Diverse Brain Myeloid Expression Profiles Reveal Distinct Microglial Activation States and Aspects of Alzheimer’s Disease Not Evident in Mouse Models

Graphical Abstract

Highlights
- Meta-analysis of purified mouse CNS myeloid cell profiles from 69 different conditions
- 7 co-regulated gene sets include one enriched in neurodegenerative disease models
- Distinct classes of activated microglia identified in Alzheimer’s mouse model
- Resources for further exploration: comprehensive Excel tables and interactive website

Data and Software Availability
- GSE89482
- GSE93179
- GSE93180
- GSE95587

In Brief
Ready to move beyond M1 and M2? In this meta-analysis of CNS myeloid cell expression profiles, Friedman et al. identify gene modules associated with diverse microglial activation states. These modules identify distinct subsets of microglia in an Alzheimer’s model and reveal aspects of the human disease not apparent in mouse models.

Graphical Abstract

Tunable Population-level Response

Diverse Activation States of Individual Microglia

Resting

Proliferation

Interferon-Related

LPS-Related

Neurodegeneration-Related

Monoglia

Macrophage

Neutrophil, Monocyte

... Other states or mixtures of states

Authors
Brad A. Friedman, Karpagam Srinivasan, Gai Ayalon, ..., Joshua S. Kaminker, Marcel P. van der Brug, David V. Hansen

Correspondence
friedman.brad@gene.com (B.A.F.), hansen.david@gene.com (D.V.H.)

In Brief
Ready to move beyond M1 and M2? In this meta-analysis of CNS myeloid cell expression profiles, Friedman et al. identify gene modules associated with diverse microglial activation states. These modules identify distinct subsets of microglia in an Alzheimer’s model and reveal aspects of the human disease not apparent in mouse models.

Resources for further exploration: comprehensive Excel tables and interactive website

Friedman et al., 2018, Cell Reports 22, 832–847
January 16, 2018 © 2017 The Author(s).
https://doi.org/10.1016/j.celrep.2017.12.066
SUMMARY

Microglia, the CNS-resident immune cells, play important roles in disease, but the spectrum of their possible activation states is not well understood. We derived co-regulated gene modules from transcriptional profiles of CNS myeloid cells of diverse mouse models, including new tauopathy model datasets. Using these modules to interpret single-cell data from an Alzheimer's disease (AD) model, we identified microglial subsets—distinct from previously reported “disease-associated microglia”—expressing interferon-related or proliferation modules. We then analyzed whole-tissue RNA profiles from human neurodegenerative diseases, including a new AD dataset. Correcting for altered cellular composition of AD tissue, we observed elevated expression of the neurodegeneration-related modules, but also modules not implicated using expression profiles from mouse models alone. We provide a searchable, interactive database for exploring gene expression in all these datasets (http://research-pub.gene.com/BrainMyeloidLandscape). Understanding the dimensions of CNS myeloid cell activation in human disease may reveal opportunities for therapeutic intervention.

INTRODUCTION

Microglia and other mononuclear phagocytes of the CNS are the primary cellular responders to CNS injury or infection (Rivest, 2000). These cells originate during embryonic development when erythromyeloid progenitor cells from the yolk sac colonize the CNS and give rise to microglial cells in parenchymal tissue and to perivascular, meningeal, and choroid plexus macrophages at the interfaces between CNS and peripheral tissues (Alliot et al., 1999; Ginhoux et al., 2010). After embryonic establishment, CNS myeloid cells are stable, long-lived populations that persist into adulthood as tissue-resident macrophages (Goldmann et al., 2016).

A common histological feature of many neurological diseases is “microgliosis,” involving changes in microglial cell morphology, marker expression, and sometimes proliferation (Boche et al., 2013; Ransohoff and Perry, 2009; Walker and Lue, 2015). However, the scope of information gained from histological studies can be limited or even misconstrued when microglial cells are classified into purported states of polarization—such as the formerly used “M1” and “M2” macrophage states—using only a handful of markers (Heppner et al., 2015; Ransohoff, 2016). Transcriptional profiling of mouse microglia acutely isolated from CNS tissue has emerged as a popular method for genome-wide surveillance of microglial activation states in diverse biological settings. Many datasets with interesting findings have been generated, but how these transcriptional profiles interrelate has not been approached systematically.

Here, we compared microglial/myeloid cell expression profiles from a wide range of CNS disease models including ischemic, infectious, inflammatory, tumorous, demyelinating, and neurodegenerative conditions, including new microglial expression profiles from two models of tau pathology. From this meta-analysis emerged modules of co-regulated genes related to proliferation, interferon response, endotoxin response, and neurodegenerative settings. We also used reference profiles of myeloid cells from peripheral tissues, different brain regions, different stages of brain development and aging, and single-cell analyses to provide further context for disease-related expression profiles.

We then used these modules to further improve our understanding of brain myeloid activation states in Alzheimer’s...
Figure 1. CNS Myeloid Gene Expression Datasets Reveal Common Co-regulated Gene Modules Reflecting Proliferative State and Other Aspects of CNS Myeloid Biology

(A) Heatmap of within-study-normalized gene expression (Z score) for the 777 genes (rows) differentially expressed in at least 7 comparisons in 18 different studies (columns). Hierarchical clustering identified 45 modules of co-regulated genes. (See Figure S1 for higher magnification.)

(B) Like (A), but only the 82 genes of module 26, which are enriched for proliferation-associated genes.

(C) Like (B), but only the samples of study GSE67858, from mice injected with PBS, lipopolysaccharide (LPS), or lymphocytic choriomeningitis virus (LCMV). Proliferation genes were induced by LCMV but not by LPS.

(D) Expression levels of two genes from the module for individual samples in the three experimental groups. As typical for this module, both genes showed elevated expression in LCMV but not LPS group. Bars above plots represent fold-change comparisons displayed in the next panel.

(E) Differential expression of each gene in the module, in LPS- or LCMV-treated animals relative to PBS. Fold changes of the four comparisons represented in (D) are shown in the indicated colors.

(legend continued on next page)
disease (AD) and AD models. First, we assessed the degree to which genes from different modules depended on Trem2 for their modulation in an AD model. Second, we probed a recent single-cell RNA sequencing (RNA-seq) dataset from the same model to learn whether our modules could identify distinct microglial activation states. Finally, we analyzed expression profiles from human neurodegenerative disease tissues—including new data from fusiform gyrus of Alzheimer’s and control patients—and observed upregulation of not only the neurodegeneration-related gene modules but also the LPS-specific and neutrophil/monocyte modules in AD tissues. We also provide a public platform for exploring gene expression in all these datasets. Our analyses identified multiple dimensions of CNS myeloid cell activation that differentially respond to specific conditions. Understanding how these dimensions can be modulated may open new opportunities for neurological disease therapy.

RESULTS

Establishing a Database of Brain Myeloid Transcriptional Responses, Including Tau Models

We searched the literature for gene expression studies of acutely isolated microglia/myeloid cells from adult mouse brains (or spinal cords). The details of purification varied from study to study, but the most common strategies were selection of CD11b+, CD11b+;CD45lo, or Cx3cr1::GFP+ cells by fluorescence-activated cell sorting (FACS). We also prioritized studies with at least 3 replicates per treatment group. From these, we considered but excluded “germ-free” (Erny et al., 2015; Matcovitch-Natan et al., 2016) and “nerve injury” (Denk et al., 2016; Jeong et al., 2016) perturbations, since these exhibited little reproducible signal and/or few significant changes (Data S7, panels 1 and 2). Our final database included 18 datasets spanning 69 different conditions and 336 individual expression profiles across a range of neurodegenerative, neoplastic, inflammatory, and infectious disease models, along with reference profiles from different developmental stages, brain regions, and myeloid cell subtypes (Data S1) (Arumugam et al., 2017; Bruttger et al., 2015; Chen-Plotkin et al., 2008; Chiu et al., 2013; Denk et al., 2016; Erny et al., 2015; Grabert et al., 2016; Hodges et al., 2006; Jeong et al., 2016; Lavin et al., 2014; Matcovitch-Natan et al., 2016; Orre et al., 2014; Poliani et al., 2015; Pyonteck et al., 2013; Rabin et al., 2010; Srinivasan et al., 2016; Szulzewsky et al., 2015; Verheijden et al., 2015; Wang et al., 2015; Webster et al., 2009; Zeisel et al., 2015; Zhang et al., 2014, 2016).

We also generated new RNA-seq data (details in Data S7 “New RNA-seq Datasets”) from Cx3cr1::GFP+ cells from the PS2APP model of AD (Richards et al., 2003) and CD11b+ cells from hippocampi of two mouse models of tauopathy (Götz et al., 2001; Yoshiyama et al., 2007). The PS2APP data resembled what we previously described in CD11b+ cells (Srinivasan et al., 2016). In the Tau models, which express either the P301L or P301S mutations associated with frontotemporal dementia and parkinsonism, we found hundreds of differentially expressed genes (DEGs) in the more aggressive P301S model at 6 months, but only a few tens of genes in the P301L model at the age of 12 months (Data S7, panel 4E). Overall, the genes upregulated in P301S tended to show slight elevation in P301L, suggesting that myeloid cells in the P301L model undergo a milder form of the same phenotypic changes that occur in the P301S model. We next considered these changes in light of the other transcriptional profiles in our database.

Each study was separately analyzed, whenever possible starting from raw data (e.g., FASTQ or CEL files) and using standardized pipelines (see Experimental Procedures). Absolute expression values were then Z score normalized within each study, and these Z scores were finally bound together across all of the studies to create one master gene expression matrix. This normalization results in relative (rather than absolute) gene expression values within each study, thus providing a statistical control for the many differences that exist between studies.

Tumor and Virus, but Not Neurodegenerative Models or LPS, Induce Proliferation Gene Module

In order to explore common types of microglial activation in an unbiased manner, we selected the genes differentially expressed in the greatest number of conditions (777 genes, ≥7 conditions, with differential expression calculated separately within each dataset). These 777 genes captured on average 36% of DEGs in the individual studies (range: 10%–76%, see Data S1). We performed hierarchical clustering on this “master matrix” to separate these genes into enough co-regulated modules (45) to allow us to categorize the major patterns of gene expression discernible by eye (Figures 1A and S1; Data S2, S3, and S4). In certain cases, we named modules of co-regulated genes based on the genes they contained or on their pattern of expression in these datasets.

One such module, module 26 (Figure 1B), consisted of 82 genes mostly associated with cell proliferation, such as Mki67, Cdk1, and Plk1. These genes showed correlated expression within the various datasets. For example, they were largely unchanged in brain myeloid cells after lipopolysaccharide (LPS) injection but robustly induced following injection of the virus LCMV (Erny et al., 2015) (Figures 1C–1E). The fold inductions of these genes likely reflect the relative proliferative state of brain myeloid cells in these different conditions (Figure 1F), from unchanged or only slightly elevated in most neurodegenerative models, to very high proliferative gene expression after virus treatment (LCMV) and during development (embryonic, perinatal).

Neurodegeneration-Related Gene Modules Reflect Different Aspects of Myeloid Activation

Beyond the proliferation module, we interpreted an additional six groupings of co-regulated genes (Figures S1, 2, 3, and 4). Three of these showed increased overall expression in multiple models.
of neurodegenerative disease: the core “neurodegeneration-related” module (composed of modules 24, 25, 36, and 37), the “interferon-related” module (module 18), and the “LPS-related” module (composed of modules 10, 12, 13, 17). It should be noted that many hundreds of genes responsive to LPS and/or virus were not elevated in neurodegenerative models and did not appear in any of our modules since they did not meet the threshold of differential expression in 7 datasets.

The interferon-related module included many well-known interferon-stimulated genes (Mostafavi et al., 2016), such as members of the Oas and Ifti families, Lsg20, and the transcription factors Lrf7 and Stat2. These genes were most highly induced in response to virus, but also high in LPS and glioma (Figures 2A–2I). The module was modestly elevated in several neurodegenerative disease models, in cerebellar relative to cortic microglia, and in microglia from aged mice.

