Developmental heterogeneity of microglia and brain myeloid cells revealed by deep single-cell RNA sequencing

Qingyun Li[1,6,8,*], **Zuolin Cheng**[2,6], **Lu Zhou**[1], **Spyros Darmanis**[3], **Norma F. Neff**[3], **Jennifer Okamoto**[3], **Gunsagar Gulati**[4], **Mariko L. Bennett**[1], **Lu O. Sun**[1], **Laura E. Clarke**[1], **Julia Marschallinger**[5], **Guoqiang Yu**[2,7], **Stephen R. Quake**[3,7], **Tony Wyss-Coray**[5,7], and **Ben A. Barres**[1,7]

[1]Department of Neurobiology, Stanford University School of Medicine, Stanford, CA 94305, USA

[2]Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Arlington, VA 22203, USA

[3]Departments of Bioengineering and Applied Physics, Stanford University and Chan Zuckerberg Biohub, Stanford, CA 94305, USA

[4]Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA

[5]Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA 94305, USA

[6]These authors contributed equally

[7]Senior author

[8]Lead Contact

Summary

[*]Correspondence: tristan.qingyun.li@gmail.com (Q.L.).

Microglia are increasingly recognized for their major contributions during brain development and neurodegenerative disease. It is currently unknown if these functions are carried out by subsets of microglia during different stages of development and adulthood or within specific brain regions. Here, we performed deep single-cell RNA sequencing (scRNA-seq) of microglia and related myeloid cells sorted from various regions of embryonic, early postnatal, and adult mouse brains. We found that the majority of adult microglia expressing homeostatic genes are remarkably similar in transcriptomes, regardless of brain region. By contrast, early postnatal microglia are more heterogeneous. We discovered a proliferative region-associated microglia (PAM) subset, mainly found in developing white matter, that share a characteristic gene signature with degenerative disease-associated microglia (DAM). Such PAM have amoeboid morphology, are metabolically active, and phagocytose newly formed oligodendrocytes. This scRNA-seq atlas will be a valuable resource for dissecting innate immune functions in health and disease.

Graphical Abstract
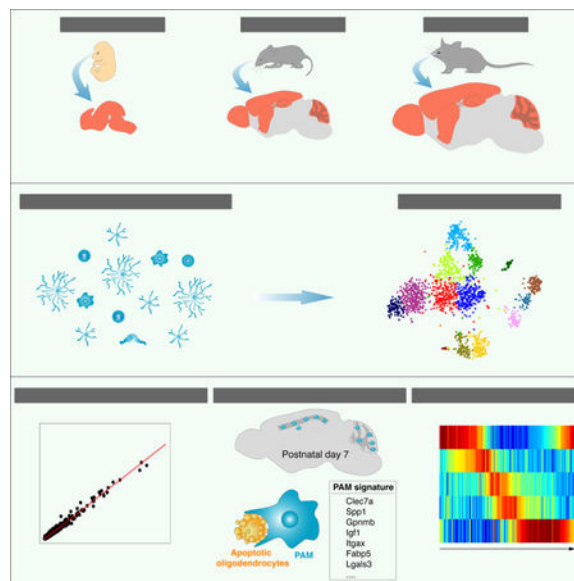
Introduction

Microglia are brain parenchymal macrophages that are implicated in numerous neurological diseases, such as Alzheimer's disease, amyotrophic lateral sclerosis, stroke, and brain tumors (Colonna and Butovsky, 2017; Prinz et al., 2011). In addition to their classical immune surveillance and scavenging functions, microglia have recently been found to actively participate in neural development by modulating neurogenesis and pruning synapses (Cunningham et al., 2013; Li and Barres, 2017; Paolicelli et al., 2011; Schafer et al., 2012; Ueno et al., 2013). Despite the importance of these multitasking cells, little is known about their molecular heterogeneity under physiological conditions and especially during development when they perform many critical non-immune functions. In addition, due to their transcriptomic resemblance to other myeloid cells which may infiltrate the brain parenchyma in disease (Goldmann et al., 2016; Prinz et al., 2017), a systematic comparison between microglia and these related immune cells remains an imperative task.

Microglia and most other tissue macrophages are long-lived, self-renewing cells that are generated by waves of erythro-myeloid progenitors in the yolk sac (Gomez Perdiguero et al., 2015; Hoeffel et al., 2015; Li and Barres, 2017). In mice, microglia migrate to the brain around embryonic day 9.5 (E9.5) and blood-brain barrier closure around E13.5 has been proposed as a mechanism to confine microglia inside the parenchyma (Ginhoux et al., 2010). Consistent with this convoluted developmental route, bulk RNA-sequencing (RNA-seq) data demonstrated a roughly step-wise differentiation program for microglia (Matcovitch-Natan et al., 2016). However, the reliance on general surface markers in these studies could overlook microglial heterogeneity, particularly potential transient populations during development, thereby underestimating developmental complexity of microglia. Furthermore, although mature microglia in different brain regions were shown to have uneven distribution with distinct morphologies (Lawson et al., 1990), which seem to correlate with region-specific expression profiles (Ayata et al., 2018; De Biase et al., 2017; Grabert et al., 2016), it remains unclear whether there are molecularly defined subtypes of microglia in the adult brain and, if so, how they are distributed across brain regions.

Here, we took an unbiased approach to investigate the heterogeneity of microglia along with other brain myeloid cells by performing deep single-cell RNA sequencing (scRNA-seq) on sorted cells across mouse brain regions and developmental stages. scRNA-seq has been proved as a powerful tool for dissecting cellular diversity from complex organs with minimal prior knowledge (Papalexi and Satija, 2018). We utilized the Smart-seq2 approach on sorted cells, due to its high sensitivity and accuracy (Svensson et al., 2017; Ziegenhain et al., 2017). In total, we sequenced 1922 cells to over 1 million raw reads per cell. Clustering analysis of this complex dataset identified 15 distinct cell populations. Two microglia clusters expressed signature genes for dividing cells, which we used to reconstruct cell cycle phases and produced phase-specific gene sets for microglia. We also found that early postnatal and adult choroid plexus macrophages were separated into distinct clusters, suggesting a developmental phenotypic switch for these particular brain resident macrophages.

Surprisingly, we found little population-wise heterogeneity among adult homeostatic microglia at the whole transcriptomic level. By contrast, we observed much higher heterogeneity in early postnatal microglia. We identified a population of proliferative region-associated microglia (PAM), that shared a transcriptional signature with degenerative disease-associated microglia (DAM) (Keren-Shaul et al., 2017; Krasemann et al., 2017). PAM mainly appeared in developing corpus callosum and cerebellar white matter around a transient period of the first postnatal week, when they engulfed newly formed oligodendrocytes. Interestingly, unlike DAM (Keren-Shaul et al., 2017; Krasemann et al., 2017), appearance of PAM did not depend on a TREM2-APOE axis, suggesting that different signals may trigger the emergence of these two microglial populations.

We created a searchable web interface for this high resolution single myeloid cell atlas as an integrated component of our widely used brain RNA-seq website: www.brainrnaseq.org. This study provided a useful resource for dissecting innate immune cell contributions to normal development and brain disease.

Results

Clustering of brain myeloid cells across developmental stages by deep scRNA-seq

To understand the spatiotemporal heterogeneity of microglia and other brain myeloid cells in an unbiased manner, we used a semi-automated Smart-seq2 platform to perform scRNA-seq on index sorted innate immune cells from the mouse brain (Figure 1A). We sequenced a total of 1922 cells to a depth of at least 1 million raw reads per cell, from embryonic day 14.5 (E14.5), postnatal day 7 (P7) and adult (P60) brain tissues. We reasoned that deeper sequencing would allow us to detect more genes compared with other scRNA-seq technologies, which could reveal even subtle differences among possible microglial subpopulations. We also analyzed non-microglia myeloid cells, due to their resemblance to parenchymal microglia and possible links to degenerative disease and injury. To uncover microglial regional heterogeneity, we isolated myeloid cells (for P7 and P60 stages) from six brain regions: cortex, cerebellum, hippocampus, striatum, olfactory bulb, and choroid plexus. We used a recently developed protocol for cell extraction to minimize microglial activation (Bennett et al., 2016). For the E14.5 stage, we gated on c-Kit⁻CD45⁺; for the P7 and P60 stages, we gated on $CD45^{low}CD11b^+$ versus $CD45^{hi}CD11b^+$ for microglia and myeloid cells, respectively (Figures 1B and S1A). Fluorescence intensities from other relevant markers, such as F4/80 (macrophage marker), CD41 (hematopoietic progenitor marker) and TMEM119 (homeostatic microglia marker), were recorded as part of the metadata in this multidimensional dataset (Figures 1A and S1A).

Using stringent criteria, 1816 cells (94.5% of 1922) passed quality control for downstream analysis, including 142 E14.5 cells, 978 P7 cells, and 696 P60 cells from various brain regions (Figures S1B–S1D; see STAR Methods). To assess the sensitivity and accuracy of the data, we analyzed the sequencing results for External RNA Controls Consortium (ERCC) Spike-In RNA, and found that our Smart-seq2 platform had nearly single molecule detection limit and high precision (R=0.98) (Figures S1E and S1F; see STAR Methods). When comparing our sequencing data for adult microglia with two other published scRNA-seq datasets (Keren-Shaul et al., 2017; Matcovitch-Natan et al., 2016), our study detected about 3 times as many genes per cell, and produced higher detection rates for over 80% of genes (Figures S1G and S1H). In summary, we generated a high-quality multidimensional scRNA-seq dataset for region-specific microglia and brain myeloid cells across developmental stages, and we incorporated these data into our widely used brain RNA-seq website (www.brainrnaseq.org).

To study the transcriptomic heterogeneity of brain myeloid cells, we carried out unsupervised clustering analysis for all 1816 cells, which gave rise to 15 distinct clusters (Figure 1C). Through differential gene expression analysis, we discovered a list of enriched genes for each cluster as potential markers (Table S1). Expression of these genes uniquely or in combinations represented individual cluster identities (Figures 1F and 1G). Based on these gene lists and other metadata, such as the developmental stage and FACS gating information for the cells from each cluster (Figures 1D and 1E), we annotated the 15 clusters into distinct cell types or states. This included 7 microglia clusters, 2 choroid plexus macrophage clusters, 3 monocyte clusters and 3 other immune or epithelial cell clusters

(Figure 1C). Of note, some clusters had less than 20 cells (e.g. cluster 14, Natural killer cells), but were robustly identified as distinct and biologically meaningful clusters, suggesting the reliability of the current technology even for small numbers of cells.

Interestingly, we observed that different clusters of microglia, choroid plexus cells or monocytes tended to be most closely juxtaposed on the tSNE plot, implicating the overall similarities between clusters of each cell type (Figure 1C). For example, neighboring clusters 0–6 were deemed "microglia-like" cells (Figures 1C and 1D). They all expressed microglial signature genes such as *P2ry12* and *Slc2a5* (Figure 1G) (Bennett et al., 2016; Zhang et al., 2014), and these cells were largely isolated from the CD45^{low} gate, whereas cells from other "non-microglia" clusters were almost exclusively from the CD45^{hi} gate (Figure 1D). We noticed the presence of many CD45^{hi} cells from the P7 stage in "microglia-like" clusters, suggesting the heterogeneity of CD45 immunophenotypes in early postnatal microglia. Furthermore, according to the developmental stages, these 7 microglia clusters were segregated into unidirectionally shifted domains on the tSNE plot, consistent with progressive changes in gene expression during microglial development (Figure 1E) (Matcovitch-Natan et al., 2016).

Postnatal transcriptional alteration of choroid plexus macrophages

We first focused on non-microglia clusters. Consistent with the bone marrow origin of circulating myeloid cells, we observed that monocyte and neutrophil clusters (clusters 8, 10, 11 and 12) comprised intermingled P7 and P60 cells within each cluster (Figures 1C and 1E). On the other hand, choroid plexus macrophages (CP MΦ) as resident macrophages at central nervous system (CNS) interfaces, have yolk-sac and bone marrow dual origins (Goldmann et al., 2016). Interestingly, CP MΦ were separated into two distinct clusters (cluster 7 and cluster 9) largely confined to their developmental stages (Figure 1E). Only in the adult CP MΦ cluster (cluster 7) did we detect high levels of MHC-II gene expression (e.g. *H2-Ab1*, *Cd74*, *H2-Aa*, *H2-Eb1*) (Figure 2A). Along with these genes involved in antigen presentation, we also found elevated expression of other immunoregulatory genes, such as *Fkbp5*, *Tsc22d3*, *Axl* and *Lilra5* in cluster 7. By contrast, postnatal CP MΦ (cluster 9) expressed many genes that function in endocytosis and intracellular trafficking of nutrients or signaling molecules (e.g. *Snx6*, *Snx2*, *Dab2*, *Cd36*, *Ap1b1*, *Mrc1*) (Figure 2A). We identified *H2-Eb1* and *Lilra5* as molecular markers for adult CP MΦ and *Clec4n* for their postnatal counterparts, and validated this selective gene expression by RNA *in situ* and immunohistochemistry (Figures 2B–2D). Our lab has previously demonstrated that microglia turn on homeostatic genes during postnatal development (Bennett et al., 2016). Here we show that CP MΦ are a second example of brain resident macrophages that undergo a postnatal transcriptional alteration.

Limited transcriptomic heterogeneity of adult homeostatic microglia

Within the 7 microglia clusters, adult microglia were almost exclusively present in cluster 0 and cluster 6 (Figures 1C–1E). Both clusters expressed similar levels of homeostatic microglial signature genes (Figure S2A), with the exception of slightly reduced *P2ry12* expression in cluster 6 (Figure 3A). Compared with cluster 0, only 27 total genes (including *P2ry12*) were differentially expressed in cluster 6, out of which half were immediate early

genes (IEG) (Table S2; Figure 3A). Therefore, we annotated cluster 6 as "IEG+" microglia. Since these immediate early genes can be rapidly induced in response to environmental stimuli, and the microglia extraction procedure is likely to trigger such an artificial response, we examined the expression of the two most up-regulated IEG genes, *Fos* and *Egr1*, on tissue sections. Indeed, adult microglia stained negative for *Fos/Egr1* transcripts or proteins, whereas their expression was readily detected in the surrounding non-microglial cells (Figures 3B, 3C, S2B and S2C). This result is consistent with recent studies showing that immediate early gene expression by microglia was only found in sorted cells but not with the RiboTag technology (Ayata et al., 2018; Haimon et al., 2018). Taken together, if we consider the contribution of IEGs to the clustering result as an experimental artifact, the vast majority of adult microglia from different brain regions appear to show the classical homeostatic phenotype with high levels of transcriptomic similarity.

Since we collected cells from different areas of the brain, these analyses imply that adult homeostatic microglia have a similar expression profile despite varying regions-of-origin. Indeed, we found that less than 20 genes were differentially expressed (FDR < 0.05) in typical adult microglia (cluster 0) between regions (Figure 3F; Table S3). To rule out a sensitivity issue of scRNA-seq, we performed region-specific bulk RNA-seq of highly purified homeostatic microglia sorted with TMEM119 antibodies, which label over 90% of microglia in the adult brain (Bennett et al., 2016). Consistent with our scRNA-seq data, samples from cortex, cerebellum, hippocampus and striatum were highly correlated (R>0.99), and individual samples did not cluster according to tissue origins, suggesting striking similarities between homeostatic microglia from different brain regions (Figures 3D and 3E). Moreover, we could not detect any differentially expressed genes (FDR < 0.05) between regions from the bulk samples (Figure 3F; Table S3). Conversely, we found that certain genes previously attributed to microglial regional heterogeneity were largely expressed by different populations of non-microglial cells from our scRNA-seq data (Grabert et al., 2016) (Figure S2D). These data suggest that classical adult microglia with homeostatic signatures (e.g. *Tmem119* and *P2ry12*), as the most dominant microglial population in the healthy brain, have only limited, if any, transcriptomic heterogeneity across brain regions.

