Cell Reports, Volume *21*

Supplemental Information

Temporal Tracking of Microglia Activation in Neurodegeneration at Single-Cell Resolution

Hansruedi Mathys, Chinnakkaruppan Adaikkan, Fan Gao, Jennie Z. Young, Elodie Manet, Martin Hemberg, Philip L. De Jager, Richard M. Ransohoff, Aviv Regev, and Li-Huei Tsai

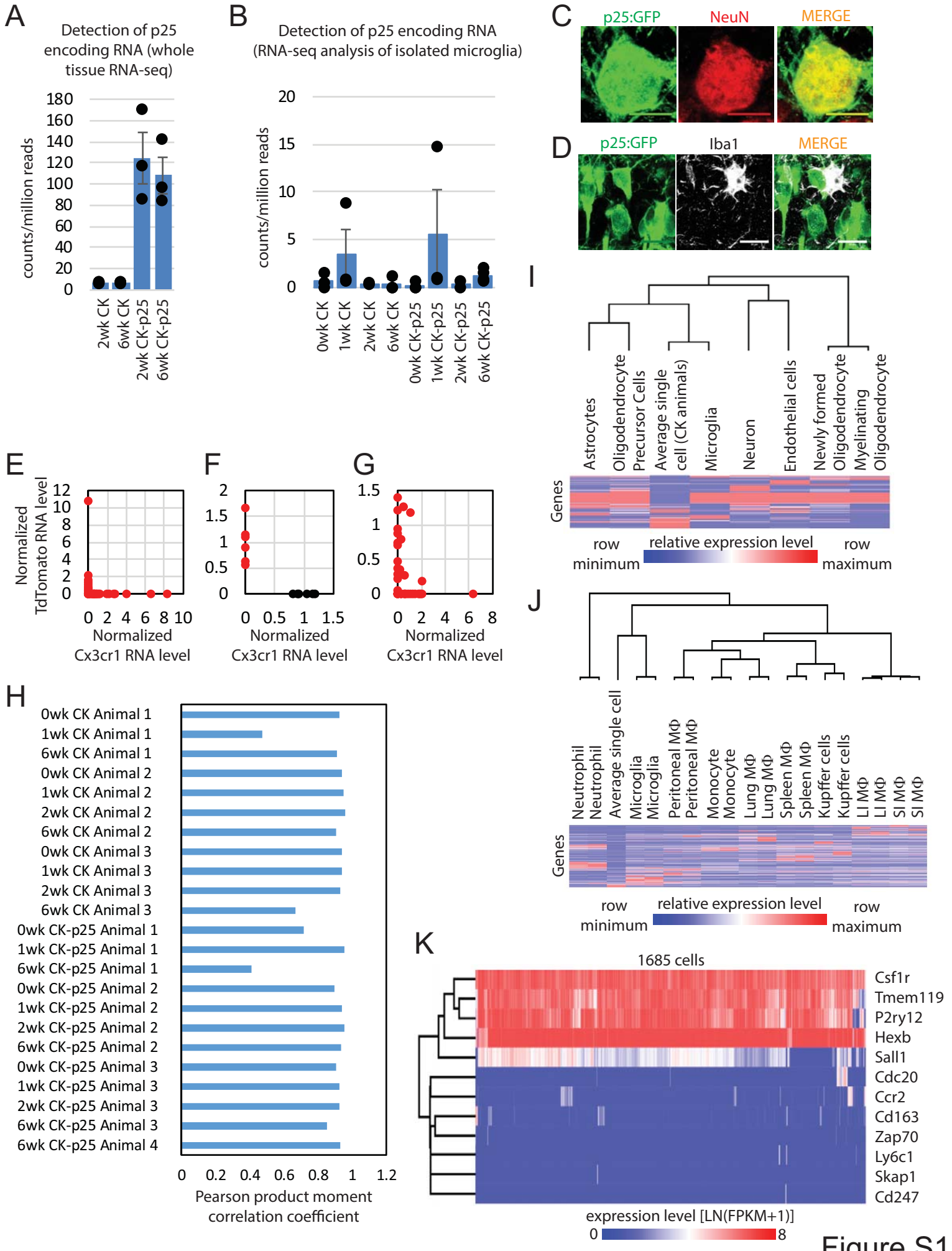

A
Detection of p25 encoding RNA (whole tissue RNA-seq)
counts/million reads
2wk CK
6wk CK
2wk CK-p25
6wk CK-p25
B
Detection of p25 encoding RNA (RNA-seq analysis of isolated microglia)
counts/million reads
0wk CK
1wk CK
2wk CK
6wk CK
0wk CK-p25
1wk CK-p25
2wk CK-p25
6wk CK-p25
C
p25:GFP
NeuN
MERGE
D
p25:GFP
Iba1
MERGE
E
Normalized TdTomato RNA level
Normalized Cx3cr1 RNA level
F
Normalized Cx3cr1 RNA level
G
Normalized Cx3cr1 RNA level
H
0wk CK Animal 1
1wk CK Animal 1
6wk CK Animal 1
0wk CK Animal 2
1wk CK Animal 2
2wk CK Animal 2
6wk CK Animal 2
0wk CK Animal 3
1wk CK Animal 3
2wk CK Animal 3
6wk CK Animal 3
0wk CK-p25 Animal 1
1wk CK-p25 Animal 1
6wk CK-p25 Animal 1
0wk CK-p25 Animal 2
1wk CK-p25 Animal 2
2wk CK-p25 Animal 2
6wk CK-p25 Animal 2
0wk CK-p25 Animal 3
1wk CK-p25 Animal 3
2wk CK-p25 Animal 3
6wk CK-p25 Animal 3
6wk CK-p25 Animal 4
Pearson product moment correlation coefficient
I
Astrocytes
Oligodendrocyte Precursor Cells
Average single cell (CK animals)
Microglia
Neuron
Endothelial cells
Newly formed Oligodendrocyte
Myelinating Oligodendrocyte
Genes
row minimum
relative expression level
row maximum
J
Neutrophil
Neutrophil
Average single cell
Microglia
Microglia
Peritoneal MΦ
Peritoneal MΦ
Monocyte
Monocyte
Lung MΦ
Lung MΦ
Spleen MΦ
Spleen MΦ
Kupffer cells
Kupffer cells
LI MΦ
LI MΦ
SI MΦ
SI MΦ
Genes
row minimum
relative expression level
row maximum
K
1685 cells
Csf1r
Tmem119
P2ry12
Hexb
Sall1
Cdc20
Ccr2
Cd163
Zap70
Ly6c1
Skap1
Cd247
expression level [LN(FPKM+1)]

Figure S1

Figure S1. Single-cell RNA-sequencing of microglia cells isolated from the hippocampus of CK-p25 animals and CK control littermates. Related to Figure 1

Bar graphs showing the expression of human p25 RNA in A) whole hippocampal tissue and B) isolated microglia cells. Average and individual data points (black dots) of two to three animals per time point and genotype is shown. Error bars show the standard error of the mean. C and D) Immunohistochemistry with anti-GFP (green), anti-NeuN (red), and anti-IBA1 (white) antibodies in the dentate