nature immunology

Resource

https://doi.org/10.1038/s41590-023-01558-2

# Exposure of iPSC-derived human microglia to brain substrates enables the generation and manipulation of diverse transcriptional states in vitro

Received: 28 April 2022

Accepted: 9 June 2023

Published online: 27 July 2023

Check for updates

Michael-John Dolan[1,2,3,20], Martine Therrien[1,2,3,20], Saša Jereb[1,2,3], Tushar Kamath[1,3], Vahid Gazestani[1], Trevor Atkeson[1], Samuel E. Marsh[2,3], Aleksandrina Goeva[1], Neal M. Lojek[1], Sarah Murphy[2], Cassandra M. White[4], Julia Joung[4,5,6,7,8], Bingxu Liu[4], Francesco Limone[1,9,10], Kevin Eggan[1,9], Nir Hacohen[4,11,12], Bradley E. Bernstein[4,13,14,15], Christopher K. Glass[16], Ville Leinonen[17], Mathew Blurton-Jones[18], Feng Zhang[4,5,6,7,8], Charles B. Epstein[4], Evan Z. Macosko[1,19] ✉ & Beth Stevens[1,2,3,4,6] ✉

Microglia, the macrophages of the brain parenchyma, are key players in neurodegenerative diseases such as Alzheimer's disease. These cells adopt distinct transcriptional subtypes known as states. Understanding state function, especially in human microglia, has been elusive owing to a lack of tools to model and manipulate these cells. Here, we developed a platform for modeling human microglia transcriptional states in vitro. We found that exposure of human stem-cell-differentiated microglia to synaptosomes, myelin debris, apoptotic neurons or synthetic amyloid-beta fibrils generated transcriptional diversity that mapped to gene signatures identified in human brain microglia, including disease-associated microglia, a state enriched in neurodegenerative diseases. Using a new lentiviral approach, we demonstrated that the transcription factor MITF drives a disease-associated transcriptional signature and a highly phagocytic state. Together, these tools enable the manipulation and functional interrogation of human microglial states in both homeostatic and disease-relevant contexts.

Microglia, the macrophages of the brain, have central roles in development, homeostasis and diseases of the central nervous system (CNS)[1]. Genetic studies have implicated microglia in late-onset Alzheimer's disease (AD), with a large fraction of AD risk genes expressed in myeloid cells[2,3]. Moreover, microglia activation and dysfunction are hallmarks of AD and other neurodegenerative disorders[4]. Single-cell RNA sequencing (scRNA-seq) has revealed that microglia exist in multiple states[5–14] that each express distinct gene signatures, including disease-associated microglia (DAMs). However, the function of these states and their impact on disease remain unknown.

As the differential expression of many genes defines microglial states, high-throughput models are needed to understand how transcriptional signatures map onto microglial function. However, there are several outstanding challenges. Many differences exist between mouse and human microglia[15], a particularly acute problem with AD risk genes[16]. Microglia also rapidly alter their gene expression upon primary

A full list of affiliations appears at the end of the paper. ✉e-mail: emacosko@broadinstitute.org; Beth.Stevens@childrens.harvard.edu

culture, presumably owing to a lack of brain-derived cues[15,17], and cell lines fail to accurately model this cell type[18]. Finally, high-throughput gene perturbation studies are not possible as microglia are resistant to conventional forms of DNA delivery[19].

Here, we addressed these issues using human stem-cell-differentiated microglia (iMGLs) by creating an in vitro platform for studying human microglial states. Exposure of iMGLs to brain-related substrates, scRNA-seq and cross-dataset integration showed that transcriptional states identified in the human brain could be recapitulated in a monolayer culture. In particular, we identified DAM-like states in vitro and demonstrated their formation was dependent on signaling through cell surface receptor TREM2, as observed in vivo[6,20]. Using a new lentiviral transduction protocol that opens up iMGLs to efficient viral manipulation, we determined that the transcription factor MITF regulates disease-associated genes and a highly phagocytic state. Finally, we demonstrated that our observations were robust by differentiating iMGLs from multiple human induced pluripotent stem cell (iPSC) lines.

## Results

### CNS substrates induce transcriptional states in iMGLs

Microglia cultured in vitro lack key transcriptional features observed in vivo[15,17]. To test whether exposure to complex CNS substrates drives the formation of microglial states in a dish, we differentiated human embryonic stem cells (H1) into iMGLs[21] and, at day 40, performed scRNA-seq on untreated iMGLs or cells incubated with synaptosomes, myelin debris, apoptotic neurons (ANs) or synthetic amyloid-beta (Aβ) fibrils for 24 h. We acquired a total of 56,454 single iMGL transcriptomes across all conditions (referred subsequently to as the iMGL dataset) (Fig. 1a, Extended Data Fig. 1a–d and Supplementary Table 1). Cells in both replicates showed minimal expression of pluripotency markers and an enrichment of microglial identity and maturity signatures (Extended Data Fig. 1e–g), confirming successful differentiation. We excluded 2.4% of the iMGL population (Extended Data Fig. 1h), as these cells showed an artifactual gene signature known to be induced by handling in single-cell preparations[22].

Applying unbiased clustering to all conditions combined, we identified 11 clusters (clusters iMGL_1 to iMGL_11) of iMGLs, which were detected across replicates (Fig. 1a and Extended Data Fig. 1d), indicating a degree of transcriptional diversity similar to that reported in human postmortem datasets[11,23–25]. We observed that exposure to synaptosomes, myelin debris, ANs or Aβ fibrils induced a shift in iMGL transcriptional signature (Fig. 1b and Extended Data Fig. 2a). Although all substrates induced phagocytosis in iMGLs (Extended Data Fig. 2b), not all treated iMGLs adopted the same states in response to the treatment (Fig. 1c,d). This indicates that iMGLs are not a homogeneous population but form diverse transcriptional states in response to different CNS substrates, as observed in vivo[1].

To understand the biological meaning of the 11 clusters observed, we performed a differential gene expression analysis (Fig. 1c and Supplementary Table 2). By leveraging previously published data[26], we classified each of the 11 clusters into established microglial states (Extended Data Fig. 3a–c)[27] and identified iMGL states similar to those identified in vivo. This included a neurodegenerative DAM state (clusters iMGL_2 and iMGL_8, based on the expression of *APOE*, *GPNMB*, *LPL* and *ABCA1*), an antigen-presenting state (clusters iMGL_3, iMGL_4 and iMGL_7, expressing *HLA-DRA* and *HLA-DRB1*), an interferon-responsive state (cluster iMGL_11, *IFIT3* and *MX1*), a proliferation state (clusters iMGL_6, iMGL_9 and iMGL_10, *MKI67* and *TOP2A*) and a homeostatic state (iMGL_5, *CX3CR1*) (Fig. 1c–e and Extended Data Fig. 3d,e). To validate the presence of these clusters, we treated iMGL with CNS substrates for 24 h and performed in situ hybridization to detect key markers of microglial states (*ABCA1* for iMGL_2, *APOE* for iMGL_8, *MKI67* for iMGL_6/9/10 and *IFIT1* for iMGL_11) (Fig. 1f), confirming that expression of these markers was present or increased after treatment. Three of these states (proliferation, antigen-presenting and DAM) consisted of multiple clusters, each expressing a large number of both unique and shared differentially expressed genes (Extended Data Fig. 3f–i). In addition, a pseudotime analysis[28] indicated that one cluster, iMGL_1, appeared to be in transition from a homeostatic to a disease-associated state (iMGL_1 to iMGL_2 and iMGL_8) (Extended Data Fig. 4a), reflecting ongoing microglial dynamics 24 h after stimulation. Thus, our data suggest that iMGLs adopt diverse transcriptional states in vitro after treatment with CNS substrates.

We next considered whether exposing iMGLs to CNS substrates resulted in context-dependent cell states. We observed many statistically significant changes in the proportions of iMGL clusters after exposure to CNS substrates (Fig. 1d, Extended Data Fig. 4b and Supplementary Table 3). Several clusters changed similarly across multiple conditions. For example, iMGL_5 was depleted in response to all treatments compared with no treatment, whereas iMLG_1 was enriched in all treated iMGLs compared with untreated iMGL (Fig. 1d and Extended Data Fig. 4c). Other clusters had substrate-specific responses; for instance, cluster iMGL_2 was broadly induced by all substrates except Aβ fibrils, whereas iMGL_8 was only induced by ANs and Aβ fibrils (Fig. 1d,e), although the change was statistically significant only in the former case. The application of in situ hybridization and immunocytochemistry for iMGL_2 and iMGL_8 cluster-enriched markers (*APOE*, *GPNMB* and *ABCA1*) (Extended Data Fig. 4d,e), indicated enrichment of these signatures in response to ANs and several other substrates. These results confirm that iMGLs form distinct cell states in response to specific CNS stimuli.

Next, we systematically compared the iMGL transcriptional states with those found in vivo, using a single-nucleus RNA sequencing (snRNA-seq) dataset of microglia from human cortical brain biopsies[29]. This dataset consisted of 54,475 single microglial nuclei from 51 individuals, classified into five microglial clusters (referred as brain biopsy (BB) clusters; Extended Data Fig. 5a–f). Integrative analysis of the brain biopsy microglia with iMGL profiles using LIGER[30] showed significant alignment between the two datasets (Fig. 2a). To quantify the alignment, we calculated for each cell the degree to which its neighbors were from one (a score of 0) or all datasets (a score of 1)[30]. This analysis resulted in an alignment score of 0.77 (Methods), suggesting extensive mixing between brain biopsy microglia and iMGL profiles. Analysis of the joint clusters generated by the LIGER integration, which we refer to as BB/iMGL, identified eight joint clusters (BB/iMGL_0 to BB/iMGL_7) (Fig. 2b), all of which were populated by both biopsy microglia and iMGLs (Fig. 2c and Extended Data Fig. 5g). Brain biopsy microglia and iMGLs contributed roughly equally to the most joint clusters (Fig. 2c), with two exceptions: BB/iMLG_5, a proliferative cluster dominated by iMGLs; and BB/iMLG_7, a small number of human-brain-specific microglia that highly expressed interferon genes but were distinct from BB/iMLG_6, the interferon-responsive joint cluster present in both datasets (Fig. 2c,d). Joint clusters mapped to established microglial states in both the iMGL and biopsy data (Fig. 2d and Extended Data Fig. 5h–k); these included BB/iMLG_4, which contained the majority of DAMs from both the iMGL (iMGL_2 and iMGL_8) and brain biopsy (BB_GPNMB_LPL) datasets (Fig. 2c,d). Joint cluster BB/iMGL_4 also contained cells from the BB_CRM_CCL3 cluster, although there were few microglia in this state (Extended Data Fig. 5l); larger datasets are needed to properly resolve such rare states[31]. Several iMGL clusters, including the homeostatic (iMGL_5) and antigen-presenting (iMGL_3, iMGL_4 and iMGL_7) states, coclustered with microglia identified as homeostatic in the brain biopsy dataset (BB_CX3CR1). We identified a gradient of antigen-presenting genes in human brain microglia from the BB_CX3CR1 cluster (Extended Data Fig. 5f), consistent with previously reported antigen-presenting states in human postmortem datasets[11,27,32], suggesting that homeostatic microglia exhibit transcriptional diversity that can be modeled in vitro. To further

https://doi.org/10.1038/s41590-023-01558-2

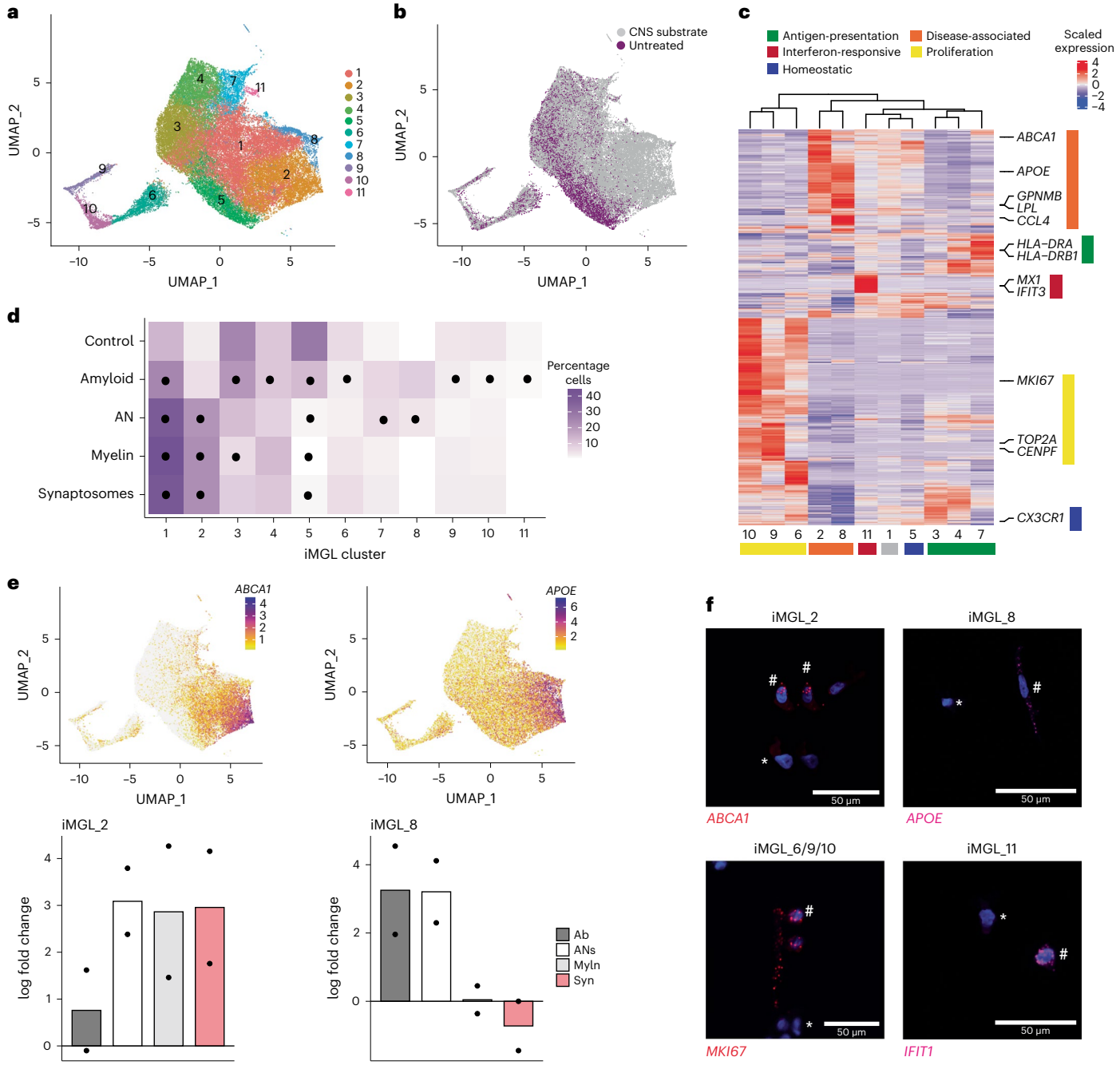



**Fig. 1 | Treatment of iMGLs with CNS substrates induces diverse transcriptional states that map to those found in vivo. a**, Uniform manifold approximation and projection (UMAP) of iMGLs that were either untreated or treated for 24 h with synaptosomes, myelin debris, synthetic Aβ fibrils or ANs (collectively referred to as CNS substrates) followed by scRNA-seq; total of 56,454 cells across two replicates, cells colored by cluster. **b**, UMAP projection as in **a**; cells colored as untreated or CNS-substrate-treated condition. **c**, Heatmap of differentially enriched genes for each cluster (iMGL_1-11) sorted by similarity and microglial states; states are labeled. **d**, Mean relative abundance of each cluster across each condition. Circles represent significant differences (adjusted $P < 0.05$; Supplementary Table 3) determined by a Dirichlet regression test for differential abundance. **e**, Marker gene expression (top) and log fold change of cluster relative to untreated (bottom) for clusters iMGL_2 (left) and iMGL_8 (right) ($n = 2$). Syn, synaptosomes; Myln, myelin debris; Ab, synthetic Aβ fibrils. **f**, Representative images of gene expression with fluorescent in situ hybridization for disease-associated (iMGL_2 and iMGL_8), proliferation (iMGL_6/9/10) and interferon-responsive (iMGL_11) states. All cells were positive for expression of *C1QA*; not shown for clarity. The hash symbol indicates a positive cell and the asterisk indicates a negative cell. Scale bar, 50 µm. See Extended Data Fig. 4 for quantifications per condition.

validate the similarity between microglial states in substrate-exposed iMGL with human states, we performed gene set enrichment analysis (GSEA)[33] between each iMGL cluster and clusters found in the brain biopsy dataset[29] or a postmortem human brain microglia dataset[24]. This analysis indicated that each iMGL state significantly overlapped with those identified in the brain biopsy dataset used above (Fig. 2e) and in the microglia dataset from ref. 24 (Extended Data Fig. 5m). Finally, we found that iMGLs exposed to substrates overlapped with some of these disease-associated states (Extended Data Fig. 5n). Together, these findings indicate that substrate-exposed iMGLs can exhibit transcriptional profiles analogous to those found in human brain microglia[27].

https://doi.org/10.1038/s41590-023-01558-2

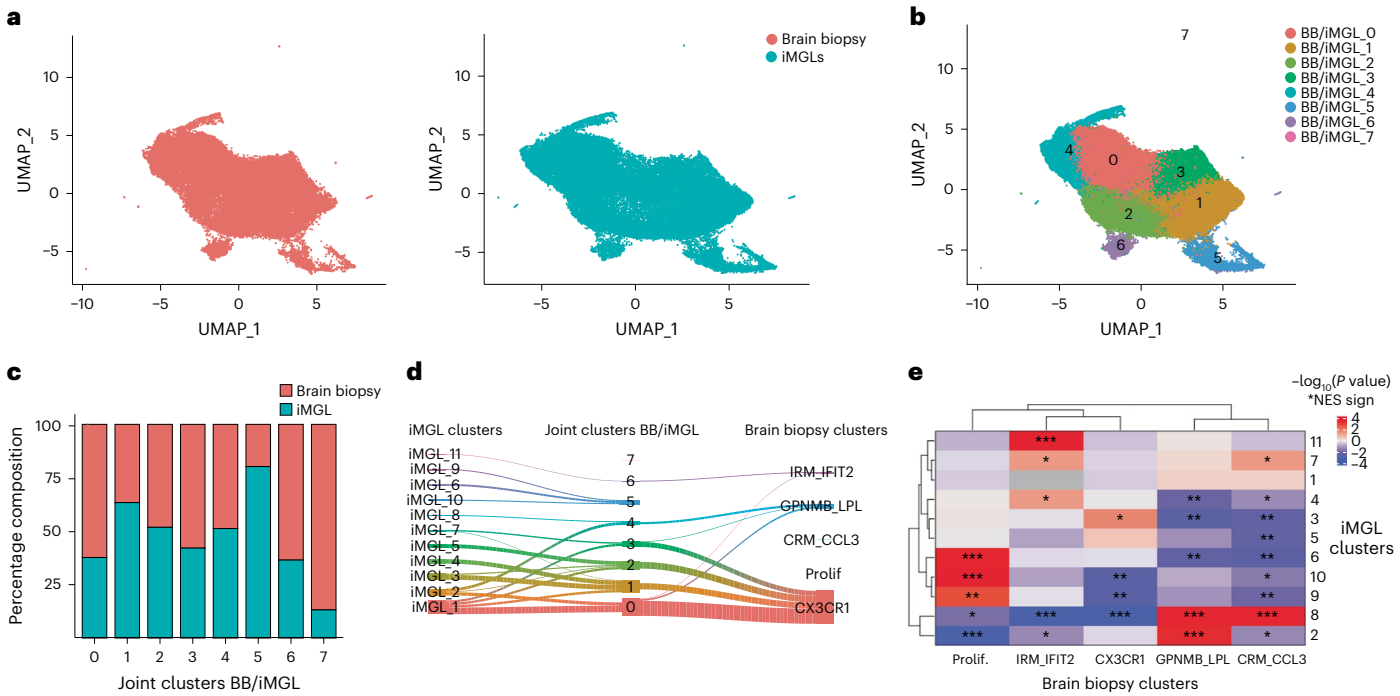



**Fig. 2 | Dataset integration of iMGL and human cortical biopsy microglia reveals analogous transcriptional states. a**, UMAP projection of integrated human brain biopsy microglia (left) and iMGL profiles (right). Cells are colored by dataset. **b**, UMAP projection of integrated human brain biopsy microglia and iMGL profiles. Cells are colored by BB/iMGL joint cluster. **c**, Proportion of cells per joint cluster from each source of data (human brain biopsy microglia or iMGLs), normalized by the proportion of total cells per dataset. **d**, River plot showing the relationship between joint clusters and the clusters defined in both the iMGL and brain biopsy datasets. Links with fewer than 200 cells have been removed for clarity. **e**, Heatmap illustrating the relative enrichment significance (as determined by fgsea; Methods) of positively enriched marker genes from each human brain biopsy cluster, within all differentially expressed genes for each iMGL cluster. Prolif, proliferative; NES, normalized enrichment score. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