Phase-specific gene expression of dividing microglia along cell cycle pseudotime

We next characterized early postnatal microglia, which were distributed among clusters 0–5 (two P7 cells in the IEG+ cluster 6 were excluded from the analysis) (Figure 4A, upper left). These included two clusters corresponding to dividing cells in two cell cycle phases (cluster 3 and 4), one embryonic-like cluster (cluster 5), two early postnatal cell-specific clusters (cluster 1 and 2), and a few cells in the adult homeostatic cluster (cluster 0) bordering cluster 2.

Microglia in the first two postnatal weeks are highly proliferative (Nikodemova et al., 2015). Previous work showed that in scRNA-seq analysis, cell cycle genes can overload the major principal components that underlie cell-to-cell variations, thereby masking other, more functionally relevant, differences between cells (Buettner et al., 2015). Indeed, many cell cycle genes ranked at the top of the first 3 principal components explaining variations in P7

microglia, and using conserved cell cycle genes for clustering, these cells were clearly segregated based on their presumable cell cycle phases (Figures S3A and S3C). To eliminate this confounding factor and unmask the underlying microglial heterogeneity, we used an established algorithm to regress out cell cycle effects followed by re-clustering of these P7 microglial cells (Figures S3B and S3C; see STAR Methods). This approach yielded three clusters: P7-C0, P7-C1 and P7-C2 (Figure 4A). Interestingly, these three clusters largely corresponded to the three dominant cell cycle-independent clusters (cluster 1, 2 and 5) found in the original analysis, and the two original cycling clusters (cluster 3 and 4) were simply redistributed among P7-C0, P7-C1 and P7-C2 following cell cycle regression (Figure 4A). These findings suggested that, irrespective of cell cycle states, there were only 3 categories of microglia at the P7 stage, and that the two major clusters (P7-C0 and P7-C1) were each comprised of non-dividing cells and their corresponding dividing cells in two cell cycle phases (Figure 4A).

It is of great interest to understand the gene regulation of microglial division because microgliosis is prevalent in disease and injury. We took advantage of the scRNA-seq data that presumably included gene expression information of individual dividing microglia at various phases of the cell cycle, and computationally reconstructed this dividing process (Figures 4B, S4A and S4B; see STAR Methods). To obtain cell cycle phase-specific gene sets for endogenous microglia, we analyzed the P7 microglia (P7-C0 and P7-C1 were analyzed separately) due to their sufficient population sizes, although we could also detect 11 dividing adult microglia in our data (Figure 1E). We used lists of conserved genes from a well-established dataset for HeLa cell lines as the "seeds" to reconstruct 5 cell cycle phases along microglial division pseudotime (Macosko et al., 2015; Whitfield et al., 2002). Then the algorithm searched for other genes correlated with phases and refined the ordering of cells based on updated phase-specific gene expression. Through a number of repetitions, lists of genes stably assigned to each phase were obtained (Figures 4B, S4B, and S4D; Table S4). Averaged expression of these gene sets showed a "wave-like" pattern along the dividing pseudotime, whereas microglial signature genes were expressed at a consistent level (Figures 4C and S4C).

In total, we identified 315 periodically expressed genes for P7-C0 cells, and 347 genes for P7-C1 cells (some genes were expressed in more than one phase) (Table S4). Excluding the seeds input, 254 (81% of 315) genes for P7-C0 and 227 (65% of 347) genes for P7-C1 were identified by the algorithm, and 183 genes overlapped in both clusters. We compared these machine-reported gene lists with four published datasets for experimentally characterized cell cycle genes (including the HeLa data we used as seeds) (Grant et al., 2013), and found that the majority of newly identified genes appeared in at least one of the four datasets (Figures 4D and S4E), suggesting the robustness of this analysis. More importantly, we identified many genes that have yet to be described as periodically expressed in dividing cells ("Novel"), although more than half of these new genes have possible links to cell proliferation based on gene function annotations (Figures 4E and S4F). For example, we discovered genes that are involved in DNA damage response and repair (*Ankle1*, *Lig1*, *Rpa1*), histone mRNA decay (*Eri1*), and genes with epigenetic functions such as replication-independent histones (*H2afz*, *H3f3b*) and chromatin modifiers (*Hmgn2*, *Dnmt1*) (Figure 4F).

Heterogeneity of early postnatal microglia revealed by scRNA-seq

To understand early postnatal microglial heterogeneity, we next examined three P7 clusters (after cell cycle regression). Interestingly, P7-C2 was a small cluster that mainly included cells clustered together with embryonic E14.5 microglia (cluster 5) in the initial analysis, suggesting the presence of primitive microglia in the postnatal brain (Figures 1C and 5A). Cells belonging to P7-C2 expressed microglial signature genes at lower levels (e.g. *Tmem119*, *Selplg*, *P2ry12*, *Tgfbr1*), and displayed up-regulation of genes in metal homeostasis (*Fth1*, *Ftl1*, *Mt1*), actin cytoskeleton dynamics (*Tmsb4x*, *Pfn1*, *Cfl1*), and ribosomal components (e.g. *Rps14*, *Rps18*, *Rpl35*, *Rps29*) (Figure 5B; Table S5).

The majority of P7 microglia fell into either P7-C0 or P7-C1 cluster. Microglia from both clusters expressed homeostatic genes, such as *Tmem119*, *P2ry12*, *Tgfbr1*, *Siglech*, and *Sall1*, but P7-C1 cells expressed them at lower levels (Figure 5C; Table S5). Interestingly, P7-C1 microglia exhibited characteristic expression of many genes that have recently been shown to be up-regulated in degenerative disease-associated microglia (DAM) (Keren-Shaul et al., 2017; Krasemann et al., 2017). These included *Spp1*, *Gpnmb*, *Igf1*, *Clec7a*, *Lpl*, *Cd9*, *Cd63*, *Lgals3*, *Fabp5*, *Itgax*, *Apoe* and *Tyrobp* (Figure 5C). When we directly compared differentially expressed genes in P7-C1 (relative to P7-C0) with those mis-regulated in DAM, we found that they were strikingly similar in that up-regulation of the disease genes was accompanied by down-regulation of the homeostatic gene cassette (Figure 5D). To obtain a better understanding of the genes underlying P7 microglial heterogeneity, we generated a gene network based on correlation between pairs of genes in all P7 microglia (Figure 5E; see STAR Methods). We highlighted the nodes with colors corresponding to the significance levels in differential gene expression analysis and cluster signatures they represented. We found that three interconnected gene modules explained P7 microglia identities and, more importantly, many DAM signature genes occupied the "hub" positions in the network.

Validation of early postnatal proliferative region-associated microglia

We were fascinated by the resemblance of the P7-C1 microglia to DAM in gene expression, and examined these early postnatal microglia in greater detail. We noticed that the top ranked marker genes, *Spp1* and *Gpnmb*, also distinguished this microglial subset from all other myeloid cell populations (Figures 6A and S5A; Table S5). Surprisingly, RNA *in situ* labeling demonstrated that expression of these two genes overlapped almost exclusively in the corpus callosum and cerebellar white matter (marked by *Mbp*+ oligodendrocytes), and that *Spp1*+*Gpnmb*+ cells were positive for the microglia marker *Cx3cr1* (Figures 6B, 6D). scRNA-seq data suggested that this P7-C1 population also had higher levels of *Igf1* and *Itgax* expression. Indeed, we detected co-expression of both transcripts in *Gpnmb*-positive cells (Figures 6C and S5B). Because these microglia mainly appear in the developing white matter of the early postnatal brain where gliogenesis is prominent, and they are also present in other neurogenic regions with cellular proliferation (see below), we named the P7-C1 population proliferative region-associated microglia (PAM).

Intriguingly, we discovered that early postnatal PAM and periventricular myeloid cells specifically expressed CLEC7A protein (Figures 6A, 6E, and S5A), which labels microglia

surrounding amyloid plaques in mouse models of Alzheimer's disease (AD) and has been described as a marker for DAM (Keren-Shaul et al., 2017; Krasemann et al., 2017). To further characterize PAM, we stained P7 microglia with the surface proteins, CLEC7A and GPNMB, and sorted the double positive cells by FACS (Figure 6F). GPNMB+CLEC7A+ cells also expressed higher levels of LILRB4 and CD63 which were part of the P7-C1 RNA signature and further established the identity of this microglial population (Figures 6A, 6F, and 6G). Additionally, scRNA-seq of microglia freshly isolated based on GPNMB and CLEC7A co-expression validated the original gene expression pattern observed with the initial P7 clustering results; the clustering algorithm also reported another GPNMB +CLEC7A+ concentrated cluster which overlapped with a portion of P7-C1 cells (Figures 6H and 6I). This new double positive cluster expressed all the markers identified in P7-C1, just at higher levels on average and in a much higher percentage of the cells (Figure 6I; Table S5). This was possibly due to the enrichment of this postnatal PAM subset at their most polarized state. Consistent with this interpretation, *Spp1* was detected in close to 100% of the double positive cells by scRNA-seq, and SPP1 only labeled a portion of CLEC7A+ cells on tissue sections, presumably representing the most typical form of postnatal PAM (Figure S5C). Interestingly, PAM also displayed distinct cytokine/chemokine secretion profiles, including *Ccl3* and *Ccl4* which have been linked to microglia in aging and AD models (Kang et al., 2018) (Figure S5D).

Because there seemed to be a gradient shift in overall gene expression from P7-C0, to P7-C1, and to P7-GPNMB+CLEC7A+ cells (Figure 6H), we combined P7 scRNA-seq data with P60-homeostatic microglia data and performed developmental pseudotime analysis to computationally reconstruct possible developmental trajectories of postnatal microglia. We discovered that one trajectory pointed towards the development of early postnatal PAM, while the other pointed towards homeostatic, mature cells (Figure 6J). These data suggest that postnatal immature microglia can either turn on genes specific for classical microglia found in the adult brain (P60-homeostatic) or assume the PAM phenotype (P7-C1 and P7-GPNMB+CLEC7A+) similar to DAM in aging and neurodegeneration (Figure 6K).

Transient appearance of PAM in developing white matter, independent of TREM2 or APOE

To investigate the temporal and spatial appearance of PAM during development, we stained brain sections from E17.5 to P60 with CLEC7A antibodies. In general, we discovered that CLEC7A+ microglia were associated transiently with the developing white matter, and found in regions of neurogenesis. During the late embryonic (E17.5) stage, CLEC7A+ microglia appeared almost exclusively in the neurogenic niches near the lateral ventricles (Figure 7A), while CLEC7A−IBA1+ cells were present throughout the brain parenchyma. Starting at P4, CLEC7A+ microglia populated the corpus callosum and cerebellar white matter (Figure 7A), and they continued to expand and peaked around P7 (Figures 6E and 7B). By P14, these white matter PAM were barely detectable (Figures 7A and 7B). Interestingly, starting at P14 and becoming more prominent during adulthood, CLEC7A+ microglia were also observed in the hippocampal dentate gyrus region, the subventricular zone and along the rostral migratory stream where adult neurogenesis and precursor migration occur (Figures 7A, 7C, and 7D). The presence of CLEC7A+ microglia in these strategic niches indicated that they might be involved in embryonic and adult neurogenesis.

Morphologically, CLEC7A$^+$ microglia in the P7 white matter were amoeboid in shape and had thicker primary branches with larger cell bodies (Figures 7E–7G). By contrast, CLEC7A$^-$microglia in the cortex were much more ramified. These CLEC7A$^+$ cells also congregated at a higher density compared with CLEC7A$^-$immature microglia (Figures 7E and 7G). Taken together, these data suggest that PAM in the developing white matter represent a specialized transient population in the first postnatal week, rather than a general immature state for all microglia.

It has been shown that transcriptional changes in degenerative disease-associated microglia depend on the TREM2-APOE pathway, and the DAM signature genes such as *Spp1*, *Gpnmb* and *Clec7a* are suppressed in disease models that also lack *Trem2* or *Apoe* (Keren-Shaul et al., 2017; Krasemann et al., 2017). Surprisingly, despite their similarity in gene expression to DAM, amoeboid *Gpnmb*$^+$*Spp1*$^+$CLEC7A$^+$ microglia were still densely populated in the P7 white matter of *Trem2* or *Apoe* knockout mice (Figures 7H and S6A). Consistent with this histology observation, scRNA-seq results showed that a subset of microglia (labeled by GPNMB and CLEC7A) in P7 *Trem2*–/– brains expressed PAM markers, even at slightly higher levels compared with wild type microglia (Figures S6B–S6F). These data suggest that, unlike DAM, early postnatal PAM do not depend on TREM2 or APOE for their polarization to occur.

Phagocytosis of newly formed oligodendrocytes by early postnatal PAM

To investigate the biological functions of early postnatal PAM, we followed up on our histological observation that CLEC7A$^+$ microglia in the developing white matter had amoeboid shapes and frequently contained pyknotic nuclei (Figures 8A and 8B; Movie S1). Amoeboid morphology has indeed been associated with a phagocytic state in microglia or macrophages (Bohlen et al., 2017; Ling and Wong, 1993). To compare the phagocytic capability between CLEC7A$^+$ and CLEC7A$^-$ microglia in the postnatal brain, we exposed acutely sectioned brain slices to pH-sensitive beads and quantified percentages of microglia that engulfed fluorescent beads based on CLEC7A expressivity and brain regions. We found that only microglia in the corpus callosum and cerebellum white matter (mostly CLEC7A+) preferentially phagocytosed beads (Figures 8C, 8D, S7A, and S7B), and this phenomenon was independent of opsonizing factors or other components in serum (Figure 8D). These results suggest that early postnatal PAM are highly phagocytic.

Since the occurrence of early postnatal PAM effectively correlates with the onset of CNS myelination when newly formed oligodendrocytes undergo massive cell death (Barres et al., 1992; Trapp et al., 1997), we hypothesized that early postnatal PAM might phagocytose some of these oligodendrocytes. Labeling oligodendrocytes with MBP in the CX3CR1-GFP mouse showed that microglia extensively interacted with star-shaped MBP$^+$ cells in the developing white matter (Figure 8E). More importantly, we observed early postnatal PAM containing cleaved Caspase-3 (cCASP3)-positive inclusions, which were MBP positive and possessed pyknotic nuclei (Figures 8E, 8F, and S7E; Movie S2). These data suggest that early postnatal PAM phagocytose dying newly formed oligodendrocytes during myelination. Interestingly, cCASP3-negative oligodendrocytes were frequently "hugged" by early postnatal PAM, raising the possibility of them actively contributing to elimination of cells

(Figure 8G). Moreover, we rarely observed juxtaposition of PDGFRA[+] oligodendrocyte precursor cells (OPCs) with early postnatal PAM (Figures 8H and S7C), suggesting that this phagocytic activity mainly targets newly formed oligodendrocytes but not their precursors. In addition, we could not find cCASP3[+] Aldh1l1-GFP labeled astrocytes, although substantial interactions between astrocytes and early postnatal PAM existed (Figures 8H, S7D, and S7F). From the scRNA-seq data, *Mbp* and *Gfap* transcripts were detected in certain percentages of P7-C1 (32/323=9.9% for *Mbp*; 5/323=1.5% for *Gfap*) and P7-GPNMB[+]CLEC7A[+] microglia (26/192=13.5% for *Mbp*; 17/192=8.9% for *Gfap*), while *Pdgfra* counts were zero in all P7 microglia (Figure 8I). Because the detection rates for P7-C0 microglia (8/274=2.9% for *Mbp*; 2/274=0.7% for *Gfap*) were more than 2 folds lower for both genes ($P<0.001$ for either gene, Chi-squared test), these differences were unlikely due to mRNA contamination during cell isolation. The simplest explanation is that scRNA-seq sufficiently captured transcripts from engulfed oligodendrocytes (*Mbp*) or astrocytes (*Gfap*) in early postnatal PAM. Taken together, these data suggest that early postnatal PAM phagocytose newly formed oligodendrocytes and perhaps to a lesser extent, astrocytes.