gyrus of CK and CK-p25 animals 6 weeks after p25 induction. E) Single-cell qPCR analysis of a mixture of microglia isolated from a wild type animal and a Cx3cr1-knockout mouse expressing TdTomato in microglia. The scatter plot shows the Cx3cr1 and TdTomato RNA levels (both normalized to the Actb RNA level) measured in each of 45 wells to which a single CD11b- and CD45-positive cell was sorted by FACS. The values shown are relative to the average Cx3cr1 and TdTomato RNA levels (both normalized to the Actb RNA level) measured in 6 wells to which 50 CD11b- and CD45-double positive cells were sorted. F) qPCR analysis of RNA isolated from 50 microglia cells isolated from a wild type animal (black) or a Cx3cr1-knockout mouse expressing TdTomato in microglia (red). The results of 6 technical replicates for each genotype are shown. The scatter plot shows Cx3cr1 and TdTomato RNA levels, normalized to the Actb RNA level. G) Single-cell qPCR analysis of a mixture of microglia isolated from a wild type animal and a Cx3cr1-knockout mouse expressing TdTomato in microglia. The scatter plot shows Cx3cr1 and TdTomato RNA levels normalized to the Actb RNA level, as measured in each of 36 wells into which two CD11b- and CD45-double positive cells were sorted by FACS. H) Pearson product moment correlation coefficients of transcript expression [log10(FPKM+1)] between the average of 95 single-cells and a bulk population of 200 cells. I) Hierarchical clustering of average microglia expression profile (CD11b- and CD45-double positive cells isolated from CK control animals at all four time points) as determined by single-cell RNA-sequencing and previously established brain cell-type specific expression profiles (Zhang et al., 2014). J) Hierarchical clustering of average microglia expression profile (CD11b- and CD45-double positive cells isolated from CK control animals at all four time points) as determined by single-cell RNA-sequencing and previously established expression profiles of monocytes and tissue-resident macrophages (Lavin et al., 2014). K) Heat map showing the expression level of selected microglia signature genes (Csf1r, Tmem119, P2ry12, Hexb, Sall1), genes preferentially expressed in peripheral immune cells (Cdc20, Ccr2, Cd163, Ly6c1), and natural killer cell and T cell signature genes (Zap70, Skap1, Cd247) across the 1685 CD11b- and CD45-positive cells analyzed in this study.