The LPS-related gene set included certain inflammatory genes such as Ikbke, Cd44, Cci5, and Tspo. They were induced highly in response to LPS and glioma, to a lesser extent in some of the neurodegenerative disease models, and hardly at all in response to virus (Figures 2J–2R). Notably, Tspo, which encodes the target of the widely “microglial activation” PET tracer 11C-PK11195 (Vivash and O’Brien, 2016), was robustly induced in response to LPS but, like many LPS-stimulated genes, showed little change in the neurodegenerative models.

The core neurodegeneration-related gene set was broadly elevated across most or all of the neurodegenerative disease models but changed little in response to either LPS or virus (Figure 3). The main grouping (modules 24 and 25)—like the interferon-related and LPS-related modules—typically showed elevated expression in the demyelination, ischemia/reperfusion, and glioma models and tended to be more abundantly expressed in peripheral, infiltrating, or perivascular macrophages than in bulk brain myeloid cells (which are mainly microglia) under normal conditions (Figure S1). The smaller grouping (modules 36 and 37) showed the peculiar combination of being normally expressed more highly in brain myeloid than peripheral/infiltrating myeloid cells, being induced to even higher levels of expression in neurodegenerative models, and typically being repressed in LPS-injected animals (Figure S1).

Of the 134 genes comprised by the core Neurodegeneration-related gene set, 101 (75%) are annotated with a Gene Ontology (GO) term associated with either the plasma membrane or extracellular space. This suggests that microglia in neurodegenerative settings change the manner in which they interact with their environment. Possible regulators include transcription factors encoded by four genes in the set: Bhlhe40, Rorg, Hif1a, and Mitf. 10 of the genes, including the cathepsins Ctsb, Ctsl, and Catz are specifically associated with the “lysosome” GO term, suggesting higher lysosomal activity in these cells. Other genes of note in this set include Gpnmb, which is genetically linked to Parkinson’s disease (Nalls et al., 2014), a biomarker for Gaucher Disease (Zigdon et al., 2015), and a known target of Mttf (Gabriel et al., 2014; Gutknacht et al., 2015); Igtf, thought to be a secreted neuroprotective factor (Dobrowolny et al., 2005; Dodge et al., 2008; Vergani et al., 1999); and Apoe, the foremost genetic determinant of late-onset AD (Liu et al., 2013).

The CD11c-encoding gene Itgax is also in the neurodegeneration-related gene set. The transcriptomes of CD11c-positive and CD11c-negative microglia isolated from mouse brains with β-amyloid pathology were recently described (Kamphuis et al., 2016). Although the raw data were not publicly available, a supplemental table contained 240 genes enriched in CD11c-positive relative to CD11c-negative microglia. 65 (49%) of our core neurodegeneration-related genes were in the table of CD11c-enriched genes. By comparison, this table included only 1 gene from our proliferation model (1%), 6 genes from the LPS-related modules (7%), and no genes from our other groupings (interferon-related, or three more below). All together, these data indicate that core neurodegeneration-related modules represent a special activation state of brain myeloid cells largely distinct from that induced by microbial challenge and characterized by altered environmental engagement and lysosomal activity.

Surprised that relatively few genes were differentially expressed in microglia from the P301L tau model (Data S7, panel 4E), we directly investigated the neurodegeneration-related gene set in the P301S and P301L models. As expected, many of the genes were robustly activated in the more severe P301S model (Figure 3J). While only a handful of the individual genes showed clear increases in P301L microglia (Figure 3K), overall the gene set was clearly elevated (compare points above the line y = x to those below it). Therefore, even in the P301L model we see a signal of this distinctive, neurodegeneration-related microglial signature.

Gene Modules Associated with Different Classes of Myeloid Cells

Modules 2, 3, 5–7, and 9 were elevated in resident brain myeloid cells relative to infiltrating (GEO: GSE63376) and peripheral (GEO: GSE63340) macrophages (Figure S1). Among these, genes of the microglia module (module 6) were unique in their specific elevation in parenchymal microglia relative to perivascular macrophages (GEO: GSE60361, Figure 4A). Some of these, such as P2ry12 (Hickman et al., 2013) and Tmem119 (Bennett et al., 2016) have already been described as distinguishing microglia from other brain-resident myeloid cells. Previously published “microglia-specific” genes (see “Sensome” (Hickman et al., 2013), “MG400” (Butovsky et al., 2014), and “Chiu MG” (Chiu et al., 2013) genes in Figure S1) also included many genes expressed generally by brain myeloid cells, not only by microglia.
Figure 3. Neurodegeneration-Related Genes Are Elevated in Myeloid Cells from Models of Neurodegenerative Disease but Not after LPS Treatment

(A–I) Like Figure 2 (A)–(I) but for the Neurodegeneration-Related gene set rather than the Interferon-Related module.

(J and K) Expression of neurodegeneration-related module genes in hMAPT-P301S (J), hMAPT-P301L (K), and control hippocampal myeloid cells. Points correspond to module genes and x and y axes represent expression levels in control and transgenic myeloid cells. Red: significant differential expression at p ≤ 0.05 and fold change ≥ 2. See also Figure S1.
Our microglia module is microglia specific only with reference to myeloid cells, since we have not accounted for expression in non-myeloid CNS cell types.

Virtually all perturbations reduced the expression of the microglia module (and the brain myeloid modules generally), with modest decreases in neurodegenerative models and pronounced reductions with LPS treatment (Figures 4A and S1). In theory, this could be due either to a change in gene expression or to partial replacement of the sorted myeloid compartment with non-microglial cells. However, in at least some of the datasets (e.g., 5XFAD and APPswe/PS1DE9β-amyloid models), the macrophage signature (see below) was also decreased, all but ruling out the most likely suspect for such a replacement. Recent parabiosis experiments confirmed that any contribution of blood-derived cells to the brain’s myeloid population is negligible in β-amyloid models (Wang et al., 2016). Therefore, the decreased expression of the microglia module in neurodegenerative models likely reflects frank cell-intrinsic transcriptional modulation.

The macrophage genes (module 45), including Mrc1 and F13a1, were elevated in perivascular, brain-infiltrating, and peripheral macrophages relative to microglia (Figure 4B). Of all the disease models tested, only glialoma showed pronounced elevation of these genes. Interestingly, the expression of the microglia and macrophage modules was inversely coordinated during brain myeloid cell development, with macrophage expression gradually reduced and microglia expression gradually increased from embryonic through perinatal to adult brains (GEO: GSE79812 in Figures 4A and 4B). By contrast, myeloid cells from cerebella, as well as from older mice of any brain region, showed the opposite pattern: slightly lower expression of the microglia module (relative to cortical or younger mice) accompanied by slightly increased macrophage module expression.

Finally, genes of the neutrophil/monocyte modules (modules 43 and 44) including Ngp and Mmp8 were identified by their elevation in neutrophils and, to a lesser extent, monocytes relative to macrophages (Figure S1; GEO: GSE63340) and other immune cell types (Heng et al., 2008) (Figure S2). Though mostly unchanged in neurodegeneration models, these genes were robustly elevated in LPS and glioma models, as well as in cerebellum, suggesting an increased abundance of neutrophils or monocytes in these conditions. This highlighted that preparations of myeloid brain cells, although dominated by parenchymal microglia, are complex mixtures of different cell subtypes unless extra measures are taken to exclude other myeloid cell types.

Having established these gene modules relating to brain myeloid subtypes and activation states, we next explored three ways the modules could be used to better understand neurodegenerative disease. First, we studied whether the different modules depended on Trem2 for their induction in an AD model. Second, we looked at whether our modules could identify unique subsets of microglia in a single-cell RNA-seq dataset from the same model. Third, we analyzed bulk tissue RNA profiles from human neurodegenerative disease samples to assess the degree to which the information from mouse models reflected brain microglia activation states observed in the human diseases.

Microglial Response to β-Amyloid Is Trem2 Dependent across Modules

Since mutations in TREM2 are among the strongest known genetic factors that elevate risk of AD, we asked whether our myeloid gene modules showed differential dependence on Trem2 for their activation. We calculated the “percent Trem2 dependence” for each DEG in the 5XFAD model (GEO: GSE65067; Figure 5A). This ranged from 0%, for genes showing similar induction in Trem2KO (knockout) and Trem2WT (wild-type) microglia, to 100%, for genes that were induced in Trem2WT but showed no induction in Trem2KO. Most genes fell between these two extremes—a diminished but not ablated response in Trem2KO animals. Notably, Apoe was among a small number
of genes, also including Cd9, in the neurodegeneration-related gene set whose fold induction in 5XFAD was Trem2 independent (Figures 5A and 5B). Interestingly, the DEGs in our proliferation, interferon-related, and LPS-related modules showed as much or greater dependence on Trem2 for their induction in 5XFAD microglia, compared to DEGs in the neurodegeneration-related gene set (Figure 5A).

The decreased expression of several genes in the microglia module, including P2ry12, also showed considerable Trem2 dependence (Figure 5B). Interestingly, the DEGs in our proliferation, interferon-related, and LPS-related modules showed as much or greater dependence on Trem2 for their induction in 5XFAD microglia, compared to DEGs in the neurodegeneration-related gene set (Figure 5A).

The decreased expression of several genes in the microglia module, including P2ry12, also showed considerable Trem2 dependence (Figure 5B), and this was true for most downregulated genes throughout the brain myeloid modules (Figure S1; GEO: GSE66926). Thus, context is essential for understanding the possible outcomes of Trem2 activity.

New Microglial Subpopulations Identified by Gene Modules

The modules we have described provided information about population-wide transcriptional changes in brain myeloid cells in various settings (Figure 5C). However, it was unclear whether these modules could be induced concurrently within individual cells or whether they represented discrete (mutually exclusive) activation states. To further validate these modules and better understand their utilization, we examined their expression in a recently published single-cell RNA-seq survey of CD45+ immune cells from the 5XFAD mouse model (Keren-Shaul et al., 2017). We recapitulated the cell clusters originally reported by the
Figure 6. Myeloid Gene Modules Represent Distinct Cell Types and Activation States
(A–I) tSNE projections of single-cell RNA-seq data from 5XFAD and control brains distinguish microglia (big cluster on top) from non-microglia cells (clusters 5, 7, 8, 9, 11, 12, 14 below), as well as activation states within the microglia (GEO: GSE98969).

(legend continued on next page)
authors (B cells, T/NK cells, granulocytes, perivascular macrophage, monocytes, resting microglia, and “disease-associated microglia” [DAM]) (Figures 6 and S3; Data S6).

As expected, the DAM cells were present almost exclusively in 5XFAD, not non-transgenic, brains (Figures 6B and B) and expressed the core neurodegeneration-related gene set (Figure 6F). We also identified other interesting clusters of microglial cells. When we probed these for expression of our modules, we were able to pinpoint unique clusters of microglia expressing the interferon-related module (cell cluster 13, Figure 6D), the proliferation module (cell cluster 16, Figure 6C), or module 8 (cell cluster 6), which consisted of the immediately early genes *Fos* and *Egr1* (Figure S3F). These cell clusters were clearly distinct from the DAM cells that expressed neurodegeneration-related genes, indicating these modules represent discrete, possibly exclusive, microglial states. Indeed, looking within each of these cell clusters for expression of the other modules revealed no apparent overlap (cell clusters 4, 13 and 16 in Figures 6C', 6D', and 6F').

The LPS-related gene set, in contrast, was not upregulated in one discrete cell cluster but rather appeared modestly increased in the DAM cluster, the cluster expressing interferon-related genes, and the cluster expressing proliferation genes, relative to resting microglia (Figures 6E and 6E'). Decreased expression of the microglial module was obvious only in the DAM cells but not in other microglial clusters (Figures 6G and 6G').