Discussion

Microglia and related brain myeloid cells are now recognized as functional modulators of developmental processes and neurodegeneration (Li and Barres, 2017). It is commonly accepted that heterogeneous myeloid populations are involved during the course of disease progression (Prinz et al., 2011). The extent of microglial heterogeneity and its functional significance in the developing and mature homeostatic brain, however, remains elusive. In this study, we used deep single-cell RNA sequencing (scRNA-seq) to address the question of microglial heterogeneity in the broad context of innate immune cells, and we systematically analyzed transcriptomes of these cells from distinct brain regions and across developmental time points. In addition, we generated cell cycle phase-specific gene sets for microglia which will not only provide a fresh perspective on our understanding of microglial proliferation, but also serve as a valuable resource for studies of mammalian cell cycle.

Our deep scRNA-seq and region-specific bulk RNA-seq data point to a lack of overt transcriptomic differences among classical adult microglia (TMEM119[+]) even across varying brain regions. These results indicate that the previously observed regional differences in adult microglia (Ayata et al., 2018) might mainly arise from genes expressed by small percentages of TMEM119[low/−] microglia (e.g. residual CLEC7A[+] cells shown in Figures 7A–7D), which were gated out in our bulk RNA-seq analysis. Consistent with this interpretation, many genes (*Apoe*, *Igf1*, *Lilrb4*, *Lyz2*, *Colec12*, *Msr1*, *Map1lc3b*) that were shown to be part of the phagocytic signature in adult cerebellar microglia (Ayata et al., 2018) were also up-regulated in PAM (Table S5). Meanwhile, other genes previously attributed to microglial regional heterogeneity could be due to differential abundance of non-microglial populations in bulk samples with the CD11b selection (Grabert et al., 2016). Furthermore, additional brain regions may uncover transcriptomic differences not present in the 4 major regions tested here. For example, microglia in the midbrain have been shown to possess distinct gene expression profiles compared with cortical microglia (De Biase et al., 2017). Microglial heterogeneity can also be represented by measurements other than gene

transcription, such as electrophysiological properties, epigenetic features or functional readouts (Ayata et al., 2018; De Biase et al., 2017).

By contrast, early postnatal microglia are much more heterogeneous. Particularly, we identified a subset of microglia, named proliferative region-associated microglia (PAM), that mimicked gene expression pattern of degenerative disease-associated microglia (DAM) (Keren-Shaul et al., 2017; Krasemann et al., 2017). The parallels between DAM and PAM support the concept that genes expressed during development are reactivated in aging and degeneration (Hong et al., 2016). Functional characterization of PAM may thus provide valuable clues for understanding pathophysiology of neurodegenerative diseases. Indeed, while early postnatal PAM engulf newly formed oligodendrocytes and possibly newborn astrocytes during development, DAM are also phagocytic and involved in phagocytosis of Aβ plaques in AD models (Keren-Shaul et al., 2017; Krasemann et al., 2017).

In addition, PAM enrich many metabolic genes including almost the entire molecular machineries for oxidative phosphorylation, glycolysis and beta oxidation (Figure S7G; Tables S6 and S7). We also found up-regulation of a large cohort of genes responsible for lysosomal acidification, lipid transport and metabolism, as well as lysosomal enzymatic functions (Table S7). This indicates high levels of energy turnover, particularly lipid metabolism, in PAM that may be required for phagocytosis of lipid-rich oligodendrocytes. Alternatively, the act of phagocytosis may drive or reinforce the expression of genes associated with lipid metabolism, lysosomal functions and phagocytosis. Future work should distinguish whether PAM represent a transcriptionally pre-disposed microglia subset favoring phagocytosis or a transient phenotypic conversion of microglia triggered by the phagocytic behavior.

We showed here that early postnatal PAM are not a general intermediate step for microglia development, but mainly transiently appear in the developing white matter. The timing of their appearance coincides with the onset of myelination, during which large numbers of newly formed oligodendrocytes die (Barres et al., 1992; Trapp et al., 1997). Overproduction of cells, balanced by controlled apoptosis, is a common theme during development of multicellular organisms, although the biological significance behind this is often unknown. Dying or dead cells need to be cleared away by professional phagocytes in order to maintain tissue homeostasis. Alternatively, these phagocytes may promote cell death to sculpt tissue structures and create functional circuits (Fuchs and Steller, 2011). It is thus possible that efficient engulfment of newly formed oligodendrocytes by PAM in the developing white matter is necessary to generate evenly spaced oligodendrocyte tiling and to make room for myelination. Besides the phagocytic function we presented here, PAM may play other roles by interacting with neural or immune cells through secreted molecules. We showed that early postnatal PAM display unique cytokine/chemokine profiles and express many trophic factors including *Igf1*, *Spp1*, *Lgals1*, and *Lgals3* (Figure S5D). Interestingly, and possibly related to our finding, a CD11c (ITGAX)-expressing microglial subset in the postnatal brain has been suggested to promote CNS myelination via secreted IGF1, which may exert its effect by supporting OPC survival (Hagemeyer et al., 2017; Wlodarczyk et al., 2017).

In summary, profiling individual mouse microglia during late embryonic and postnatal development provides a resource which complements current gene expression datasets of microglia from adult and diseased brains. Deep single-cell sequencing detected many additional genes expressed in microglia at different developmental stages and brain regions, and led to the discovery of a unique microglial subset termed PAM. Future functional studies of the identified microglial and myeloid populations will likely provide insights into the pivotal roles innate immune cells play in brain development, homeostasis and disease.

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Qingyun Li (tristan.qingyun.li@gmail.com).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mouse Models—All animal studies were performed with approval from the Stanford Administrative panel on Laboratory Animal Care in accordance with institutional and national regulations. All mice appeared healthy, received regular monitoring from veterinary and animal care staff, and were not involved in prior procedures or testing. Animals were housed in a non-barrier facility with 12-hour light/dark cycles at 23+/− 2°C, in disposable ventilated cages with no more than 5 animals per cage. Animals were provided water and standard chow *ad libitum*. Cages and bedding were changed every 2 weeks. Male mice were used in all experiments.

All sequencing experiments and histology validation (RNAscope or IHC) were performed using C57BL/6 mice of defined ages (or timed pregnant mice for harvesting E14.5 embryos) ordered from Charles River. *Trem2*–/− mice (C57BL/6J-Trem2em2Adiuj/J, 027197) and *Apoe*–/− mice (B6.129P2-Apoetm1Unc/J, 002052) were purchased from Jackson Laboratories. Their genotypes were confirmed according to instructions from Jackson Laboratories, and P7 pups were harvested from breeding cages for scRNA-seq or histology experiments. For the WT control matching the genetic background of *Trem2*–/− or *Apoe*–/− in Figures 7H and S6B-S6F, P7 C57BL/6J mice were ordered from Jackson Laboratories (000664). CX3CR1-GFP mice (B6.129P-*Cx3cr1tm1Litt*/J, 005582) were purchased from Jackson Laboratories and heterozygotes were used for the experiments. Aldh1l1-EGFP mice maintained in our laboratory were used to visualize astrocytes (Chung et al., 2013).

METHOD DETAILS

General considerations for experimental design

<u>Replication:</u> For the scRNA-seq data, we pooled samples from multiple mice to generate the gene expression atlas followed by histology validation. For the bulk RNA-seq data, we included 3 biological replicates for each brain region in our analysis. All of the histology experiments (RNA *in situ* or immunohistochemistry) were performed on at least 3 individual animals, and repeated at least twice.

Strategy for randomization and/or stratification: This paper did not include treatment studies, so we encountered no occasion to randomize a cohort of subjects. For sequencing or histology experiments, randomly chosen wild type or mutant mice that were housed under the same condition were used.

Blinding: Processing of RNA-seq data and clustering analysis were performed blinded, and cell identities were unblinded when combining the meta data into the analysis pipeline. Genotype information was blinded when analyzing PAM presence in the wild type, *Trem2*–/– or *Apoe*–/– mutant mouse brains.

Sample-size estimation and statistical method of computation: The number of cells for scRNA-seq and the number of replicates for bulk RNA-seq were empirically determined based on previous studies (Darmanis et al., 2015; Tabula Muris et al., 2018; Zhang et al., 2014). The used sample-size in scRNA-seq allowed us to robustly identify cell populations as small as 20 cells. Statistical methods used for RNA-seq analysis and quantification were described in each related section and detailed in "quantification and statistical analysis".

Inclusion and exclusion criteria of any data or subjects: For scRNA-seq data analysis, we applied 3 statistical criteria in the quality control workflow to exclude cells that had low sequencing reads (poor RNA recovery), low numbers of detected genes (low complexity of libraries), or low ERCC correlation coefficients (insufficient accuracy) using the methods described below. For bulk RNA-seq, all 3 replicates for each region were prepared identically and included in the analysis. They generated sequencing results with similar read counts and numbers of detected genes. For histology analysis, animals were randomly chosen to performed the experiments, and all prepared samples were examined yielding consistent results.

Tissue dissection and cell isolation—To isolate cells for scRNA-seq and bulk RNA-seq, P60 mice or pregnant mice with E14.5 embryos were euthanized with CO_2, and P7 mice were decapitated directly. For the E14.5 stage, whole brains were dissected out from timed embryos, and collected into 1ml PBS with 1% fetal bovine serum (added 10ul Ambion DNase I, 2ul RNase inhibitor) on ice. The sample was passed through a 26G needle 5–10 times to create single cell suspension. Cells were spun down (500g) at 4˚C for 5min, and suspended in 1ml PBS for FACS staining. For P7 and P60 stages, cortex, cerebellum, hippocampus, striatum, olfactory bulb and choroid plexus were dissected out. For P7 GPNMB/CLEC7A double sorting in Figure 6F, only the cerebellum was dissected and used. For P7 microglia comparing *Trem2*–/– with WT in Figure S6B, the whole brain was used. Tissues from the same region were pooled from multiple mice (3–8 depending on the ages and regions) into a 6cm petri dish with 200ul cold medium A on ice (15mM HEPES, 0.5% glucose in 1 X Hanks' Balanced Salt Solution (HBSS without phenol red)). Microglia (myeloid cells) extraction was carried out following a published protocol (Bennett et al., 2016). The whole procedure was done on ice with cold buffers. Briefly, tissues were chopped with razor blade into fine pieces and then transferred into a 5ml douncer containing 5ml medium A (added 200ul 12500 units/mL DNase I and 10ul recombinant RNase inhibitor). Tissues were dounced with 6–10 full strokes until no visible chunks were present

and then filtered into 50ml falcon tubes using 70um strainers to obtain single cell suspension. Cells were washed with medium A and resuspended in 1.8ml MACS buffer with 3.6ul RNase inhibitor (sterile-filtered 0.5% BSA, 2mM EDTA in 1 X PBS). To remove myelin, cells were incubated with 200ul myelin removal beads (MACS Miltenyl Biotec 130–096-433) for 10 min, and loaded onto LD columns (Miltenyi Biotec 130–042-901) assembled on a MACS magnet stand. Cells in the flow through were collected and washed for standard FACS staining.

Single cell index sorting—To stain the cells for sorting, 5ul mouse Fc block (BD Pharmingen 553142, 1:60) was added into each sample and incubated for 5 min on ice. Then the cells were incubated with antibodies conjugated with fluorophores or primary antibodies for 10 min at room temperature (RT) on shaker. After wash, if necessary, samples were stained with secondary antibodies for 10 min at room temperature and followed by wash. Cells were resuspended in 300ul FACS buffer (sterile-filtered 1%FCS, 2mM EDTA, 25mM HEPES in 1XPBS) with DNase I (1:100, Qiagen 79254) and RNase inhibitor (1:500) for single cell index sorting.

Cell sorting/flow cytometry analysis was done on the cell sorter (BD InFlux) at the Stanford Shared FACS Facility. All events were gated with the consecutive gates: (1) forward scatter-area (FSC-A)/side scatter-area (SSC-A) (2) Trigger Pulse Width/ FSC (3) Live-Dead negative (Green for E14.5; Far Red for P7 and P60 (Thermal Fisher Scientific L34960); Propidium Iodide for P7 GPNMB/CLEC7A double sorting) (4) P7/P60: $CD45^{low}CD11b^+$, $CD45^{hi}CD11b^+$; E14.5: $c\text{-}Kit^-CD45^+$; P7 GPNMB/CLEC7A: $GPNMB^+CLEC7A^+$(gated on $CD45^{low}CD11b^+$). We drew the gates to include single, live myeloid cells without bias, while excluding debris or doublets (from FSC-A/SSC-A and Trigger Pulse Width/FSC gating). Single cells were index sorted into 96-well plates containing 4uL lysis buffer (4U Recombinant RNase Inhibitor (Takara Bio 2313B), 0.05% Triton X-100, 2.5mM dNTP mix (Thermo Fisher Scientific R0192), 2.5uM Oligo-dT30VN (5′-AAGCAGTGGTATCAACGCAGAGTACT$_{30}$VN-3′, ERCC Spike-ins (Thermo Fisher Scientific 4456740) diluted at 1:2.4×10^7). Plates were briefly vortexed, spun down and snap frozen on dry ice. Plates were stored at −80°C freezer until library preparation.

Antibodies used for FACS sorting/flow cytometry: Rat anti-mouse CD117 (c-Kit), APC (clone 2B8) (ThermoFisher Scientific 17–1171-82, 1:200); Rat anti-mouse CD45, PE-Cy7 (clone 30-F11) (Thermo Fisher Scientific 25–0451-82,1:300); Rat anti-mouse CD11b, Brilliant Violet 421 (clone M1/70) (BioLegend 101236, 1:300); Rat anti-mouse F4/80, Per-CP/Cy5.5 (clone BM8) (BioLegend 123128,1:100); Rat anti-mouse CD41, PE (BioLegend 133906,1:200); Rabbit anti-mouse TMEM119 (clone 106–6) (abcam ab210405, 1:400); Goat anti-rabbit Alexa 488 (Thermal Fisher Scientific 11034, 1:300); Rat anti-mouse GPNMB, eFluor 660 (clone CTSREVL) (ThermoFisher Scientific 50–5708-82, 1:20); Rat anti-mouse CD369 (DECTIN-1/CLEC7A) (clone RH1) (BioLegend 144302, 1:50); Mouse anti-rat IgG1 Antibody, FITC (clone MRG1–58) (BioLegend 407406, 1:200); Armenian Hamster anti-mouse CD85k, PE (H1.1) (BioLegend 144904, 1:100); Rat anti-mouse CD63, PE/Cy7 (clone NVG-2) (BioLegend 143910, 1:100).