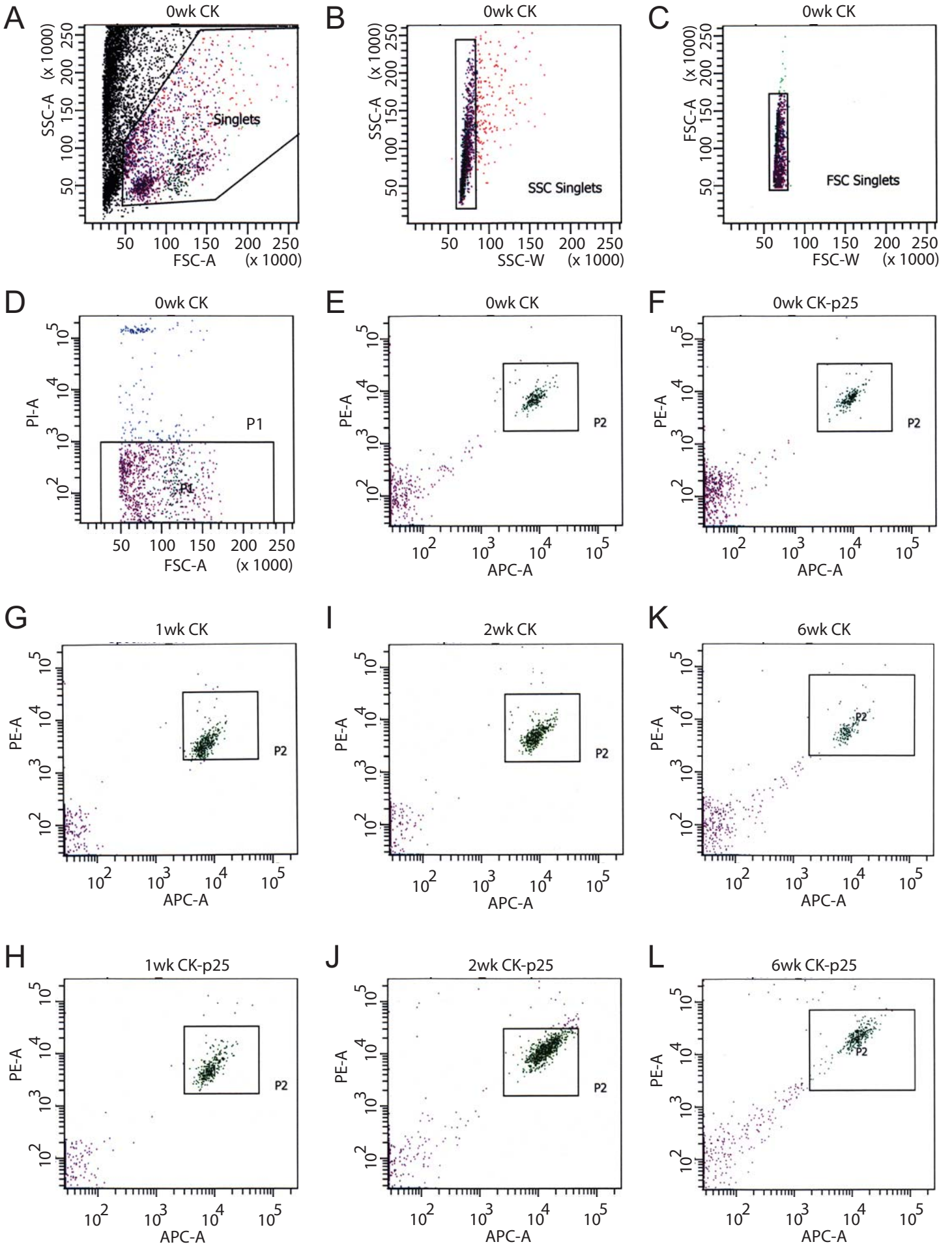

A
0wk CK
SSC-A (x 1000)
FSC-A (x 1000)
Singlets
B
0wk CK
SSC-A (x 1000)
SSC-W (x 1000)
SSC Singlets
C
0wk CK
FSC-A (x 1000)
FSC-W (x 1000)
FSC Singlets
D
0wk CK
PI-A
FSC-A (x 1000)
P1
P1
E
0wk CK
PE-A
APC-A
P2
F
0wk CK-p25
PE-A
APC-A
P2
G
1wk CK
PE-A
APC-A
P2
I
2wk CK
PE-A
APC-A
P2
K
6wk CK
PE-A
APC-A
P2
H
1wk CK-p25
PE-A
APC-A
P2
J
2wk CK-p25
PE-A
APC-A
P2
L
6wk CK-p25
PE-A
APC-A
P2

Figure S2

Figure S2. FACS gating strategy for the isolation of CD11b- and CD45-positive cells. Related to Figure 1

FACS scatter plots showing the gating strategy for isolating CD11b- and CD45-positive cells. Panels A to E show the entire gating tree for one representative sample. A) Intact cells were separated from most of the debris based on forward scatter area (FSC-A) and side scatter (SSC-A). B) and C) Events that could represent more than one cell were excluded based on the side scatter width (SSC-W) (B) and the forward scatter width (FSC-W) (C). D) Propidium iodide-negative cells (viable cells) were selected. E) CD11b-APC and CD45-PE-positive cells were selected and sorted into 96-well plates. Panels E) to L) show the selection of CD11b-APC- and CD45-PE-positive cells of a representative sample for each of the four time points and two genotypes analyzed in this study.