## Human DAM state can be induced in vitro

DAMs are enriched in both human patients with AD[23,24] and mouse models of neurodegeneration[6,34], and this state depends on TREM2 receptor signaling[6,20,25,34]. To further examine the generation of DAM-like iMGL clusters using our in vitro platform, we first compared the gene expression signatures of DAM-like iMGLs with the DAM signature observed in the brain biopsy dataset. We compiled a list of DAM-associated genes by comparing DAM-like iMGLs (iMGL_2 and cluster iMGL_8 versus all other iMGL clusters) or human brain biopsy DAMs (BB_GPNMB_LPL versus all other brain biopsy clusters) with all the other clusters from the respective datasets. Comparison of these gene lists indicated a significant shared transcriptional signature between iMGL and brain biopsy DAMs (Fig. 3a). These shared genes included *ABCA1*, *APOE*, *GPNMB*, *LPL* and *TREM2* (Supplementary Table 4), confirming our GSEA results in which iMGL_2 and iMGL_8 were enriched for genes found in brain biopsy DAMs (BB_GPNMB_LPL) (Fig. 2e). To further evaluate the similarities between in vivo and iMGL DAMs, we examined the iMGL and brain biopsy microglia dataset integration. We observed that iMGL_2 and iMGL_8 cells were statistically significantly enriched in the same joint clusters (BB/iMGL_0 and BB/iMGL_4) as in vivo DAMs (BB_GPNMB_LPL) (Fig. 3b–d and Extended Data Fig. 6a). These cells, regardless of dataset source, shared a transcriptional signature (LIGER metagene) enriched for genes involved in neurodegeneration and amyloid pathology, including *GPNMB*, *MYO1E*, *ABCA1*, *CD9* and *APOE* (Fig. 3d and Supplementary Table 5)[6,23,24]. These results indicate that CNS substrate exposure can induce a DAM transcriptional state in vitro.

To understand the role of pathology in the formation of the DAM state, we integrated iMGL single-cell profiles with single-cell transcriptomic datasets from mouse[6] or xenotransplanted human microglia[26] exposed to amyloid plaques in an AD mouse model (5xFAD). Dataset integration indicated alignment of iMGL DAMs with 5xFAD mice microglia (Fig. 3e–g and Extended Data Fig. 6b–d, alignment score 0.58) and 5xFAD xenotransplanted iMGLs (Extended Data Fig. 6e–h, alignment score 0.83). We observed colocalization of iMGL DAMs, especially iMGL_8, with mouse DAMs and demonstrated significant overlap in gene signature using a hypergeometric test (Extended Data Fig. 6b). We also observed colocalization and shared gene signatures of iMGL_2 and iMGL_8 with xenotransplanted DAMs (Extended Data Fig. 6e–h). These results indicate a key role for amyloid pathology in DAM formation.

Mouse and human DAM transcriptomic profiles fall on a continuum[6,24]. We ordered the AN-exposed iMGLs, which included both iMGL DAM clusters (12,217 cells, $n = 2$), on a pseudotime trajectory based on their transcriptional profiles[28]. This showed a transition from cluster iMGL_1 to iMGL_2 to iMGL_8 (Extended Data Fig. 4a), suggesting that CNS substrate stimuli differentially drove iMGLs along a disease-associated trajectory, explaining the existence of multiple DAM-like iMGL clusters.

Mouse and human microglia depend on TREM2 receptor signaling to induce DAM signatures[6,20,34]. *TREM2* messenger RNA (mRNA) was highly expressed in iMGL_2 and iMGL_8 (Extended Data Fig. 1e), and TREM2 protein expression was elevated in iMGLs treated with ANs (Extended Data Fig. 7a). Glycoprotein NMB (*GPNMB*) specifically labeled both iMGL_2 and iMGL_8 DAM clusters (Fig. 3g) and was also enriched in BB_GPNMB_LPL microglia in the human biopsy data (Fig. 3d and Extended Data Fig. 5e) and published postmortem studies[6,20,23,24,26]. Immunocytochemistry and quantitative real-time PCR (rtPCR) indicated that exposure to ANs drove an increase in GPNMB in iMGLs compared with untreated cells (Fig. 3h,i and Extended Data Fig. 4e). The AN-induced or myelin-debris-induced increase in expression of *GPNMB*, *APOE* and *ABCA1* was lost in an iMGL line that lacked TREM2 (ref. 20) compared with the isogenic control (Fig. 3i and

https://doi.org/10.1038/s41590-023-01558-2

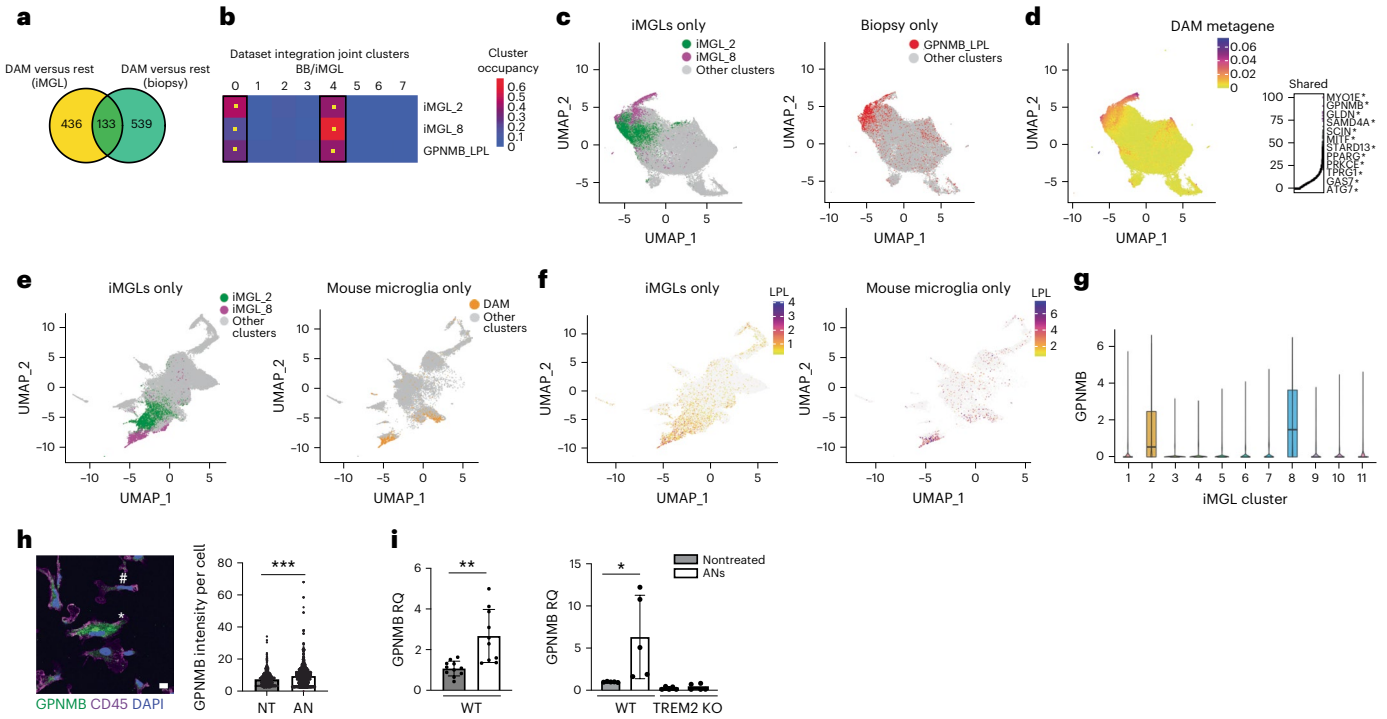



**Fig. 3 | DAMs generated in vitro are similar to those found in human and in mouse in vivo models. a**, Venn diagram showing overlap between DAM signature iMGL and brain biopsy ($P = 3.06 \times 10^{-68}$, hypergeometric test). **b**, Percentage distribution of cells from iMGL_2, iMGL_8 and BB_GPNMB_LPL over joint clusters BB/iMGL. Yellow square indicates significant enrichment using binomial test, $P < 0.05$. **c**, UMAP projection (as in Fig. 2) colored by cluster identity. **d**, UMAP projection (as in **c**) showing the metagene common to both iMGL and brain biopsy DAMs (left) and top genes (right). Asterisk labels genes in the 'AD1' DAM cells identified in ref. 24. **e**, UMAP projection of LIGER integration of iMGL and microglia from the 5xFAD mouse model, cells colored by cluster: iMGL_2 (green), iMGL_8 (magenta) or mouse microglia with DAMs (orange). **f**, UMAP projection (as in **e**) showing expression of *LPL* in iMGLs (left) and mouse microglia (right). **g**, Violin plot of *GPNMB* mRNA expression plotted by iMGL cluster; $n = 2$ replicates per condition, all five conditions pooled. Boxes show the first and third quartiles of the data with a line marking the median. Whiskers mark values closest to 1.5 times the interquartile range; no outliers are plotted. **h**, Immunocytochemistry of GPNMB expression in iMGLs treated with ANs or PBS. Left, representative image showing GPNMB (green) and CD45 (magenta, microglia marker). Asterisk indicates a positive cell, and hash indicates a negative cell. Right, quantitative analysis (two-tailed *t*-test, $^{***}P < 0.001$; at least 500 cells were counted per condition, data combined from four replicates, mean and s.d. are shown). Scale bar, 50 µm. **i**, Quantitative rtPCR of *GPNMB* expression in H1 iMGLs (left) and TREM2-deficient or isogenic iMGLs (right) treated with ANs or PBS. For H1 iMGL, two-tailed *t*-test, $P = 0.0009$, $n > 4$, mean and s.d. of relative quantification (RQ) are shown; for TREM2-deficient or isogenic iMGL, two-tailed *t*-test, $P = 0.0427$, $n > 4$, mean and s.d. are shown. KO, knockout; WT, wild type. $^{*}P < 0.05$, $^{**}P < 0.01$.

Extended Data Fig. 7b,c), indicating that the DAM-like state in iMGLs is dependent on TREM2. Treatment of iMGL with ANs and cytochalasin D did not lead to induction of *GPNMB* (Extended Data Fig. 7d), indicating that the expression of *GPNMB* is dependent on phagocytosis. *GPNMB* was not induced by *Escherichia coli* (Extended Data Fig. 7e), suggesting that DAM formation is specific to CNS substrates. Together, these data suggest that iMGL DAMs share characteristics with human and mouse DAMs and are formed specifically in response to CNS-derived stimuli.

## MITF is a driver of the DAM signature and phagocytosis

To identify transcription factors that could be key regulators of DAM states, we performed ATAC-seq (assay for transposase-accessible chromatin using sequencing) on both untreated and AN-exposed iMGLs to identify regions with changes in chromatin and in accessibility of transcription factor binding sites. Treatment with ANs increased the number of differentially accessible chromatin regions in iMGLs compared with phosphate-buffered saline (PBS)-treated iMGLs (Fig. 4a). HOMER[35] was then used to identify transcription factor binding motifs enriched in the regions with increased accessibility, nominating transcription factors that may potentially regulate gene expression in response to ANs, including PU.1, MAFB, EGR1 and MITF (Supplementary Table 6).

To further identify regulators of microglial states, we performed a gene regulatory network analysis on iMGL single-cell profiles using SCENIC[36]. This identified the differential expression of transcription factors and their putative direct targets (together termed regulons) (Extended Data Fig. 8a and Supplementary Table 7). Analysis of the enriched regulons in clusters iMGL_1, iMGL_2 and iMGL_8 indicated that whereas some were shared across these three clusters (CEBPD, TCF4), several were differentially expressed in each cluster (ATF3 regulon in iMGL_8, REB1 in iMGL_2) (Extended Data Fig. 8a). Although we focused on DAMs, this analysis revealed potential regulatory networks underlying all iMGL clusters (Supplementary Table 7). Combining the transcription factors nominated by both chromatin accessibility and SCENIC analyses, we identified 16 putative regulators of the DAM state, including MITF, MAFB and EGR2 (Fig. 4b and Supplementary Table 6). We chose to focus on MITF as expression of this gene is upregulated in microglia of patients with AD[23,24], and this transcription factor has been proposed as a regulator of the neurodegenerative signature[37]. Correspondingly, MITF mRNA is highly upregulated in DAM clusters iMGL_2 and iMGL_8 (Fig. 4c), and *MITF* expression scaled with the disease-associated trajectory identified in iMGLs (Extended Data Fig. 4a).

To functionally assess MITF in iMGL, we developed a lentiviral approach to transduce differentiated microglia, which are resistant to DNA delivery[19]. Transduction of monocytes with lentiviruses is facilitated by the SIV-encoded protein Vpx[38,39], which degrades SAMHD1, a restriction factor that prevents reverse transcription of lentiviral RNA[40,41]. Co-delivery of Vpx packaged in virus-like particles (VLPs) led to ~89% transduction in iMGLs, compared with ~4% of iMGLs with

https://doi.org/10.1038/s41590-023-01558-2

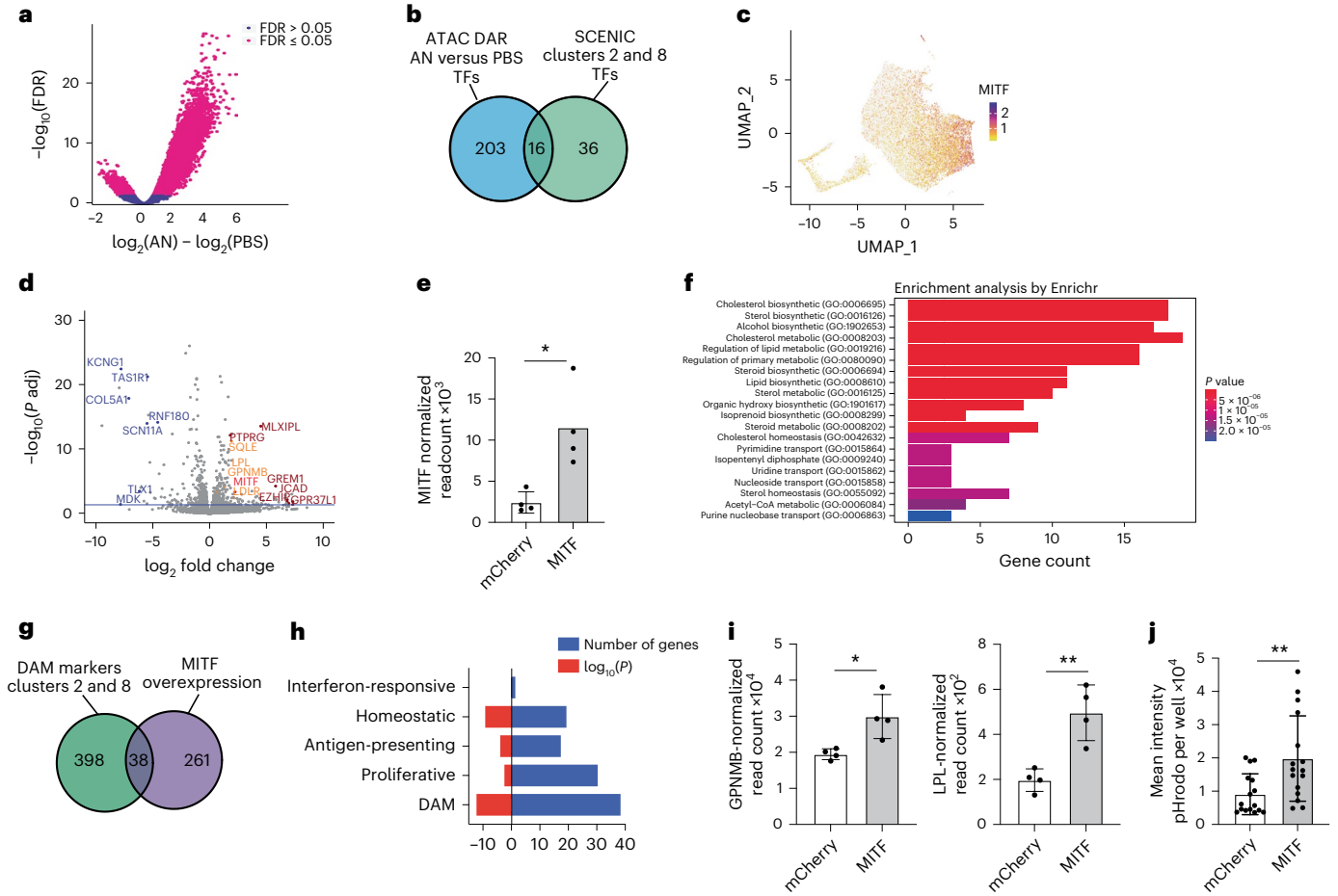



**Fig. 4 | MITF is a key DAM regulator and driver of phagocytosis. a**, Volcano plot of differentially accessible peaks according to ATAC-seq of iMGLs exposed to ANs compared with untreated. Peaks with false discovery rate (FDR) ≤ 0.05 are shown in magenta. **b**, Intersection of transcription factors nominated by ATAC-seq and SCENIC analysis. **c**, UMAP projection of iMGL dataset (as in Fig. 1a) showing *MITF* expression in iMGLs treated with CNS substrates. **d**, Volcano plot of differentially expressed genes (DEG) in the iMGLs overexpressing MITF ($n = 4$) compared with those overexpressing mCherry ($n = 4$). Blue indicates top downregulated genes, dark red indicates top upregulated genes and orange indicates genes involved in DAM signature and lipid metabolism. Blue line represents adjusted FDR ≤ 0.05. *P* adj, adjusted *P* value. **e**, Quantification of *MITF* expression in iMGLs transduced with mCherry-expressing and MITF-expressing lentivirus (two-tailed *t*-test, $P = 0.0129$, mean and s.d. are shown; $n = 4$ for each condition). **f**, Bar graph of gene ontology analysis for MITF-overexpression DEG. Bar length represents the number of genes, and shading represents statistical significance. **g**, Venn diagram of overlap between DEG in iMGL samples transduced with MITF-expressing versus mCherry-expressing lentivirus, and DEG in iMGL_02 and iMGL_8; hypergeometric test, $P = 1.098 \times 10^{-12}$. **h**, Comparison of MITF-overexpression genes with iMGL states; numbers of genes and *P* values from the hypergeometric test are shown. **i**, Quantification of *GPNMB* expression (left, two-tailed *t*-test, $P = 0.0159$, $^{*}P < 0.05$) and *LPL* expression (right, two-tailed *t*-test, $P = 0.0042$, $^{**}P < 0.01$) in RNA sequencing samples as in **d**. **j**, Mean intensity of pHrodo-conjugated AN in iMGLs transduced with either MITF or mCherry (two-tailed *t*-test, $P = 0.0051$, $^{**}P < 0.01$; $n > 5$ for each condition). In all bar graphs, error bars represent s.d.

lentivirus alone (Extended Data Fig. 8b). Vpx VLPs improved transduction of iMGLs derived from five stem cell lines (Extended Data Fig. 8c), indicating that Vpx has a robust effect regardless of the iMGL source. We performed scRNAseq and found that no-lentivirus (control) iMGLs and Vpx$^{+}$ lentivirus-exposed iMGLs clustered together (Extended Data Fig. 8d) and had similar expression of key microglia markers (Extended Data Fig. 8e); however, fewer lentivirus-transduced iMGLs were proliferating (11.9%) compared with no-lentivirus iMGLs (24.8%) (Extended Data Fig. 8f). Although lentiviral transduction significantly increased the expression of 156 interferon-stimulated genes (Supplementary Table 8), assessment of a representative subset (*IFI6*, *IFITM3*, *ISGM3*) showed downregulation of these genes over 7 and 12 days compared with 2 days after transduction (Extended Data Fig. 8g–i). Thus, this transduction method is an efficient tool to manipulate gene expression in iMGLs while minimally affecting their transcriptional profile.