Single and bulk RNA-seq library preparation—For scRNA-seq, sequencing libraries were prepared following the published Smart-seq2 protocol with the aid of liquid handling robotics (Picelli et al., 2014). Briefly, plates with sorted cells were thawed on ice and incubated at 72°C for 3 min in order to anneal RNAs to the Oligo-dT30VN primer. After that, 6ul reverse transcription mixture (95U SMARTScribe™ Reverse Transcriptase (100U/ul, Clontech 639538), 10U RNase inhibitor (40U/ul), 1XFirst-Strand buffer, 5mM DTT, 1M Betaine, 6mM MgCl2, 1uM TSO (Exiqon, Rnase free HPLC purified)) was added into each well, and RT was performed at 42°C for 90 min, followed by 70°C, 5 min. To amplify cDNA, 15ul PCR amplification mix (1X KAPA HIFI Hotstart Master Mix (Kapa Biosciences KK2602), 0.1uM ISPCR Oligo (AAGCAGTGGTATCAACGCAGAGT), 0.56U Lambda Exonuclease (5U/ul, New England BioLabs M0262S)) was added, and the following PCR program was used: (1) 37°C 30 min; (2)95°C 3 min; (3) 21 cycles of 98°C 20 sec, 67°C 15 sec, 72°C 4 min; (4) 72°C 5 min. Amplified cDNA samples were then purified with PCRClean DX beads (0.7:1 ratio, Aline C-1003–50), and resuspended in 20ul EB buffer. cDNA quality was examined with a Fragment Analyzer (AATI, High Sensitivity NGS Fragment Analysis Kit:1 bp - 6000 bp) following the manufacture's instruction. Samples with sufficient amount of cDNA content (>0.05ng/ul) and normal peaks on the quantification graphs were retained for library preparation. To make libraries, all samples were first diluted down to 0.15ng/ul (only if higher than 0.15ng/ul) in 384-well plates using Mantis Liquid Handler (Formulatrix) and Mosquito X1 (TTP Labtech) with customized scripts. Nextera XT DNA Sample Prep Kit (Illumina FC-131–1096) was used at 1/10 of recommendation volume, with the help of a Mosquito HTS robot for liquid transfer. Specifically, tagmentation was done in 1.6ul (1.2ul Tagment enzyme mix, 0.4ul diluted cDNA) at 55°C, 10 min. Neutralization buffer was added 0.4ul per well and incubated at room temperature for 5 min to stop the reaction. Then 0.8ul Illumina Nextera XT 384 Indexes (0.4ul each, 5uM from 4 sets of 96 indexes) and 1.2ul PCR master mix were added to amplify whole transcriptomes using the following program: (1) 72°C 3 min; (2) 95°C 30 sec; (3) 10 cycles of 95°C 10 sec, 55°C 30 sec, 72°C 1 min; (4) 72°C 5 min. Libraries from a single 384 plate were pooled together in an Eppendorf tube and purified twice with PCRClean DX beads. The quality and concentrations of the final mixed libraries were measured with Bioanalyzer and Qubit, respectively, before Illumina Nextseq sequencing (Illumina FC-404-2002).

For bulk RNA-seq, TMEM119$^+$ cells from P60 brain regions were sorted (SONY SH800) into a 1.5ml Eppendorf tube containing 350ul RLT buffer (3000 cells per sample). RNA was extracted using RNeasy Micro Kit (Qiagen 74004) and eluted with 14ul nuclease-free water. To make cDNA, Smart-Seq v4 ultra low input RNA kit (Clontech 634890) was used following manufacture's instruction (9.5ul RNA input). cDNA was amplified for 9 cycles, and cDNA concentrations were measured with Qubit. Libraries were prepared using Nextera XT DNA Sample Prep Kit (380pg cDNA per sample) following the standard protocol. Nextera XT 24 Indexes (Illumina FC-131–1001) were used during whole transcriptomes amplification (13 cycles). Library quality was assessed with Bioanalyzer and Qubit. Nextseq sequencing was performed to a depth of $(1.6\pm0.33) \times 10^7$ raw reads per sample and $(1.07\pm0.24) \times 10^7$ mapped reads per sample. This led to the detection of 11727±563 genes,

comparable to 11407±2177 genes detected in P60 microglia from our previously published dataset (Bennett et al., 2016).

Processing of scRNA-seq raw data—Prinseq (Schmieder and Edwards, 2011) was first used to filter sequencing reads shorter than 30 bp (-min_len 30), trim the first 10 bp at the 5'-end (-trim_left 10) of the reads, trim reads with low quality from the 3'-end (-trim_qual_right 25) and remove low complexity reads (-lc_method entropy, -lc_threshold 65). Then, Trim Galore was applied to trimmed the Nextera adapters (--stringency 1). The remaining reads were aligned to the mm10 genome by calling STAR (Dobin et al., 2013) with the following options: --outFilterType BySJout, --outFilterMultimapNmax 20, --alignSJoverhangMin 8, --alignSJDBoverhangMin 1, --outFilterMismatchNmax 999, --outFilterMismatchNoverLmax 0.04, --alignIntronMin 20, --alignIntronMax 1000000, --alignMatesGapMax 1000000, --outSAMstrandField intronMotif. Picard was then used to remove the duplicate reads (VALIDATION_STRINGENCY=LENIENT, REMOVE_DUPLICATES=true). Finally, the aligned reads were converted to counts for each gene by using HTSeq (-m intersection-nonempty, -s no) (Anders et al., 2015).

Quality control for scRNA-seq data—To filter out cells with low sequencing quality, three criteria, namely the number of total reads, the number of total detected genes and correlation coefficient between ERCC spike-ins input and corresponding read counts, were each evaluated based on the distributions of the data. The distribution of the total reads (in logarithmic scale) was fitted by a truncated Cauchy distribution, and data points in two tails of the estimated distribution were considered as outliers and eliminated. Fitting and elimination were then applied to the remaining data. This process was run iteratively until the estimated distribution became stable. The threshold was set to the value where the cumulative distribution function of the estimated distribution reaches 0.05. Similarly, cells with small numbers of detected genes, and poor correlation coefficients for ERCC (low sequencing accuracy) were dropped. After filtering, 1816 cells, out of 1922, were retained.

scRNA-seq sensitivity and accuracy assessment—We used a previously published algorithm to estimate sensitivity and accuracy of scRNA-seq data based on ERCC spike-ins with known concentrations (Svensson et al., 2017). To estimate sensitivity, for each cell, binomial logistic regression was used to fit the actual numbers of ERCC spike-in molecules and the responses (detected or not) and then the number of ERCC molecules (detection limit) corresponding to 0.5 detection probability was determined. Distribution of detection limit for all cells that passed quality control (QC) was plotted in Figure S1E, which indicates the overall sensitivity level of the assay. To estimate accuracy, Pearson correlation coefficient between the actual numbers of spike-in molecules used as input and the corresponding average reading counts across all cells (passed QC) was calculated and shown in Figure S1F.

Clustering analysis of scRNA-seq data—The Seurat package was used to perform unsupervised clustering analysis on scRNA-seq data (Macosko et al., 2015). Briefly, gene counts for cells that passed QC were normalized to the total expression and log-transformed, and then highly variable genes were detected (y.cutoff=0.5). Depending on the analysis, cell cycle effect could be regressed out using the ScaleData function. Using highly variable

genes as input, principal component analysis was performed on the scaled data in order to reduce dimensionality. Statistically significant principal components were determined by using the JackStrawPlot function. These principal components were used to compute the distance metric, which then generated cell clusters. Non-linear dimensional reduction (tSNE) was used to visualize clustering results. Differentially expressed genes were found using the FindAllMarkers (or FindMarkers) function that ran Wilcoxon rank sum tests.

DE gene analysis on regional microglia—To identify differentially expressed (DE) genes of adult microglia in each of the four regions (cerebellum, hippocampus, striatum, and cortex), we performed DE gene analysis (one versus rest) on both scRNA-seq and bulk RNA-seq data by using edgeR (Robinson et al., 2010). For scRNA-seq data, genes expressed (CPM>2) in at least 5% of total cells were tested by the likelihood ratio method for DE gene identification. For bulk RNA-seq data, genes that were detected (CPM>2) in at least 2 samples were used to perform quasi-likelihood F-test to identify DE genes across regions.

Cell cycle regression on scRNA-seq data—We used Seurat package to regress out cell cycle effects in scRNA-seq data when analyzing P7 microglia heterogeneity. Briefly, each cell was assigned with a G2M score and an S score, based on its expression of G2/M and S phase markers. Then for each gene, the relationship between the gene expression and the G2M and S scores across all cells was modeled. A "corrected" expression matrix, determined by the scaled residuals of this model was obtained and used for downstream analysis.

Pseudotime analysis of dividing microglia

Work flow to generate pseudotime order: This algorithm is based on a published method with modifications (Macosko et al., 2015). The same analysis was conducted on P7-C0 and P7-C1 microglia separately. Simply, 5 conserved gene sets from studies on HeLa cell lines that are periodically expressed in 5 cell cycle phases (G1/S, S, G2/M, M and M/G1) were curated (see below) and used as the initial seeds in order to discover microglia-specific cell cycle genes. Cells that had significantly different expression levels (ANOVA test) for the 5 curated gene sets were selected as "dividing cells" for pseudotime ordering, and the rest were temporarily labeled as "non-dividing cells". Dividing cells were ordered based on expression of these phase-specific genes as previously described (Macosko et al., 2015). Then new genes that were expressed by cells non-randomly distributed along the order and correlated with average expression of gene seeds specific to any one or two adjacent phases, were assigned with a phase label and added to the corresponding seeds gene list (see below). After this, using updated seeds as input, the whole process was run iteratively, that is "cell selection"-"dividing cell ordering"-"seeds updating with new genes", until the cell ordering was stable (the correlation between current ordering and previous ordering larger than 0.9) for more than ten times. During the iteration, the initial gene seeds were always kept in the updated list for the next round to avoid random drifting by noise. Once cell ordering became stable, genes that were frequently identified (larger than five times) in the last ten iterations were saved as final cell cycle phase-specific genes. The final gene lists were then used again to determine dividing cells versus G0 cells (cells expressing genes of all phases at high levels were manually excluded), and the final pseudotime order of dividing cells.

Initial gene seeds curation: Five sets of conserved genes, each specific to a single phase of the cell cycle (G1/S, S, G2/M, M and M/G1), were selected from a published dataset (Macosko et al., 2015). A gene with low correlation (R<0.3) between its expression level (log2(FPKM+1)) and the average expression of its own gene set was discarded from the list. To ensure the phase label for each remaining gene is applicable to microglia, correlations between its expression and the average expression of each gene set were calculated, and the gene was assigned to the gene set that had the maximal correlation with it. The correlation examination and re-assignment were run iteratively until no more genes were re-assigned. Such 5 gene sets were used as initial gene seeds.

New phase-specific genes to update seeds: For each gene, a sliding window with a certain size (length of ordered cells/10) was used to find out the windows with maximal average expression and minimal average expression along the cell order. The difference between the maximal and minimal average expression was calculated as the observed difference. Then the order of the cells was randomly permutated, and the same sliding window operation was performed to obtain a permutation difference. Permutation was repeated for 1.0e+05 times to generate a distribution of permutation differences, with which the observed difference was compared to calculate a *p*-value. Genes with adjusted *p*-values less than 0.05 (non-randomly expressed along the cell order) were retained. To determine which phase(s) an identified gene belongs to, 10 scenarios of phase assignment were considered: five single phases and five dual phases (any two consecutive phases). Correlations between the gene expression and average expression of gene seeds corresponding to a single phase or two consecutive phases were calculated, and the phase(s) yielding the maximal correlation was assigned to the gene. Genes were ignored if they had poor correlations (R<0.3) with all of the ten scenarios mentioned above.

Gene network construction for P7 microglia—A gene network was built to reveal the relationship between different genes in P7 cells. To this end, genes that contributed most to cellular heterogeneity were chosen and correlations between these genes were shown on a network graph with features explained below. Specifically, genes that were expressed in at least 3% of total cells and contributed to the first 15 principal components underlying sources of variances were used for analysis. After quantile normalization for each gene's expression, graphical lasso was used to explore the relationship between these genes (Friedman et al., 2008). Graphical lasso estimated a sparse precision matrix (i.e., inverse covariance matrix) of the variables (genes) by using a lasso (L1) penalty. Such precision matrix revealed the conditional relationships between genes, upon which a gene network was built. In the network, each vertex was a gene, and the weight of each edge was the absolute value of the corresponding entry in the precision matrix. If two genes were nonadjacent (value of the corresponding entry in the precision matrix was 0), they were conditionally independent, given all the other genes. For better visualization, the width of each edge in the graph was proportional to its weight. Edges with small weights were not shown to simply the illustration. If the value of an entry in the estimated covariance matrix was positive, the corresponding edge was a solid line, indicating a positive correlation between the two genes. A negative value for an entry was translated to a dashed line, indicating a negative correlation. The area of each vertex (gene) was proportional to the total

weights of all edges connecting to it. To illustrate how genes in the network contributed to cell clustering, genes were painted with different colors based on whether they were differentially up-regulated in any P7 cluster and, if so, in which cluster(s). Genes shown in the network that were significantly up-regulated in a single P7 cluster were colored according to a predefined colormap (P7-C0=red; P7-C1=blue; P7-C2=green). Genes significantly up-regulated in two clusters were mapped to a new colormap, which included mixed colors of the two corresponding clusters (e.g. P7-C1&P7-C2= turquoise). The shades and tints of colors were determined by the logarithmic transform of adjusted p-values. Genes not differentially expressed were in grey.

Developmental pseudotime analysis—Sequencing reads for 648 P7 microglia (excluding P7-C2), 218 P60 homeostatic microglia (excluding 3 cells with detectable *Clec7a* expression) and 133 P7-GPNMB⁺CLEC7A⁺ microglia were analyzed together to build developmental pseudotime (Figure 6J) using the Monocle package (Trapnell et al., 2014). First, clustering analysis was performed with Seurat after cell cycle regression as described above. Then the Seurat object was imported into Monocle, which used statistical models to find out differentially expressed (DE) genes based on the clustering result. Top 350 DE genes (q<0.01) were selected for cell ordering. Monocle then used reversed graph embedding (DDRTree method) to do dimensionality reduction. Finally, Monocle performed manifold learning in order to generate a tree-like structure reflecting the developmental trajectories from one cell state to others.

GO term analysis—PANTHER Classification System was used to perform Gene Ontology term analysis (Mi et al., 2010). The input included 782 genes up-regulated in P7-GPNMB⁺CLEC7A⁺ microglia (compared with P7-C0 immature microglia), and statistical overrepresentation test was run using default settings. Fold enrichment values and false discovery rate (–log10 FDR) for the enriched GO-Slim Biological Process (BP) terms and GO-Slim Cellular Component terms were plotted. FDR was calculated by Fisher's Exact with FDR multiple test correction.