Other than the DAM cells, most of the microglial clusters showed no clear difference in cell numbers between 5XFAD and non-transgenic brains, with the exception of one other cluster (Figure 6B'). Cluster 13 cells expressing the interferon-related module were roughly twice as abundant in 5XFAD brains (0.9%) than in controls (0.4%), and we also observed a similar ~doubling of this population using single-cell RNA-seq in the PS2APP model (unpublished data). Our gene sets therefore aid in interpreting these complex single-cell data and highlight that interferon-related activation occurs independently and in parallel to neurodegeneration-related activation.

**Comparisons of Module Expression between AD Tissues and Mouse Models**

The difficulties of post-mortem tissue acquisition have, to date, limited the availability of sorted cell expression data in human neurodegenerative disease, so we examined the expression of our gene modules in bulk tissues. We sequenced RNA from frozen specimens of fusiform gyrus of 33 neurologically normal controls and 84 autopsy-confirmed Alzheimer’s cases, possibly confounded with age, which itself was not well controlled between AD and control cohorts (Figure S4). We also explored expression profiles in datasets from other neurodegenerative disease bulk tissues, including spinal cord anterior horn pool of sporadic amyotrophic lateral sclerosis (ALS) patients (Rabin et al., 2010), frontal cortex from PGRN mutant FTLD patients (Chen-Plotkin et al., 2008), and caudate nucleus from Huntington’s disease (HD) patients (Hodges et al., 2006).

Because signals from bulk tissue can be dominated by changes in the relative abundance of different cell types (Srinivasan et al., 2016), we first examined changes in cellular composition, using cell-type-specific gene sets as a proxy for CNS cell-type abundance. We created gene sets for the major CNS cell types (Data S5) from mouse and human expression data (Tasic et al., 2016; Zhang et al., 2016) and then scored these gene sets in bulk tissue expression profiles to analyze cell-type enrichment in each sample. Compared to mouse models, a greater range of cell-type variability was observed within both neurologically normal and diseased human tissues. Nonetheless, significant changes in the distribution of cell-type scores were associated with various diseases (Data S7, panels 5 and 6). For example, both excitatory and GABAergic neuron scores were significantly lower in most of the disease conditions relative to their controls (see Data S7, panel 7 for GABAergic gene set in AD). All diseases were associated with higher myeloid scores, although the effect was surprisingly modest in the AD datasets. Astrocyte scores were typically higher in neurodegenerative tissues but only in one of the two AD datasets (Data S7, panel 8).

Keeping in mind the variability in cellularity, we next examined the expression of the mouse brain myeloid gene modules in the RNA expression data from human neurodegenerative and mouse model bulk tissues. Interestingly, all seven modules were at least modestly elevated in bulk tissue from PS2APP and/or SOD1G93A mouse models (Data S7, panel 9), even though some were not elevated or, in the case of the microglia gene module, even slightly lower in purified myeloid cells from these transgenic models (Figure 5C). This highlights the challenges of untangling changes in cellularity from changes in expression using bulk tissue expression data.

In human neurodegenerative tissues, many of the myeloid activation gene sets were modestly elevated in the disease conditions (Data S7, panel 9). However, compared to the cell-type marker gene sets, genes within in the myeloid activation modules were not as well correlated (e.g., compare Figures S5C and S5F to corresponding panels of Data S7, panels 7 and 8). This was not surprising since the microglia gene module, even slightly lower in purified myeloid cells from these transgenic models (Figure 5C). This highlights the challenges of untangling changes in cellularity from changes in expression using bulk tissue expression data.

In human neurodegenerative tissues, many of the myeloid activation gene sets were modestly elevated in the disease conditions (Data S7, panel 9). However, compared to the cell-type marker gene sets, genes within in the myeloid activation modules were not as well correlated (e.g., compare Figures S5C and S5F to corresponding panels of Data S7, panels 7 and 8). This was not surprising since the microglia gene module, even slightly lower in purified myeloid cells from these transgenic models (Figure 5C). This highlights the challenges of untangling changes in cellularity from changes in expression using bulk tissue expression data.

Thus, trying to assess myeloid activation states in bulk tissue profiles using our gene modules was confounded by at least two
Figure 7. Orthologs of Core Neurodegeneration-Related Genes Identified in Mouse Myeloid Cells Are Elevated in AD Fusiform Gyrus
(A) Z score expression of mouse genes in the core neurodegeneration-related module in a representative dataset of sorted mouse CNS myeloid cells. Genes are shown in the same order as their human orthologs in (B). Genes without human orthologs, and genes with high neuronal expression are omitted (see Experimental Procedures). The entire set is elevated in purified cells.

(B and C) Z score expression of human orthologs of core neurodegeneration-related genes on the rows, and samples from “myeloid-balanced” control or AD tissue (fusiform gyrus, B, or temporal cortex, C) on the columns (see Experimental Procedures and Figure S6). Selected genes labeled; full list in Data S4 (“Neurodegeneration” in the “Myeloid Activation (Coarse)” column). Below, gene set score for each sample (color scale not shown).

(legend continued on next page)
factors—artificial elevation due to increased abundance of myeloid cells and obfuscation due to gene expression in non-myeloid cells. We next tried to correct for these confounding factors in AD datasets in order to more accurately assess whether our myeloid gene modules were in fact elevated in AD tissues.

We noticed that almost half of the genes in the neurodegeneration-related module were enriched in AD bulk tissue, many were essentially unchanged, and a smaller number surprisingly showed lower expression in the AD cohorts (Figures S5A and SSD). However, many of the “down” genes, but few of the “unchanged” or “up” genes, showed enriched expression in neurons relative to other cell types (Figures S5G and SSH). Since the decreased abundance of these mRNAs in bulk tissue likely reflected neuronal loss rather than reduced microglial expression, we excluded neuron-enriched genes from our analysis (Data S7, panel 10).

Next, since many of the remaining genes in the neurodegeneration-related gene set are enriched in myeloid cells (Figures S5G and SSH), we created subsets of the control and AD cohorts with similar myeloid content (Figure S6). Even in these myeloid-balanced datasets, the neurodegeneration-related module scores were significantly higher in AD samples than controls (Figures 7A–7F). Therefore, elevated neurodegeneration-related scores in AD cohorts reflected, at least in part, frank transcriptional activation, and the myeloid reaction observed in mouse models also likely occurs in human patients.

Following the same analysis for the other myeloid gene modules—removing neuron-enriched genes and examining myeloid-balanced cohorts—we found that some modules were still elevated in bulk AD tissue (Figures S7A–S7G). In particular, the LPS-related and neutrophil/monocyte modules showed higher expression in both the fusiform gyrus and temporal cortex datasets (Figures 7G and 7H). Since many genes in the LPS-related modules are somewhat elevated in neurodegenerative models, we tested whether genes elevated only in myeloid cells from LPS-treated animals, but not in any neurodegenerative models or following LCMV infection, also showed elevated expression in AD tissues. Even these “LPS-specific” genes (and, to a lesser extent, an analogous set of “LCMV-specific” genes) were elevated in bulk patient tissue (Figures 7I, S7H, and S7I). Although it is not possible to confidently deduce whether this signal in bulk tissue RNA arises from altered microglial gene expression or from increased presence of peripheral myeloid cells (see expression of LPS-related module in peripheral macrophages, Figure 2J), these results suggest an important difference between myeloid activation or recruitment in AD patient brains compared to existing mouse models of neurodegenerative disease.

**Resources for Further Data Mining and Visualization**

To enable others to explore these data, we have assembled an Excel file (Data S2) giving all gene annotations from all the figures, including those derived from other sources (e.g., GWAS hits and Sensome genes), as well as those developed in this manuscript (e.g., gene set memberships), average expression levels in each experimental group of every dataset, and statistics for every differential expression analysis. Each column is “filter-ready,” enabling further mining of these data. Each row corresponds to a human gene (with mouse orthologs as relevant). Data S3 contains similar data, but organized with one row per mouse gene. We also provide two smaller files, containing just the myeloid activation modules (Data S4), and just cell-type markers (Data S5).

We have also built an interactive website at http://research-pub.bio.com/BrainMyeloidLandscape. The website provides reports for each gene and for each study. The gene reports include an overview of differential expression results across all of the studies, followed by expression plots showing the gene’s expression levels across samples in each study. The study reports include plots showing the genome-wide differential expression results. This should be a user-friendly go-to resource for scientists and enthusiasts interested in brain myeloid gene expression.

**DISCUSSION**

We have compared the genome-wide transcriptional responses of brain myeloid cells obtained from diverse models of neurodegenerative disease, aging, viral infection, inflammatory stimulus, ischemic injury, demyelinating disorder, and brain tumor growth. From these profiles, we identified modules of genes that show similar response in multiple settings, and we have highlighted the proliferative, interferon-related, LPS-related, and core neurodegeneration-related modules. Using these modules to analyze a published single-cell RNA-seq dataset, we recognized clusters of activated microglia, showing that the interferon-related, proliferation, and core neurodegeneration-related modules represent independent activation states while the LPS-related gene set was enriched in all three activated clusters. As the identification of these modules required the genes in a given set to be differentially regulated in multiple comparisons, we also performed targeted analyses identifying LPS-specific and LCMV-specific genes. All of these gene sets, as well as the individual datasets, can be further explored using the web resource and supplemental tables we have provided.

While transcriptional responses in various disease settings were diverse, one change that was almost universal in all comparisons was the downregulation of most genes whose expression in microglia/brain myeloid cells distinguishes these cells from myeloid cells/macrophages in peripheral tissues (see modules 2, 3, 5–7, and 9 in Figure S1). It is curious that with any perturbation—even normal aging—the genes that set microglia apart from other tissue macrophages and are presumably involved in CNS-specialized microglial functions show reduced expression. Looking at broad trends across modules 10–25
mice robustly upregulate these genes in response to evidence suggest that in fact the neurodegeneration-related activative models has pathogenic consequences, certain lines of many assume the microglial activation in mouse neurodegenerative settings. This is an important question as it has been shown that the microglial activation in mouse neurodegenerative models is protective, whereas microglia from wild-type mice robustly upregulate these genes in response to β-amyloid pathology (Wang et al., 2015) or cuprizone-induced demyelination (Poliani et al., 2015), the response is not attenuated in Trem2-deficient mice. Correlated with this attenuated response, Trem2-deficient mice had lower outcomes: increased phosphorylated tau and axonal dystrophy in the 5XFAD and APPPS1-21 β-amyloid models (Wang et al., 2016; Yuan et al., 2016) and persistent demyelination and axonal dystrophy after withdrawal from prolonged cuprizone treatment (Poliani et al., 2015). These examples of protective, Trem2-dependent microglial activation in neurodegenerative settings are consistent with human genetic evidence indicating that TREM2 hypomorphism or deficiency is associated with increased AD incidence or Nasu-Hakola disease (which includes loss of myelin), respectively (Guerreiro et al., 2013; Jonsson et al., 2013; Paloneva et al., 2002). Further preclinical experiments will be necessary to understand how activation or suppression of the neurodegeneration-related response may alter the course of disease at different stages, whether this response is Trem2 dependent in models of tau-driven or Sod1-driven pathology, and whether Trem2 function is protective or detrimental in those models.

Our analysis of bulk AD tissues (Figures 7 and S7) suggested both similarities and differences with existing mouse models. On the one hand, the core “neurodegeneration-related” module genes were elevated in these tissues, suggesting that this type of activation is common to both mouse and human. However, genes of the “neutrophil/monocyte” module and the specially prepared “LPS-specific” module were also elevated in bulk human tissues. This suggested that human AD could involve more classical inflammatory signaling and/or peripheral immune cell infiltration than is apparent in expression profiles from mouse neurodegenerative disease models. β-amyloid pathology is known to prime microglia for augmented inflammatory response to systemic infection (Sly et al., 2001), which is clinically associated with accelerated cognitive decline in AD patients (Holmes et al., 2009). The housing of laboratory mice in pathogen-free conditions results in a naive immune system with less innate immune activation (Tao and Reese, 2017), possibly reducing the transmission of inflammatory signals from the periphery into the CNS (Perry and Holmes, 2014). It is also tempting to associate the lack of classical inflammatory gene expression in mouse models of β-amyloid pathology with the lack of overt neurodegeneration in those models, both in contrast to human AD tissues. Further studies on human tissue with more refined technologies—profiling purified brain myeloid cells or nuclei as a population and at the single-cell level—will clarify the extent to which these phenotypes actually occur in disease and guide attempts to model them in preclinical research.