To understand MITF function, we transduced iMGL with a lentiviral vector expressing either MITF or mCherry (as a control) and performed gene expression and functional assays. On average, 79% of control lentivirus-transduced iMGLs were mCherry$^{+}$ (Extended Data Fig. 8j). Bulk RNA-seq of iMGLs transduced with either MITF[42] or mCherry (as control) indicated that MITF induced the differential expression of 670 genes compared with mCherry; of these, 300 genes were upregulated and 396 genes were downregulated (Fig. 4d and Supplementary Table 9), including *MITF*, confirming overexpression (Fig. 4e). Gene ontology analysis showed enrichment of cytokine signaling and cholesterol metabolism, pathways upregulated in the DAM transcriptional signature[6,23,24,34], in MITF-transduced iMGLs compared with mCherry$^{+}$ iMGLs (Fig. 4f). We observed a significant overlap between 38 genes upregulated in DAM-like iMGL_2 and iMGL_8 and MITF-transduced iMGLs (Fig. 4g and Supplementary Table 9), including *GPNMB* and *LPL* (Fig. 4i). Comparison of the overlap between the MITF-induced genes with iMGL markers of homeostatic, DAM, interferon-responsive, proliferative and antigen-presenting states highlighted a stronger overlap (as determined by hypergeometric test and number of genes) between

https://doi.org/10.1038/s41590-023-01558-2

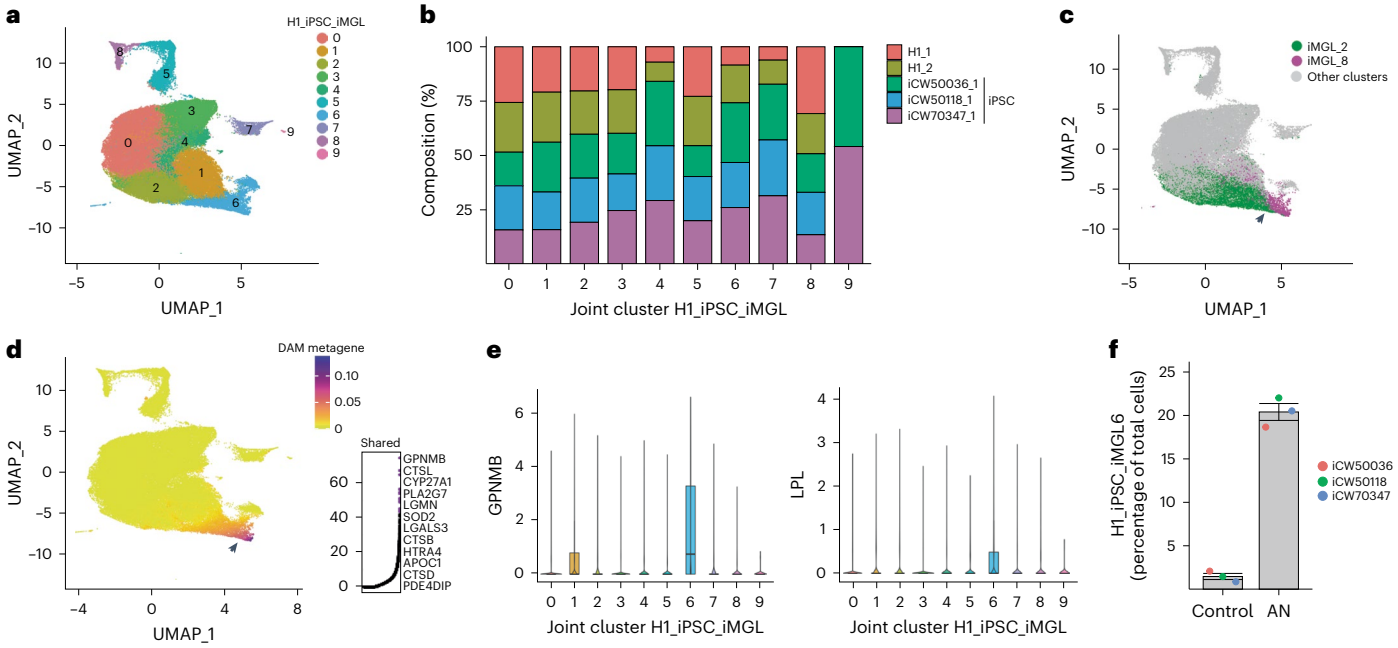



**Fig. 5 | Multiple iPSC-line-derived iMGLs also exhibit transcriptional diversity and DAM induction. a**, UMAP projection of integrated H1 iMGL and iPSC iMGL profiles; cells are colored by identity of H1_iPSC_iMGL joint clusters. **b**, Proportion of cells per H1_iPSC_iMGL joint cluster for each dataset (either iMGL H1 replicates or three separate iPSC lines differentiated into iMGLs). **c**, UMAP projection as in **a**. Green, iMGL_2; magenta, iMGL_8. **d**, UMAP projection (as in **a**) of the shared metagene common to both datasets in cluster 6. Right, top constituent genes of this shared factor. **e**, Violin plots showing expression of GPNMB and LPL across joint clusters in **a**. Boxes show the first and third quartiles of the data, with a line marking the median. Whiskers mark values closest to 1.5 times the interquartile range; no outliers are plotted. **f**, Percentages of nontreated (control) and AN-treated cells in joint cluster 6, from iPSC-derived iMGL only (two-tailed *t*-test, $P = 0.00471$; $n = 3$; mean and s.e.m. are shown). In **c** and **d**, the black arrow highlights joint cluster 6.

MITF-induced genes and the DAM state compared to other microglia states (Fig. 4h). An increase in phagocytosis has been described in neurodegenerative disease models[4], and we investigated the functional impact of MITF overexpression. MITF-transduced iMGLs exhibited increased phagocytosis of pHrodo-labeled ANs compared with mCherry$^+$ iMGLs (Fig. 4j), suggesting that MITF regulates a phagocytic state in microglia. These observations identify MITF as a transcription factor that potentially regulates the neurodegenerative and phagocytic transcriptional signature in DAMs.

### Microglial state diversity is observed across multiple iPSC lines

A key advantage of stem-cell-derived models is the ability to perform parallel experiments on iPSC lines derived from multiple patients. To test whether our findings were broadly applicable, we differentiated iMGLs from three independent iPSC lines (iCW50118, iCW500036 and iCW70437) derived from healthy individuals with APOE 3/3 status (as previously defined[43]) and no *TREM2* mutations. We exposed these iMGL ANs or PBS, and performed scRNA-seq. This analysis generated 41,655 iPSC-derived single-cell transcriptomes, which we integrated with our iMGL dataset (here referred to as the H1 dataset) using LIGER (Fig. 5a,b and Supplementary Table 10). We found a high level of similarity between the two datasets (alignment score, 0.973; Fig. 5b and Extended Data Fig. 9a–c). This integrative analysis produced ten joint clusters (H1_iPSC_iMGL0-9). Most joint clusters were well distributed between H1-derived and iPSC-derived iMGLs, with the exception of H1_iPSC_iMGL9; this was only present in iMGLs from two iPSC lines (Fig. 5b), which contained rare (100 of 41,655) cells enriched for ribosomal genes (for instance, *RPS18*). Further analysis indicated similar transcriptional states to those identified previously as determined by marker gene expression (Extended Data Fig. 9d–h) and module score calculation (Extended Data Fig. 10). Based on this analysis, DAM clusters iMGL_2 and iMGL_8 mapped to H1_iPSC_iMGL_6 (Fig. 5a,c,d), which corresponded to a metagene consisting of DAM signature genes (Fig. 5d and Extended Data Fig. 10c) including canonical marker genes *GPNMB* and *LPL* (Fig. 5e). We found that iPSC-derived iMGLs from joint cluster H1_iPSC_iMGL_6 were significantly increased as a proportion of total iMGLs following exposure to ANs versus no treatment (Fig. 5f). This indicates that microglia differentiated from iPSCs also exhibit the DAM state in the presence of ANs and highlights the generalizable nature of the iMGL platform.

## Discussion

Here, we provide a toolbox for the study of human microglial states, including a single-cell gene expression resource for in vitro iMGL. We found that challenge with CNS substrates produced a large degree of transcriptional diversity and that the iMGL states induced in vitro mapped to those found in the human brain. Exposure to some CNS substrates was sufficient to induce a TREM2-dependent DAM state that was analogous to those found in both the human and mouse brain[6,20,44], suggesting that addition of easily sourced substrates can produce models of human microglial states in vitro. We also demonstrated that these features are shared over multiple iPSC lines, which will enable investigators to analyze microglia from a full spectrum of patients and diseases, and prioritize human cell lines for in-depth functional characterization[45].

One key finding was that iMGLs existed in distinct transcriptional states in vitro even in the absence of stimulation, and that upon stimulation with a specific substrate, these states did not converge to a single transcriptional signature. This was similar to in vivo observations that microglia in similar locations could adopt distinct transcriptional states[5] and suggests that heterogeneous response to substrates is common in microglia. However, microglial profiling ex vivo represents only a snapshot of the transcriptional diversity at a particular time after treatment and may not reflect the entire dynamic process. While the dynamics and plasticity of microglial states after a stimulus remain

unclear[14]; this in vitro platform will enable future longitudinal and imaging studies to address this important question.

An additional challenge is that microglia, unlike other brain cell types, are not amenable to viral transduction. The development of genetically modified iPSCs is low throughput and costly, and deletion of essential genes can affect differentiation[46]. To further enable scalable experiments in iMGLs, we developed a lentiviral protocol that broadly transduced iMGLs in the last phase of their differentiation, inducing only modest perturbation of the core microglial transcriptional signature. We applied this technology to interrogate the function of MITF, a transcription factor that we identified as a putative regulator of DAM, which is also upregulated in AD[23,24,37]. We found that MITF overexpression drives a subset of the DAM signature and increased phagocytosis. Although the number of overlapping genes between microglia with MITF overexpression and those in the DAM state was relatively modest, gene ontology analysis of all MITF-driven genes converged on known pathways upregulated in the DAM transcriptional signature across species[6,23,24,34], suggesting that MITF is a potential regulator of the neurodegenerative-associated signature and function[23,24,37]. By allowing facile genetic modification of iMGLs, this technology has the potential to enable the identification of key regulators of microglial states and functions[14,27,47] and the rapid application of molecular tools.

Although we focused on the DAM signature, the data provided here will enable the dissection of many other microglial states, such as interferon-responsive or proliferative microglia[14,27]. While our analyses suggested that iMGLs adopted states analogous to those found in vivo, we caution that this does not mean the signatures described here were identical. For example, although iMGLs and human brain biopsy DAMs shared many aspects of their gene signatures, there were also differences. Similarly, modeling baseline homeostatic states in vitro is notoriously challenging[15,17], and further optimization will be needed to enhance the in vitro microglial homeostatic signature. We envision this resource as a tool for investigators to examine the function of human microglial diversity in a high-throughput manner. These studies could then be augmented and validated with lower-throughput approaches that recapitulate the in vivo environment, such as organoid or chimeric models[16,26,32,48,49].

Combining substrate exposure, scRNA-seq, integrative analyses, epigenetics and viral vectors will allow manipulation of the genes that define microglial states and rapidly assay their impact on microglial functions. The use of additional substrates (for example, different forms or species of amyloid, tau and synuclein) may enable the modeling of microglial states found in other disorders. This platform also allows the generation of large numbers of cells for unbiased characterization methods such as proteomics or epigenetics, which are challenging in human tissue because of the small number of microglia that can be obtained from a single individual. Finally, combining iMGLs with neuron and/or astrocyte cocultures[50] could enable the study of noncell-autonomous phenotypes and neuron–glia immune interactions. In conclusion, the in vitro approach described here will enable a broad characterization of human microglial states and bridge the gap between the transcriptomic profile and functional role of these states in neuroimmune interactions in health and disease.

## Online content

Any methods, additional references, Nature Portfolio 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/s41590-023-01558-2.

## References


1. Salter, M. W. & Stevens, B. Microglia emerge as central players in brain disease. *Nat. Med.* **23**, 1018–1027 (2017).
2. 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).
3. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
4. Hickman, S., Izzy, S., Sen, P., Morsett, L. & El Khoury, J. Microglia in neurodegeneration. *Nat. Neurosci.* **21**, 1359–1369 (2018).
5. 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).
6. Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**, 1276–1290.e17 (2017).
7. Dorman, L. C. et al. A type I interferon response defines a conserved microglial state required for effective phagocytosis. Preprint at *bioRxiv* https://doi.org/10.1101/2021.04.29.441889 (2023).
8. Li, Q. et al. Developmental heterogeneity of microglia and brain myeloid cells revealed by deep single-cell RNA sequencing. *Neuron* **101**, 207–223.e10 (2019).
9. Young, A. M. H. et al. A map of transcriptional heterogeneity and regulatory variation in human microglia. *Nat. Genet.* **53**, 861–868 (2021).
10. Mathys, H. et al. Temporal tracking of microglia activation in neurodegeneration at single-cell resolution. *Cell Rep.* **21**, 366–380 (2017).
11. Olah, M. et al. Single cell RNA sequencing of human microglia uncovers a subset associated with Alzheimer's disease. *Nat. Commun.* **11**, 6129 (2020).
12. Masuda, T. et al. Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution. *Nature* **566**, 388–392 (2019).
13. Belhocine, S. et al. Context-dependent transcriptional regulation of microglial proliferation. *Glia* **70**, 572–589 (2022).
14. Paolicelli, R. C. et al. Microglia states and nomenclature: a field at its crossroads. *Neuron* **110**, 3458–3483 (2022).
15. Gosselin, D. et al. An environment-dependent transcriptional network specifies human microglia identity. *Science* **356**, eaal3222 (2017).
16. Mancuso, R. et al. Stem-cell-derived human microglia transplanted in mouse brain to study human disease. *Nat. Neurosci.* **22**, 2111–2116 (2019).
17. Bohlen, C. J. et al. Diverse requirements for microglial survival, specification, and function revealed by defined-medium cultures. *Neuron* **94**, 759–773.e8 (2017).
18. He, Y. et al. RNA sequencing analysis reveals quiescent microglia isolation methods from postnatal mouse brains and limitations of BV2 cells. *J. Neuroinflammation* **15**, 153 (2018).
19. Maes, M. E., Colombo, G., Schulz, R. & Siegert, S. Targeting microglia with lentivirus and AAV: recent advances and remaining challenges. *Neurosci. Lett.* **707**, 134310 (2019).
20. McQuade, A. et al. Gene expression and functional deficits underlie TREM2-knockout microglia responses in human models of Alzheimer's disease. *Nat. Commun.* **11**, 5370 (2020).
21. Abud, E. M. et al. iPSC-derived human microglia-like cells to study neurological diseases. *Neuron* **94**, 278–293.e9 (2017).
22. Marsh, S. E. et al. Dissection of artifactual and confounding glial signatures by single-cell sequencing of mouse and human brain. *Nat. Neurosci.* **25**, 306–316 (2022).
23. Smith, A. M. et al. Diverse human astrocyte and microglial transcriptional responses to Alzheimer's pathology. *Acta Neuropathol.* **143**, 75–91 (2021).

24. Gerrits, E. et al. Distinct amyloid-β and tau-associated microglia profiles in Alzheimer's disease. *Acta Neuropathol.* **141**, 681–696 (2021).
25. Sayed, F. A. et al. AD-linked R47H-TREM2 mutation induces disease-enhancing microglial states via AKT hyperactivation. *Sci. Transl. Med.* **13**, eabe3947 (2021).
26. Hasselmann, J. et al. Development of a chimeric model to study and manipulate human microglia in vivo. *Neuron* **103**, 1016–1033.e10 (2019).
27. Chen, Y. & Colonna, M. Microglia in Alzheimer's disease at single-cell level. Are there common patterns in humans and mice? *J. Exp. Med.* **218**, e20202717 (2021).
28. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
29. Gazestani, V. H. et al. Early Alzheimer's disease pathology in human cortex is associated with a transient phase of distinct cell states. Preprint at *bioRxiv* https://doi.org/10.1101/2023.06.03.543569 (2023).
30. Welch, J. D. et al. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873–1887.e17 (2019).
31. Liddelow, S. A., Marsh, S. E. & Stevens, B. Microglia and astrocytes in disease: dynamic duo or partners in crime? *Trends Immunol.* **41**, 820–835 (2020).
32. Mancuso, R. et al. A multi-pronged human microglia response to Alzheimer's disease Aβ pathology. Preprint at *bioRxiv* https://doi.org/10.1101/2022.07.07.499139 (2022).
33. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
34. Zhou, Y. et al. Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat. Med.* **26**, 131–142 (2020).
35. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
36. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
37. Friedman, B. A. et al. Diverse brain myeloid expression profiles reveal distinct microglial activation states and aspects of Alzheimeras disease not evident in mouse models. *Cell Rep* **22**, 832–847 (2018).
38. Goujon, C. et al. Characterization of simian immunodeficiency virus SIVSM/human immunodeficiency virus type 2 Vpx function in human myeloid cells. *J. Virol.* **82**, 12335–12345 (2008).
39. Liu, B. et al. An efficient lentiviral CRISPRi approach to silence genes in primary human monocytes. Preprint at *bioRxiv* https://doi.org/10.1101/2020.12.23.424242 (2020).
40. Laguette, N. et al. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* **474**, 654–657 (2011).
41. Hrecka, K. et al. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature* **474**, 658–661 (2011).
42. Joung, J. et al. A transcription factor atlas of directed differentiation. *Cell* **186**, 209–229.e26 (2023).
43. Serrano-Pozo, A., Das, S. & Hyman, B. T. APOE and Alzheimer's disease: advances in genetics, pathophysiology, and therapeutic approaches. *Lancet Neurol.* **20**, 68–80 (2021).
44. Krasemann, S. et al. The TREM2-APOE pathway drives the transcriptional phenotype of dysfunctional microglia in neurodegenerative diseases. *Immunity* **47**, 566–581.e9 (2017).
45. Sandoe, J. & Eggan, K. Opportunities and challenges of pluripotent stem cell neurodegenerative disease models. *Nat. Neurosci.* **16**, 780–789 (2013).
46. Wu, M. et al. Conditional gene knockout and reconstitution in human iPSCs with an inducible Cas9 system. *Stem Cell Res.* **29**, 6–14 (2018).
47. Deczkowska, A. et al. Disease-associated microglia: a universal immune sensor of neurodegeneration. *Cell* **173**, 1073–1081 (2018).
48. Fagerlund, I. et al. Microglia-like cells promote neuronal functions in cerebral organoids. *Cells* **11**, 124 (2021).
49. Popova, G. et al. Human microglia states are conserved across experimental models and regulate neural stem cell responses in chimeric organoids. *Cell Stem Cell* **28**, 2153–2166.e6 (2021).
50. Bassil, R. et al. Improved modeling of human AD with an automated culturing platform for iPSC neurons, astrocytes and microglia. *Nat. Commun.* **12**, 5220 (2021).