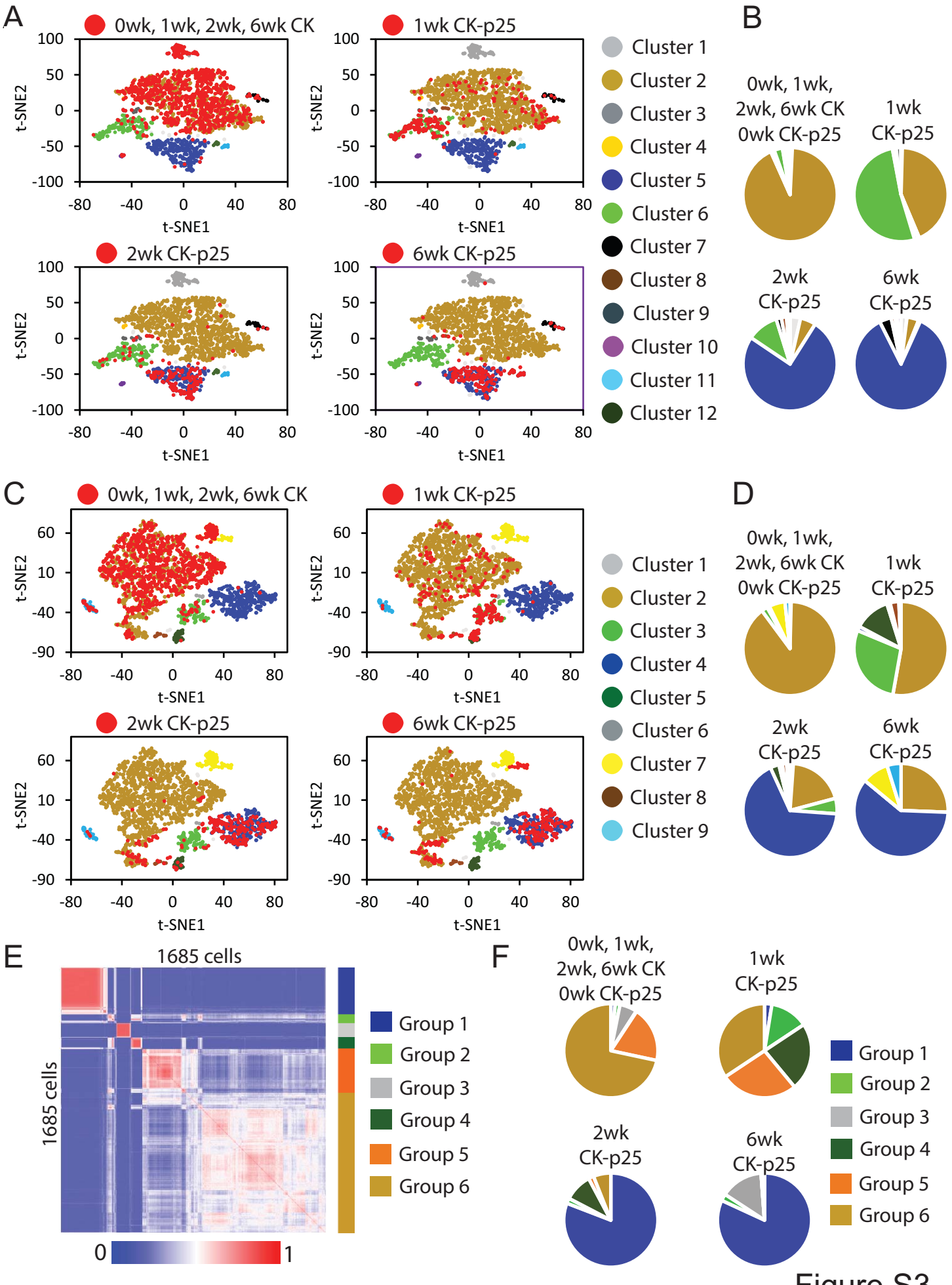

A
0wk, 1wk, 2wk, 6wk CK
1wk CK-p25
2wk CK-p25
6wk CK-p25
t-SNE1
t-SNE2
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
Cluster 10
Cluster 11
Cluster 12
B
0wk, 1wk, 2wk, 6wk CK 0wk CK-p25
1wk CK-p25
2wk CK-p25
6wk CK-p25
C
0wk, 1wk, 2wk, 6wk CK
1wk CK-p25
2wk CK-p25
6wk CK-p25
t-SNE1
t-SNE2
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cluster 9
D
0wk, 1wk, 2wk, 6wk CK 0wk CK-p25
1wk CK-p25
2wk CK-p25
6wk CK-p25
E
1685 cells
1685 cells
0
1
Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
F
0wk, 1wk, 2wk, 6wk CK 0wk CK-p25
1wk CK-p25
2wk CK-p25
6wk CK-p25
Group 1
Group 2
Group 3
Group 4
Group 5
Group 6

Figure S3

Figure S3. Non-linear dimensionality reduction and consensus clustering reveal multiple distinct and disease stage-specific microglia cell states. Related to Figure 2

A) t-SNE plots showing a two-dimensional representation of global gene expression profile relationships among 1685 CD11b- and CD45-positive cells. t-SNE was performed after reducing the effects of confounders by normalizing the gene expression measures (Gaublomme et al., 2015). Cells isolated from CK control animals (at all four time points) and CK-p25 animals 1, 2, and 6 weeks after p25 induction (1wk, 2wk, and 6wk, respectively) are highlighted in red in the individual panels. B) Pie charts showing the percentage of cells found in the clusters identified in (A). Cells were grouped by genotype and time point. Clusters are colored as in (A). C) t-SNE plot showing a two-dimensional representation of global gene expression profile relationships among 2070 CD11b- and CD45-positive cells. Gene expression profiles were determined using RSEM and cells with less than 300 genes detected were excluded (retaining 2070 cells). Cells isolated from CK control animals (at all four time points) and 1wk, 2wk, and 6wk CK-p25 animals are highlighted in red in the individual panels. D) Pie charts showing the percentage of cells found in the clusters identified in (C). Cells were grouped by genotype and time point. Clusters are colored as in C. E) Consensus matrix as generated by Single Cell Consensus Clustering (SC3). The matrix indicates how often each pair of cells was assigned to the same cluster by the different parameter combinations as indicated by the color bar at the bottom of the panel. A value of 1 (red) indicates that the cells were always assigned to the same cluster whereas a value of 0 (blue) indicates that the cells were never assigned to the same cluster. F) Pie charts showing the percentage of cells found in the groups identified in (E). Cells were grouped by genotype and time point.