Two notable caveats of the human datasets deserve mention. First, the human expression data are from end-stage tissue and may not inform the pathogenic mechanisms of earlier disease stages. Second, stress-induced changes in inflammatory gene expression may occur post-mortem while cellular energy stores and temperature permit, and microglia from neurodegenerative tissues may be “primed” for an augmented inflammatory response (Perry, 2010). These caveats for human expression data are somewhat insurmountable given current technological and ethical constraints; thus, addressing these questions will likely require better preclinical models of human disease. Understanding the dimensions of brain myeloid cell activation defined herein, and learning how to manipulate them, may lead to novel therapeutic approaches for human neurological disease.

**EXPERIMENTAL PROCEDURES**

Further details and an outline of resources used in this work can be found in Supplemental Experimental Procedures.

**Mice**

All protocols involving animals were approved by Genentech’s Institutional Animal Care and Use Committee, in accordance with guidelines that adhere to and exceed state and national ethical regulations for animal care and use in research. For GSE89482, Cx3cr1::GFP (Jackson Labs #006582) mice were crossed to our PS2APP colony, and microglia from the cortex of Cx3cr1<sup>GFPy</sup>PS2APP<sup>negative</sup> mice (“Non-transgenic”) and Cx3cr1<sup>GFPy</sup>PS2APP<sup>homozygous</sup> (“PS2APP”) mice were compared. Hippocampal microglia were collected from hMAPT-P301L<sup>homozygous</sup> and hMAPT-P301S<sup>homozygous</sup> mice and their non-transgenic littermate controls, for GSE93179 and GSE93180, respectively. At 6 (P301L), 12 (P301L), or 14–15 (PS2APP) months of age, animals were perfused and processed in control and transgenic pairs using two BD FACSAria sorters simultaneously.

**Patient Bulk Tissue**

Frozen fusiform gyrus tissue blocks and pathology/clinical reports, including age, sex, diagnosis, and Braak stage, were obtained from the Banner Sun Health Research Institute Brain and Body Donation Program in accordance with institutional review boards and policies at both Genentech and Banner Sun Health Research Institute. RNA was extracted from approximately 300 μg of frozen sections from each tissue block as described (Benitez et al., 2014) and standard polyA-selected Illumina RNA-seq analysis was performed as described (Srivivasan et al., 2016) on samples with RNA integrity (RIN) scores at least 5 and post-mortem intervals (PMIs) no greater than 5 hr.

**Gene Expression Analysis**

Relevant gene expression datasets were identified using the criteria in Results by a combination of searches on GEO and PubMed databases with terms such as “microglia” and “neurodegeneration” and also naturally discovered as we followed the literature and presentations at scientific symposia. Datasets were processed and Z score normalized separately, and then Z scores were combined into a master matrix for hierarchical clustering, from which gene modules were defined.

**Statistical Methods**

Differential expression statistics were calculated using limma, voom+limma, DESeq2, or Mann-Whitney tests, as described in Supplemental Experimental
Procedures. Unpaired t tests were used to compare gene set scores and immunohistochemistry (IHC) intensities, as described in figure legends.

DATA AND SOFTWARE AVAILABILITY

The accession number for the RNA-seq datasets reported in this paper are Gene Expression Omnibus: GSE89482, Cortical Cx3cr1-GFP cells in 14- to 15-month PS2APP model; GSE93179, Hippocampal CD11b cells in Tau-P301L model; GSE93180, Hippocampal CD11b cells in Tau-P301S model; GSE95587, Alzheimer’s Patient Fusiiform Gyrus (whole tissue).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, an seven data files and can be found with this article online at https://doi.org/10.1016/j.celrep.2017.12.066.

ACKNOWLEDGMENTS

This work benefited from data assembled by the ImmGen consortium. We thank Mary Chen, Ainhoa Etxeberria, Kevin Huang, Dorothée Nickles, and Jesse Hanson for helpful feedback on the manuscript. We are grateful to the Banner Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona for the provision of human brain tissue, and to Thomas Beach for helpful discussion relating to case selection. The Brain and Body Donation Program has been supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026 National Brain and Tissue Resource for Parkinson’s Disease and Related Disorders), the National Institute on Aging (P30 AG19610 Arizona Alzheimer’s Disease Core Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimer’s Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson’s Disease Consortium) and the Michael J. Fox Foundation for Parkinson’s Research.

AUTHOR CONTRIBUTIONS

B.A.F. conceived of and coordinated the study, in collaboration with D.V.H. B.A.F. performed all analysis except as indicated below. B.A.F. and D.V.H. wrote the manuscript. J.S.K. advised on bioinformatics analysis and contributed to RNA-seq studies, G.A., W.J.M., H.L., and H.N. performed IHC studies of tau models, K.S., D.V.H., J.L.L., and Z.M. performed cell sorting and RNA-seq studies of tau models (GEO: GSE93179 and GSE93180), P.C.G.H. and M.P.v.d.B. performed RNA-seq study of Alzheimer’s bulk tissue, and M.A.H. and Y.C. performed initial analysis and quality control of these data (GEO: GSE95587). J.L.L. import and performed initial analysis of temporal cortex Alzheimer’s microarrays (GEO: GSE15222).

DECLARATION OF INTERESTS

All authors are current or former employees of Genentech, Inc.

REFERENCES

Alliot, F., Godin, I., and Pessac, B. (1999). Microglia derive from progenitors, originating from the yolk sac, and which proliferate in the brain. Brain Res. Dev. Brain Res. 117, 145–152.

Arumugam, T.V., Manzano, S., Furtado, M., Biggins, P.J., Hsieh, Y.-H., Gelderblom, M., MacDonald, K.P., Salimova, E., Li, Y.-I., Korn, O., et al. (2017). An atypical role for the myeloid receptor Mincle in central nervous system injury. J. Cereb. Blood Flow Metab. 37, 2098–2111.

Beach, T.G., Adler, C.H., Sue, L.I., Serrano, G., Shill, H.A., Walker, D.G., Lue, L., Roher, A.E., Dugger, B.N., Mauroot, C., et al. (2015). Arizona Study of Aging and Neurodegenerative Disorders and Brain and Body Donation Program. Neuropathology 35, 354–389.

Benitez, B.A., Jin, S.C., Guerreiro, R., Graham, R., Lord, J., Harold, D., Sims, R., Lambert, J.-C., Gibbs, J.R., Bras, J., et al. 3C Study Group; EADI consortium; Alzheimer’s Disease Genetic Consortium (ADGC); Alzheimer’s Disease Neuroimaging Initiative (ADNI); GERA Consortium (2014). Missense variant in TREML2 protects against Alzheimer’s disease. Neurobiol. Aging 35, 1510.e19–1510.e26.

Bennett, M.L., Bennett, F.C., Liddleow, S.A., Ajami, B., Zamanian, J.L., Fehm-off, N.B., Mulinyawe, S.B., Bohlen, C.J., Adil, A., Tucker, A., et al. (2016). New tools for studying microglia in the mouse and human CNS. Proc. Natl. Acad. Sci. USA 113, E1738–E1746.

Boche, D., Perry, V.H., and Nicoll, J.A. (2013). Review: activation patterns of microglia and their identification in the human brain. Neuropath. Appl. Neurobiol. 39, 3–18.

Bruttger, J., Karram, K., Wörtge, S., Regen, T., Marini, F., Hoppmann, N., Klein, M., Blank, T., Yona, S., Wolf, Y., et al. (2015). Genetic cell ablation reveals clusters of local self-renewing microglia in the mammalian central nervous system. Immunity 43, 92–106.

Butovsky, O., Jedrychowski, M.P., Moore, C.S., Cialic, R., Lanser, A.J., Gabrie, G., Koeglsperger, T., Baker, D., Wu, P.M., Doykan, C.E., et al. (2014). Identification of a unique TGF-β-dependent molecular and functional signature in microglia. Nat. Neurosci. 17, 131–143.

Chen-Potkin, A.S., Geser, F., Plotkin, J.B., Clark, C.M., Kwong, L.K., Yuan, W., Grossman, M., Van Deehin, V.M., Trojanowski, J.Q., and Lee, V.M-Y. (2008). Variations in the progranulin gene affect global gene expression in frontotemporal lobar degeneration. Hum. Mol. Genet. 17, 1349–1362.

Chiu, I.M., Morimoto, E.T.A, Goodarzi, H., Liao, J.T., O’Keefe, S., Phatnani, H.P., Muratet, M., Carroll, M.C., Levy, S., Tavazoie, S., et al. (2013). Neurodegeneration-specific gene-expression signature of acutely isolated microglia from an amyotrophic lateral sclerosis mouse model. Cell Rep. 4, 385–401.

Denk, F., Crow, M., Dicangelas, A., Lopes, D.M., and McMahon, S.B. (2016). Persistent alterations in microglial enhancers in a model of chronic pain. Cell Rep. 15, 1771–1781.

Dobrowolsny, G., Giacinti, C., Pelosi, L., Nicoletti, C., Winn, N., Barberi, L., Molinaro, M., Rosenthal, N., and Musaro, A. (2005). Muscle expression of a local IGF-1 isoform protects motor neurons in an ALS mouse model. J. Cell Biol. 168, 193–199.

Dodge, J.C., Haidet, A.M., Yang, W., Passini, M.A., Hester, M., Clarke, J., Roskelley, E.M., Treleaven, C.M., Rizo, L., Martin, H., et al. (2008). Delivery of AAV-IGF-1 to the CNS extends survival in ALS mice through modification of aberrant glial cell activity. Mol. Ther. 16, 1056–1064.

Emry, D., Hrabé de Angelis, A.L., Jatin, D., Wieghofer, P., Staszewski, O., David, E., Keren-Shaul, H., Mahlaokt, I., Jakobshagen, K., Buch, T., et al. (2015). Host microbiota constantly control maturation and function of microglia in the CNS. Nat. Neurosci. 18, 965–977.

Gabriel, T.L., Tol, M.J., Ottenhof, R., van Roomen, C., Claessen, N., Gabriel, T.L., Tol, M.J., Ottenhof, R., van Roomen, C., Claessen, N., Hooibrink, B., de Weijer, B., Seroie, M.J., Argmann, C., et al. (2014). Lysosomal stress in obese adipose tissue macrophages contributes to MITF-dependent Gpnmb induction. Diabetes 63, 3310–3323.

Ghosh, F., Greter, M., Leboeuf, M., Nandi, S., See, P., Gokhan, S., Mehler, M.F., Conway, S.J., Ng, L.G., Stanley, E.R., et al. (2010). Fate mapping analysis reveals that adult microglia derive from primitive macrophages. Science 330, 841–845.

Goldmann, T., Wieghofer, P., Jordão, M.J.C., Prutek, F., Hagermeyer, N., Frenzel, K., Amann, L., Staszewski, O., Kierdorf, K., Krueger, M., et al. (2016). Origin, fate and dynamics of macrophages at central nervous system interfaces. Nat. Immunol. 17, 797–805.
Götz, J., Chen, F., Barrett-Lee, R., and Nitsch, R.M. (2001). Tau filament formation in transgenic mice expressing P301L tau. J. Biol. Chem. 276, 529–534.

Grabert, K., Michoel, T., Karavolos, M.H., Clohisey, S., Baillie, J.K., Stevens, M.P., Freeman, T.C., Summers, K.M., and McColl, B.W. (2016). Microglial brain region-dependent diversity and selective regional sensitivities to aging. Nat. Neurosci. 19, 504–516.