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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023

[1]Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Boston Children's Hospital, F.M. Kirby Neurobiology Center, Boston, MA, USA. [3]Harvard Medical School, Boston, MA, USA. [4]Broad Institute of MIT and Harvard Cambridge, Cambridge, MA, USA. [5]Department of Biological Engineering, MIT, Cambridge, MA, USA. [6]Howard Hughes Medical Institute, Boston, MA, USA. [7]Department of Brain and Cognitive Science, MIT, Cambridge, MA, USA. [8]McGovern Institute for Brain Research at MIT, Cambridge, MA, USA. [9]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. [10]Leiden University Medical Center, LUMC, Leiden, the Netherlands. [11]Department of Medicine, Center for Cancer Research, Massachusetts General Hospital, Boston, MA, USA. [12]Department of Medicine, Harvard Medical School, Boston, MA, USA. [13]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA. [14]Ludwig Center at Harvard, Harvard Medical School, Boston, MA, USA. [15]Departments of Cell Biology and Pathology, Harvard Medical School, Boston, MA, USA. [16]Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA. [17]Department of Neurosurgery, Kuopio University Hospital and Institute of Clinical Medicine – Neurosurgery, University of Eastern Finland, Kuopio, Finland. [18]Department of Neurobiology and Behavior, Sue and Bill Gross Stem Cell Research Center, UCI Institute for Memory Impairments and Neurological Disorders, Institute for Immunology, University of California, Irvine, CA, USA. [19]Massachusetts General Hospital, Department of Psychiatry, Boston, MA, USA. [20]These authors contributed equally: Michael-John Dolan, Martine Therrien.
✉e-mail: emacosko@broadinstitute.org; Beth.Stevens@childrens.harvard.edu

## Methods

### Embryonic stem cell and iPSC lines

All stem cell work was reviewed and approved by the Broad Institute Office of Research Subject Protection. Unless otherwise stated, H1 embryonic stem cells (WiCell) were used. The iPSC lines were obtained from the California Institute of Regenerative Medicine (CIRM) hPSC Repository funded by CIRM: CW50118, CW50008, CW50065, CW500036 and CW70437. The cell lines chosen for scRNA-seq (CW50118, CW500036 and CW70437) were confirmed as APOE 3/3 status with no *TREM2* mutation by Sanger sequencing. In addition, individuals did not exhibit cognitive decline at the time of collection, which was after 70 years of age for all lines (Supplementary Table 10), suggesting that these lines should represent healthy individuals. The TREM2-knockout and isogenic control iPSC lines were as previously characterized[20] and were obtained from M. Blurton-Jones. These lines were derived from cell line AICS-0036-006 from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research; thus, they have not undergone the standard quality control of the Repository.

### iMGL differentiation

iMGLs were differentiated as previously described[21]. Briefly, iPSC or embryonic stem cells were cultured in Essential 8 (E8) (Thermo Fisher Scientific) media on six-well plates coated with Matrigel (Corning). When confluent, cells were dissociated using Accutase (Stem Cell Technologies), centrifuged for 5 min at 300*g* and counted using trypan blue (Thermo Fisher Scientific). Next, 200,000 cells per well were resuspended in E8 containing 10 µM Y-27632 ROCK inhibitor (Selleckchem) in a low-adherence six-well plate (Corning). For the first 10 days, cells were cultured in HPC medium (50% IMDM Thermo Fisher Scientific), 50% F12 (Thermo Fisher Scientific), ITSG-X 2% v/v (Thermo Fisher Scientific), L-ascorbic acid 2-phosphate (64 µg ml$^{-1}$, Sigma), monothioglycerol (400 mM, Sigma), poly(vinyl) alcohol (10 mg ml$^{-1}$, Sigma), GlutaMAX (1×, Thermo Fisher Scientific), chemically defined lipid concentrate (1×, Thermo Fisher Scientific) and nonessential amino acids (Thermo Fisher Scientific)). At day 0, embryoid bodies were gently collected, centrifuged at 100*g* and resuspended in HPC medium supplemented with 1 µM ROCK inhibitor, FGF2 (50 ng ml$^{-1}$, Thermo Fisher Scientific), BMP4 (50 ng ml$^{-1}$, Thermo Fisher Scientific), Activin-A (12.5 ng ml$^{-1}$, Thermo Fisher Scientific) and LiCL (2 mM, Sigma), then incubated in a hypoxic incubator (5% $O_2$, 5% $CO_2$, 37 °C). On day 2, cells were gently collected, and the medium was changed to HPC medium supplemented with FGF2 (50 ng ml$^{-1}$, Thermo Fisher Scientific) and VEGF (50 ng ml$^{-1}$, PeproTech), before cells were returned to the hypoxic incubator. On day 4, cells were gently collected and the medium was changed to HPC medium supplemented with FGF2 (50 ng ml$^{-1}$, Thermo Fisher Scientific), VEGF (50 ng ml$^{-1}$, PeproTech), TPO (50 ng ml$^{-1}$, PeproTech), SCF (10 ng ml$^{-1}$, Thermo Fisher Scientific), IL6 (50 ng ml$^{-1}$, PeproTech) and IL3 (10 ng ml$^{-1}$, PeproTech); then, cells were incubated in a normoxic incubator (20% $O_2$, 5% $CO_2$, 37°C). On days 6 and 8, 1 ml of day 4 media was added to each well. On day 10, cells were collected, counted using trypan blue and frozen in Cryostor (Sigma Aldrich) in aliquots of 300,000–500,000 cells.

To start iMGL differentiation, cells were thawed, washed 1× with PBS and plated at 100,000–200,000 cells per well in a six-well plate coated with Matrigel in iMGL media ((DMEM/F12 (Thermo Fisher Scientific), ITSG (2% v/v, Thermo Fisher Scientific), B27 (2% v/v, Thermo Fisher Scientific), N2 (0.5% v/v, Thermo Fisher Scientific), monothioglycerol (200 mM, Sigma), GlutaMAX (1×, Thermo Fisher Scientific), nonessential amino acids (1×, Thermo Fisher Scientific)) supplemented with M-CSF (25 ng ml$^{-1}$, PeproTech), IL-34 (100 ng ml$^{-1}$, PeproTech) and TGFB-1 (50 ng ml$^{-1}$, PeproTech). Cells were fed every 2 days and replated at day 22. On day 30, cells were collected and replated in iMGL media supplemented with M-CSF (25 ng ml$^{-1}$, PeproTech), IL-34 (100 ng ml$^{-1}$, PeproTech), TGFB-1 (50 ng ml$^{-1}$, PeproTech), CD200 (100 ng ml$^{-1}$, VWR) and CX3CL1 (100 ng ml$^{-1}$, PeproTech) to a final concentration of 40,000 cells cm$^{-2}$. Cells were used at day 40 for functional and transcriptomic assays. iMGL differentiation was assessed at day 10 (expression of CD43 and CD45) and day 40 (expression of CD45, CD11B, P2RY12, CX3CR1 and DAPI for viability) by flow cytometry (see below for protocol). During the quality control on day 40, we observed similar protein expression among the replicates, with the exception of CX3CR1, which had lower expression in replicate 1 (Extended Data Fig. 1a). At no point were iMGLs exposed to serum of any kind. Throughout the differentiation, cell viability was >90% as measured by trypan blue.

### Flow cytometry

For sample processing, iMGLs were detached using cold PBS then resuspended in fluorescence-activated cell sorting (FACS) buffer (PBS containing 2% bovine serum albumin and 0.05 mM EDTA). Samples were incubated for 15 min in human Fc block (BD Biosciences), followed by 1 h staining with conjugated antibodies (see below) at 4 °C. Samples were washed three times with FACS buffer and resuspended in 500 µl of FACS buffer for flow cytometry. Samples were run on a CytoFLEX S analyzer (Beckman Coulter) until at least 2,000 cells have been recorded. For analysis, cells were identified according to the following gating: (1) cell versus debris (FSC-A versus SSC-A); (2) singlets (FSC-A versus FSC-H), and antibody-specific gating based on a negative control sample. Antibodies for staining iMGLs were CD45-FITC (BioLegend), CD11B-APC-750 (BioLegend), P2RY12-PB450 (BioLegend), Cx3CR1-PrCp (BioLegend) 1/500 for all.

For phagocytosis, cells were treated with substrates labeled with pHrodo Green (see below) for 24 h. iMGLs were then washed with PBS and collected in FACS buffer (PBS containing 2% bovine serum albumin and 0.05 mM EDTA). Samples were then directly run on a CytoFLEX S analyzer (Beckman Coulter) until at least 2,000 cells had been recorded. For analysis, cells were identified according to the following gating: (1) cell versus debris (FSC-A versus SSC-A); (2) singlets (FSC-A versus FSC-H); (3) viability (based on DAPI stain). Phagocytosis was quantified as the mean fluorescence intensity of pHrodo for each sample. Each experiment was performed using multiple independent differentiations, and each dot in the corresponding graph represents a biological replicate.

All analyses were performed using FlowJo v.9, with Prism9 for statistical analysis.

### CNS substrate isolation and iMGL treatment

Unless otherwise stated, iMGLs were treated with CNS substrates for 24 h before transcriptomic, flow cytometry or RNAscope analysis.

Synaptosomes were prepared from rodent brains as described previously[51]. Briefly, C57BL/6J mice were euthanized with $CO_2$; then, brains were dissected and homogenized in HEPES-buffered sucrose (0.32 M, 5 mM HEPES, pH 7.4). The resulting homogenate was spun at 800–1,200*g* to separate the nuclear fraction. A further spin at 15,000*g* was carried out to generate crude synaptosomes, and 3.1 µg cm$^{-2}$ was used for experiments.

Myelin was isolated from C57BL/6J mice. Animals were transcardially perfused with cold HBSS, and whole brains were manually homogenized in RPMI. Samples were applied to a Percoll gradient and, after a 30-min spin at 500g, the top layer was collected. Myelin was washed twice with water, and 3.1 µg cm$^{-2}$ was used for experiments.

ANs were generated using either SH-SY5Y cells or iNeurons generated as previously described[44], using neurogenin-2 (NGN2) for 14 days. Briefly, stem cells were grown in StemFlex (Stem Cell Technologies) and grown on plates coated with Matrigel (Corning) plates at 37 °C and 5% $CO_2$. Cells were infected with TetO-Ngn2-Puro and with rtTA lentiviral particles in StemFlex medium with 1 µM ROCK inhibitor Y-27632 for 24 h; they were then passaged, and differentiation was started when cells reached 70–80% confluence. For the first 2 days, cells were grown in N2 media (DMEM/F12 (Life Technologies, 11320-033), N2 supplement

(0.5% v/v, Gibco), 1× GlutaMAX (Gibco), 0.1 mM nonessential amino acids (Gibco), 0.5% glucose, doxycycline hyclate (2 μg $ml^{-1}$)). On day 0, N2 media was supplemented with 0.1 μM LDN, 5 μM XAV and 10 μM SB; and on day 2, the cells were fed with N2 media with no supplement. On day 3, cells were replated to 100,000 cells per well in a six-well plate and transferred to neurobasal media (Neurobasal (Gibco), 1× B27 (Gibco), 1× GlutaMAX (Gibco), 0.1 mM nonessential amino acid (Gibco), 0.5% glucose, doxycycline hyclate (2 μg $ml^{-1}$)) supplemented with 10 ng $ml^{-1}$ CNTF, 10 ng $ml^{-1}$ BDNF and 10 ng $ml^{-1}$ GDNF. Cells were fed every other day until day 14. Mature cells were submitted to ultraviolet radiation (500 J $m^{-2}$ using an ultraviolet cross-linker), gently collected 24 h later, washed in PBS and counted. On day 40, 35,000 ANs per $cm^2$ were added to the iMGL for each experiment.

For conjugation to pHrodo (Red SE or Green, STP Ester; Thermo Fisher Scientific), labeling was done according to the manufacturer's protocol. Briefly, substrates were incubated with 1 ml pHrodo per 1 mg substrate for 2 h at room temperature, protected from light, washed ten times using PBS and frozen in 5% dimethyl sulfoxide. Only synaptosomes, myelin or ANs were conjugated.

Amyloid fibrils were prepared from fluorescently labeled Aβ (Beta-Amyloid (1-42) HiLyte Fluo, AS-60479-01, Anaspec). Aβ peptides corresponding to amino acids 1–42 were dissolved in sterile water to 100 μg $ml^{-1}$ concentration, vortexed thoroughly and shaken in a 37 °C incubator for 4 h. Then, the Aβ peptides were incubated for 5 days at 37 °C to fibrilize. Amyloid fibrils were then mixed and delivered to cells at 5 μg $ml^{-1}$. Fresh amyloid fibrils were created for each experiment.

Phagocytosis of substrates was tested by flow cytometry. Although we observed variability in phagocytosis for amyloid fibrils among replicates, all other substrates were readily and consistently phagocytosed (Extended Data Fig. 3b).

### Sample preparation for scRNA-seq

For single-cell sequencing experiments, iMGLs at day 40 were treated with myelin, synaptosomes, ANs or amyloid fibrils for 24 h (see above). Cells were washed with warm media to remove floating cells and detached in PBS on ice. Cells were then centrifuged and resuspended in cold PBS at 1,000 cells per μl.

For the single-cell sequencing experiment comparing control versus lentivirus-transduced cells, iMGLs were transduced with a non-targeting (NT) guide RNA–Cas9–mCherry lentivirus (see below for more details on lentiviral vector, production and exposure) on day 35. Cells were exposed to the virus overnight, and 100% of the medium was changed the next morning. Cells were then cultured for 7 more days and fed on a regular schedule. On day 7, mCherry expression was confirmed, and all media were removed. Cells were washed with warm media to remove floating cells and detached in PBS on ice. Cells were then centrifuged and resuspended in PBS.

After resuspension and counting, cells were loaded into the 10x Chromium V3 system (10x Genomics). Reverse transcription and library generation were performed according to the manufacturer's protocol. Owing to a counting error, the number of cells loaded for replicate 1 of the H1 scRNA-seq dataset was approximately threefold lower than that for replicate 2, necessitating a data integration strategy for batch correction (see below). Sequencing was performed on a NovaSeq S2 (Illumina).

### Quantitative real-time PCR

iMGLs on day 40 were treated for 24 h and collected using RLT buffer. RNA extraction was done using an RNease Plus mini kit (Qiagen) according to the manufacturer's protocol. For quantitative rtPCR, a TaqMan RNA-to-Ct 1-step kit (Thermo Fisher Scientific) was used according to the manufacturer's protocol using the following TaqMan probes (Thermo Fisher Scientific): GAPDH (HS02786624_G1) and GPNMB (HS01095669_M1). Quantification was done using the $2^{-\Delta\Delta CT}$ method[52]. Each biological sample was measured in triplicate, and the average for each biological sample is shown in the corresponding graph. Statistics were analyzed using Prism9 software, and graphs show individual cells with mean and s.d. Two-tailed *t*-tests were used to determine significance.

### Immunocytochemistry, imaging and quantification

iMGLs plated on Matrigel-coated coverslips were fixed in 4% paraformaldehyde (PFA) for 15 min, followed by permeabilization (0.2% Triton-X in PBS) for 10 min and blocking (5% normal donkey serum in 0.025% Triton-X/PBS) for 30 min. iMGLs were incubated with primary antibodies overnight (concentrations below) and washed three times for 10 mins with 0.025% Triton-X/PBS. Then, cells were incubated with a secondary antibody (Thermo Fisher Scientific) for 60 min at room temperature. Following three washes for 10 min each with 0.025% Triton-X/PBS, coverslips were mounted on slides using ProLong Gold antifade (Thermo Fisher Scientific).

The primary antibodies used and their concentrations were as follows: TREM2 (R&D systems, AF1828-SP) 1/100, GPNMB (Cell Signaling, E1YZJ) 1/500, APOE clone 6B9 (Helmholtz Antibody Collection) (1/100). More information on the APOE antibody can be found at https://www.alzforum.org/alzantibodies/apoe-clone-6b9. The secondary antibodies used and their concentrations were: donkey anti-rabbit 488 (Thermo Fisher Scientific) 1/500, goat anti-rabbit 594 (Thermo Fisher Scientific) 1/500.

Imaging was performed on an Andor CSU-X spinning disk confocal system coupled to a Nikon Eclipse Ti microscope equipped with an Andor iKon-M camera. Images were acquired using a ×60 oil objective (Nikon). All images shown are representative images taken from at least three independent experiments. Statistics were analyzed using Prism9 software, and graphs show individual cells with mean and s.d. Two-tailed *t*-tests were used to determine significance.

### In situ hybridization RNAscope and quantification

RNAscope (Advanced Cell Diagnostics) was carried out according to the manufacturer's protocol for cultured adherent cells on coverslips. Briefly, iMGLs plated on Matrigel-coated coverslips were fixed for 30 min using 4% PFA and dehydrated (5 min 50% EtOH, 5 min 70% EtOH, 2× 5 min 100% EtOH) and kept at −20 °C until use. When ready for treatment, cells were subjected to the RNAscope Multiplex Fluorescent V2 Assay (Advanced Cell Diagnostics), and probes against human APOE (catalog no. 433091), ABCA1 (catalog no. 432291) and C1QA (catalog no. 485451-C2) were used at the recommended concentrations. TSA Plus fluorophores (Perkin Elmer) were used at 1:1500 concentration.

Imaging was performed on an Andor CSU-X spinning disk confocal system coupled to a Nikon Eclipse Ti microscope equipped with an Andor iKon-M camera. Images were acquired using a ×60 oil objective (Nikon). C1QA was used to identify regions of interest to identify individual cells. As microglia change shape and can become ameboid in response to stimuli, quantification of intensity could not be normalized to the cell's area; therefore, only the fluorescence intensity of the target probe was measured using Fiji[53]. Statistics were analyzed using Prism9 software, and graphs show individual cells with mean and s.d. Two-tailed *t*-tests were used to determine significance.

### Production of lentivirus and Vpx VLPs

Lentivirus and Vpx VLPs were produced by transfecting HEK293T cells using TransIT-LT1 reagent (Mirus). For lentivirus, the following plasmids were transfected (amounts per well for a six-well plate): 1.6 μg mCherry-expressing and Cas9-expressing lentiviral genome plasmid pXPR_BRD044 (obtained from the Broad Institute Genome Perturbation Platform), 0.4 μg pCMV-VSV-G (Addgene no. 8454) and 1 μg psPAX2 (Addgene no. 12260). For Vpx VLPs, 0.4 μg pCMV-VSV-G and 2.6 μg pSIV3 Vpx plasmids were transfected. We achieved identical results with both an in-house Vpx plasmid (pSIV3) and pSIV-D3psi/delta env/delta Vif/delta Vpr[54] (Addgene no. 132928). Media were

changed to OptiMEM 18 h after transfection (2 ml per well in a six-well plate). Two days after transfection, viral supernatant was harvested and centrifuged at 500*g* for 10 min to remove any cells in the supernatant. Clarified supernatant was concentrated 10× by incubation with Lenti-X Concentrator (Takara) overnight (owing to the size of the viral genome, the virus had to be concentrated to achieve sufficient titer). The mixture was spun down the next day per the manufacturer's instructions and resuspended in base iMGL media. Vpx VLPs were not concentrated. Lentivirus was titered using Lenti-X p24 Rapid Titer Kit (Takara). Both lentivirus and Vpx VLPs were flash frozen, stored at −80 °C and thawed on ice before use.

### Lentivirus transduction for FACS, single-cell sequencing and quantitative rtPCR time-course study of the effects of lentiviral transduction of iMGLs

For FACS, iMGLs were seeded in a 96-well plate. To transduce iMGLs, we added 53 ng of p24 per 10,000 cells of lentivirus and 10 µl per 10,000 cells of Vpx. Lentivirus and Vpx VLPs were added on day 35. The next day, all media were removed and replaced with fresh media, which were then changed every 2 days. On day 42, cells were detached as described above and FACS was performed to quantify mCherry expression from lentivirus.