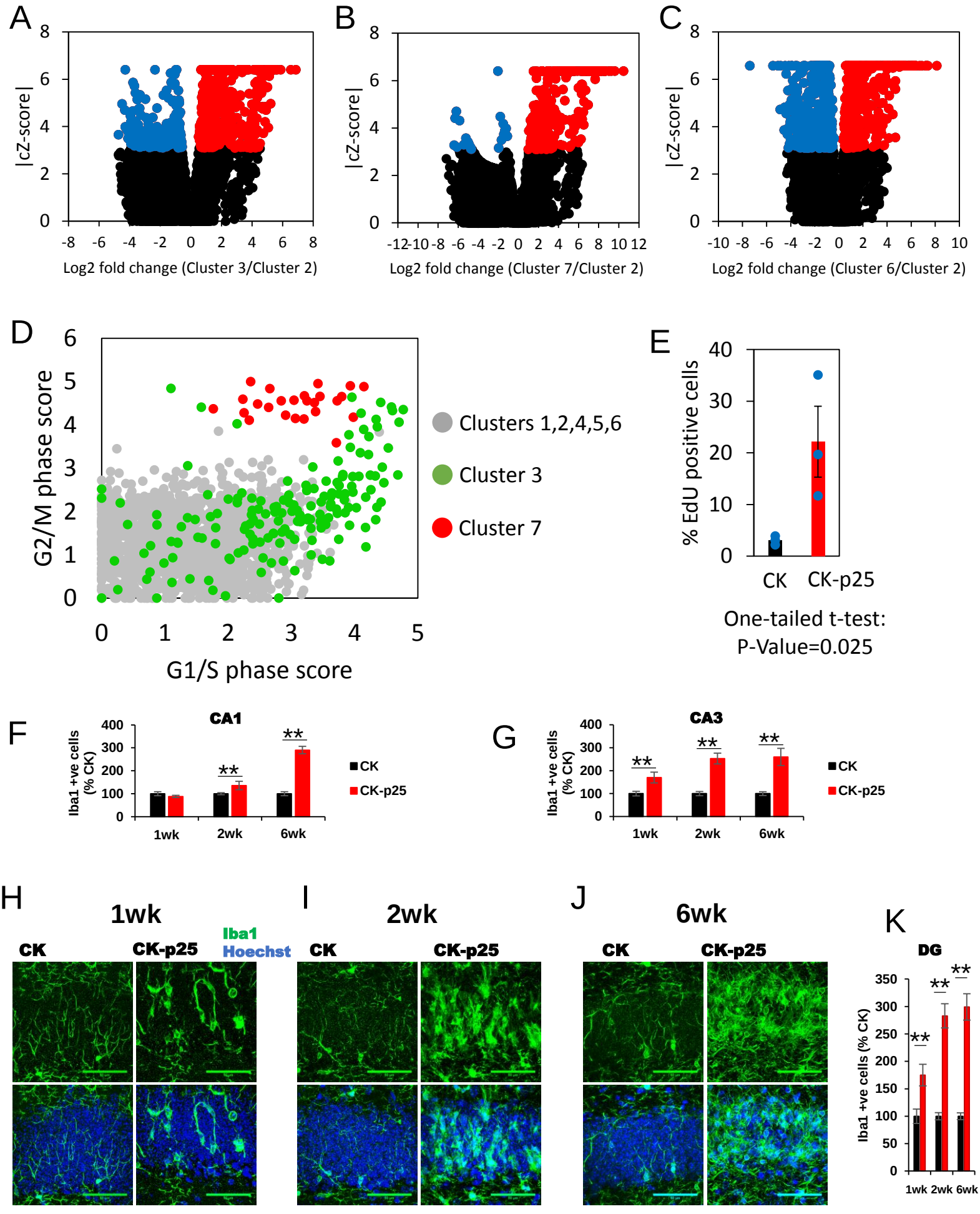
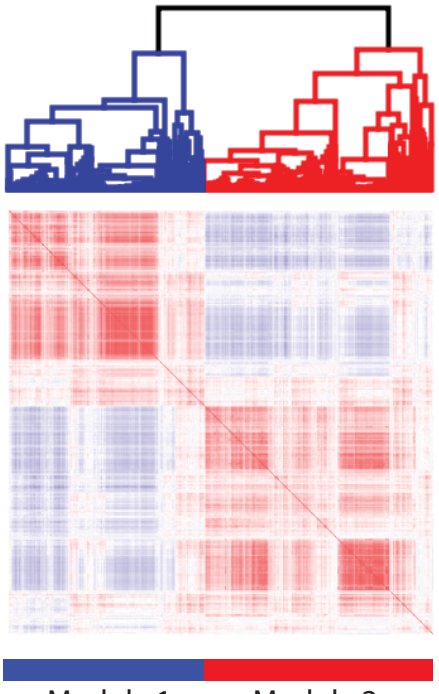