Guerreiro, R., Wijesinghe, D., Bras, J., Carrasquillo, M., Rogaeva, E., Majounie, E., Cruchaga, C., Sassi, C., Kauwe, J.S.K., Yoonkin, S., et al.; Alzheimer Disease Neuroimaging Initiative (2013). TREM2 variants in Alzheimer’s disease. Nat. Engl. J. Med. 368, 117–127.

Gutknecht, M., Geiger, J., Joas, S., Dörfel, D., Salih, H.R., Müller, M.R., Grünbach, F., and Rittig, S.M. (2015). The transcription factor MITF is a critical regulator of GPNMB expression in dendritic cells. Cell Commun. Signal. 13, 19.

Hamerman, J.A., Jarjoura, J.R., Humphrey, M.B., Nakamura, M.C., Seaman, W.E., and Lanier, L.L. (2006). Cutting edge: inhibition of TLR and FcR responses by triggering receptor expressed on myeloid cells (TREM)-2 and DAP12. J. Immunol. 177, 2051–2055.

Heng, T.S.P., and Painter, M.W.; Immunological Genome Project Consortium (2008). The Immunological Genome Project: networks of gene expression in immune cells. Nat. Immunol. 9, 1091–1094.

Heppner, F.L., Ransohoff, R.M., and Becher, B. (2015). Immune attack: the immune cells. Nat. Rev. Immunol. 16, 358–372.

Hickman, S.E., Kingery, N., Osurni, T.K., Borowsky, M.L., Wang, L.C., Means, T.K., and El Khoury, J. (2013). The microglial sensome revealed by direct RNA sequencing. Nat. Neurosci. 16, 1896–1905.

Hodges, A., Strandus, A.D., Aragaki, A.K., Cruzhaga, C., Sassi, C., Kauwe, J.S.K., Yoonkin, S., et al.; Alzheimer Disease Neuroimaging Initiative (2013). TREM2 variants in Alzheimer’s disease. N. Engl. J. Med. 368, 117–127.

Keren-Shaul, H., Spinrad, A., Weiner, A., Matcovitch-Natan, O., Dvir-Sternhell, S., Rivest, S. (2009). Regulation of innate immune responses in the brain. Nat. Rev. Immunol. 9, 119–145.

Kraemer, H.R., Korczyn, A.D., and Raskind, M.W. (1999). Transcriptional profiling of CD11c-positive microglia accumulating around amyloid plaques in a mouse model for Alzheimer’s disease. Biochem. Biophys. Acta 1862, 1847–1860.

Keren-Shaul, H., Spinrad, A., Weiner, A., Matcovitch-Natan, O., Dvir-Sternhell, S., Ulland, T.K., David, E., Baruch, K., Lara-Astaiso, D., Toth, B., et al. (2017). A unique microglia type associated with restricting development of Alzheimer’s disease. Cell 169, 1276–1290.e17.

Lavin, Y., Winter, D., Blecher-Gonen, R., David, E., Keren-Shaul, H., Merad, M., Jung, S., and Amit, I. (2014). Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. Cell 159, 1312–1326.

Li, C.-C., Liu, C.C., Kanekiyo, T., Xu, H., and Bu, G. (2013). Apolipoprotein E and Alzheimer disease: risk mechanisms and therapy. Nat. Rev. Neuro. 10, 115–118.

Matcovitch-Natan, O., Winter, D.R., Giladi, A., Vargas Aguilar, S., Spinrad, A., Sarrazin, S., Ben-Yehuda, H., David, E., Zelada Gonzalez, F., Perrin, P., et al. (2016). Microglia development follows a stepwise program to regulate brain homeostasis. Science 353, aad8670.

Mostafavi, S., Yoshida, H., Moodley, D., LeBoit, H., Rothamel, K., Raj, T., Ye, C.J., Chevrier, N., Zhang, S.-Y., Feng, T., et al.; Immunological Genome Project Consortium (2016). Parsing the interferon transcriptional network and its disease associations. Cell 164, 564–578.
cell taxonomy revealed by single cell transcriptomics. Nat. Neurosci. 19, 335–346.

Turnbull, I.R., Gilfillan, S., Cell, M., Aoshi, T., Miller, M., Piccio, L., Hernandez, M., and Colonna, M. (2008). Cutting edge: TREM-2 attenuates macrophage activation. J. Immunol 177, 3520–3524.

Vergani, L., Losa, M., Lesma, E., Di Giulio, A.M., Torsello, A., Müller, E.E., and Gorlo, A. (1999). Glycosaminoglycans boost insulin-like growth factor-1-promoted neuroprotection: blockade of motor neuron death in the wobbler mouse. Neuroscience 93, 565–572.

Verheijden, S., Beckers, L., Casazza, A., Butovsky, O., Mazzone, M., and Baes, M. (2015). Identification of a chronic non-neurodegenerative microglia activation state in a mouse model of peroxisomal β-oxidation deficiency. Glia 63, 1606–1620.

Vivash, L., and O’Brien, T.J. (2016). Imaging microglial activation with TSPO PET: Lighting up neurologic diseases? J. Nucl. Med. 57, 165–168.

Walker, D.G., and Lue, L.-F. (2015). Immune phenotypes of microglia in human neurodegenerative disease: challenges to detecting microglial polarization in human brains. Alzheimer’s Res. Ther. 7, 56.

Wang, Y., Cella, M., Mallinson, K., Ulrich, J.D., Young, K.L., Robinette, M.L., Gilfillan, S., Krishnan, G.M., Sudhakar, S., Zinselmeyer, B.H., et al. (2015). TREM2 lipid sensing sustains the microglial response in an Alzheimer’s disease model. Cell 160, 1061–1071.

Wang, Y., Ulland, T.K., Ulrich, J.D., Song, W., Tzaferis, J.A., Hole, J.T., Yuan, P., Mahan, T.E., Shi, Y., Gilfillan, S., et al. (2016). TREM2-mediated early microglial response limits diffusion and toxicity of amyloid plaques. J. Exp. Med. 213, 667–675.

Webster, J.A., Gibbs, J.R., Clarke, J., Ray, M., Zhang, W., Holmans, P., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., et al.; NACC-Neuropathology Group (2009). Genetic control of human brain transcript expression in Alzheimer disease. Am. J. Hum. Genet. 84, 445–458.

Yoshiyama, Y., Higuchi, M., Zhang, B., Huang, S.-M., Iwata, N., Saio, T.C., Maeda, J., Suhara, T., Trojanowski, J.Q., and Lee, V.M.-Y. (2007). Synapse loss and microglial activation precede tangles in a P301S tauopathy mouse model. Neuron 53, 337–351.

Yuan, P., Condello, C., Keene, C.D., Wang, Y., Bird, T.D., Paul, S.M., Luo, W., Colonna, M., Baddeley, D., and Grutzendler, J. (2016). TREM2 haploinsufficiency in mice and humans impairs the microglia barrier function leading to decreased amyloid compaction and severe axonal dystrophy. Neuron 90, 724–739.

Zeil, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Jurés, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science 347, 1138–1142.

Zhang, B., Gaiteri, C., Bodea, L.-G., Wang, Z., McElwee, J., Podtelezhnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., et al. (2013). Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer’s disease. Cell 153, 707–720.

Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O’Keeffe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Rudierisch, N., et al. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J. Neurosci. 34, 11929–11947.

Zhang, Y., Sloan, S.A., Clarke, L.E., Caneda, C., Plaza, C.A., Blumenthal, P.D., Vogel, H., Steinberg, G.K., Edwards, M.S.B., Li, G., et al. (2016). Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. Neuron 89, 37–53.

Zigdon, H., Savidor, A., Levin, Y., Meshcheriakova, A., Schiffmann, R., and Futterman, A.H. (2015). Identification of a biomarker in cerebrospinal fluid for neuronopathic forms of Gaucher disease. PLoS ONE 10, e0120194.
Supplemental Information

Diverse Brain Myeloid Expression Profiles Reveal
Distinct Microglial Activation States and Aspects
of Alzheimer's Disease Not Evident in Mouse Models

Brad A. Friedman, Karpagam Srinivasan, Gai Ayalon, William J. Meilandt, Han Lin, Melanie A. Huntley, Yi Cao, Seung-Hye Lee, Patrick C.G. Haddick, Hai Ngu, Zora Modrusan, Jessica L. Larson, Joshua S. Kaminker, Marcel P. van der Brug, and David V. Hansen
Figure S1. CNS myeloid gene expression datasets reveal common co-regulated gene modules reflecting proliferative state and other aspects of CNS myeloid biology, Related to Figures 1-4. This figure shows the same heatmap as Figure 1A but with much greater detail and additional annotations. The rows correspond to the 777 genes differentially expressed in the greatest number of conditions (≥ 7) in the database. Each column represents a different sample in one of the studies in the database. The samples are grouped by study, and then further ordered by sample meta-data, with normal adult brain/cortical microglia/myeloid cells always preceding the experimental conditions (e.g., transgenic, treated, peripheral cells, perivascular macrophages, infiltrating macrophages, aged mice, or cerebellar myeloid cells). Full descriptions and references for all the studies are available in Supplemental Data 1, with these studies listed as being used in this manuscript for “Mouse Myeloid Activation Gene Clustering”. The heatmap shows Z-score gene expression, normalized separately within each study.

To the left of this central heatmap are additional gene annotations. The first two columns show Parkinson’s (Nalls et al., 2014) and Alzheimer’s disease (Lambert et al., 2013) GWAS hits (labeled with gene symbols). The next three are cell type marker sets derived from Barres lab sorted cell RNA-Seq (Zhang et al., 2014, 2016) (human, GSE73721, and mouse, GSE52564), and Allen Brain Atlas single-cell RNA-Seq (Tasic et al., 2016) (GSE71585). These demonstrate that many but not all of the genes are selectively expressed in normal myeloid cells. The last three columns show previously published sets of genes that were proposed to be specific to microglia (Butovsky et al., 2014; Chiu et al., 2013; Hickman et al., 2013). (The MG400 set includes both microglial markers and a few controls that were included in a nanostring panel, since we were unable to distinguish between these based on the original manuscript or annotations available in GEO (GPL18002).) These three gene sets are enriched in the “Microglia” gene module, but also include genes in other modules.
Experimental conditions are labeled above the sample groups, and broad study categories are also indicated for each study. Finally, the seven selected “super-groups” discussed extensively in the text are indicated on the right of the heatmap. The “Brain Myeloid” super-group is also indicated, although it is mentioned only briefly in the text and not included in the Supplemental files.