For single-cell sequencing, 200,000 iMGLs were seeded in a six-well plate, where one well corresponded to one replicate for single-cell sequencing, and treated with lentivirus and Vpx VLPs as described above. Controls were treated the same as lentivirus in terms of media changes. Seven days after virus addition, cells were detached with cold PBS on ice for 10 min and spun at 300*g* for 5 min. Cells were then resuspended in a small volume of PBS and counted. Approximately 20,000 cells were used as input for each single-cell sequencing replicate.

For the time-course study of expression of interferon-stimulated genes after exposure to lentivirus, iMGLs were seeded into 24-well plates. On day 35, cells were cotransduced with lentivirus and Vpx VLPs. RNA was harvested on days 37, 42 or 42. To harvest RNA, iMGL media were removed and RLT Plus buffer from an RNeasy kit (Qiagen) was added directly to the wells. RNA was purified following the manufacturer's instructions. For quantitative rtPCR, a TaqMan RNA-to-Ct 1-step kit (Thermo Fisher Scientific) was used according to the manufacturer's protocol using the following TaqMan probes (Thermo Fisher Scientific): GAPDH (HS02786624_G1), IFI6 (Hs00242571_m1), IFITM3 (Hs03057129_s1) and ISG15 (Hs01921425_s1). Data were analyzed using CFX Maestro software (BioRad). For all quantitative rtPCR experiments, biological replicates are shown. Error bars represent s.d.

### ATAC-seq library preparation and sequencing

ATAC-seq libraries were prepared essentially as described[55], with the following modifications. For each sample, 10,000 viable frozen cells were used per library construction. We optimized the number of PCR cycles by stopping PCR after five cycles, taking an aliquot from the partially amplified library, and performing quantitative PCR for 25 cycles. We visually examined the quantitative PCR amplification profile to determine the number of cycles required to reach one-third of the plateau and extended the original PCR by this number of additional cycles. Libraries were sequenced on an Illumina NextSeq platform using a 75-cycle NextSeq 500 high-output V2 kit (read 1: 36 cycles; index 1: eight cycles; index 2: eight cycles; read 2: 36 cycles).

### MITF-overexpression lentivirus

For MITF overexpression, lentiviral vectors contained either MITF (NM_198159.3) or mCherry under the EF1α promoter[42]. Lentivirus was produced as described above. At day 35, iMGLs were infected with MITF-expressing or mCherry-expressing lentivirus, and the media were changed the following day. At day 42, iMGLs were treated for functional or transcriptomic characterization.

### Bulk RNA-seq library preparation and sequencing

iMGLs were collected using RLT buffer (Qiagen), and RNA extraction was done using an RNeasy Plus mini kit (Qiagen) according to the manufacturer's protocol. Quantification was performed and RNA integrity was assessed using an RNA 600 Pico chip on a Bioanalyzer (Agilent). RNA (5 ng) was used as input for library constructions. Libraries were constructed using a NEBNext Poly(A) mRNA magnetic isolation module (NEB E7490) in combination with NEBNext Ultra RNA library prep for Illumina (NEB E7530) following the manufacturer's instructions. Libraries were quantified using an Agilent high-sensitivity DNA chip and sequenced on NextSeq (Illumina). Samples were sequenced in four flow chambers, and reads from each chamber were merged to generate a final fastq final for each sample.

### scRNA-seq preprocessing, quality control and general analysis

Cell Ranger (v.3.1, 10x Genomics) was used for demultiplexing, barcode preprocessing, generation of fastq files, alignment (to GRCh38-2020-A/GENCODE v32/Ensembl 98) and counting of unique molecular identifiers (UMIs). Owing to differences in sequencing depth between replicates (see above), we applied Cell Ranger's aggr function to all samples to merge and normalize by sequencing depth.

All downstream analyses (see below for specific analyses) were performed in R using the LIGER (v.0.5) and Seurat (v.2.3.4 or v.3.2.1) packages[30,56]. Lentivirus-exposed iMGLs were excluded from this analysis, and analysis of this experiment is described below. All data and metadata (replicate and condition) were incorporated into a single object, and the percentage of mitochondrial RNA per cell was calculated. We only considered cells with the following characteristics: (1) number of genes: 2,000–7,000; (3) number of UMIs: 500–6,000; (3) percentage mitochondrial RNA 0–0.2%. This yielded 57,837 single iMGL cells.

Cells of interest were subsetted and converted to a LIGER object for integration. To integrate the two replicates, we considered the replicate (1 or 2) as the LIGER dataset variable and applied the LIGER pipeline for dataset integration ($k$ = 20, lambda = 5). Clustering was performed using the louvainCluster function (clustering resolution = 0.7), followed by merging of clusters with fewer than 45 unique differentially expressed genes. Clusters were merged based on hierarchical clustering of iMGL clusters by differential gene expression. A small population of cells activated by manual handling (ExAM signature, see below for score calculation, 1116 cells) were removed from the final dataset. In addition, another small set of cells of unknown identity, which had low microglial identity scores, were removed (267 cells). These two filtration steps removed a total of 1383 (2.4%) of the 57,837 cells, resulting in a final dataset of 56,454 (Fig. 1). A UMAP embedding was calculated using LIGER's runUMAP function. The alignment was evaluated with the CalcAlignment and CalcAgreement LIGER functions. All gene expression, metagene and violin plots were produced using LIGER functions and modified using ggplot2. Cell cycle scoring was performed using the CellCycleScoring function in Seurat.

### Module score for microglia identity, maturity and artifactual activation of microglia (ExAMs)

The module scores for testing for microglial identity and maturity were created using published gene lists[15,21]. For human microglial identity[21], the following genes were used: P2RY12, *GPR34*, *C1QA*, *CX3CR1*, *CABLES1*, *BHLHE41*, *TREM2*, *OLFML3*, *PROS1*, *APOE*, *SLCO2B1*, *SLC7A8*, *PPARD* and *CRYBB1*. For microglial maturity, the top 30 enriched genes identified in mature human microglia[15] were used: *SPP1*, *CD74*, *ACTB*, *C3*, *FTL*, *FOS*, *CSF1R*, *B2M*, *C1QC*, *C1QB*, *PSAP*, *A2M*, *ITM2B*, *LAPTM5*, *CTSB*, P2RY12, *C1QA*, *SLCO2B1*, *RGS1*, *APOE*, *CCL4L2*, *RNASET2*, *NEAT1*, *CX3CR1*, *DUSP1*, *SAT1*, *ZFP36*, *CD81*, *HLA-B* and *HLA-DRA*. The module score for testing for artifactual gene expression was created using genes previously identified as being upregulated during single-cell isolation and cell handling[22]. These genes were: *RGS1*, *HIST2H2AA1*, *HIST1H4I*, *NFKBIZ*, *KLF2*, *JUNB*, *DUSP1*, *CCL3*, *HSPA1A*, *HSP90AA1*, *FOS*, *HSPA1B*, *JUN*, *JUND*,

*NFKBID*, *GEM*, *CCL4*, *IER5*, *TXNIP*, *HIST1H2BC*, *ZFP36*, *HIST1H1C*, *EGR1*, *ATF3* and *RHOB*. In all cases, the module score was implemented in Seurat using the AddModuleScore function with a control size of 25.

### Module score for cluster identity

The module score for cluster identity was calculated using only positively enriched genes from a published xenotransplanted iMGL single-cell study[26]. The module score was implemented in Seurat using the AddModuleScore function with a control size of 100. Larger and smaller control sizes were tested with no observed impact on results (data not shown).

### Single-cell differential expression analysis and UpSet plots

To identify differentially expressed genes per LIGER cluster, we used the MAST package (v.1.12.0)[57] and included the number of UMIs and number of genes as covariates. Only genes expressed in more than 10% of cells per cluster were considered. Cell gene expression was then scaled, centered and averaged per cluster before heatmap generation. Heatmaps were generated using the ComplexHeatmap R package. To calculating the number of unique differentially expressed genes, UpSet plots were created for each state using all differentially expressed genes with the UpSetR R package.

### Monocle trajectory analysis

For trajectory analysis of AN-exposed iMGLs, normalized gene expression data were aligned by replicate using Batchelor[58], which was chosen owing to compatibility with the Monocle3 pipeline. Dataset integration across replicates with Batchelor gave similar results to those obtained with LIGER and Seurat (data not shown). The preprocessing, dimensionality reduction and pseudotime ordering of cells were performed using Monocle3 (ref. 28), and cluster labels from the initial LIGER analysis (see above) were used for analysis and plotting.

### Cluster proportions, differential abundance test and fold change analysis

For plotting cell proportions, datasets were downsampled to the same size randomly using the groupdata2 or LIGER R packages. The percentage contribution of each condition to a cluster by replicate was then calculated and averaged. To determine statistical significance of cluster abundances across conditions, we used counts (not proportions) and implemented Dirichlet regression, a multivariate test that accounts for overall composition per sample, using the DirichletReg R package (https://github.com/maiermarco/DirichletReg), v.0.7, and a previously published workflow[59]. Fold change values were calculated by replicate, relative to the control condition, and a $\log_2$ transformation for final log fold-change plots was reported.

### GSEA of iMGL clusters and other datasets

GSEA was run using the fgsea package[60] (v.1.12.0). Genes for each iMGL cluster were ordered by log(fold change). Pathways were defined as positively differentially enriched genes identified by differential expression analysis in other studies: Hasselmann et al.[26], Gerrits et al.[24] and the meta-analysis described in Gazestani et al.[29]. Data were plotted using the ComplexHeatmap R package[61].

### LIGER integration of iMGLs with different datasets

For large dataset integrations (that is, those with large numbers of cells, individuals or both) we used LIGER because of both its scalability and its performance on large datasets. For analysis of the human cortical biopsy microglia, the final microglia object, cluster labels and differentially expressed genes from the Gazestani et al. object were used. Nonmicroglial cells, including myeloid and peripheral macrophages, were excluded before analysis. For analysis of the mouse DAM dataset, the data were reprocessed and analyzed using Seurat (v.3.2.1), and cluster identities (in particular, the identification of DAM microglia) were determined based on differential gene expression and comparison with the source paper (for instance, upregulation of *Lpl*, *Clec7a*, *Trem2*, *Itgax*, *Cd9* and *Axl*). For xMG analysis, the cluster assignments from the original paper were used.

All cross-context (xenotransplanted iMGLs) and cross-species (human in vivo and mouse in vivo) dataset integrations were performed using the LIGER package[30]. Post-QC (quality control) iMGL count data were used for all alignments. For each alignment, the optimal *k* and lambda values were identified using the appropriate LIGER functions. For each alignment, the data, parameters and dataset variable arrangement were as follows.

| Data | Reference | Parameters | Dataset variable |
|---|---|---|---|
| Human cortical brain biopsy microglia | Gazestani et al.[29] | *k*=20, lambda=10, clustering resolution=0.3 | iMGL replicates (untreated and all CNS substrates), individual patients (51 datasets) |
| Mouse in vivo: DAMs | Keren-Shaul et al.[6] | *k*=20, lambda=5 | iMGL replicates, mouse microglia as one dataset |
| Xenotransplanted iMGLs: xMGs | Hasselmann et al.[26] | *k*=20, lambda=5 | iMGL replicates, xMG in WT, xMG in 5xFAD |

Note: LIGER alignment of H1 iMGLs with iPSC-derived iMGLs (Fig. 5) is described below. For the human cortical brain biopsy dataset integration, two small clusters of doublets (with marker genes for other brain cell types) were identified and removed before downstream analysis.

These alignments were evaluated with the CalcAlignment and CalcAgreement LIGER functions. All gene expression plots and violin plots were produced using LIGER functions and modified using ggplot2.

### Hypergeometric test comparing human and iMGL DAM gene expression

Before performing the hypergeometric test, both iMGL and in vivo human DAM genes were identified. DAM genes in iMGLs were determined by taking the union of differentially expressed genes for iMGL_02 and iMGL_8 (see Single-cell differential expression analysis and UpSet plots above). We used the differentially expressed genes from the author's analysis of the cortical biopsy dataset (Gazestani et al.[29]). Only positively differentially expressed genes were used. The hypergeometric test was run on intersecting genes using the dhyper function from the Stats R package (v.3.6.3). The background number of genes for each dataset was calculated based on genes expressed in more than 1% of cells.

### Hypergeometric test comparing mouse and iMGL DAM gene expression

Hypergeometric test was performed as above, with the following exceptions. Mouse DAM genes were identified from a published dataset[6], and mouse microglia background genes were determined from the same dataset as genes expressed in more than 1% of cells. Only positively differentially expressed genes and those with a fold change greater than 0.2 were used (matching our fold change threshold). To compare the background and DAM genes across datasets, the babelgene R package (v.2.9) was used (https://github.com/igordot/babelgene). The hypergeometric test was run on intersecting genes using the dhyper function from the Stats R package (v.3.6.3).

### Cluster occupancy test

To test the statistical significance of the coclustering of iMGL_2, iMGL_8 and human DAMs (BB_GPNMB_LPL), a binomial test was used

to determine whether cells from each group were enriched in joint clusters 8 and 15 relative to other clusters. This test used raw numbers rather than percentages. The binom.test function from the Stats R package (v.3.6.3) was used for this analysis.

### SCENIC transcription factor analysis

The post-QC, normalized iMGL gene expression matrix was used as a starting point for analysis with SCENIC[36] v.1.1.2-2. Cells from the same LIGER cluster were randomly averaged into pseudocells of 20, an approach that has previously been shown to yield more robust results[62]. GENIE3 (v.1.4.3) was used to identity transcription factor–gene coexpression networks, followed by regulon analysis (runSCENIC_2_createRegulons) and regulon scoring per cell (runSCENIC_3_scoreCells) to create a regulon expression matrix for all cells. For the final list of significant regulons, a differential expression test was performed with the presto R package (v.1, wilcoxauc function) (https://github.com/immunogenomics/presto), and results were filtered by adjusted $P < 0.01$ and area under the curve >0.6. Regulons were scaled and centered for plotting using the pheatmap package (https://github.com/raivokolde/pheatmap).

### Lentivirus-exposed versus control iMGLs

Sequencing data were processed with Cell Ranger 3.1.0 (10x Genomics). Fastq files were mapped to a modified reference genome (GRCh38-2020-A/GENCODE v32/Ensembl 98) containing the mCherry gene. Count data were obtained using Cell Ranger's count function. The count data were further analyzed using Seurat 3.2.2 (ref. 56). Count matrices from four conditions were merged, filtered (only cells with 2,000–7,000 genes, 500–60,000 reads and less than 20% mitochondrial RNA were kept) and normalized using standard log-normalization. Integration was performed using the FindIntegrationAnchors and IntegrateData functions. The CellCycleScoring function was used to analyze the number of cycling cells. To analyze the remaining gene expression changes in the dataset, cycling cells were removed (only cells in G1 phase were kept), the data were aggregated and differentially expressed genes were identified using the Deseq2 1.30.1 package[63] with default parameters (except for adjusted $P = 0.01$). Enriched gene ontology terms were annotated with the enrichR package using database GO_Biological_Process_2015 (ref. 64).

### Analysis of iPSC-derived iMGLs and dataset integration

For analysis of lines CW50118, CW500036 and CW70437, single-cell gene expression matrices were first filtered for low-quality cells. As above, we only considered cells with the following characteristics: (1) number of genes: 2,000–7,000; (2) number of UMIs: 500–6,000, (3) percentage mitochondrial RNA 0–0.2%. After this cell-filtration step, cells were merged with the post-QC CNS-substrate-exposed H1 iMGL scRNA-seq dataset above (Fig. 1). To integrate the two replicates, we considered the H1 iMGL replicate (1 or 2) and the iPSC stem cell line as the LIGER dataset variable and applied the LIGER pipeline for dataset integration ($k = 20$, lambda = 6). Clustering was performed using the louvainCluster function (clustering resolution = 0.3). A UMAP embedding was calculated using LIGER's runUMAP function. The alignment was evaluated with the CalcAlignment LIGER function. Two small clusters were enriched for neuronal genes and specific to the AN-treated conditions for all lines. These were considered to be doublets and removed before visualization and downstream analyses. In addition, a small population of cells enriched for the ExAM module score were removed. All gene expression, metagene and violin plots were produced using LIGER functions and modified using ggplot2.

### Module score iMGL state markers

To examine the similarity of H1-derived and iPSC-derived iMGL states, module scores for each H1 iMGL cluster (Fig. 1) were created. These module scores were calculated for each cell in the iPSC-derived iMGL dataset only using positive differentially expressed genes identified above for each cluster in the H1-derived iMGL dataset (see Single-cell differential expression analysis and UpSet plots). In all cases, the module score was implemented in Seurat using the AddModuleScore function with a control size of 100.

### Bulk RNA-seq analysis

Reads were aligned to the human genome (hg19) using Picard (https://broadinstitute.github.io/picard/). Read counts were obtained using Featurecounts[65], ComBat-seq was used to adjust for batch effects[66], genes with fewer than 50 read counts across samples were removed from analyses and DESeq2 was used for differential expression[63] with Wald test adjusted $P < 0.05$. For the hypergeometric test, the background number of genes for each dataset was calculated based on genes expressed in more than 1% of cells (13,193 genes).

### ATAC-seq data processing

Reads were aligned to the human reference genome (hg19) with the BWA aligner[67]. Duplicates were removed using Picard tools (MarkDuplicates) (https://broadinstitute.github.io/picard/), and peaks were called using HOMER 4.11.1 (ref. 35) in 'histone' mode. Diffbind[68] was used to identify differentially represented regions between conditions.

### Statistics and reproducibility

Each experiment was done on multiple independent differentiations and all conditions were included in each differentiation. Sample sizes were based on standards in the field, and experimental treatment groups were randomly assigned.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

All iMGL data have been deposited on Terra, including raw and Cell Ranger outputs of iMGL (H1 and CW50118, CW500036 and CW70437) scRNA-seq, fastq and bam files of iMGL untreated and treated with ANs for ATAC-seq, and fastq and bam files of MITF-overexpression and mCherry control bulk RNA sequencing. Summary level data are available at https://app.terra.bio/#workspaces/Stevenslab/public_iMGL-datasets. Raw data are available via managed access at DUOS (https://duos.org); ID: DUOS-000151. Any additional data are available from the corresponding authors.

## Code availability

Code is available from the corresponding authors.

## References


51. Lehrman, E. K. et al. CD47 protects synapses from excess microglia-mediated pruning during development. *Neuron* **100**, 120–134.e6 (2018).
52. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408 (2001).
53. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
54. Kim, K. et al. Cyclophilin A protects HIV-1 from restriction by human TRIM5α. *Nat. Microbiol.* **4**, 2044–2051 (2019).
55. Xu, H. et al. Transcriptional atlas of intestinal immune cells reveals thatneuropeptide α-CGRP modulates group 2 innate lymphoid cell responses. *Immunity* **51**, 696–708.e9 (2019).
56. Stuart, T. et al. Comprehensive integration of single cell data. *Cell* **177**, 1888–1902.e21 (2019).
57. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).

58. Haghverdi, L., Lun, 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.* **36**, 421–427 (2018).
59. Smillie, C. S. et al. Intra- and inter-cellular rewiring of the human colon during ulcerative colitis. *Cell* **178**, 714–730.e22 (2019).
60. Korotkevich, G. et al. Fast gene set enrichment analysis. Preprint at *bioRxiv* https://doi.org/10.1101/060012 (2021).
61. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
62. Suo, S. et al. Revealing the critical regulators of cell identity in the mouse cell atlas. *Cell Rep.* **25**, 1436–1445.e3 (2018).
63. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
64. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
65. 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).
66. Zhang, Y., Parmigiani, G. & Johnson, W. E. *ComBat-seq*: batch effect adjustment for RNA-seq count data. *NAR Genom. Bioinform.* **2**, lqaa078 (2020).
67. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
68. Ross-Innes, C. S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).