RNAscope—Mouse brains were rapidly dissected out and immediately embedded in OCT (optimal cutting temperature) compound (Tissue-Tek). The fresh frozen sections (sagittal) were prepared with cryostat at 12um thickness and RNA *in situ* was performed using RNAscope technology (Advanced Cell Diagnostics) following the manufacturer's protocol (RNAscope Multiplex Fluorescent Assay). Briefly, after fixation and dehydration, the slides were treated with protease IV for 20 min at room temperature. RNAscope probes were hybridized for 2 hours at 40 °C and then fluorescence staining was done through a signal amplification system. Pre-designed Probes (ACD) against the following mRNA were used: *Cx3cr1 (314221), Gpnmb (489511), Spp1 (435191), Mbp (451491), Igf1 (443901), H2-Eb1 (509081), Lilra5 (514461), Itgax (311501), Tmem119 (472901), Fos (316921), Egr1 (423371).*

Immunohistochemistry—Mice were transcardially perfused with PBS followed with 4% para-formaldehyde (PFA). Dissected brains were fixed with PFA overnight at 4°C and then cryopreserved with 30% sucrose for 24 hours at 4°C. Tissues were embedded in the optimal

cutting temperature (OCT) compound (Tissue-Tek). Free floating sagittal sections (50um) were stored in 6-well plates with PBS and then blocked and permeabilized with 10% serum/ 0.2% Triton-X in PBS for 15 min at room temperature. Sections were incubated with primary antibodies at 4°C overnight followed by secondary a ntibody staining for 2 hours at room temperature. After washes, the sections were transferred onto glass slides and mounted in Vectashield with DAPI (Vector Laboratories H-1200). Images were acquired using Zeiss LSM 880 inverted confocal and Zeiss Axio Imager fluorescence microscopy, and imported into Fiji (Schindelin et al., 2012) for quantification or Imaris for 3D reconstruction.

The following primary antibodies were used: Goat polyclonal anti IBA1 (Abcam ab5076, 1:500); Sheep polyclonal anti c-FOS (Abcam ab6167, 1:200); Rabbit monoclonal anti EGR1 (clone 15F7) (Cell Signaling Technology 4153, 1:50); Rabbit polyclonal anti cleaved CASPASE-3 (Asp175) (Cell Signaling Technology 9661, 1:200); Rat neutralizing monoclonal anti-mDECTIN-1-IgG (clone R1–4E4) (Invivogen mabg-mdect, 1:30); Rat monoclonal anti Myelin Basic Protein (clone 12) (Abcam ab7349, 1:100); Rat monoclonal anti-DECTIN 2 (clone D2.11E4) (Abcam ab34724, 1:100); Rabbit monoclonal anti PDGF Receptor α (D1E1E) (Cell Signaling Technology 3174S, 1:500); Goat polyclonal anti Osteopontin/OPN (R&D Systems AF808, 1:13). The following secondary antibodies were used: Donkey anti-goat IgG (H+L) cross-adsorbed, Alexa Fluor 488 (Thermo Fisher Scientific 11055, 1:1000); Donkey anti-rabbit IgG (H+L), Alexa Fluor 488 (Thermo Fisher Scientific 21206, 1:1000); Donkey anti-sheep DyLight 594 (Jackson Immuno Research Laboratories 713–586-147, 1:200); Donkey anti-rabbit IgG (H+L) Highly cross adsorbed, Alexa Fluor 594 (Thermo Fisher Scientific 21207, 1:200); Donkey anti-rat IgG (H+L) highly cross-adsorbed, Alexa Fluor 594 (Thermo Fisher Scientific 21209, 1:200); Donkey anti-rabbit IgG (H+L) highly cross-adsorbed, Alexa Fluor 647 (Thermo Fisher Scientific 31573, 1:200); Donkey anti-rat IgG (H+L) AffiniPure, Alexa Fluor 647 (Jackson Immuno Research Laboratories 712–605-153, 1:200).

Phagocytosis assay by slice culturing system—P7 mice were decapitated and dissected brains were immediately put in pre-cooled culture medium with serum (65% MEM (Sigma #M2279); 10% horse serum; 25% HBSS; 6.5 mg/ml glucose; 2mM L-Glutamine; 1% Pen/Strep), or serum-free microglial culture medium (DMEM/F12, 1% Pen/Strep; 2 mM L-glutamine; 5 ug/mL N-acetyl cysteine; 5 ug/mL insulin; 100 ug/mL apo-transferrin; 100 ng/mL sodium selenite; 2 ng/mL human TGFb2; 100 ng/mL murine IL-34; 1.5 ug/mL ovine wool cholesterol; 10 ug/mL heparan sulfate; 0.1 ug/ml oleic acid; 0.001 ug/ml gondoic acid) (Bohlen et al., 2017). The entire procedure was done on ice with pre-cooled solutions until culturing. Sagittal sections were prepared with vibratome (Leica VT1000S) at 250um thickness and then transferred to insert wells (Millicell Cell Culture Insert, 30 mm, Millipore #PICM03050) on a 6-well plate with medium. The plate with the sections was incubated for 1 hour in the incubator (37°C, 5% CO2) before p Hrodo™ Red Zymosan Bioparticles (Thermo Fisher Scientific P35364) were added at 0.5mg/ml to cover the whole sections (about 150ul each). Then the plate was incubated for 4 hours for phagocytosis (37°C, 5% CO2). After washes with PBS, sections were fixed with 4% PFA for 30 min at room temperature and then blocked with 10% serum/0.2% Triton-X in PBS for 15 min at room temperature. Immunohistochemistry was performed as described above.

3D reconstruction of confocal images—Confocal z-stacks were imported into Imaris software (Bitplane). Regions of interest were isolated to make IsoSurfaces for each channel. To quantify interactions between early postnatal PAM and other neural cell types, whole sections were 3D reconstructed, and overlapping volumes between early postnatal PAM and corresponding cell types were obtained with the function of distance transformation. The percentage of the overlapping volume over the volume of a given cell (MPB$^+$, Aldh1l1-GFP $^+$ or PDGFRA$^+$) was calculated to represent the level of interaction between this cell and the early postnatal PAM. Substantial interaction was defined as 30% overlapping volume in the cell. Similar results were observed with higher percentages.

QUANTIFICATION AND STATISTICAL ANALYSIS

To find out marker genes for each cluster in scRNA-seq data (Tables S1, S2, S5), differentially expressed genes were determined using the Find All Markers (or FindMarkers) function in Seurat package with Wilcoxon rank sum tests (adjusted P<0.05). DE genes for microglia (expressing typical homeostatic markers) from different brain regions (one region versus rest, Table S3) were determined by the likelihood ratio method (adjusted P<0.05) for the scRNA-seq data and quasi-likelihood F-test (adjusted P<0.05) for the bulk RNA-seq data with edgeR package. To calculate Pearson correlation for microglia gene expression from different brain regions in bulk RNA-seq, mean expression of a gene from 3 biological replicates for each region was used, and the pairwise analysis was performed in R. For GO term analysis (Table S6), PANTHER overrepresentation test was performed with the online software (version 13.1, released 2018–02-03), and statistical significance (P<0.05) was determined by Fisher's Exact with FDR multiple test correction.

To quantify RNA *in situ* fluorescence intensity of *Egr1* and *Fos* for microglia and non-microglial cells (30 cells each), a presentative cortical region on a 12um tissue section was selected, and fluorescence intensity from each channel was measured with Fiji by subtracting background signals from area integrated intensity. To quantify EGR1 and FOS protein levels in microglia and non-microglial cells (30 cells each), similar measurements were performed except using a 50um IHC staining section. To quantify the numbers of CLEC7A$^+$ microglia in corpus callosum or cerebellum from different developmental ages, we counted cells on 50um sagittal brain sections (cutting through similar anatomical planes) with Fiji (n=3 sections for each stage), and summarized the results with mean ± SEM. To compare the morphological and density differences between CLEC7A$^+$ and CLEC7A $^-$populations, we used Fiji to quantify each perimeter on 50um sagittal brain sections (n=10 cells for morphology data, and n=3 sections for density data). We summarized the results with mean ± SEM. Statistical significance (P<0.05) was determined by one-way ANOVA followed by pairwise t-test with Bonferroni correction. To do quantification in the phagocytosis assay, we counted the total number of cells (satisfying each condition) from 3 randomly chosen fields on each 50um sagittal section, and performed analysis using 5 sections per condition. The data were presented as mean ± SEM, and statistical significance (P<0.05) was determined by two-way ANOVA followed by Tukey's multiple comparisons test. To quantify the percentage of other neural cells interacting with CLEC7A$^+$ microglia, we first defined interaction as >30% of cellular volume overlapping with the CLEC7A (or CX3CR1-GFP) signal on 50um sagittal sections (n=3 sections for each cell type) calculated

by Imaris 3D reconstruction, and then counted the number of cells that fulfilled such criterion. The data were presented as mean ± SEM, and statistical significance ($P<0.05$) was determined by one-way ANOVA followed by pairwise t-test with Bonferroni correction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Anders S, Pyl PT, and Huber W (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. [PubMed: 25260700]

Ayata P, Badimon A, Strasburger HJ, Duff MK, Montgomery SE, Loh YE, Ebert A, Pimenova AA, Ramirez BR, Chan AT, et al. (2018). Epigenetic regulation of brain region-specific microglia clearance activity. Nature neuroscience 21, 1049–1060. [PubMed: 30038282]

Barres BA, Hart IK, Coles HS, Burne JF, Voyvodic JT, Richardson WD, and Raff MC (1992). Cell death and control of cell survival in the oligodendrocyte lineage. Cell 70, 31–46. [PubMed: 1623522]

Bennett ML, Bennett FC, Liddelow SA, Ajami B, Zamanian JL, Fernhoff NB, Mulinyawe SB, Bohlen CJ, Adil A, Tucker A, et al. (2016). New tools for studying microglia in the mouse and human CNS. Proceedings of the National Academy of Sciences of the United States of America

Bohlen CJ, Bennett FC, Tucker AF, Collins HY, Mulinyawe SB, and Barres BA (2017). Diverse Requirements for Microglial Survival, Specification, and Function Revealed by Defined-Medium Cultures. Neuron 94, 759–773 e758. [PubMed: 28521131]

Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, Teichmann SA, Marioni JC, and Stegle O (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat Biotechnol 33, 155–160. [PubMed: 25599176]

Butovsky O, Jedrychowski MP, Moore CS, Cialic R, Lanser AJ, Gabriely G, Koeglsperger T, Dake B, Wu PM, Doykan CE, et al. (2014). Identification of a unique TGF-beta-dependent molecular and functional signature in microglia. Nature neuroscience 17, 131–143. [PubMed: 24316888]

Chung WS, Clarke LE, Wang GX, Stafford BK, Sher A, Chakraborty C, Joung J, Foo LC, Thompson A, Chen C, et al. (2013). Astrocytes mediate synapse elimination through MEGF10 and MERTK pathways. Nature 504, 394–400. [PubMed: 24270812]

Colonna M, and Butovsky O (2017). Microglia Function in the Central Nervous System During Health and Neurodegeneration. Annu Rev Immunol

Cunningham CL, Martinez-Cerdeno V, and Noctor SC (2013). Microglia regulate the number of neural precursor cells in the developing cerebral cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 33, 4216–4233. [PubMed: 23467340]

Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM, Hayden Gephart MG, Barres BA, and Quake SR (2015). A survey of human brain transcriptome diversity at the single cell level.

Proceedings of the National Academy of Sciences of the United States of America 112, 7285–7290. [PubMed: 26060301]

De Biase LM, Schuebel KE, Fusfeld ZH, Jair K, Hawes IA, Cimbro R, Zhang HY, Liu QR, Shen H, Xi ZX, et al. (2017). Local Cues Establish and Maintain Region-Specific Phenotypes of Basal Ganglia Microglia. Neuron 95, 341–356 e346. [PubMed: 28689984]

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. [PubMed: 23104886]

Friedman J, Hastie T, and Tibshirani R (2008). Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9, 432–441. [PubMed: 18079126]

Fuchs Y, and Steller H (2011). Programmed cell death in animal development and disease. Cell 147, 742–758. [PubMed: 22078876]

Ginhoux F, Greter M, Leboeuf M, Nandi S, See P, Gokhan S, Mehler MF, Conway SJ, Ng LG, Stanley ER, et al. (2010). Fate Mapping Analysis Reveals That Adult Microglia Derive from Primitive Macrophages. Science 330, 841–845. [PubMed: 20966214]

Goldmann T, Wieghofer P, Jordao MJ, Prutek F, Hagemeyer 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

Gomez Perdiguero E, Klapproth K, Schulz C, Busch K, Azzoni E, Crozet L, Garner H, Trouillet C, de Bruijn MF, Geissmann F, et al. (2015). Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. Nature 518, 547–551. [PubMed: 25470051]

Grabert K, Michoel T, Karavolos MH, Clohisey S, Baillie JK, Stevens MP, Freeman TC, Summers KM, and McColl BW (2016). Microglial brain region-dependent diversity and selective regional sensitivities to aging. Nature neuroscience 19, 504–516. [PubMed: 26780511]

Grant GD, Brooks L, 3rd, Zhang X, Mahoney JM, Martyanov V, Wood TA, Sherlock G, Cheng C, and Whitfield ML (2013). Identification of cell cycle-regulated genes periodically expressed in U2OS cells and their regulation by FOXM1 and E2F transcription factors. Mol Biol Cell 24, 3634–3650. [PubMed: 24109597]

Hagemeyer N, Hanft KM, Akriditou MA, Unger N, Park ES, Stanley ER, Staszewski O, Dimou L, and Prinz M (2017). Microglia contribute to normal myelinogenesis and to oligodendrocyte progenitor maintenance during adulthood. Acta Neuropathol

Haimon Z, Volaski A, Orthgiess J, Boura-Halfon S, Varol D, Shemer A, Yona S, Zuckerman B, David E, Chappell-Maor L, et al. (2018). Re-evaluating microglia expression profiles using RiboTag and cell isolation strategies. Nat Immunol 19, 636–644. [PubMed: 29777220]

Hoeffel G, Chen J, Lavin Y, Low D, Almeida FF, See P, Beaudin AE, Lum J, Low I, Forsberg EC, et al. (2015). C-Myb(+) erythro-myeloid progenitor-derived fetal monocytes give rise to adult tissue-resident macrophages. Immunity 42, 665–678. [PubMed: 25902481]

Hong S, Dissing-Olesen L, and Stevens B (2016). New insights on the role of microglia in synaptic pruning in health and disease. Current opinion in neurobiology 36, 128–134. [PubMed: 26745839]

Kang SS, Ebbert MTW, Baker KE, Cook C, Wang X, Sens JP, Kocher JP, Petrucelli L, and Fryer JD (2018). Microglial translational profiling reveals a convergent APOE pathway from aging, amyloid, and tau. J Exp Med

Keren-Shaul H, Spinrad A, Weiner A, Matcovitch-Natan O, Dvir-Szternfeld R, Ulland TK, 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 e1217. [PubMed: 28602351]

Krasemann S, Madore C, Cialic R, Baufeld C, Calcagno N, El Fatimy R, Beckers L, O'Loughlin E, Xu Y, Fanek Z, et al. (2017). The TREM2-APOE Pathway Drives the Transcriptional Phenotype of Dysfunctional Microglia in Neurodegenerative Diseases. Immunity 47, 566–581 e569. [PubMed: 28930663]

Lawson LJ, Perry VH, Dri P, and Gordon S (1990). Heterogeneity in the distribution and morphology of microglia in the normal adult mouse brain. Neuroscience 39, 151–170. [PubMed: 2089275]

Li Q, and Barres BA (2017). Microglia and macrophages in brain homeostasis and disease. Nat Rev Immunol

Ling EA, and Wong WC (1993). The origin and nature of ramified and amoeboid microglia: a historical review and current concepts. Glia 7, 9–18. [PubMed: 8423067]

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 161, 1202–1214. [PubMed: 26000488]

Matcovitch-Natan O, Winter DR, 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.

Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, and Thomas PD (2010). PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. Nucleic Acids Res 38, D204–210. [PubMed: 20015972]

Nikodemova M, Kimyon RS, De I, Small AL, Collier LS, and Watters JJ (2015). Microglial numbers attain adult levels after undergoing a rapid decrease in cell number in the third postnatal week. J Neuroimmunol 278, 280–288. [PubMed: 25468773]

Paolicelli RC, Bolasco G, Pagani F, Maggi L, Scianni M, Panzanelli P, Giustetto M, Ferreira TA, Guiducci E, Dumas L, et al. (2011). Synaptic pruning by microglia is necessary for normal brain development. Science 333, 1456–1458. [PubMed: 21778362]

Papalexi E, and Satija R (2018). Single-cell RNA sequencing to explore immune cell heterogeneity. Nat Rev Immunol 18, 35–45. [PubMed: 28787399]

Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, and Sandberg R (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc 9, 171–181. [PubMed: 24385147]

Prinz M, Erny D, and Hagemeyer N (2017). Ontogeny and homeostasis of CNS myeloid cells. Nat Immunol 18, 385–392. [PubMed: 28323268]

Prinz M, Priller J, Sisodia SS, and Ransohoff RM (2011). Heterogeneity of CNS myeloid cells and their roles in neurodegeneration. Nature neuroscience 14, 1227–1235. [PubMed: 21952260]

Robinson MD, McCarthy DJ, and Smyth GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. [PubMed: 19910308]

Schafer DP, Lehrman EK, Kautzman AG, Koyama R, Mardinly AR, Yamasaki R, Ransohoff RM, Greenberg ME, Barres BA, and Stevens B (2012). Microglia sculpt postnatal neural circuits in an activity and complement-dependent manner. Neuron 74, 691–705. [PubMed: 22632727]

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods 9, 676–682. [PubMed: 22743772]

Schmieder R, and Edwards R (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863–864. [PubMed: 21278185]

Svensson V, Natarajan KN, Ly LH, Miragaia RJ, Labalette C, Macaulay IC, Cvejic A, and Teichmann SA (2017). Power analysis of single-cell RNA-sequencing experiments. Nat Methods 14, 381–387. [PubMed: 28263961]

Tabula Muris, C., Overall, c., Logistical, c., Organ, c., processing, Library, p., sequencing, Computational data, a., Cell type, a., Writing, g., et al. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature 562, 367–372. [PubMed: 30283141]

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, and Rinn JL (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 32, 381–386. [PubMed: 24658644]

Trapp BD, Nishiyama A, Cheng D, and Macklin W (1997). Differentiation and death of premyelinating oligodendrocytes in developing rodent brain. The Journal of cell biology 137, 459–468. [PubMed: 9128255]

Ueno M, Fujita Y, Tanaka T, Nakamura Y, Kikuta J, Ishii M, and Yamashita T (2013). Layer V cortical neurons require microglial support for survival during postnatal development. Nature neuroscience 16, 543–551. [PubMed: 23525041]

Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO, et al. (2002). Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell 13, 1977–2000. [PubMed: 12058064]

Wlodarczyk A, Holtman IR, Krueger M, Yogev N, Bruttger J, Khorooshi R, Benmamar-Badel A, de Boer-Bergsma JJ, Martin NA, Karram K, et al. (2017). A novel microglial subset plays a key role in myelinogenesis in developing brain. EMBO J 36, 3292–3308. [PubMed: 28963396]

Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, Phatnani HP, Guarnieri P, Caneda C, Ruderisch N, et al. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 34, 11929–11947. [PubMed: 25186741]

Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, Leonhardt H, Heyn H, Hellmann I, and Enard W (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol Cell 65, 631–643 e634. [PubMed: 28212749]
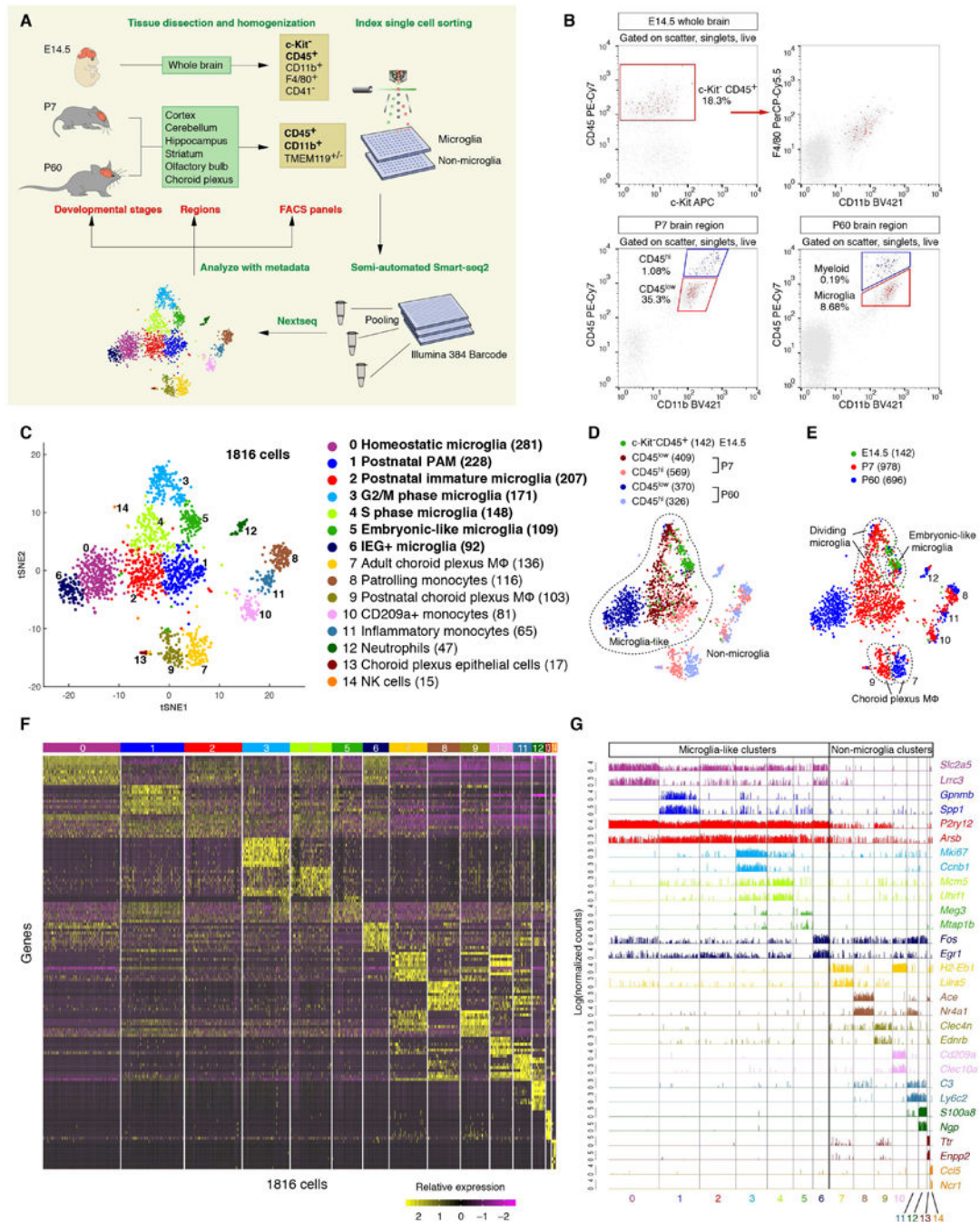
Figure 1. Clustering of brain myeloid cells across developmental stages by deep scRNA-seq.
(A) Schematic graph showing the experimental design. c-Kit⁻CD45⁺ (bold) was the sorting gate for E14.5 cells, and CD45⁺CD11b⁺ (bold) was for P7 and P60 cells. Other surface markers were recorded as metadata. (B) Representative FACS plots showing the cells sequenced (red or blue). (C) tSNE plot showing 15 clusters and population annotations with microglia-like clusters in bold and the number of cells in parentheses. (D) Overlaying FACS gating information onto the same tSNE plot in (C). The vast majority of P60 microglia are CD45^{low}, while P7 microglia are comprised of both CD45^{low} and CD45^{hi} cells. Almost all

cells from non-microglia clusters are CD45[hi]. (E) Overlaying developmental stage information onto the same tSNE plot in (C). (F) Heatmap showing the top 20 markers (or all markers if less than 20) for each of the 15 clusters. (G) Bar plots showing gene expression levels of two representative markers for each cluster. See also Figure S1, Table S1.
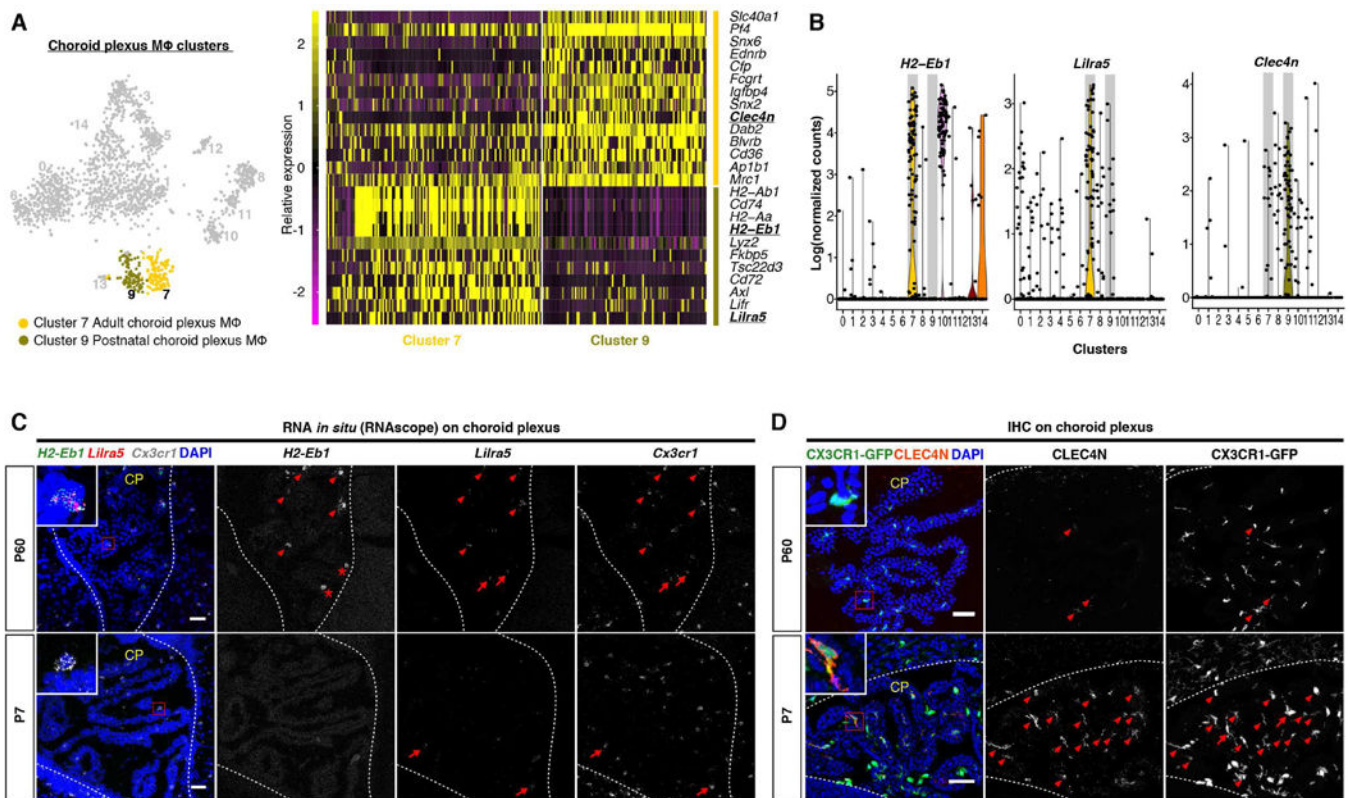
Figure 2. Postnatal transcriptional alteration of choroid plexus macrophages (CP MΦ).
(A) tSNE plot (left) and heatmap (right) highlighting two stage-specific CP MΦ clusters and differentially expressed genes. Genes for validation in (C) and (D) are underlined and in bold. (B) Violin plots showing that *H2-Eb1* and *Lilra5* are highly expressed by adult CP MΦ (cluster 7), and *Clec4n* is highly enriched in early postnatal CP MΦ (cluster 9). (C) RNA *in situ* validation of abundant *H2-Eb1* and *Lilra5* expression in P60 CP MΦ (*H2-Eb1*+*Lilra5*+*Cx3cr1*+, arrow heads) but not in P7 cells. *H2-Eb1*− *Cx3cr1*+*Lilra5*+ (arrows) cells were occasionally seen in both stages. *H2-Eb1*+*Lilra5*− *Cx3cr1*− (asterisks) are non-myeloid cells. (D) Immunostaining validation of more abundant CLEC4N expression in P7 CP MΦ (CLEC4N+CX3CR1-GFP+, arrow heads). CLEC4N−CX3CR1-GFP+ (arrow) cells also exist in P7. CP, choroid plexus. Dashed lines demarcate ventricle borders. Scale bars in (C) and (D) are 50um.
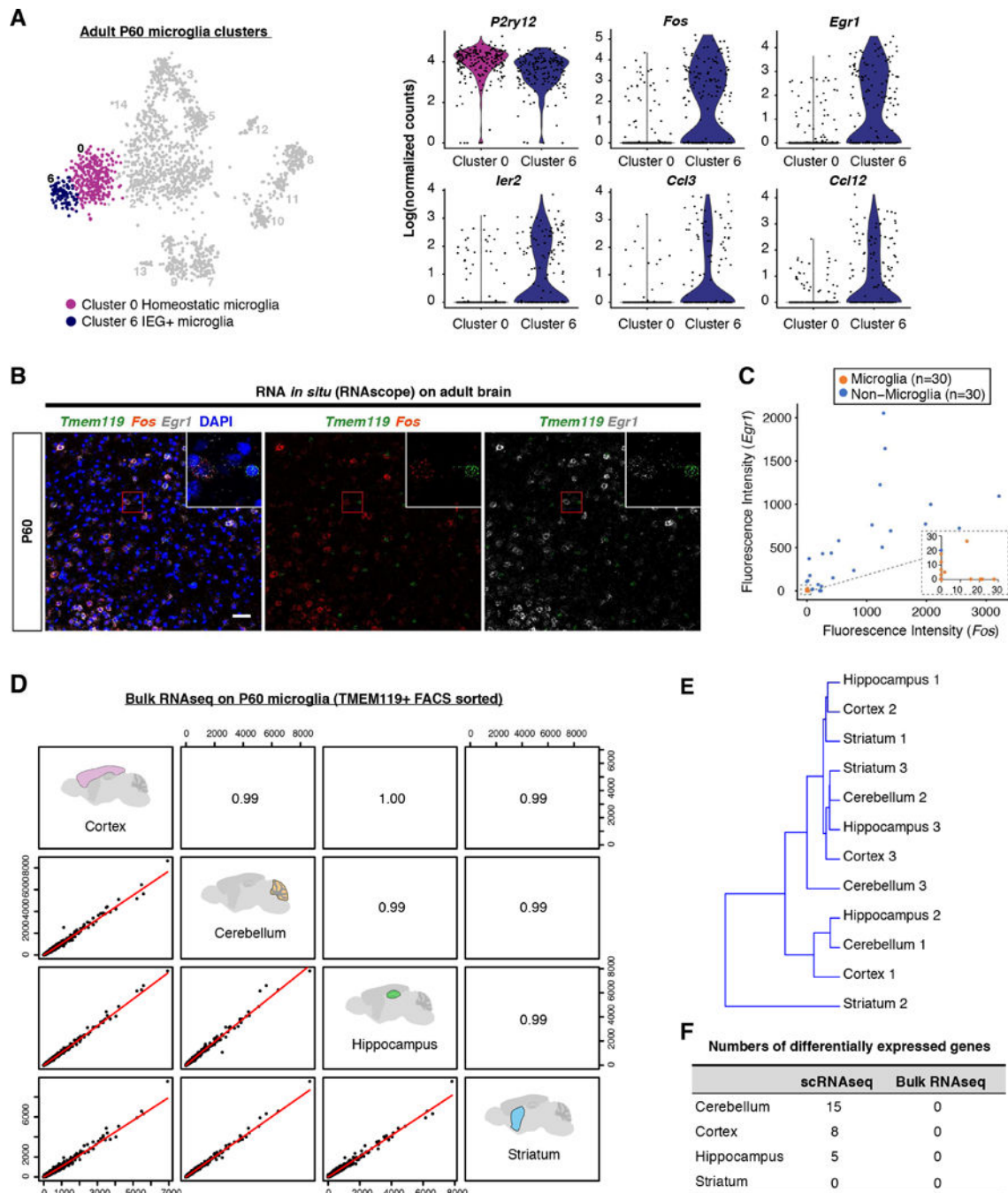
Figure 3. Limited transcriptomic heterogeneity of adult homeostatic microglia across brain regions.