Supplemental Data 3 is an Excel file enabling further mining of these data. It contains one row for each gene in this figure, with columns giving the gene meta-data, clustering results, and expression and differential expression statistics across all of these datasets. (Supplemental Data 2 has the same content except that it is organized by human genes rather than mouse genes.)
Figure S2. Genes in the “Neutrophil/Monocyte” modules are selectively expressed in neutrophils and monocytes relative to other peripheral immune cells, Related to Figure 4C. (A) Heatmap showing genes in module across immune cell populations from ImmGen (Heng et al., 2008). Each column represents average expression taken across multiple replicates of a particular immune cell population. (B) Mean-centered gene expression plot (“W-plot” on ImmGen website) showing relative gene expression values for each gene in the set within each cell type. Both plots were made on the ImmGen website, http://rstats.immgen.org/MyGeneSet/.
Figure S3. Single-cell clusters represent distinct cell types and activation states, Related to Figure 6. Like Figure 6, but small sets of 1-3 genes were manually selected to highlight the identities of the various cell clusters.
Figure S4. Relationship between age, microglial content, and disease status in three bulk Alzheimer’s disease tissue gene expression datasets, Related to Figure 7. (A,D,G) Age of samples (in years) stratified by disease status and sex in GSE95587 (fusiform gyrus, this study), GSE15222 (temporal cortex (Webster et al., 2009)) and GSE44770 (prefrontal cortex (Zhang et al., 2013)). The latter shows significant differences in age between controls and AD patients, including patients much younger than seen in the other datasets. (B,E,H) Myeloid content score (based on gene expression data, see Methods) stratified by sex and disease status. (C,F,I) Correlation of age and myeloid content. (I) shows that myeloid content increases steadily as patient age increases from <50 through >90 years old. The relative paucity of young (<70 years old) donors in the first two datasets probably explains the reduced correlation seen there.
Figure S5. Orthologs of Neurodegeneration-Related genes identified in mouse myeloid cells are elevated in Alzheimer’s disease fusiform gyrus, however this signal is not seen in a microarray study of temporal cortex, Related to Figure 7. (A) Heatmap showing human orthologues of mouse myeloid Neurodegeneration-Related genes on the rows, and samples from control or Alzheimer’s disease bulk fusiform gyrus tissue on the columns (GSE95587). Color indicates Z-score. See Supplemental Data 4 for the full list of genes (annotated with “Neurodegeneration” in the “Myeloid Activation (Coarse)” column). Samples within control and AD groups are ordered by first principal component, and genes are ordered by correlation with overall gene set score. Below, gene set score for each sample (color scale not shown, however red indicates higher score and blue indicates lower score). Horizontal lines visually separate groups of genes based on their differential expression in AD bulk tissue. (B) Neurodegeneration-Related gene set scores in control and AD cohorts. (C) Empirical cumulative distribution function (ECDF) showing the distribution of correlations between (log-scale) expression of individual genes and gene-set scores. Compare to Panels 7C, 7F, 8C and 8F of Supplemental Data 7. (D-F) Like A-C, but for the microarray study (GSE15222 (Webster et al., 2009)). Even though there is broad agreement with the RNA-Seq data from panels A-C, the signal is essentially absent. There appear to be more “noisy” genes, which might be explained by lower sensitivity of this microarray platform relative to RNA-Seq. (G, H) Relative expression of these genes in human and mouse sorted cells (like Panels 7G-H and 8G-H of Supplemental Data 7) shows that the genes anti-correlated with the overall signal are enriched for neuronal expression, especially in the human sorted cells (panel G). The gray colors for the bottom few genes in panels A, D and G indicate mouse genes without annotated human orthologues. Calculations for module enrichment in this figure are prior to any analytical adjustments for altered cellularity in AD tissues. See Supplemental Data 7 Panel 10 for exclusion of genes preferentially expressed in neurons, and see Figure 7 for final calculations using a subset of samples with balanced myeloid content (established in Figure S6).
**Figure S6. Myeloid-balancing.** Related to Figure 7. (A) Left two groups, myeloid gene set score for control (CON Orig.) and AD (AD Orig.) samples from RNA-Seq dataset. There is a modest but statistically significant elevation in myeloid score. These two distributions are the same data as shown in Supplemental Data 7 Panel 6, in the “Myeloid” row for the “Fusiform gyrus” Control and AD cohorts. Right two groups, random subsets of similar myeloid score were selected (see Methods). The absolute difference (Δ) and statistical significance (P) between the groups are both effectively eliminated by this procedure. (B) Similar to panel A, but for the temporal cortex dataset. Again, the left panels are the same data as shown in Supplemental Data 7 Panel 6 in the “Myeloid” row, but for the “Temporal Cortex” dataset.
Figure S7. Many myeloid activation modules are elevated in bulk AD patient tissue, Related to Figure 7. (A-G) Like Figure 7D,E,G,H,I, gene Set scores (y-axis) of myeloid activation modules in control and AD patient tissue (points). GSE95587 (fusiform gyrus dataset) on left, GSE15222 (temporal cortex dataset) on right. (H, I) Like A-G, but for LPS-Specific and LCMV-Specific gene sets (see Methods, and genes are given in “Immune-Specific” column of Supplemental Data 2). Calculations for module enrichment in this figure are after analytical adjustments for decreased neuronal content and increased myeloid content in AD tissues.
Supplemental Experimental Procedures

Gene Expression Analysis

When possible, raw data (Affymetrix .CEL files, Agilent .txt files, or RNA-Seq FASTQ files) were downloaded from public repositories (Gene Expression Omnibus, Short Read Archive or ArrayExpress). Affymetrix and Agilent microarrays, and RNA-Seq datasets were analyzed as described (Srinivasan et al., 2016), except as specified below. Briefly, microarray analyses resulted in matrices of log2-scale probe intensities and RNA-Seq analyses resulted in matrices of size-factor-normalized RPKMs (nRPKM), which were then stabilized and log2-transformed with the transformation log2(nRPKM + 1). These log2-scale gene expression values, from microarray or RNA-Seq, were Z-score transformed and used as input into the heatmap clustering. Differential expression for microarray studies was performed with limma (Ritchie et al., 2015), and either DESeq2 (Love et al., 2014) or voom+limma (Law et al., 2014) was used for RNA-Seq studies, as indicated in the Supplemental files.

A few of the datasets required slightly different handling than previously described. GSE74615 was an Agilent (2-color) microarray datasets. However, instead of using a universal reference RNA in the “reference” channel, the authors included test samples in both channels. We treated the two channels as separate expression profiles. We controlled for their correlation by including an “array” term in the linear model used for differential expression.

GSE62710 and GSE79812 are RNA-Seq projects that use a special 3’-end library strategy, GSE60361 is a single-cell RNA-Seq project, and GSE15222 is an Illumina microarray project. Since we did not already have existing bioinformatics pipelines for these types of data we downloaded the author’s gene expression matrices provided by GEO (or, for GSE60361, directly from the authors’ website, https://storage.googleapis.com/linnarsson-lab-www-blobs/blobs/cortex/expression_mRNA_17-Aug-2014.txt).

For GSE62710 and GSE79812 the provided normalized expression values were already set away from 0, with a minimum value of 0.5, so were not further offset before log2 transformation. limma was used for differential expression analysis of these projects.

For GSE60361 we further normalized the author-provided UMI counts by the total UMIs for each cell, generating a norm_mRNA_mol statistic. As described above for other RNA-Seq projects, these values were offset 1 unit before log2 transformation. Our analysis focused only on those cells that the authors identified as being myeloid cells (level1class="microglia"). The authors further subdivided these as perivascular macrophages (level2class="Pvm1" or "Pvm2") or parenchymal microglia (level2class="Mgl1" or "Mgl2"). 12 microglia cells with expression of the macrophage marker Mrc1 higher than 5 norm_mRNA_mol were also set aside from this analysis as we found that these cells tended to express most macrophage markers and few of the microglial markers. These cells had the following IDs: 1772060226_D02, 1772058177_A07, 1772062109_E05, 1772062128_C02, 1772060224_H07, 1772067094_C07, 1772067096_G06, 1772062226_H12, 1772060900_B06, 1772062118_E06, 1772062109_C05 and 1772062128_B03. This left 21 microglial cells and 65 perivascular macrophages. For differential expression analysis between these groups, we first filtered out low-expressed genes, next performed Mann-Whitney tests on the norm_mRNA_mol statistic, and finally adjusted P-values by the Benjamini-Hochberg method. The Log2-fold-change was defined simply as the Log2-fold-change between the median expressions of the cells in the two groups. For the purposes of hierarchical clustering to define the gene modules each cell was treated individually, on par with samples from the other bulk expression datasets (rather than combining into an aggregate signature).

Finally, for GSE15222, the author-provided gene expression matrix were further normalized with the justvsn function from the vsn R package (Huber et al., 2002) before use in this analysis.
Hierarchical Clustering to create gene modules

After Z-score normalizing the (possibly offset and) log2-transformed expression matrices within each project, the genes most commonly differentially expressed across all projects were used to generate the comprehensive heatmap in Figure S1 and perform hierarchical clustering. “Differentially expressed” for the purposes of this analysis was considered adjusted P-value ≤ 0.05 and fold-change at least 1.5 (up or down). Clustering was performed with R’s hclust algorithm using the Euclidean distance metric.

To select the roughly 30 genes to display for each dataset in Figures 2-4 we calculated, for each comparison, the average log2 fold-change for the genes in the set. We then ranked the genes in the set by the correlation with this average profile, and showed the top genes.

Cell Type Markers

We used three different datasets to create lists of CNS cell-type markers: Allen Brain Atlas single-cell RNA-Seq (Tasic et al., 2016) (“ABA”), and RNA-Seq of sorted mouse (Zhang et al., 2014) and human (Zhang et al., 2016) cells from the Barres lab (“Barres mouse” and “Barres human”). We made several adjustments to the author-provided cell type classes as follows, and also excluded certain samples. For ABA the adjustments were based on the “primary type” and described below. For Barres mouse we set aside both the “newly formed oligodendrocyte” samples and the “whole cortex” samples. The reason for this is that our approach for calling cell type markers excludes genes observed at moderate expression levels in other “cell types”, so these samples would result in loss of otherwise good markers. For Barres human we set aside any samples that came from diseased tissue (Sclerotic Hippocampus, Tumor Periphery or Tumor Core), and also set aside the 6 fetal samples, so that we only analyzed normal adult cells.

Next, for all datasets, we focused only on annotated protein coding genes, and applied up to three different filters to identify cell type markers. First, we required that a gene be expressed at a minimum level in a minimum number of the samples (or cells) of the given cell type. For ABA the minimum expression level was 1 nRPKM in at least 60% of the cells of the given cell type. For both Barres datasets the cutoff was at least 3 nRPKM, either in all (mouse, in which there were only 2 samples per cell type) or in at least 80% (human, in which some cell types had as many as 12 samples) of the samples.

The second filter was that if a gene is to be considered a cell type marker it should not be expressed in any of the other cell types. This filter was applied in turn to each of the other cell types. For example, for ABA, the filter was that, for each of the other cell types, the expression had to be less than 1 nRPKM in 75% of the cells.

For the Barres datasets this filter was not used. Instead, an enrichment ratio was calculated as (expression in cell type of interest) / (expression in other cell types), and was required to be at least 10 relative to all other cell types. The numerator was defined as the minimum expression in the cell type of interest + 0.3. For each of the other cell types, the denominator was defined for the mouse as the maximum expression + 0.3, and for human at the 80th percentile of expression in the other cell types + 0.3.

The final gene sets for all cell types from each of the three datasets can be found in Supplemental Data 2, 3 and 5.

Analysis of Allen Brain Atlas Single-Cell RNA-Seq

“primary type” provided by authors was aggregated into six main cell types as follows:

| Cell type group for this study | “Primary type”s from authors | Number of cells |
|-------------------------------|-------------------------------|-----------------|
| Astro (astrocyte)             | Astro Gja1                    | 43              |
| Endo (endothelial)            | Endo Myl9                     | 26              |
|                               | Endo Tbc1d4                   |                 |
| Type                  | Cells | Omitted |
|----------------------|-------|---------|
| **Excitatory (neurons)** | L2/3 Otof, L2 Ngb, L4 Sparc11, L5a Batf3, L4 Scnn1a, L5a Syt17, L6b Trh, L6b Mup5, L4 Arf5, L5 Hsd11b1, L5a Fam5c, L6a Plcx3, L6 Syt17, L6a Mgp, L6 Car12, L5 Ucma, L5 Chrna6, L5b Samd3, L5b Cd113 | 609 |
| **GABAergic (neurons)** | Vip Mybpc1, Vip Cxcl14_Car4, Vip Chat, Ndnf Cxcl14, Vip Gpc3, Vip Sneg, Ndnf Car4, Sst Th, Sst Myh8, Pvalb Gpx3, Sst Cdk6, Pvalb Wt1, Pvalb Tacr3, Sst Cbhn4, Sneg, Igtp, Pvalb Cpne5, Pvalb Rsopo2, Cd34, Sst Tacstd2, Sst Chodl, Pvalb Obox3, Pvalb Tpbg | 664 |
| **Micro (microglia)** | Micro Ctss | 22 |
| **Oligo (oligodendrocytes)** | Oligo Opalin | 30 |
| **OPC (oligodendrocyte precursor cells)** | OPC Pdgfra | 22 |
| **Omitted*** |  | 262 |
| **Total** |  | 1678 |

* Cells with an annotated “secondary type” were omitted from the analysis in this study, as well as 8 cells with the primary type of “Oligo/OPC”, which had an intermediate phenotype between the precursor and mature cells.
**Gene Set Scores**

For a set of genes, the “gene set score” $s_j$ was calculated for each sample $j$ in a project. We started with the log2-scale expression matrix described above, which we call $E$ here (with $e_{ij}$ representing the expression of gene $i$ in sample $j$). First, the centered gene expression matrix $C$ was calculated as

$$c_{ij} = e_{ij} - \frac{1}{n_s} \sum_{j} e_{ij}$$

Where $n_s$ represents the number of samples in the dataset. $c_{ij}$ represents the log2-fold-change in expression of gene $i$ in sample $j$ relative to that gene’s average log2 expression in all samples in the project. The gene set score for sample $j$ (notated here as $s_j$) is then defined as the column-average of the matrix $C$, or

$$s_j = \frac{1}{n_g} \sum_{i} c_{ij}$$

where $n_g$ is the number of genes in the gene set.

