Cluster Nuc4 showed similarities with the "*in vitro* 2" cluster, suggesting it could be a cluster of activation, however all 5 overlapping genes were mitochondrial genes. Cluster Nuc3 markers *RPS12*, *TPT1*, *FTL*, *RPS18* and *EEF1A1* also appeared as markers of "*in vitro* 2".

We performed similar analyses using more markers, however we found that introducing more markers resulted in nuclei markers overlapping with more than one cell cluster. We also noticed that introducing more markers resulted in overlaps between markers of the cellular clusters with each other. Selecting 40 markers allowed us to align cellular and nuclear clusters in an almost one-to-one fashion (see Supplementary Fig. 1e).

Overall, cytokine clusters and macrophage clusters were recovered well using single nucleus methods, however, differences between other microglial subpopulations were not convincingly recovered.