(A) tSNE plot (left) and violin plots (right) highlighting two P60 microglia clusters, and examples of differentially expressed genes. (B) RNA *in situ* showing negative expression of *Fos* and *Egr1* in microglia (*Tmem119*+), and positive signals from surrounding non-microglia cells. (C) Quantification of fluorescence signals in (B). n=30 cells for each. (D) Pearson correlation between bulk RNA-seq samples from P60 brain regions. (E) Dendrogram showing hierarchical clustering of bulk RNA-seq samples, irrelevant of regions. (F) Numbers of differentially expressed genes between homeostatic microglia from different

regions by scRNA-seq and bulk RNA-seq. Scale bar in (B) is 50um. See also Figure S2, Table S2, Table S3.
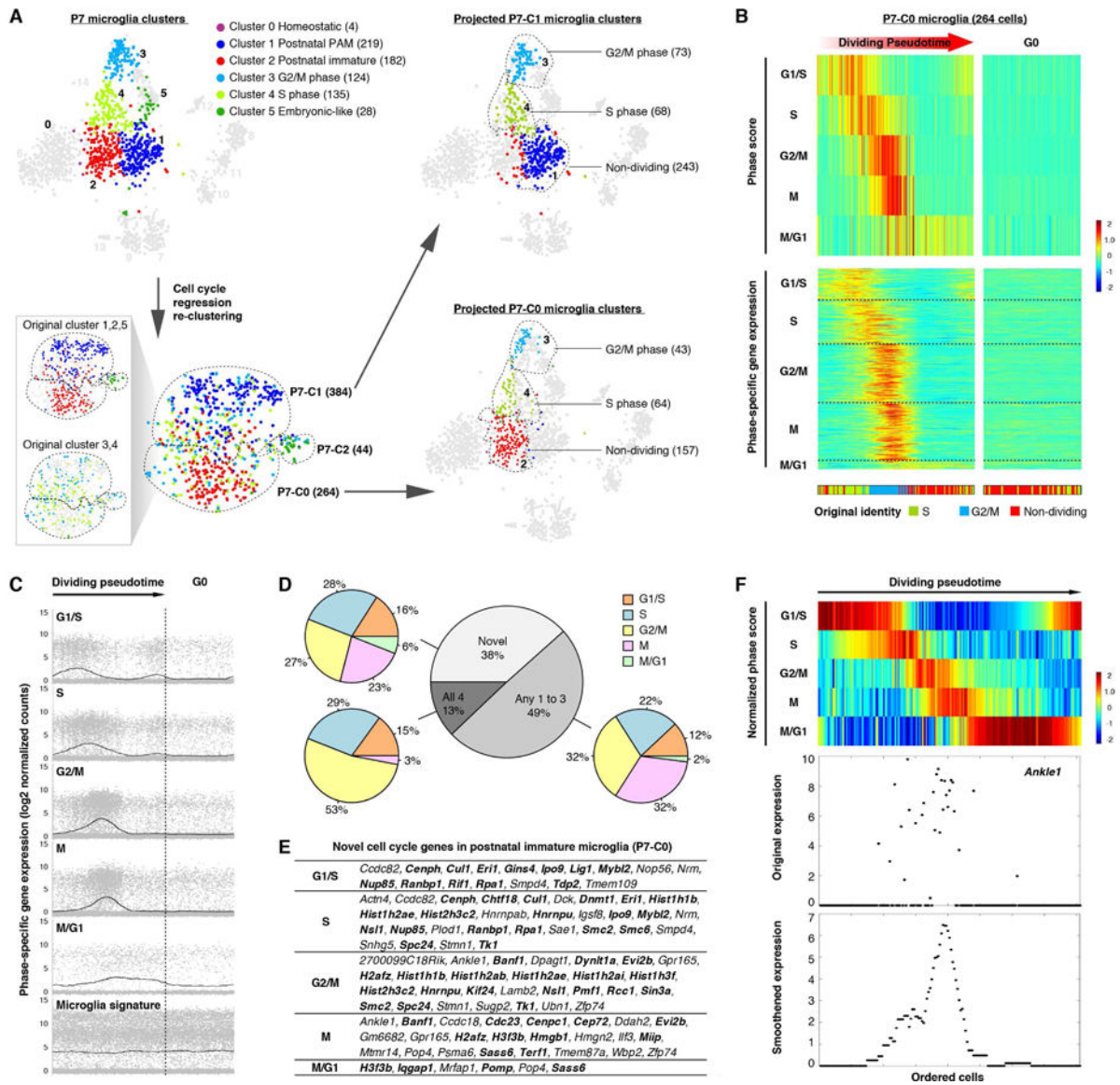
Figure 4. Phase-specific gene expression of dividing microglia (P7-C0 cluster) along cell cycle pseudotime.

(A) tSNE plots demonstrating cell cycle regression and re-clustering of P7 microglia. Cells in all tSNE plots are color coded exactly the same way as in Figure 1C. Numbers of cells in each cluster are given in parentheses. Two tSNE plots in the box (lower left) demonstrate consistency of two analyses (before and after cell cycle regression) and distribution of the original dividing clusters (cluster 3 and 4) into P7-C0, P7-C1 and P7-C2 following cell cycle regression. (B) – (F) showing analysis for the 264 P7-C0 microglia identified in (A). (B) Heatmap (upper panel) showing pseudotime ordering of P7-C0 microglia based on raw phase scores. Each column is a cell, and each row denotes raw scores for a specific cell cycle phase. G0 cells have no dominant phase scores for any phase and were not ordered. Heatmap (lower panel) showing expression levels of identified phase-specific genes. Each column is a cell and each row is a gene. The original identities (bottom bar) are color coded as in (A). (C) Dot plots showing expression levels of phase-specific genes along microglia dividing

pseudotime. Genes for each phase are plotted in separate graphs with each dot representing the level of expression for a given gene in a given cell. Curves show average expression of all genes assigned to a phase along dividing pseudotime. Expression of microglial signature genes (Butovsky et al., 2014) is shown at the bottom. (D) The gray pie chart showing overlaps of phase-specific genes identified by the algorithm compared with four published cell cycle gene sets (Grant et al., 2013). Colored pie charts are breakdowns of the genes from each category based on the phase assignment. (E) Table showing gene names identified as "Novel" in (D). Genes that may play a role in cell division are in bold. (F) Heatmap (upper panel) showing pseudotime ordering of P7-C0 microglia by normalized phase scores. The bottom panels show *Ankle1* as an example for its expression dynamics along dividing pseudotime. Each dot is a single cell. The smoothened expression is the average expression of *Ankle1* computed with a fixed window size (length of ordered cells/10). See also Figure S3, Figure S4, Table S4.
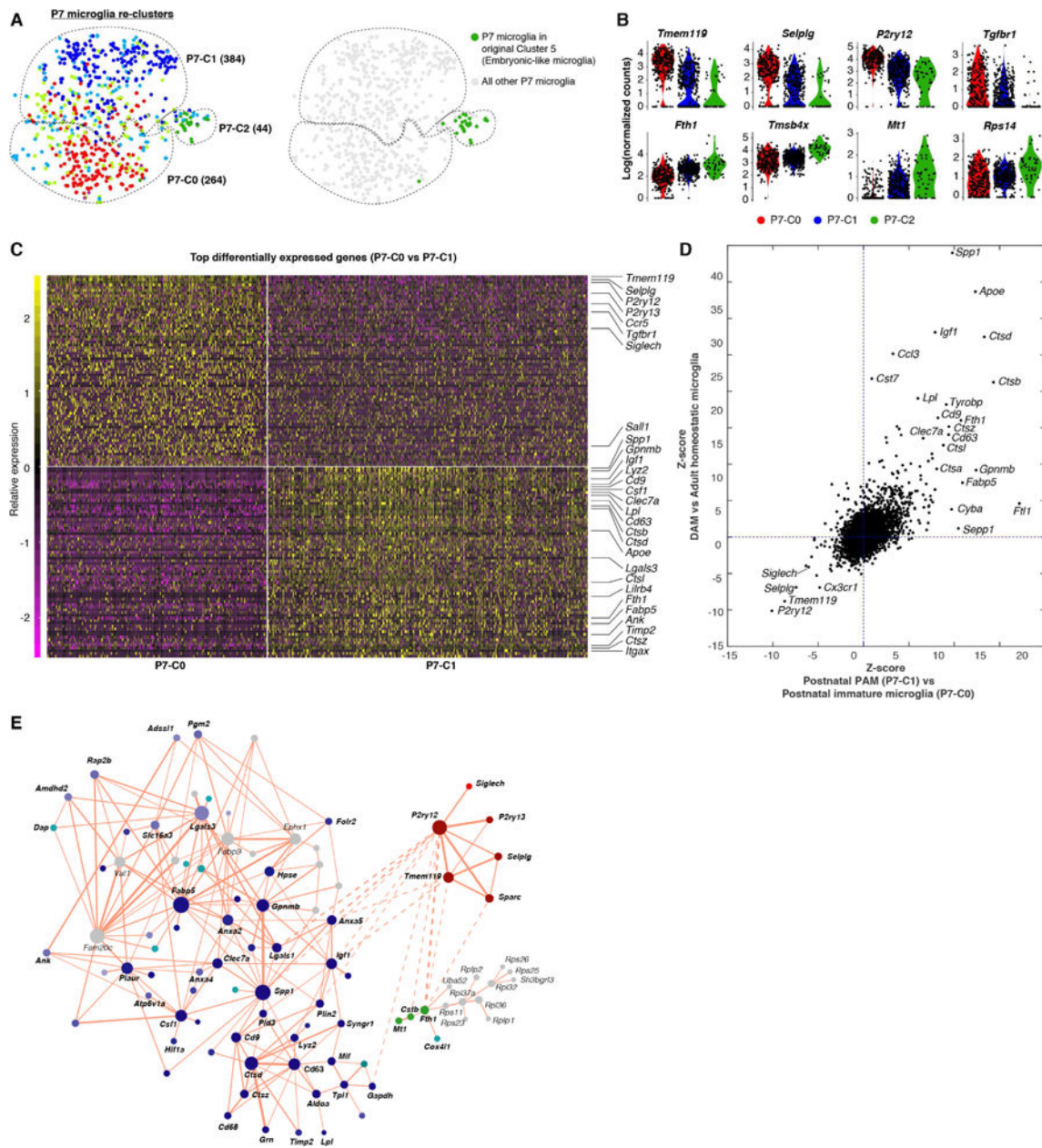
Figure 5. Heterogeneity of early postnatal microglia revealed by scRNA-seq.
(A) Re-clustering of P7 microglia (as in Figure 4A) to highlight that P7-C2 are mainly embryonic-like cells (cluster 5) in the original clustering result. (B) Violin plots showing some differentially expressed genes between P7-C2 and the other two clusters. (C) Heatmap showing top 70 differentially expressed genes between P7-C0 and P7-C1. P7-C0 express higher levels of homeostatic genes and P7-C1 enrich many disease-associated genes. (D) Comparison of gene expression changes in DAM (relative to homeostatic microglia) with the changes in P7-C1 (relative to P7-C0) showing similar sets of up-and down-regulated genes in two cases. (E) Gene network showing correlated gene modules underlying cluster identities of P7 microglia (See STAR Methods). Each gene is colored based on its