Several figures (Supplemental Data 7 Panels 7, 8 and 10; Figure S5 and Figure 7) visualize gene sets in bulk tissue. These show genes on the rows, and samples from control or Alzheimer’s disease tissue on the columns. Color indicates Z-score. Samples within control and AD groups are ordered by first principal component, and genes are ordered by correlation with overall gene set score. Below the heatmaps, gene set scores are shown for each sample (color scale not shown, however red indicates higher score and blue indicates lower score).

**Trem2 Dependence Analysis**

We calculated the “percent Trem2-dependence” of each gene in the 5XFAD model (GSE65067; Figure 5A). Focusing only on genes induced in the 5XFAD model in Trem2 wildtype animals (at the relaxed cutoff of $P \leq 0.1$), those which were also induced to the same extent (or more) in Trem2 knockouts were considered to be 100% Trem2 independent (equivalently, 0% Trem2 dependent). Those which were not induced at all (or even suppressed) in the Trem2 knockouts were considered to be 100% Trem2 dependent, and the percentage of Trem2 dependence was determined for other 5XFAD-induced genes based on the ratio of fold-inductions in Trem2 knockout and wildtype microglia, according to these formulas:

$$LFC_{WT} = \log_2 \left( \frac{Trem2^{WT};5XFAD}{Trem2^{WT};Non-Tg} \right)$$

$$LFC_{KO} = \log_2 \left( \frac{Trem2^{KO};5XFAD}{Trem2^{KO};Non-Tg} \right)$$

Percent Trem2 Dependence =

$$\begin{cases} 
100 \times \frac{LFC_{WT} - LFC_{KO}}{LFC_{WT}} & \text{if } 0 \leq LFC_{KO} \leq LFC_{WT} \\
100 & \text{if } LFC_{KO} \leq 0 \\
0 & \text{if } LFC_{WT} \leq LFC_{KO}
\end{cases}$$

For the microglial module, which is down-regulated in 5XFAD microglia, we considered instead genes which were down-regulated. The ratios for $LFC_{WT}$ and $LFC_{KO}$ were similarly inverted (this matters for the sub-domain definitions on the right, but not in the values of the sub-functions themselves).

**Myeloid Balancing**

In order to balance the myeloid content of two groups (here Alzheimer’s and control) we started with the calculated myeloid gene set scores $s_j$ for all of the samples. The full range of scores across either group was then split into 20 equal-width bins. Then, the number of samples within each group and bin was counted,
with \(n_{bg}\) representing the number the samples in bin \(b \in \{1, \ldots, 20\}\) in group \(g = 1\) or \(2\). We defined the set of “balanceable” bins \(B \subseteq \{1, \ldots, 20\}\) as those bins \(b\) with \(n_{bg} \geq 1\) for both \(g = 1\) and \(2\) (that is, it has samples from both groups). The remaining bins, and the samples they contain, are considered “unbalanceable” and set aside for this analysis. We then calculate the fraction of balanceable samples within each of the balanceable bins, separately for each group:

\[
f_{bg} = \frac{n_{bg}}{\sum_{b \in B} n_{bg}}
\]

Finally, we compare \(f_{b1}\) and \(f_{b2}\) for each balanceable bin. If they are unequal (which they usually are), then group with a larger fraction of samples in this bin is down-sampled by randomly including each sample with a probability equal to the ratio of these fractions (that is, the smaller of \(f_{b1}/f_{b2}\) and \(f_{b2}/f_{b1}\)). The outcome of this procedure is shown in Figure S6.

**GO Analysis**

The follow GO terms were used in the description of the neurodegenerative modules:

| Category     | GO ID     | GO Term                                      | GO Definition                                                                                                                                 |
|--------------|-----------|----------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| Membrane     | GO:0016021| integral component of membrane              | The component of a membrane consisting of the gene products and protein complexes having at least some part of their peptide sequence embedded in the hydrophobic region of the membrane. |
|              | GO:0005886| plasma membrane                              | The membrane surrounding a cell that separates the cell from its external environment. It consists of a phospholipid bilayer and associated proteins.       |
|              | GO:0009897| external side of plasma membrane             | The leaflet of the plasma membrane that faces away from the cytoplasm and any proteins embedded or anchored in it or attached to its surface.       |
| Extracellular| GO:0070062| extracellular exosome                        | A membrane-bounded vesicle that is released into the extracellular region by fusion of the limiting endosomal membrane of a multivesicular body with the plasma membrane. Extracellular exosomes, also simply called exosomes, have a diameter of about 40-100 nm. |
|              | GO:0031012| extracellular matrix                         | A structure lying external to one or more cells, which provides structural support for cells or tissues; may be completely external to the cell (as in animals and bacteria) or be part of the cell (as in plants). |
|              | GO:0005615| extracellular space                          | That part of a multicellular organism outside the cells proper, usually taken to be outside the plasma membranes, and occupied by fluid.             |
| Transcription factor | GO:0003700| transcription factor activity, sequence-specific DNA binding | Interacting selectively and non-covalently with a specific DNA sequence in order to modulate transcription. The transcription factor may or may not also interact selectively with a protein or macromolecular complex. |
GO:0000981 | RNA polymerase II transcription factor activity, sequence-specific DNA binding | Interacting selectively and non-covalently with a specific DNA sequence in order to modulate transcription by RNA polymerase II. The transcription factor may or may not also interact selectively with a protein or macromolecular complex.

Lysosome | GO:0005764 | lysosome | A small lytic vacuole that has cell cycle-independent morphology and is found in most animal cells and that contains a variety of hydrolases, most of which have their maximal activities in the pH range 5-6. The contained enzymes display latency if properly isolated. About 40 different lysosomal hydrolases are known and lysosomes have a great variety of morphologies and functions.

### Immune-Specific Gene Sets

For Figure 7 and Figure S7 we generated “LPS-Specific” and “LCMV-Specific” gene sets based on the sorted mouse myeloid data sets. “LPS-Specific” was defined as genes that were significantly induced (adjusted \( P \leq 0.05 \), any fold-change > 1) in both LPS studies (GSE75246 and GSE67858), but not significantly changed (adjusted \( P > 0.05 \)) in response to LCMV (GSE67858), in any of the \( \beta \)-amyloid models (GSE89482, GSE65067 or GSE7461), tau models (GSE93179 or GSE93180) or SOD1 (GSE43366). Similarly, “LCMV-Specific” were those genes induced by LCMV (GSE67858), but not in either of the LPS datasets or any of these six neurodegenerative model datasets. The full list of these genes can be found in the “Immune-Specific” column of Supplemental Data 2 and 3.

### Mice

All protocols involving animals were approved by Genentech’s Institutional Animal Care and Use Committee, in accordance with guidelines that adhere to and exceed state and national ethical regulations for animal care and use in research. For GSE89482, Cx3cr1::GFP (Jackson Labs #005582) mice were crossed to our PS2APP colony, and microglia from the cortex of Cx3cr1\(^{GFP/+}\);PS2APP\(^{-}\) mice (“Non-transgenic”) and Cx3cr1\(^{GFP/+}\);PS2APP\(^{homozygous}\) (“PS2APP”) mice were compared. Hippocampal microglia were collected from hMAPT-P301L\(^{homozygous}\) and hMAPT-P301S\(^{homozygous}\) mice and their non-transgenic littermate controls, for GSE93179 and GSE93180, respectively. At 6 (P301S), 12 (P301L), or 14-15 (PS2APP) months of age animals were perfused and processed in control and transgenic pairs using two BD FACSAria sorters simultaneously. Cells were sorted and RNA-Seq libraries prepared using Ovation RNA-Seq System V2 (NuGEN) as previously described (Srinivasan et al., 2016), except cells were sorted live when the GFP marker was used. For P301S, of 9 transgenic animals that were processed for RNA-Seq, 3 did not show any microglial transcriptional responses and were excluded from this study. Average sequencing depths were 28 million (GSE89482), 25 million (GSE93179) and 30 million (GSE93180) total reads. Average percent of reads aligning to exons was 20% (GSE89482), 11% (GSE93179) and 8% (GSE93180), typical values for our NuGEN-amplified RNA-Seq.

### Analysis of Single-Cell RNA-Seq Dataset GSE98969

Matrix of UMI counts was downloaded from GEO. Gene symbols were mapped, where possible, to our internal gene annotation. For unmapped genes, unique aliases were considered. 95% of total UMIs were maintained through this procedure, which enabled us to easily compare gene expression to other data sets. Cells with less than 500 resulting UMIs were discarded. Normalized UMIs were then calculated based on this “derived” count matrix, dividing the counts for each sample by the factor totalCounts/medianTotalCounts. The 13,586 CD45+ sorted cells from 6 month old Trem2-WT mice of the “5XFAD” cohort (including C57BL/6 controls) were analyzed further, including cells from whole-brain preparations as well as cortical and cerebellar preparations. “Detected” genes were selected as those with
normalized UMI ≥ 1.5 in at least 68 cells (0.5%). Principal components and tSNE dimensional reductions (using Rtseq, perplexity=30, theta=0.5 using Rtseq (Krijtje, 2017)) were performed on the matrix of Log2(normalized UMI + 1). tSNE coordinates were used for visualization in Figures 6 and S3. The first 20 principal components were then used to generate sample clusters using the function from FindClusters() (with k.param=20 and k.scale=20) from the Seurat package (Butler and Satija, 2017; Satija et al., 2015).

For Figure 6B′ and Figure S3B′, the statistic “5XFAD Enrichment” was calculated as follows. Let \( n_{cg} \) denote the number of cells in cluster \( c \) of genotype \( g \), and let \( n_g = \sum_c n_{cg} \) denote the total number of cells of genotype \( g \). The cell numbers were converted to fraction for each genotype as \( f_{cg} = n_{cg} / n_g \). Then “5XFAD Enrichment” for cell cluster \( c \) was defined as the ratio of these fractions, \( f_{c,5XFAD} / f_{c,57BL/6} \).

**Patient Bulk Tissue RNA-Seq**

Frozen fusiform gyrus tissue blocks and pathology/clinical reports, including age, sex, diagnosis and Braak stage, were obtained from the Banner Research Institute. RNA was extracted from approximately 300 µg of frozen sections from each tissue block as described (Benitez et al., 2014) and standard polyA-selected Illumina RNA-Seq analysis was performed as described (Srinivasan et al., 2016) on samples with RNA integrity (RIN) scores at least 5 and post-mortem intervals (PMI) no greater than 5 hours. Samples were sequenced to an average depth of 35 million reads. Although the distribution of PMIs did not significantly differ between Control and AD samples, Control samples had somewhat higher RIN scores. Using DESeq2 linear models we analyzed the dependence of gene expression RIN score and PMI. Although very few (<10) genes showed any expression dependence on PMI, tens to hundreds showed some dependence on RIN score. However, when we compared differential expression results in models with and without RIN scores (~Diagnosis versus ~Diagnosis + RIN) we found them to be very similar. We also examined the correlation of Myeloid cell type score, as well as activation scores for the seven main gene sets described in this manuscript. None of these scores, which average over many genes, were found to depend on RIN score. Therefore we did not include RIN scores directly in any of the analyses in this manuscript. RIN scores and PMIs are available as sample characteristics from the GEO website.