## Acknowledgements

We thank all members of the Stevens and Macosko laboratories for helpful discussion when preparing the manuscript. We also thank the members of the International Neuroimmune Consortium for their insightful conversations and support in this project. We thank C. Haass (German Center for Neurodegenerative Diseases (DZNE)), S. Lammich (DZNE) and all members of the Monoclonal Antibody Core Facility at Helmholtz Munich who helped to validate and share the APOE antibody. This work was supported in part by the Alzheimer's Association (ADSF-21-831114-C; B.S.), Stanley Center for Psychiatric Research (B.S.), Broad Institute (V2F; B.S.), Cure Alzheimer's Fund (B.S., M.B.-J. and C.K.G.), OneMind/Alzheimer's Foundation of America (B.S.) and Howard Hughes Medical Institute (B.S.). M.-J.D. is an Open Philanthropy Project Awardee of the Life Sciences Research Foundation. M.T. is a recipient of Fonds de recherche du Québec and Canadian Institutes of Health Research postdoctoral fellowships. T.K. is supported by the F30 (National Institutes of Health (NIH) National Institute on Aging) F30AG069446-01. V.L. is funded by Kuopio University Hospital VTR Fund, Academy of Finland (grant number 339767) and Sigrid Juselius Foundation. Isogenic TREM2-knockout iPSC lines were generated by the University of California, Irvine Alzheimer's Disease Research Center iPS cell core supported by National Institute of Health (US) P30-AG066519. Experiments using the GFP-expressing iPSC line AICS-0036 were made possible through the Allen Cell Collection, available from Coriell Institute for Medical Research.

## Author contributions

M.T., S.J., M.-J.D. and B.S. designed the study. M.T., S.J. and M.-J.D. performed experiments, with help from T.A., N.M.L., S.M., C.M.W. and C.B.E. M.-J.D., S.J. and M.T. analyzed the results. Assistance was provided by T.K., V.G., S.E.M., A.G. and E.Z.M. M.-J.D., M.T. and S.J. wrote the manuscript with input from all coauthors. J.J., F.L., F.Z., M.B.-J. and B.L. provided reagents. V.L., T.K. and E.Z.M. provided human tissue and snRNA-seq data. N.H., M.B.-J., B.L., K.E., C.B.E., B.E.B., C.K.G. and F.Z. advised on methodology. E.Z.M. and B.S. supervised the project.

## Competing interests

K.E. is cofounder of Q-State Biosciences, Quralis and Enclear Therapies, and is group vice president at BioMarin Pharmaceutical. M.B.-J. is a coinventor of patent application WO/2018/160496, related to the differentiation of pluripotent stem cells into microglia, and cofounder of NovoGlia Inc. B.S. serves on the SAB and is also a minor shareholder in Annexon Biosciences and Neumora Therapeutics.

## Additional information

**Extended data** is available for this paper at https://doi.org/10.1038/s41590-023-01558-2.

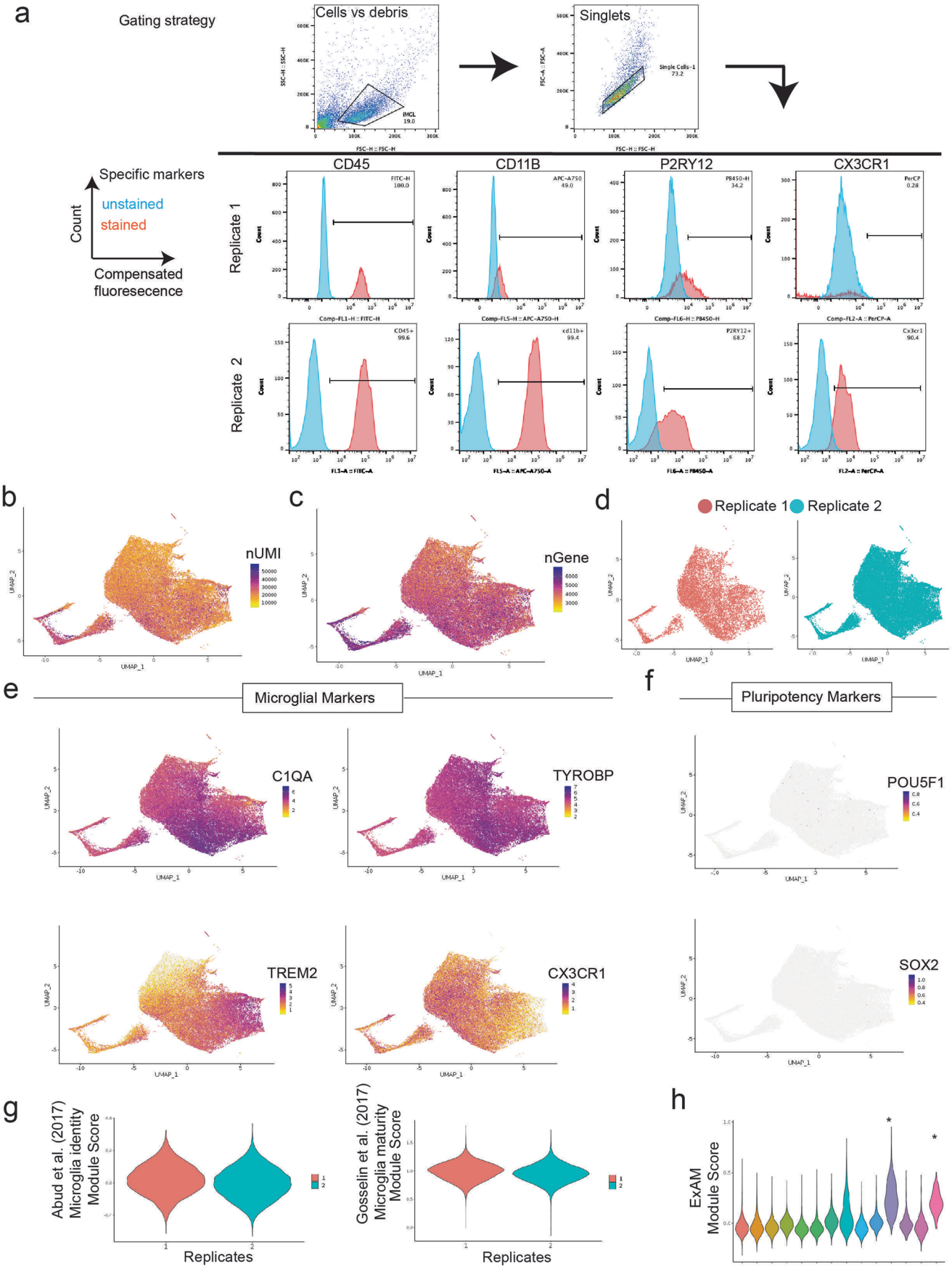
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41590-023-01558-2.

**Correspondence and requests for materials** should be addressed to Evan Z. Macosko or Beth Stevens.

**Peer review information** *Nature Immunology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Ioana Visan, in collaboration with the *Nature Immunology* team.

**Reprints and permissions information** is available at www.nature.com/reprints.

https://doi.org/10.1038/s41590-023-01558-2

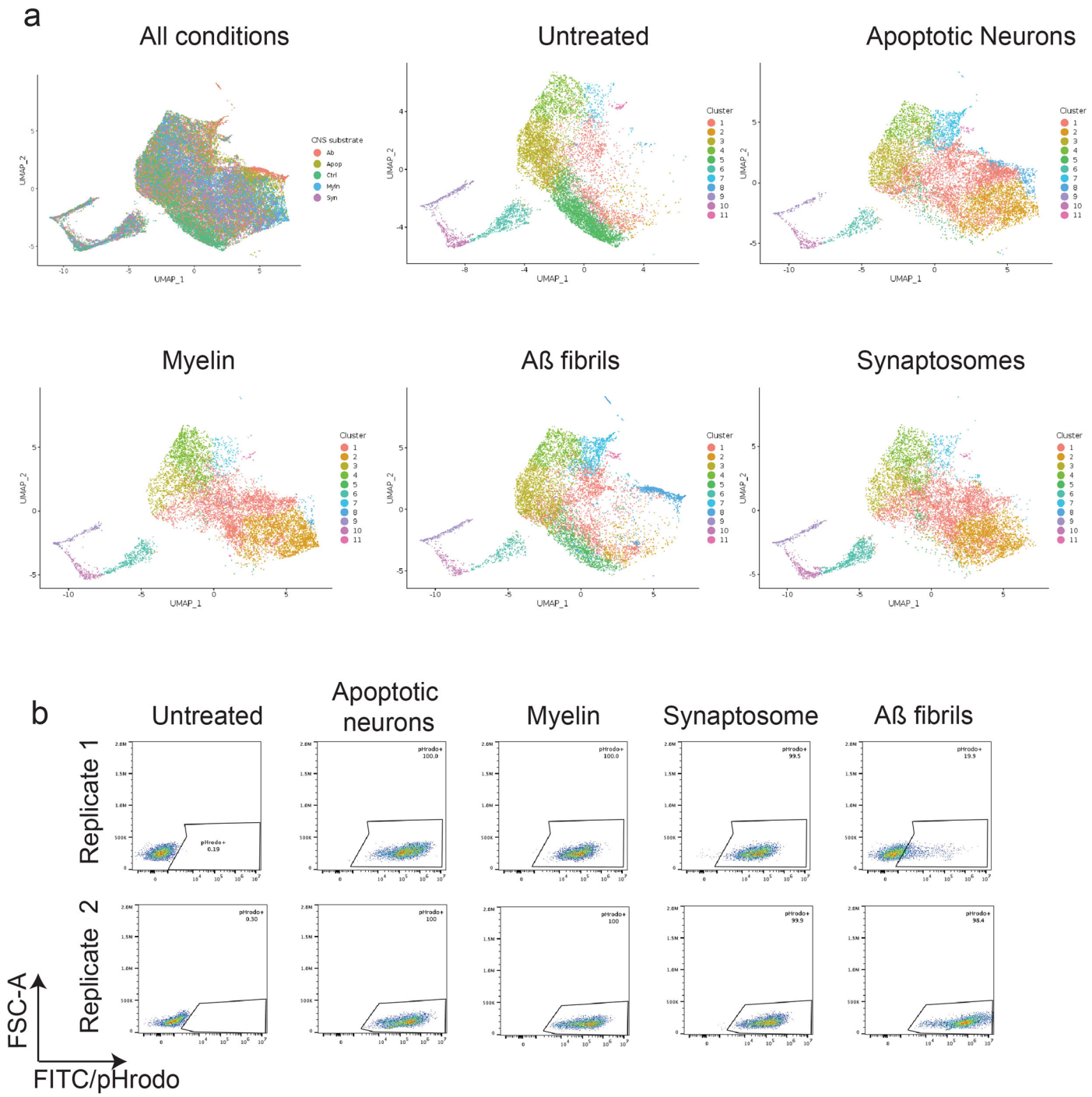



**Extended Data Fig. 1 | See next page for caption.**

https://doi.org/10.1038/s41590-023-01558-2

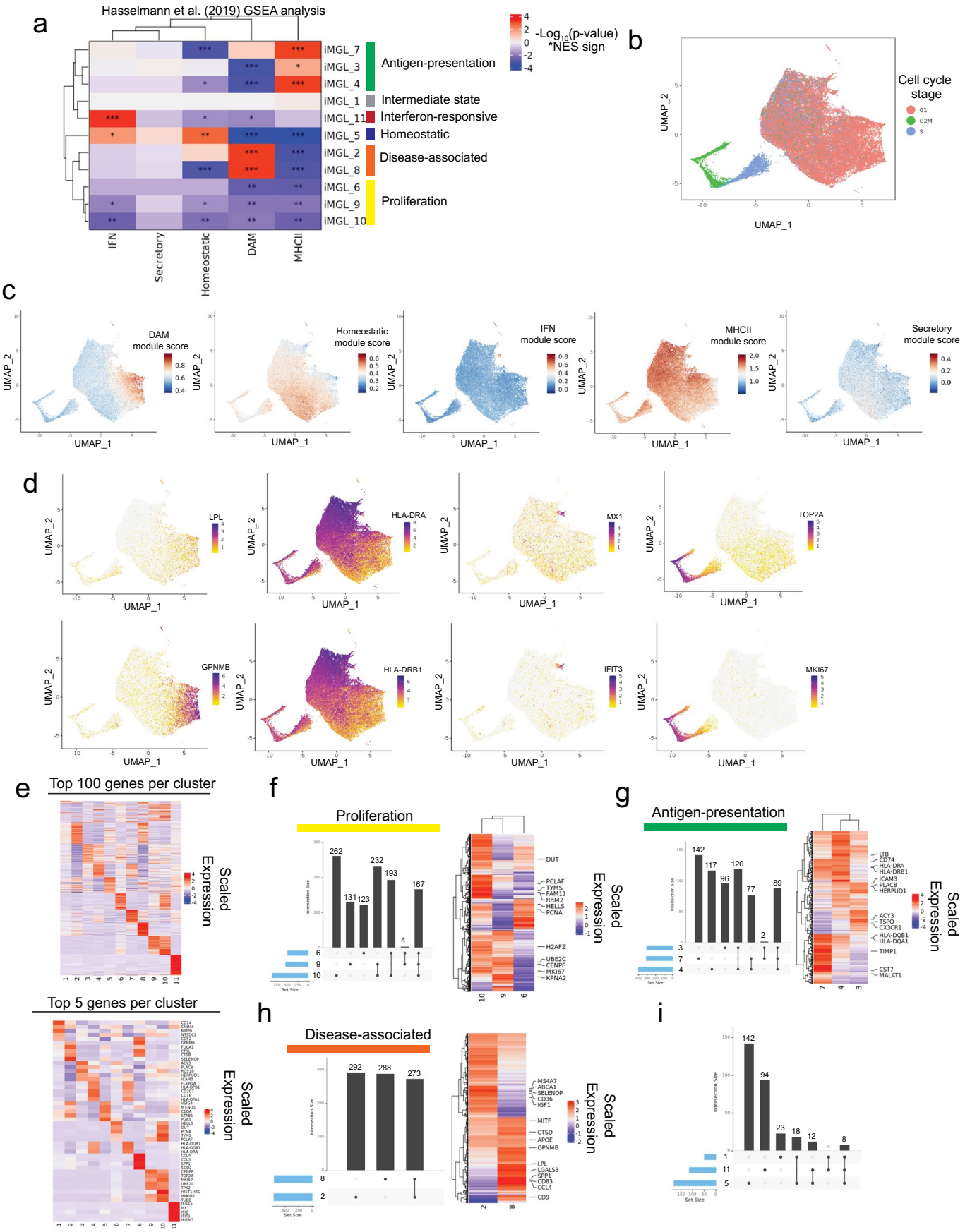
**Extended Data Fig. 1 | Single-cell RNAseq of iMGLs and quality control.** a) Flow cytometry gating strategy and plots illustrating expression of microglial protein markers from the two independent differentiations used for the scRNAseq dataset. Cells underwent either flow cytometry or single-cell RNAseq. b-c) UMAP projection (as in Fig. 1a) of quality control metrics for iMGL dataset highlighting the number of unique molecular identifiers (nUMI, b) and the number of genes (nGene, c). d) UMAP projection (as in Fig. 1a), cells colored by replicate. e) UMAP projection (as in Fig. 1a) showing expression of four key microglial markers. f) UMAP projection (as in Fig. 1a) showing expression of two stem cell markers. g) Module scores for microglia identity and maturity per cell, plotted by replicate h) Module scores for microgliaExAM signature[22] per cells plotted per iMGL cluster. Two high ExAM clusters were removed(*). Note n = 2 independent differentiations for B-F.

https://doi.org/10.1038/s41590-023-01558-2




**Extended Data Fig. 2 | Single-cell RNAseq data and phagocytosis of substrates for each condition. a)** UMAP projection of iMGL dataset for each condition (n = 2 independent differentiations per condition), cells are colored by conditions. b) Flow cytometry measurement of pHrodo-488/FITC for each CNS substrate exposed to the iMGL dataset. FSC=forward scatter. FITC/pHrodo=phagocytosis of substrate.

https://doi.org/10.1038/s41590-023-01558-2

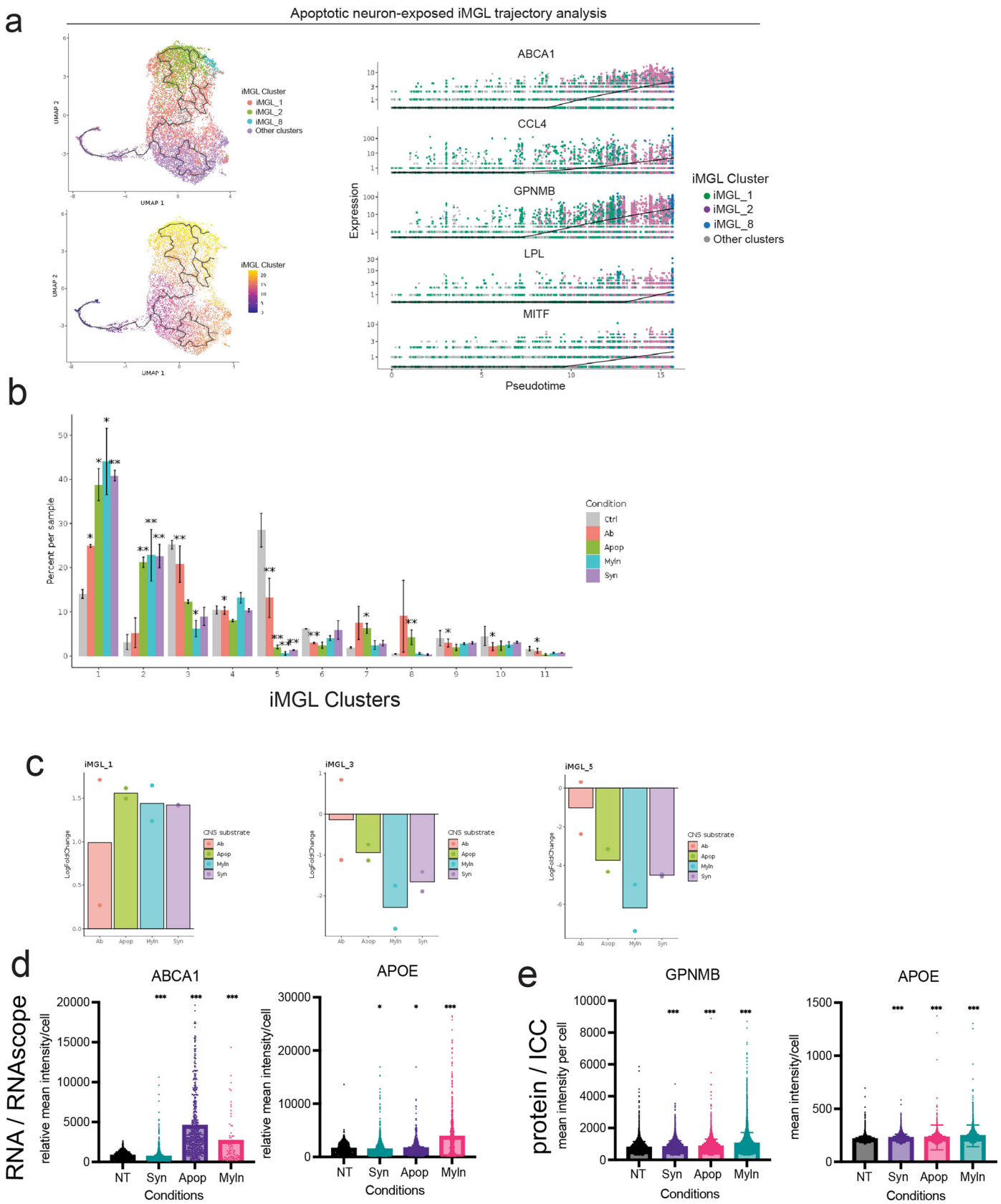



**Extended Data Fig. 3 | See next page for caption.**

https://doi.org/10.1038/s41590-023-01558-2

**Extended Data Fig. 3 | Annotation and gene expression signatures of iMGL datasets.** A) Heatmap illustrating the relative GSEA enrichment significance of positively enriched marker genes from each microglial cluster from xenotransplanted iMGLs *in vivo*[26], within all differentially expressed genes for each iMGL cluster. b) UMAP projection of iMGL dataset, cells colored by cell cycle phase (G1, G2M and S) c) UMAP projection of iMGL dataset, cells colored by module scores of transcriptional signature identified in xenotransplanted iMGLs *in vivo*[26]. d) UMAP projection of iMGL dataset, cells colored by expression of *LPL*, *GPNMB* (iMGL_2 and iMGL_8), *HLA-DRA*, *HLA-DRB1* (marker genes for iMGL_3, iMGL_4 and iMGL_7), *MX1*, *IFIT3* (marker genes for iMGL_11), *TOP2A* or *MKI67* (marker genes for iMGL_6, iMGL_9, iMGL_10). e) Heatmap of the top 100 (top) or top 5 (bottom) differentially expressed genes for each iMGL cluster. f-h) Upset plots (left) summarizing the number of unique and shared differentially expressed genes for each cluster within each microglial state identified and heatmaps (right) showing the expression of positively enriched genes for the constituent clusters of each state. i) Upset plots (as in g) for clusters iMGL_1, iMGL_5 and iMGL_11.