differential expression levels among three P7 clusters. Solid lines are for positive correlation and dashed lines are for negative correlation. See also Table S5.
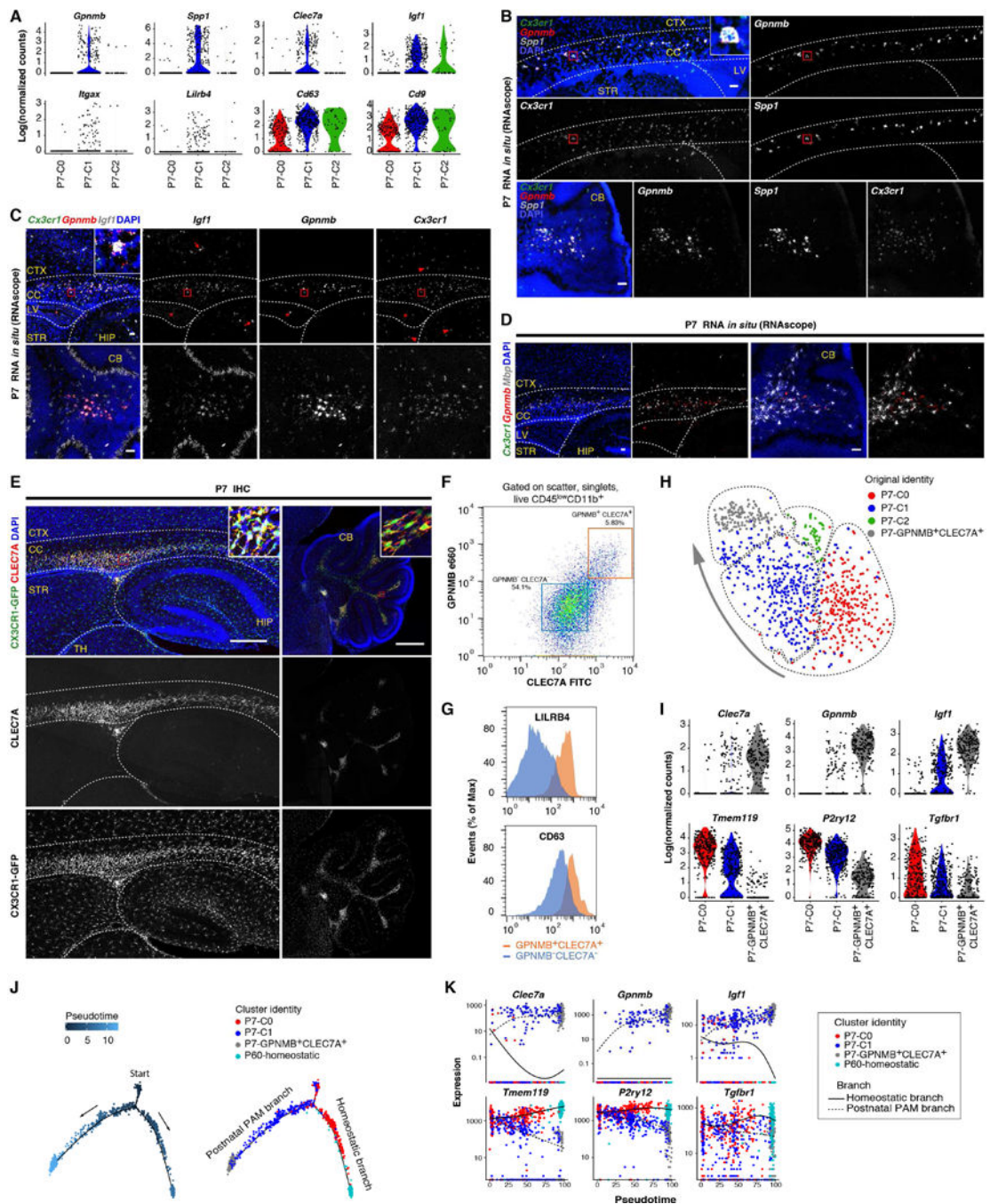
Figure 6. Validation of early postnatal proliferative region-associated microglia.
(A) Violin plots showing some top up-regulated genes in P7-C1 (PAM) compared with the other two clusters. (B) RNA *in situ* (RNAscope) showing that *Spp1* and *Gpnmb* signals mainly overlap in the corpus callosum (CC, inset) and white matter region of cerebellum (CB), and they also overlap with the microglia marker *Cx3cr1*. (C) RNA *in situ* showing that *Gpnmb⁺Cx3cr1⁺* microglia are positive for *Igf1* in CC (inset) and CB white matter. Microglia in the cortex (CTX), hippocampus (HIP) and striatum (STR) are negative for *Gpnmb* and *Igf1* (arrow heads). *Gpnmb⁺Cx3cr1⁺Igf1⁺* cells in the lateral ventricle (LV)

(asterisk) and *Igf1*+*Cx3cr1*− neural cells (arrows) are also labeled. (D) RNA *in situ* showing that *Gpnmb*+*Cx3cr1*+ microglia are intermingled with *Mbp*+ oligodendrocytes in the developing white matter. (E) Immunohistochemistry showing CLEC7A expression by P7-C1 (PAM) in the developing white matter and near ventricles. TH: thalamus. (F) FACS plot showing P7 cerebellar GPNMB+CLEC7A+ cells (orange box) isolated for scRNA-seq. (G) Histograms showing higher levels of LILRB4 and CD63 surface expression in GPNMB+CLEC7A+ microglia compared with GPNMB−CLEC7A−cells. (H) tSNE plot showing clustering result (dashed circles) after combining GPNMB+CLEC7A+ cells with the originally sequenced P7 microglia. The arrow indicates gradual changes of transcriptomes from postnatal immature state (P7-C0) towards more polarized GPNMB+CLEC7A+ state. (I) Violin plots showing further up-or down-regulation of differentially expressed genes in P7-GPNMB+CLEC7A+ compared with P7-C1 microglia by scRNA-seq. (J) Pseudotime analysis of P7 microglia together with P7-GPNMB+CLEC7A+ and P60 homeostatic microglia showing developmental trajectories from P7-C0/P7-C1 mixed starting point to P7-GPNMB+CLEC7A+ early postnatal PAM branch (via P7-C1) and to P60 homeostatic branch (via P7-C0). Each dot is a cell. (K) Gene expression dynamics for two trajectories in (J) along developmental pseudotime, when the PAM branch gradually turns on disease-associated genes and down-regulates homeostatic genes. Scale bars: 50um in (B)-(D), 500um in (E). See also Figure S5, Table S5.
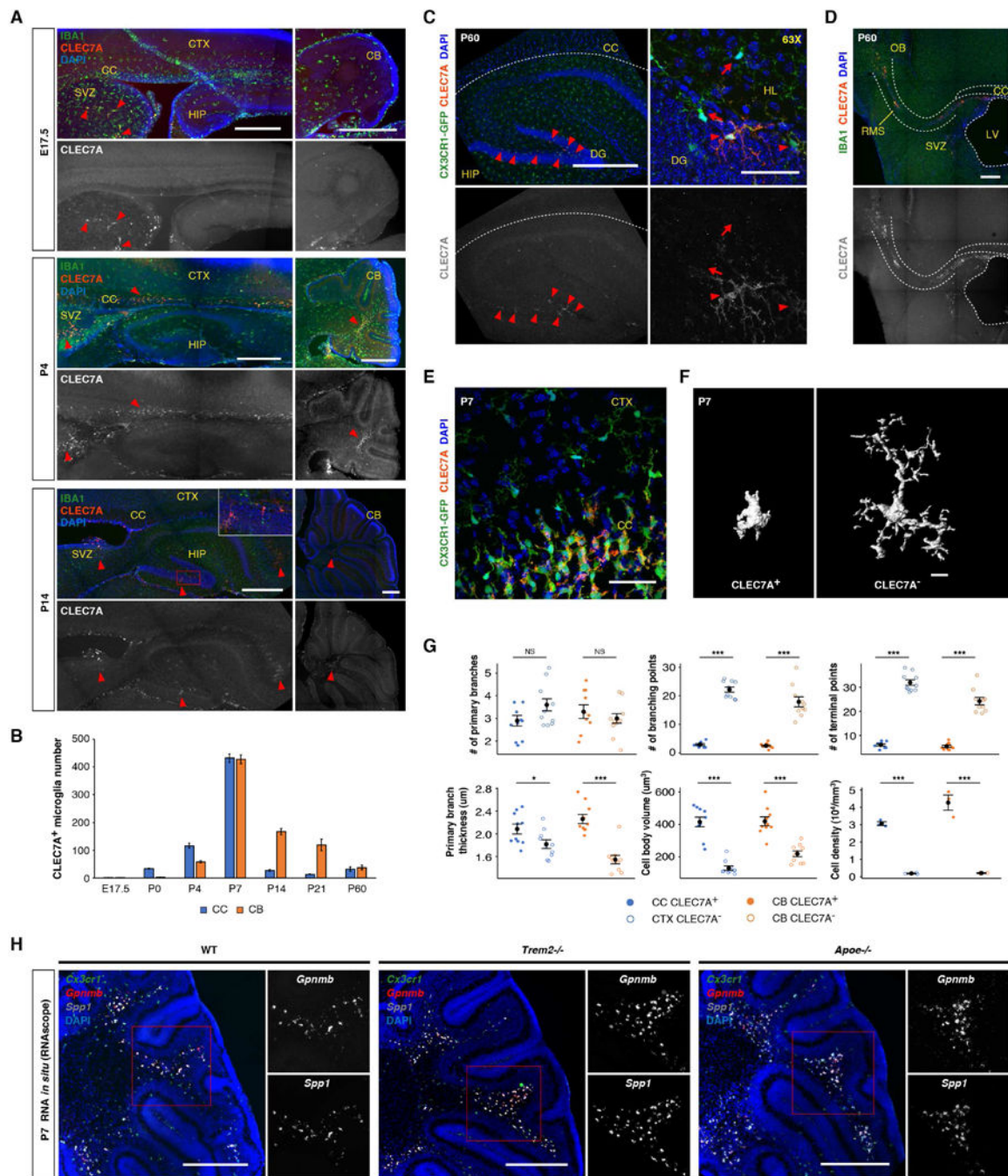
Figure 7. Transient appearance of PAM in developing white matter, independent of TREM2-APOE regulation.

(A) Immunohistochemistry showing CLEC7A+ microglia (arrow heads) from different developmental stages. Inset shows CLEC7A+ microglia in hippocampal dentate gyrus at P14. SVZ: subventricular zone. (B) Quantification of CLEC7A+ microglia across developmental stages. n=3 sections for each stage. (C) At P60, CLEC7A+ microglia are present in hippocampal dentate gyrus (DG, arrow heads). Arrows point to CLEC7A−cells. HL: hilus. (D) At P60, CLEC7A+ microglia are also present in SVZ and rostral migratory stream (RMS). OB: olfactory bulb. (E) 63X confocal images showing differences in

morphology and density between CLEC7A$^+$ (in CC) and CLEC7A$^-$(in CTX) microglia at P7. (F) 3D reconstruction of representative P7 microglia. (G) Quantification showing differences in morphology (n=10 cells each) and density (n=3 sections each) between CLEC7A$^+$ and CLEC7A$^-$microglia at P7. *** P<0.001, ** P<0.01, * P<0.05. (H) RNA *in situ* showing the presence of *Gpnmb$^+$Spp1$^+$* microglia in *Trem2*–/– or *Apoe*–/– cerebellum at P7. Scale bars: 500um in (A), left panels in (C), (D) and (H); 50um in (E) and the right panels of (C); 10um in (F). Data are represented as mean ± SEM in (B) and (G). See also Figure S6.
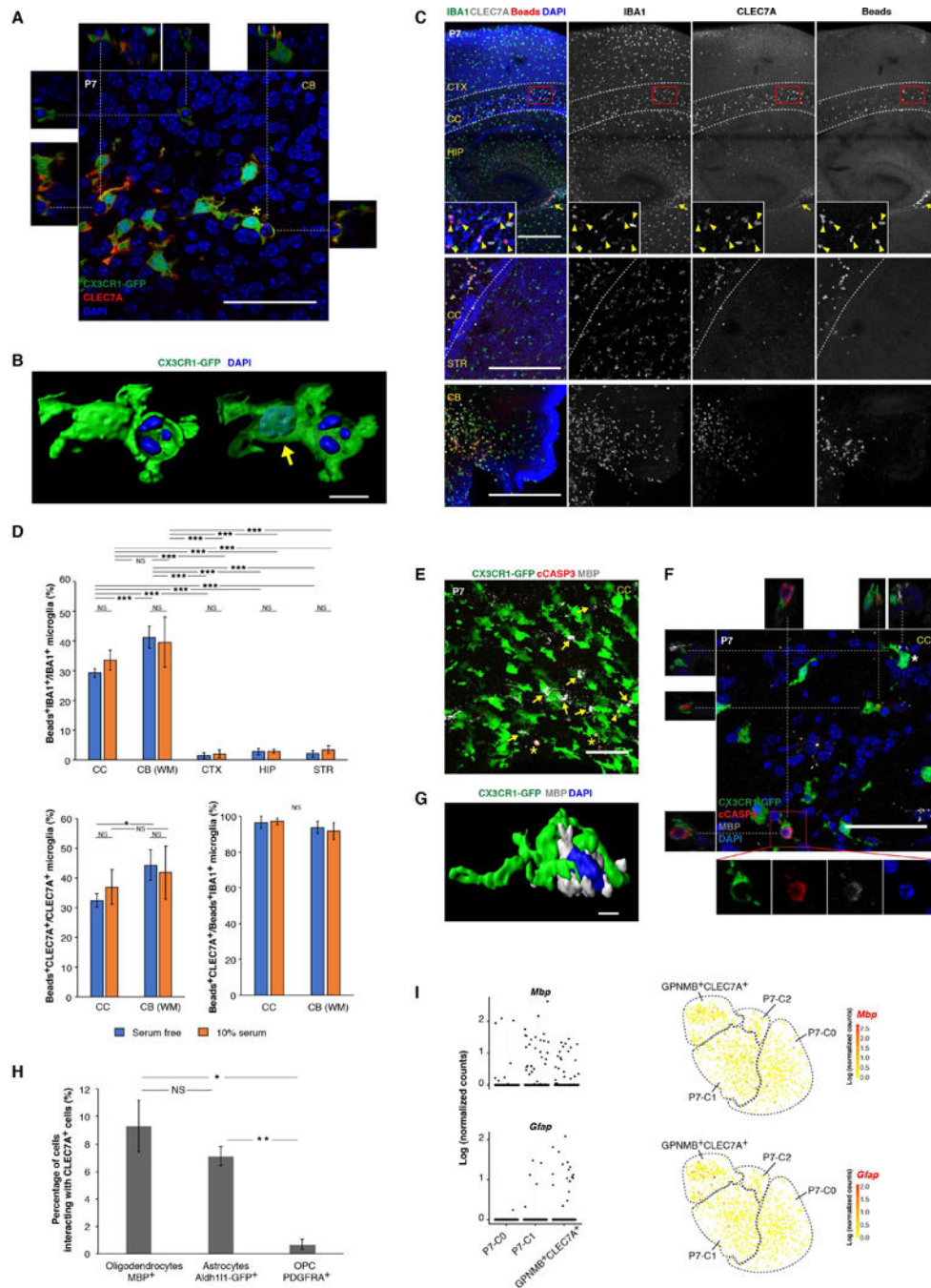
Figure 8. Phagocytosis of newly formed oligodendrocytes by early postnatal PAM.
(A) Single optical section of a confocal image (full stack in Movie S1) showing engulfment of nuclei by CLEC7A+ microglia in the P7 cerebellar white matter. Images from the X-Z and Y-Z axes are shown on the top and sides, respectively. (B) 3D reconstruction of a CLEC7A+ microglia (asterisk in (A)) engulfing a pyknotic nucleus. Arrow points to the microglia nucleus in transparent rendering of the CX3CR1-GFP channel. (C) Phagocytosis on slice culture sections showing that pH-sensitive beads were engulfed by CLEC7A+ microglia in CC (arrow heads) and CB white matter but rarely in other regions. Arrow points

to beads-eaten CLEC7A$^+$ cells near ventricles. (D) Quantification of the phagocytosis assay in (C). n=5 sections (3 fields for each section). *** *P*<0.001, ** *P*<0.01, * *P*<0.05. WM: white matter. (E) Immunostaining of P7 corpus callosum (CC) showing interactions between early postnatal PAM and MBP$^+$ oligodendrocytes (arrows). Asterisks label cCASP3$^+$ oligodendrocytes. Arrow head points to cCASP3$^+$ inclusion in microglia. (F) Single optical section of a confocal image showing engulfment of cCASP3$^+$ oligodendrocytes (inset) by CLEC7A$^+$ microglia in the P7 CC. Images from the X-Z and Y-Z axes are shown on the top and left, respectively. The microglial cell (asterisk) physically contacting an oligodendrocyte is 3D reconstructed in (G). (H) Quantification for percentage of each cell type interacting with CLEC7A$^+$ cells in CC. n=3 sections each. ** *P*<0.01, * *P*<0.05. (I) *Mbp* and *Gfap* transcripts in different P7 populations by scRNA-seq. tSNE plots (same as in Figure 6H) on the right highlight cells that have detectable *Mbp* or *Gfap* expression. Scale bars: 50um in (A), (E) and (F); 500um in (C); 10um in (B) and (G). Data are represented as mean ± SEM in (D) and (H). See also Figure S7, Table S6, Table S7, Movie S1, Movie S2.