**Mouse Brain Immunohistochemistry**

Immunohistochemistry (Supplemental Data 7 Panel 4A-D) was performed as described (Lee et al., 2016) using Phosho-tau (Ser202, Thr205) monoclonal antibody (AT8, ThermoFisher #MN1020, lot #LB1412200) or anti-CD68 (clone FA-11, Serotec #MCA1957). Whole slide brightfield images were acquired using the Leica SCN400 or Hamamatsu Nanozoomer system. Regions of interest consisting of the hippocampus were manually drawn and positive pixel area and integrated intensity analysis was performed using Matlab as previously described (Lee et al., 2016).

**Resource Table**

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| **Antibodies**      |        |            |
| AlexaFlour 488-conjugated anti-NeuN | Millipore | Cat#MAB377X |
| PE-conjugated anti-GFAP | BD Biosciences | Cat#561483 |
| APC-conjugated anti-CD11b | BD Biosciences | Cat#561690 |
| rat anti-CD68 (clone FA-11) | Serotec | Cat#MCA1957 |
| Phospho-Tau (Ser202, Thr205) Monoclonal Antibody (AT8) | ThermoFisher | Cat#MN1020; Lot# LB1412200 |
| **Biological Samples** |        |            |
| Alzheimer’s disease and control Fusiform Gyrus | Banner Research Institute | https://www.bannerhealth.com/research |
| **Chemicals, Peptides, and Recombinant Proteins** |        |            |
| Accutase | Millipore | Cat#SCR005 |
### Critical Commercial Assays

| Assay | Manufacturer | Catalog Number |
|-------|--------------|---------------|
| Ovation RNA-Seq System V2 | Nugen | Cat#7102-A01 |
| TruSeq RNA Sample Preparation Kit v2 | illumina | Cat#RS-122-2001 |

### Deposited Data (see Supplemental Data 1 for more details)

| Condition | Cells | Method | Publication | GEO Accession |
|-----------|-------|--------|-------------|---------------|
| Cortical Cx3cr1-GFP cells in 14-15mo PS2APP model (RNA-Seq) | | This paper | GEO: GSE89482 |
| Cortical CD11b cells in 7/13mo PS2APP model (RNA-Seq) | | Pubmed: 27097852 | GEO: GSE75431 |
| WT and TREM2-deficient microglia in 5xFAD model (MoGene) | | Pubmed: 25728668 | GEO: GSE65067 |
| Cortical microglia in APPswe/PS1dE9 model (WMG4x44kv2) | | Pubmed: 25002035 | GEO: GSE74615 |
| Hippocampal CD11b cells in Tau-P301L model (Mouse RNA-Seq) | | | GSE93179 |
| Hippocampal CD11b cells in Tau-P301S model (Mouse RNA-Seq) | | | GSE93180 |
| Spinal cord CD11b cells in SOD1-G93A model (Mouse RNA-Seq) | | Pubmed: 23850290 | GEO: GSE43366 |
| Brain microglia in Mfp2-deficient mice (MoGene) | | Pubmed: 25846981 | GEO: GSE66420 |
| High-resolution transcriptome analysis reveals neuropathic pain gene-expression signatures in spinal microglia after nerve injury (MoGene) | | Pubmed: 26761385 | GEO: GSE60670 |
| Injury-induced enhancer remodelling in spinal microglia (Mouse RNA-Seq) | | Pubmed: 27184839 | GEO: GSE71133 |
| Microglia response to transient ischemic injury (MoGene) | | Pubmed: 27492949 | GEO: GSE77986 |
| WT and TREM2-deficient microglia in response to cuprizone-mediated demyelination (MoGene) | | Pubmed: 25893602 | GEO: GSE66926 |
| Cortical CD11b cells from LPS-injected mice (Mouse RNA-Seq) | | Pubmed: 27097852 | GEO: GSE75246 |
| Commensal microbiota constantly control maturation and function of microglia in the central nervous system (Mouse RNA-Seq) | | Pubmed: 26030851 | GEO: GSE62710 |
| Microglia from LPS-injected or LCMV-infected mice (MoGene) | | Pubmed: 26030851 | GEO: GSE67858 |
| Glioma-associated CD11b cells and naive control cells, mouse (MoGene) | | Pubmed: 25658639 | ArrayExpress E-MTAB-2660 |
| Glioma-associated macrophages with CSF-1R inhibition, mouse (Mouse430A) | | Pubmed: 24056773 | GEO: GSE37475 |
| Resident, repopulating, and infiltrating brain myeloid cells (Mouse RNA-Seq) | | Pubmed: 26163371 | GEO: GSE68376 |
| Myeloid cells from blood, brain, and peripheral tissues (Mouse RNA-Seq) | | Pubmed: 25480296 | GEO: GSE63340 |
| Single-cell RNA-seq of mouse cerebral cortex | | Pubmed: 25700174 | GEO: GSE60361 |
| The regional mouse microglial transcriptome during aging (Mouse4302) | | Pubmed: 26780511 | GEO: GSE62420 |
| Embryonic, perinatal, and adult microglia (Mouse RNA-Seq) | | Pubmed: 27338705 | GEO: GSE79812 |
| PS2APP model cortex (whole tissue) (Mouse RNA-Seq) | | Pubmed: 27097852 | GEO: GSE75357 |
| SOD1(G93A) model spinal cord, (whole tissue) (Mouse4302) | | Pubmed: 23139902 | GEO: GSE18597 |
| Alzheimer's Patient Fusiform Gyrus (whole tissue) (Human RNA-Seq) | | This paper | GEO: GSE95587 |
| Alzheimer's Patient Temporal Cortex (whole tissue) (illuminahumanv1) | | Pubmed: 19361613 | GEO: GSE15222 |
| ALS patient anterior horn remnant following removal of motor neurons by LCM (HuExon) | | Pubmed: 19864493 | GEO: GSE18920 |
| Frontal cortex from FTD progranulin mutant patients (whole tissue) (hgu133a2) | | Pubmed: 18223198 | GEO: GSE13162 |
| Huntington's disease caudate nucleus (whole tissue) (HGU133A+B) | | Pubmed: 16467349 | GEO: GSE3790 |
| Purified cell types from human brain specimens (Human RNA-Seq) | | Pubmed: 26687838 | GEO: GSE73721 |
| Study                                                                 | Publication ID   | GEO Accession   |
|----------------------------------------------------------------------|-----------------|----------------|
| An RNA-Seq transcriptome and splicing database of neurons, glia, and vascular cells of the cerebral cortex (Mouse RNA-Seq) | Pubmed: 25186741 | GEO: GSE52564  |
| Adult mouse cortical cell taxonomy by single cell transcriptomics    | Pubmed: 26727548 | GEO: GSE71585  |
| Single cell RNA-seq identifies a unique microglia type associated with Alzheimer's disease (single-cell RNA-Seq) | Pubmed: 28602351 | GEO: GSE98969  |

| Experimental Models: Organisms/Strains                                   |
|--------------------------------------------------------------------------|
| PS2APP Mice                                                              | Richards et al., 2003 | N/A            |
| Cx3cr1::GFP Mice                                                         | Jackson Labs        | #005582        |
| hMAPT-P301L Mice (pR5-183)                                               | Götz et al., 2001   | N/A            |
| hMAPT-P301S Mice (PS19)                                                  | Yoshiyama et al., 2007 | N/A          |

| Software and Algorithms                                                  |
|--------------------------------------------------------------------------|
| HTSeqGenie                                                               | https://www.bioconductor.org/packages/release/bioc/html/HTSeqGenie.html | N/A            |
| Other bioconductor packages, including limma, vsn, DESeq2 and dependencies | https://www.bioconductor.org/                                      | N/A            |
| ImmGen website gene set analysis tool                                    | http://rstats.immgen.org/MyGeneSet/                               | N/A            |

| Other                                                                    |
|--------------------------------------------------------------------------|
| Interactive Website                                                      | This paper | http://research-pub.gene.com/BrainMyeloidLandscape |
Supplemental References

Benitez, B.A., Jin, S.C., Guerreiro, R., Graham, R., Lord, J., Harold, D., Sims, R., Lambert, J.-C., Gibbs, J.R., Bras, J., et al. (2014). Missense variant in TREML2 protects against Alzheimer’s disease. Neurobiol. Aging 35, 1510.e19–e26.

Butler, A., and Satija, R. (2017). Integrated analysis of single cell transcriptomic data across conditions, technologies, and species. bioRxiv 164889.

Butovsky, O., Jedrychowski, M.P., Moore, C.S., Cialic, R., Lanser, A.J., Gabriely, G., Koeglsperger, T., Dake, B., Wu, P.M., Doykan, C.E., et al. (2014). Identification of a Unique TGF-β Dependent Molecular and Functional Signature in Microglia. Nat. Neurosci. 17, 131–143.

Chiu, I.M., Morimoto, E.T.A., Goodarzi, H., Liao, J.T., O’Keeffe, S., Phatnani, H.P., Muratet, M., Carroll, M.C., Levy, S., Tavazoie, S., et al. (2013). A neurodegeneration-specific gene-expression signature of acutely isolated microglia from an amyotrophic lateral sclerosis mouse model. Cell Rep. 4, 385–401.

Heng, T.S.P., Painter, M.W., and Immunological Genome Project Consortium (2008). The Immunological Genome Project: networks of gene expression in immune cells. Nat. Immunol. 9, 1091–1094.

Hickman, S.E., Kingery, N.D., Ohsumi, T.K., Borowsky, M.L., Wang, L., Means, T.K., and El Khoury, J. (2013). The microglial sensome revealed by direct RNA sequencing. Nat. Neurosci. 16, 1896–1905.

Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A., and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformat. Oxf. Engl. 18 Suppl 1, S96–S104.

Krijthe, J. (2017). Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation.

Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease. Nat. Genet. 45, 1452–1458.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15, R29.

Lee, S.-H., Pichon, C.E.L., Adolfsson, O., Gafner, V., Pihlgren, M., Lin, H., Solanoy, H., Brendza, R., Ngu, H., Foreman, O., et al. (2016). Antibody-Mediated Targeting of Tau In Vivo Does Not Require Effector Function and Microglial Engagement. Cell Rep. 16, 1690–1700.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Nalls, M.A., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M., et al. (2014). Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson’s disease. Nat. Genet. 46, 989–993.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. 33, 495–502.
Srinivasan, K., Friedman, B.A., Larson, J.L., Lauffer, B.E., Goldstein, L.D., Appling, L.L., Borneo, J., Poon, C., Ho, T., Cai, F., et al. (2016). Untangling the brain’s neuroinflammatory and neurodegenerative transcriptional responses. Nat. Commun. 7, 11295.

Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat. Neurosci. 19, 335–346.

Webster, J.A., Gibbs, J.R., Clarke, J., Ray, M., Zhang, W., Holmans, P., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., et al. (2009). Genetic control of human brain transcript expression in Alzheimer disease. Am. J. Hum. Genet. 84, 445–458.

Zhang, B., Gaiteri, C., Bodea, L.-G., Wang, Z., McElwee, J., Podtelezhnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., et al. (2013). Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer’s disease. Cell 153, 707–720.

Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O’Keeffe, S., Phatnani, H.P., Guarneri, P., Caneda, C., Rudersich, N., et al. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J. Neurosci. Off. J. Soc. Neurosci. 34, 11929–11947.

Zhang, Y., Sloan, S.A., Clarke, L.E., Caneda, C., Plaza, C.A., Blumenthal, P.D., Vogel, H., Steinberg, G.K., Edwards, M.S.B., Li, G., et al. (2016). Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. Neuron 89, 37–53.