https://doi.org/10.1038/s41590-023-01558-2

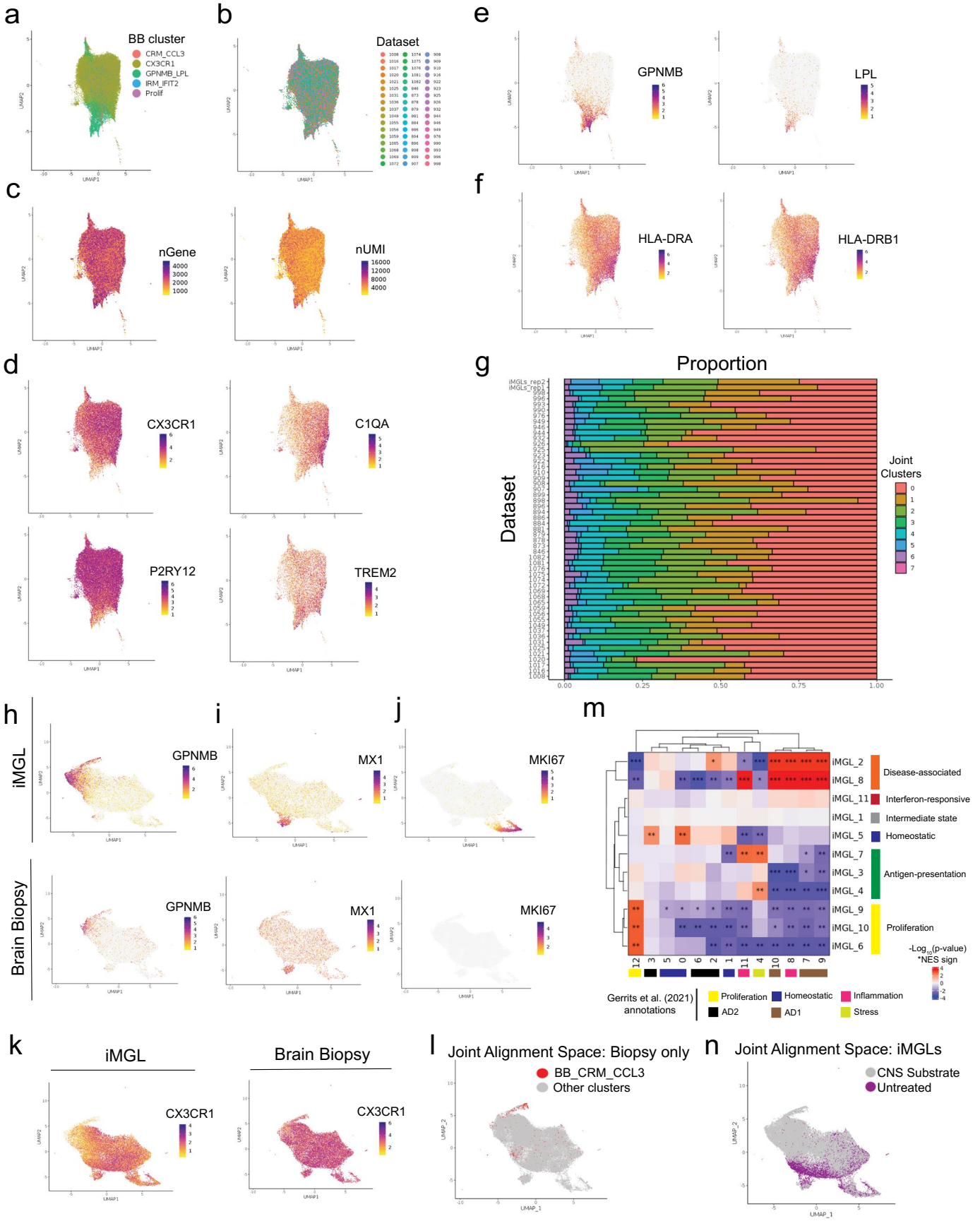



**Extended Data Fig. 4 | See next page for caption.**

https://doi.org/10.1038/s41590-023-01558-2

**Extended Data Fig. 4 | iMGL clusters by trajectory, substrate exposure and validation with in situ hybridization and ICC.** a) Single cell profiles of apoptotic neuron-exposed iMGLs (n = 2) in UMAP space (left) and ordered by pseudotime (right) for clusters iMGL_1, iMGL_2 and iMGL_8. b) Barchart of percentage composition of cells per condition per cluster. Statistical significance determined by Dirichlet regression. * = p-value < 0.05 and ** = p-value < 0.01. c) Fold change of iMGL_1,iMGL_2 and iMGL_4 relative to the untreated control condition d) Quantification of fluorescent *in situ* hybridization of *ABCA1* and *APOE* by mean intensity per cell. *ABCA1*: NT vs Syn p < 0.0001, NT vs Apop p < 0.0001, NT vs Myln p < 0.0001. *APOE*: NT vs Syn p = 0.076, NT vs Apop p = 0.01786, NT vs Myln p < 0.0001. e) Quantification immunocytochemistry of *GPNMB* and *APOE* by mean intensity per cell. *GPNMB*: NT vs Syn p < 0.001, NT vs Apop p < 0.0001, NT vs Myln p < 0.0001. *APOE*: NT vs Syn < 0.001, NT vs Apop p < 0.0001, NT vs Myln <0.0001. For d-e) >500 cells were counted by conditions across 4 biological replicates. NT= Not treated, Syn=synaptosomes, Myln=myelin debris, Ab=synthetic Aβ fibrils, Apop= apoptotic neurons.

https://doi.org/10.1038/s41590-023-01558-2

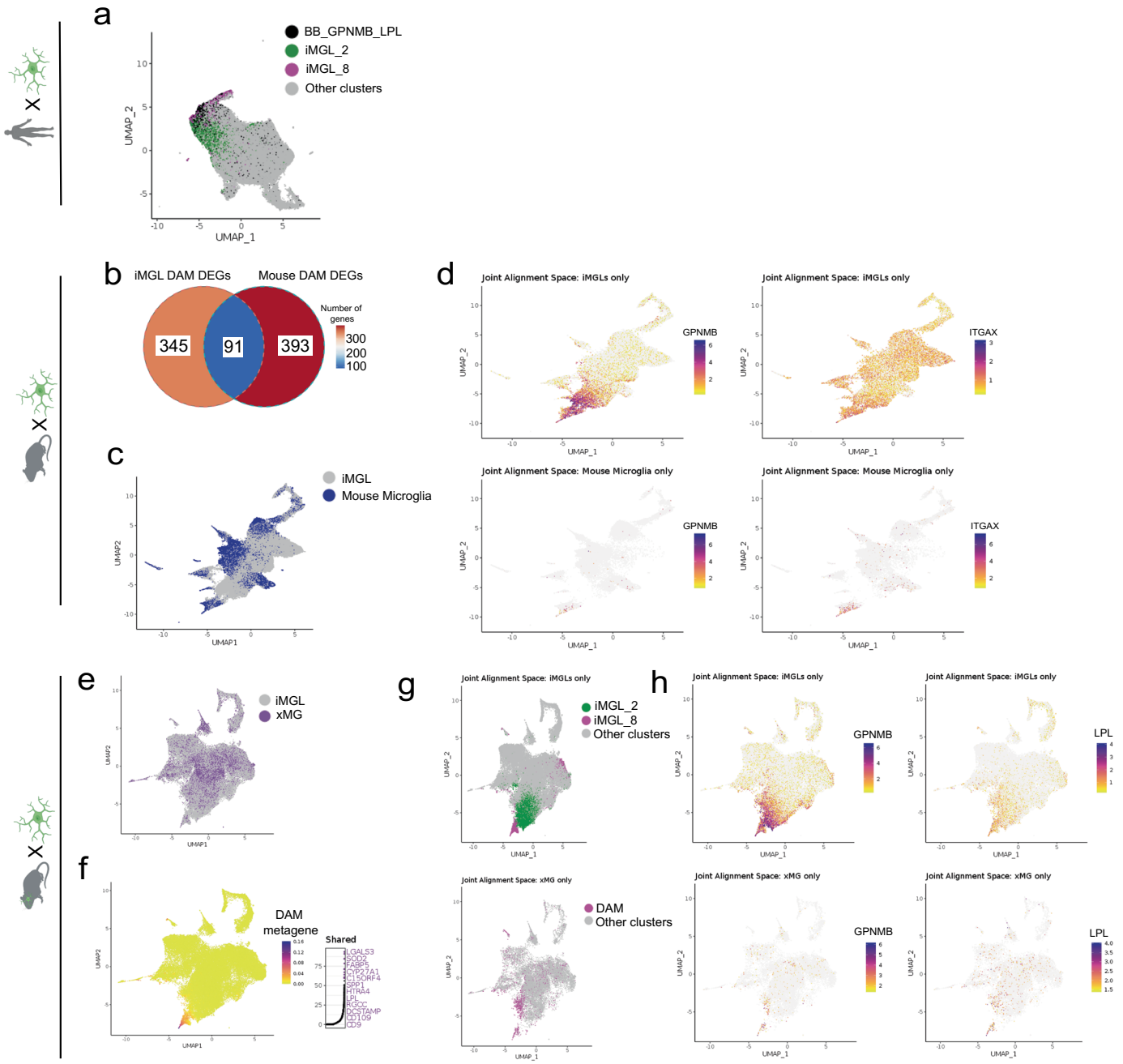



**Extended Data Fig. 5 | See next page for caption.**

https://doi.org/10.1038/s41590-023-01558-2

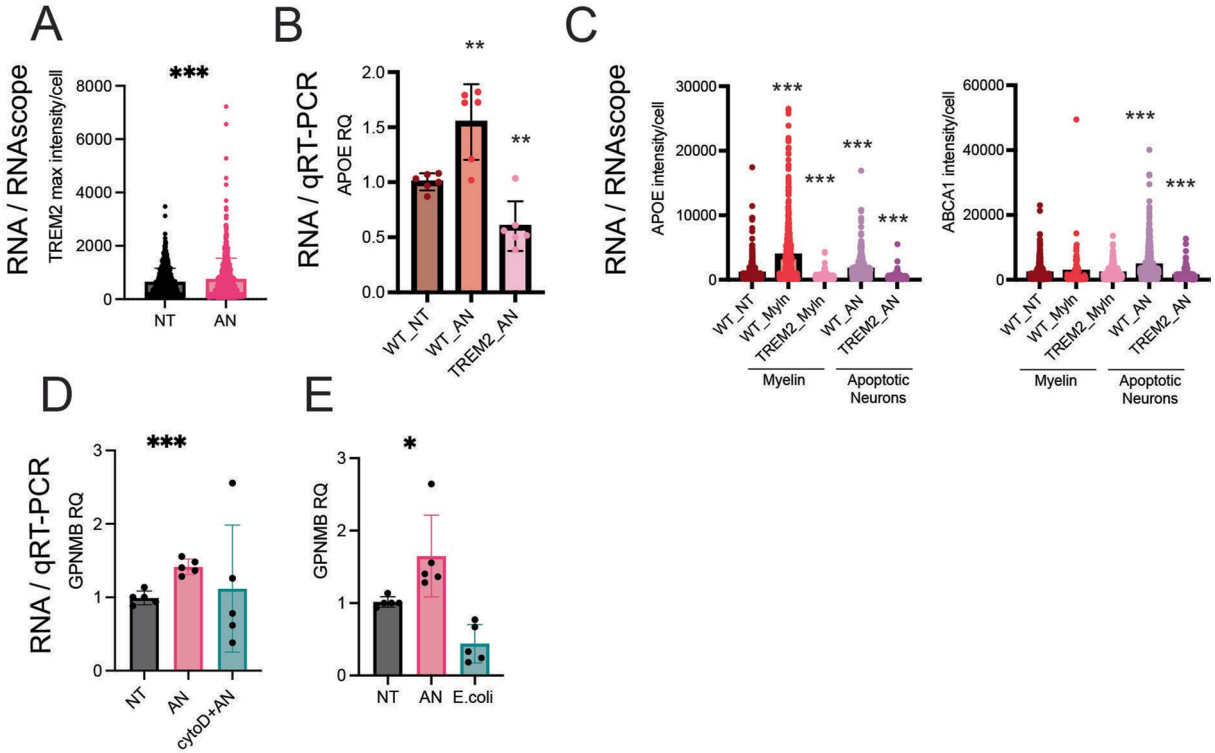
**Extended Data Fig. 5 | Summary of the human cortical biopsy microglial dataset and LIGER alignment with iMGL dataset.** a) UMAP projection of cluster annotations of human brain biopsy (BB dataset) single dataset from ref. 29. b) UMAP projection, cells colored by patient. c) UMAP projection (as in a) number of genes (nGene, left) and number of unique molecular identifiers (nUMI, right). d-f) UMAP projection (as in a), cells colored by expression of *CX3CR1*, *C1QA*, P2RY12, *TREM2* (d), *GPNMB* and *LPL* (e), *HLA-DRA* and *HLA-DRB1* (f). g) Proportion of cells across joint clusters for each dataset (iMGL or brain biopsy dataset). h-k) UMAP projection LIGER integration of iMGL or brain biopsy dataset, cells colored by datasets (iMGL top, brain biopsy bottom) for expression of *GPNMB* (h), *MX1* (i), *MKI67* (j), *CX3CR1* (k, iMGL left, brain biopsy right). l) UMAP projection (as in h), brain biopsy cells only with the BB_CRM_CCL3 state highlighted. m) Heatmap of the relative enrichment significance of positively enriched marker genes from each *in vivo* human microglial cluster from ref. 24, within differentially expressed genes of each iMGL cluster (GSEA analysis). n) UMAP projection (as in h) of cells from iMGL dataset only colored by substrates or untreated.

https://doi.org/10.1038/s41590-023-01558-2



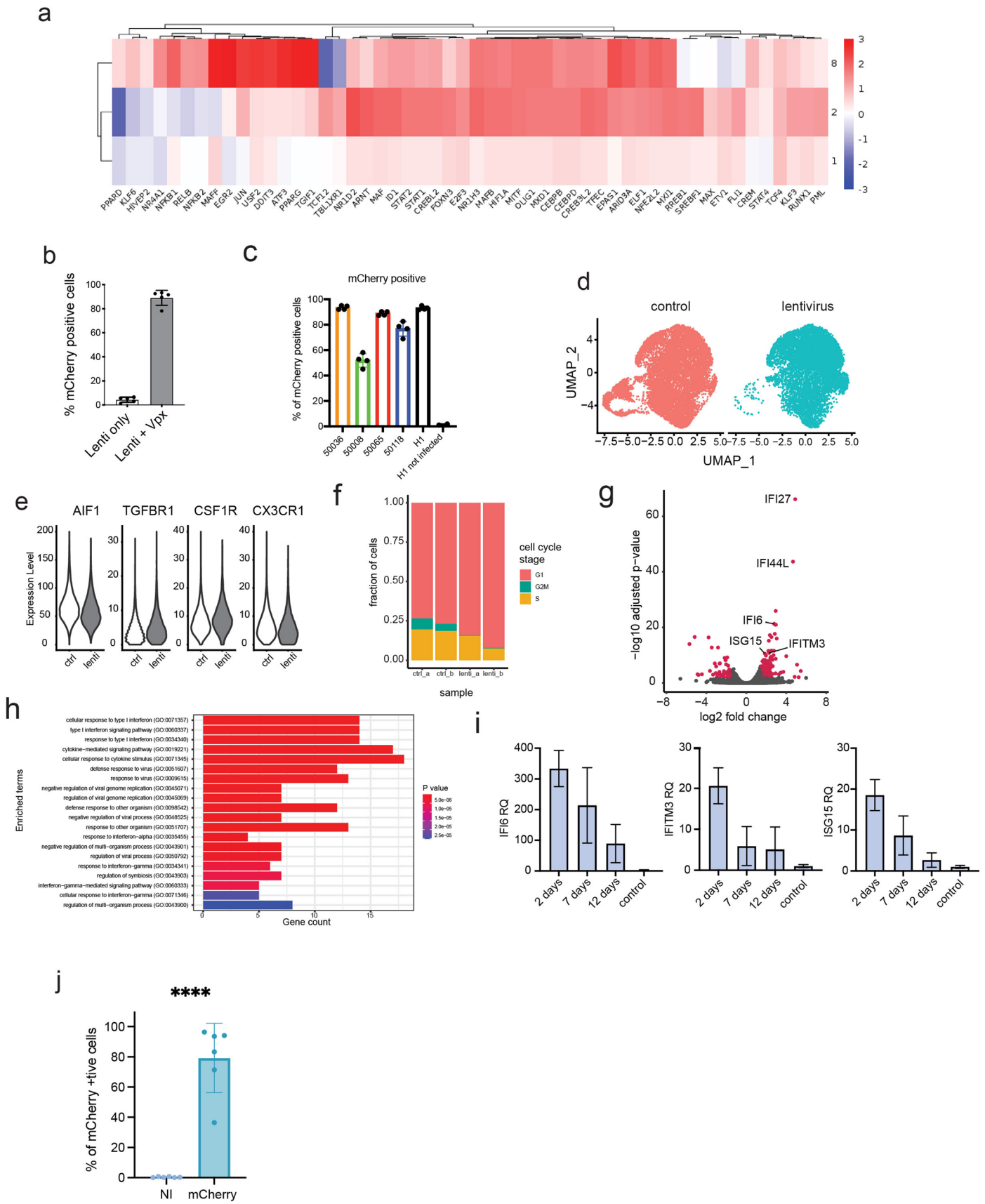

**Extended Data Fig. 6 | Dataset integration of iMGLs with mouse AD and xenotransplanted human microglia.** a) UMAP projection of LIGER integration of iMGL or brain biopsy dataset, cells colored by identity to cluster iMGL_2 (green), iMGL_8 (magenta) highlighted or BB_GPNMB_LPL cluster (black). b) Comparing overlap of iMGL DAM (iMGL_2 and iMGL_8) enriched genes with those enriched in murine DAMs in the 5xFAD genotype[6] (p value: 5.41e-19, hypergeometric test). c) UMAP projection of LIGER integration of iMGL datasets with *in vivo* mouse microglia (from wild-type and 5xFAD genotypes). Cells colored by dataset. d) UMAP projection (as in c) of iMGLs (top) or mouse microglia (bottom) cells colored for expression of *GPNMB* (left) and *ITGAX* (right). e) UMAP projection LIGER integration of iMGLs and xenotransplanted iMGLs. Cells colored by dataset. f) UMAP projection (as in e) colored by expression of shared factor DAM metagene. Right, top constituent genes of this shared factor. g) UMAP projection (as in e), of iMGLs (top) or xenograft microglia (bottom) colored by identity to iMGL_2, iMGL_8 or DAM. h) UMAP projection (as in e), of iMGLs (top) or xenograft microglia (bottom) cells colored for expression of *GPNMB* (left) and *LPL* (right).

https://doi.org/10.1038/s41590-023-01558-2




**Extended Data Fig. 7 | Dependence of iMGL DAM formation on TREM2 expression and substrate.** a) Quantification of relative intensity of TREM2 antibody stain determined by immunocytochemistry (p-value < 0.0001). b) rtPCR of APOE expression in WT and TREM2 KO iMGL (APOE WT_NT vs WT_AN p-value < 0.0035, APOE WT_AN vs TREM2_AN p-value < 0.0002). ABCA1 WT_NT vs WT_MYLN p-value = 0.074, c) APOE, ABCA1 expression level measured by RNAscope. (APOE WT_NT vs WT_MYLN p-value < 0.0001, APOE WT_MYLN vs TREM2_MYLN p-value < 0.0001, APOE WT_NT vs WT_AN p-value < 0.0001, APOE WT_AN vs TREM2_AN p-value < 0.0001, ABCA1 WT_NT vs WT_MYLN p-value = 0.0740, ABCA1 WT_MYLN vs TREM2_MYLN p-value = 0.0754, ABCA1 WT_NT vs WT_AN p-value < 0.0001, ABCA1 WT_AN vs TREM2_AN p-value < 0.0001). For b-c: At least 500 cells were counted by conditions across 4 biological replicates d) rtPCR of GPNMB mRNA in iMGLs untreated or exposed to AN or AN+ cytochalasin D (p-value < 0.0001) e) rtPCR of GPNMB mRNA in iMGLs untreated, treated with AN or E. coli (p-value = 0.0374).

https://doi.org/10.1038/s41590-023-01558-2

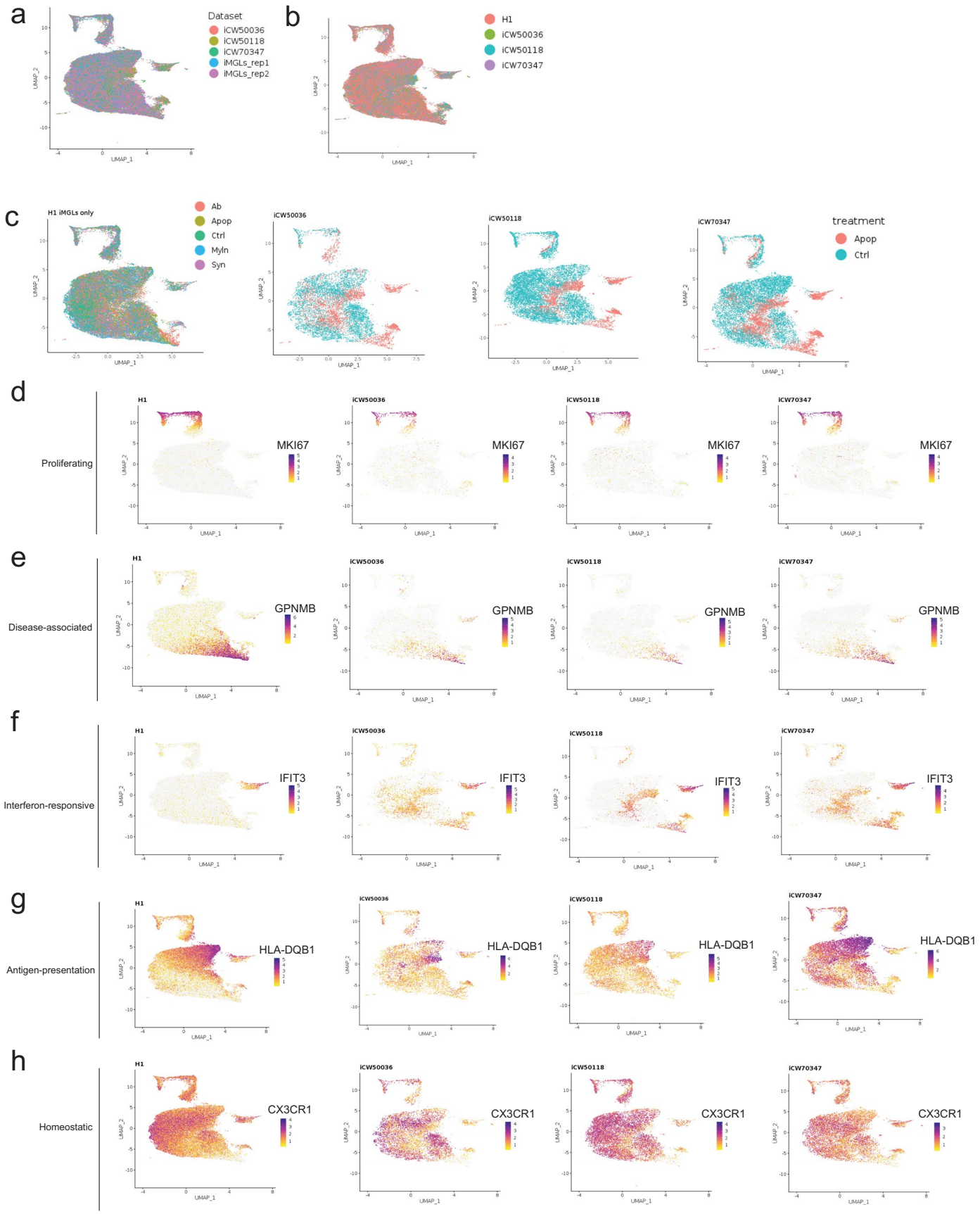



**Extended Data Fig. 8 | See next page for caption.**

https://doi.org/10.1038/s41590-023-01558-2

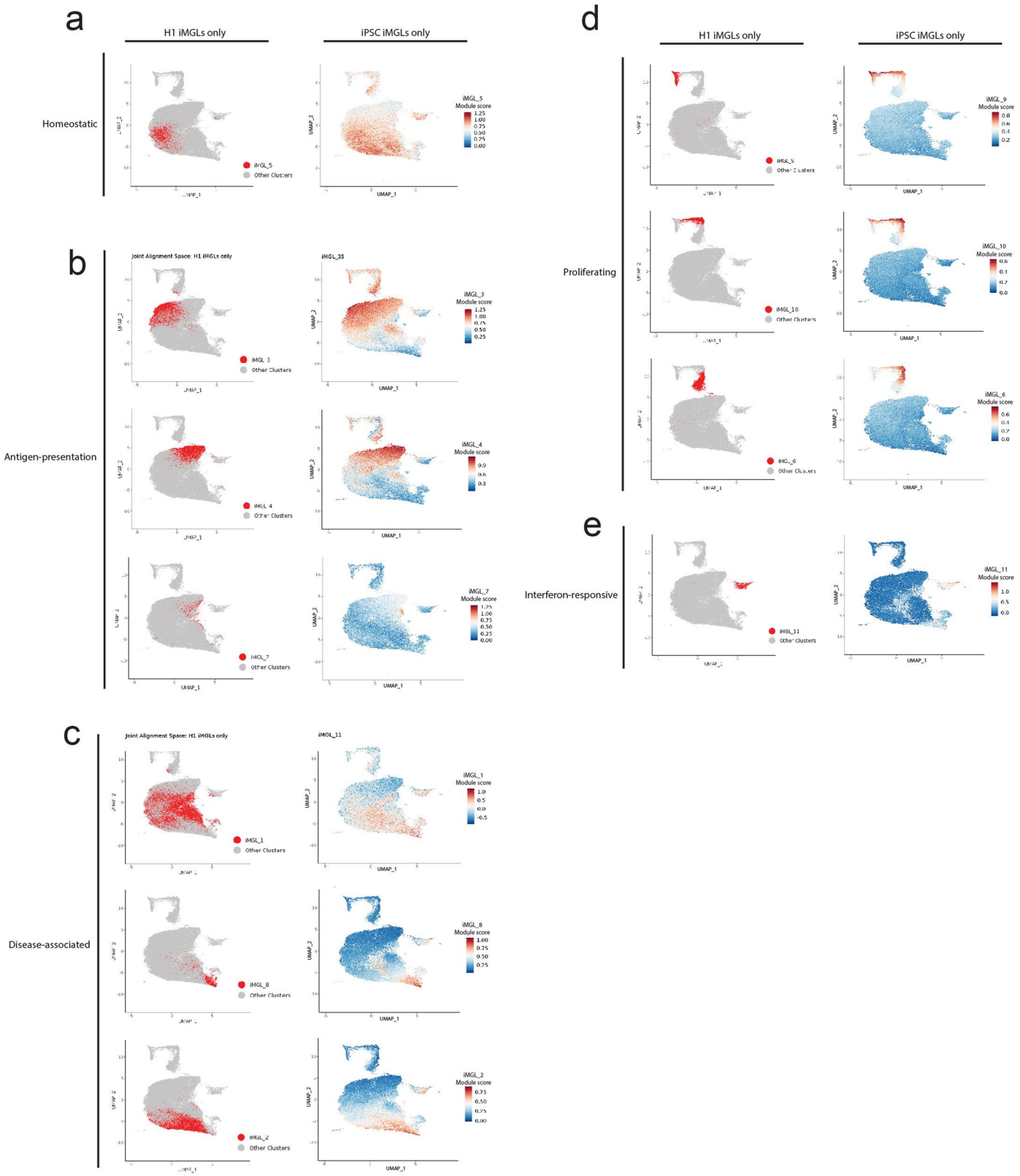
**Extended Data Fig. 8 | Lentivirus-mediated genetic modification of iMGLs.** a) Heatmap of TFs identified by SCENIC DAM iMGL clusters (iMGL_1, iMGL_2 and iMGL_8), plotted by a scaled and centered area under the curve (AUC) for each regulon b) % of mCherry positive cells in iMGLs transduced with lentivirus only and lentivirus with VPX virus-like particles, determined by flow cytometry (p-value = 0.0079, two-tailed Mann-Whitney test). c) Transduction efficiency of iMGLs transduced from multiple cell lines using co-transduction strategy with VPX. Vector is non-targeting Cas9 control co-expressing mCherry and positive cells determined by flow cytometry d) UMAP projection of single-cell sequencing data from two untreated and lentivirus transduced iMGL differentiations, cells colored by treatment. e) Expression of AIF1, TGFBR1, CSF1R and CX3CR1 in the control and lentivirus samples. (AIF1 p-value = 0.463, TGFBR1 p-value = 0.999, CSF1R p-value = 0.999 and CX3CR1 p-value = 0.2193, Wald test with multiple hypothesis correction was performed with DESeq2). f) Distribution of cells by cell cycle stage in the control and lentivirus transduced samples. g) Volcano plot of differentially expressed genes between control and lentivirus transduced samples. Genes highlighted in red have adjusted p-value < 0.01 and log2 fold change < 1.5. Cycling cells were excluded prior to analysis. h) Gene ontology analysis of differentially expressed genes between control and lentivirus transduced samples. i) rtPCR time-course analysis of expression of three interferon-induced genes in iMGLs after lentivirus transduction. j) Percentage of mCherry-positive cells determined by flow cytometry, the number of cells transduced successfully (**** = p-value < 0.0001) (NI = non-infected). Error bars represent standard deviation.

https://doi.org/10.1038/s41590-023-01558-2




**Extended Data Fig. 9 | Comparison of H1 and iPSC-derived iMGLs revealed similar transcriptional signatures.** a) UMAP projection of LIGER dataset integration of iMGL dataset and iPSC-derived iMGLs. Cells colored by dataset. b) UMAP projection (as in a), cells colored by line. c) UMAP projection (as in a), cells separated by line identity and colored by treatment condition. d-h) UMAP projection (as in a), cells separated by line identity and colored by expression of *MKI67 (d)*, *GPNMB (e)*, *IFIT3 (f)*, *HLA-DQB1 (g)* or *CXCR3 (h)*.

https://doi.org/10.1038/s41590-023-01558-2




**Extended Data Fig. 10 | H1-derived and iPSC-derived iMGLs share similar transcriptional states, as demonstrated by module scores.** UMAP projection of LIGER dataset integration of iMGL dataset and iPSC-derived iMGLs. Cells colored by iMGL cluster identity (left) or expression of corresponding module score (right) for homeostatic (a), antigen-presentation (b), disease-associated (c) proliferating (d), interferon-responsive (e) states.

nature portfolio

Corresponding author(s): Beth Stevens

Last updated by author(s): Jan 9, 2023

# Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

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

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*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 | Data was aquired using the following softwares:<br>qRTPCR software BioRad CFX Manager 3.1<br>flow cytometry: CytExpert 2.5.0.77<br>imaging: NIS elements 4.4, Perkin Elmer Harmony V4.9 |
| Data analysis | data analysis was done using standard code and details description and parameters used can be found in the Methods section of the manuscript. packages used are listed below<br>flow cytometry analysis: FloJo V9<br>Image analysis: FIJI2.3.0, Perkin Elmer Harmony V4.9<br>statistics: Prism v9<br>qRT-PCR analysis CFX Maestro software (BioRad)<br>single cell analysis: Cellranger (version 3.1, 10X Genomics), LIGER (version 0.5), Seurat (either version 2.3.4 or 3.2.1), MAST package, DirichletReg R package, fgsea package, Monocle3, SCENIC3, ComplexHeatmap R package, “UpSetR” R package, Stats R package, babelgene R package, presto R package, enrichR, DESEQ2, Picard,<br>bulk RNA-seq:Featurecounts2.0.3, ComBat-seq3.36.0, DESeq21.28.1<br>ATAC-seq: Picard tools, HOMER 4.11.1, Diffbind 2.16.2 |

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All iMGL data is deposited on Terra, including raw and Cell Ranger output of iMGL (H1 and CW50118, CW500036 and CW70437) single cell RNA-sequencing, fastq and bam files of iMGL untreated and treated with apoptotic neurons for ATAC-seq and fastq and bam files of MITF overexpressing and mCherry control bulk RNA-sequencing. Raw data is available via managed access at DUOS https://www.duos.org; ID: DUOS-000151. Any additional data and code is available from the corresponding authors. Terra link: https://app.terra.bio/#workspaces/Stevenslab/public_iMGLdatasets

## Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

| | |
|---|---|
| Reporting on sex and gender | NA |
| Population characteristics | NA |
| Recruitment | NA |
| Ethics oversight | NA |

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

# Field-specific reporting

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

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

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

# Life sciences study design

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

| | |
|---|---|
| Sample size | For single cell analysis , ATAC-sseq and bulk RNA-seq , sample size was determine based on standards in the field (see PMID: 35953545 for example) . For immunocytochemistry and in situ hybridization at least 50 cells were measured across 2 biological replicates for each conditions. For qRT-PCR, at least 4 biological replicates were done for each conditions. All experiments were done across multiple differentiations and all conditions were done on each differentiation. |
| Data exclusions | No data was discarded except for data filtered during dataset quality control (applies to all single-cell/single-nuclei experiments), which was performed according to field standard techniques as described in methods.<br>single cell exclusion criteria 1) Number of Genes: 2000-7000, 2) Number of UMIs: 500-6000, 3) Percentage mitochondrial RNA 0-0.2%. |
| Replication | replication was performed for all experiments using at least 2 biological replicates. All experiments were done across multiple differentiations, all conditions were done on each differentiation, all attempt to replicate the data were succesfull |
| Randomization | for all experiments, all experimental groups were done simultaneously. Samples were randomly allocated in experimental group |
| Blinding | Investigators were not blinded. All analysis were applied to all samples without adjustment for conditions or genotypes |

# 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 |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |

## Methods

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

# Antibodies

**Antibodies used**

flow cytometry: CD45-FITC (BioLegend #368508), CD11B-APC-750 (BioLegend #982616), P2RY12-PB450 (BioLegend #392105), Cx3CR1-PrCp (BioLegend #341614)
ICC: TREM2 (R&D systems, AF1828-SP) 1/100, GPNMB (Cell Signaling, E1YZJ), Donkey anti-Rabbit 488 (Thermo Fisher Scientific) 1/500, Goat anti-Rabbit 594 (Thermo Fisher Scientific) 1/500, APOE clone 6B9 (Helmholtz Antibody Collection) (1/100)

**Validation**

APOE validation can be found https://www.alzforum.org/alzantibodies/apoe-clone-6b9

Biolegends and TREM2 antibodies were validated by the manufacturer and used previously in PMID: 28426964
Biolegend antibodies validation from manufacturers: To ensure they are both specific and sensitive, we validate our antibodies through a variety of methods including:Testing on multiple cell and tissue types with a variety of known expression levels.Validation in multiple applications as a cross-check for specificity and to provide additional clarity for researchers. Comparison to existing antibody clones.
Using cell treatments to modulate target expression, such as phosphatase treatment to ensure phospho-antibody specificity.

Cells snailing antibodies were previously validated by the manufacturer
Cell signaling : To ensure our antibodies will work in your experiment, we adhere to the Hallmarks of Antibody Validation™, six complementary strategies that can be used to determine the functionality, specificity, and sensitivity of an antibody in any given assay. CST adapted the work by Uhlen, et. al., ("A Proposal for Validation of Antibodies." Nature Methods (2016)) to build the Hallmarks of Antibody Validation, based on our decades of experience as an antibody manufacturer and our dedication to reproducible science.

# Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

**Cell line source(s)**

The following iPSC lines used in Figure 6 and Figure S12 were obtained from the CIRM hPSC Repository funded by the California Institute of Regenerative Medicine (CIRM): CW50118, CW50008, CW50065, CW500036 and CW70437.
H1 embryonic stem cell line was obtained from WiCell
The TREM2 knockout and isogenic control iPSC lines were previously characterized(McQuade et al. 2020). These lines were derived from cell line AICS-0036-006 from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research
HEK293T cells were purchased from ATCC

**Authentication**

All cell lines were authenticated in 2020 using genotyping analysis (Fluidigm FPV5) performed at the Broad Institute

**Mycoplasma contamination**

all iPSC and iMGL cultures are routinely tested for mycoplasma contamination using MycoAlert Mycoplasma detection kit from Lonza (LT07-118 ) and tested negative

**Commonly misidentified lines** (See ICLAC register)

no misidentified lines were used in this study

# Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

**Laboratory animals**

C57/Bl6 older than 28 days but no older than 90 days were used in this

**Wild animals**

NA

**Reporting on sex**

NA

**Field-collected samples**

All experiments were approved by the Broad Institute IACUC (Institutional Animal Care and Use Committee) and conducted in

| | |
|---|---|
| Field-collected samples | accordance with the NIH Guide for the Care and Use of Laboratory Animals. Mice were housed at AAALAC-approved facilities on a 12-hour light/dark cycle, with food and water available ad libitum |
| Ethics oversight | Broad Institute IACUC (Institutional Animal Care and Use Committee) |

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

# Flow Cytometry

## Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

| | |
|---|---|
| Sample preparation | iMGLs were detached using cold PBS, then resuspended in FACS buffer (PBS containing 2%BSA and 0.05mM EDTA). Samples were incubated for 15 mins in human Fc block (BD Biosciences) followed by 1h staining with conjugated antibodies (see below) at 4C. Samples were washed 3x with FACS buffer and resuspended in 500ul of FACS buffer for flow cytometry |
| Instrument | CytoFLEX S analyzer (Beckman Coulter) |
| Software | All analyses were performed using FlowJo V10 and statistical analysis performed using Prism9. |
| Cell population abundance | shown in each Figure S1A, S3B |
| Gating strategy | all flow cytometry experiments ( S1A-B, S3A) were gated based on cell vs debris ( FCS-A vs SSC-A), singlets (FSC-A vs FSC-H), live (DAPI negative), specific gatings for antibodies was based on negative unstained sample and for phagocytosis sample without phagocytosis substrate. |

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.