A
B
C
|cZ-score|
Log2 fold change (Cluster 3/Cluster 2)
Log2 fold change (Cluster 7/Cluster 2)
Log2 fold change (Cluster 6/Cluster 2)
D
G2/M phase score
G1/S phase score
Clusters 1,2,4,5,6
Cluster 3
Cluster 7
E
% EdU positive cells
CK
CK-p25
One-tailed t-test:
P-Value=0.025
F
CA1
Iba1 +ve cells (% CK)
1wk
2wk
6wk
CK
CK-p25
G
CA3
H
1wk
I
2wk
J
6wk
CK
CK-p25
Iba1
Hoechst
K
DG
Iba1 +ve cells (% CK)
1wk 2wk 6wk
CK
CK-p25

Figure S4

Figure S4. Single-cell differential expression analysis reveals hundreds of genes regulated during cellular reprogramming of microglia in response to neurodegeneration. Related to Figure 3

A) Volcano plot showing the fold change in gene expression in Cluster 3 microglia compared to Cluster 2 microglia. Genes whose expression was significantly altered (|cZ| > 3.0902) are highlighted in red (up-regulated genes) or blue (downregulated genes). cZ: expression difference Z-score corrected for multiple hypothesis testing using the Holm procedure (Kharchenko, Silberstein, & Scadden, 2014). B) Volcano plot showing the fold change in gene expression in Cluster 7 microglia compared to Cluster 2 microglia. Data points are colored as described in A. C) Volcano plot showing the fold change in gene expression in cells of Cluster 6 compared to cells of Cluster 2. Data points are colored as described in A. D) Scatter plot showing the average expression of a core set of 43 G1/S phase and 55 G2/M phase genes across the microglia of Clusters 1 to 7. Data points are colored by cluster. E) Click-iT Plus EdU cell proliferation assay. Bar graph showing the percentage of EdU positive microglia as determined by flow cytometry. The individual data points (three animals per genotype) are shown in blue. Error bars show the standard error of the mean. F and G) Bar plots showing the percent of Iba1-positive cells in hippocampal CA1 (F) and CA3 (G) as compared to CK control. The quantification is based on 3 to 7 CK animals and 3 to 4 CK-p25 animals, with 2 sections per animal. H, I, and J) Immunohistochemistry with anti-Iba1 (green) antibody in the dentate gyrus of 1 week- (1wk), 2 week- (2wk), and 6 week-induced (6wk) CK and CK-p25 animals. K) Quantification of immunostaining shown in H-J: Number of Iba1-positive cells indicated as percent compared to CK control. The quantification is based on 3 to 7 CK animals and 3 to 4 CK-p25 animals, with 2 sections per animal.

A Cluster 3 (early response)



B Cluster 3 (early response)

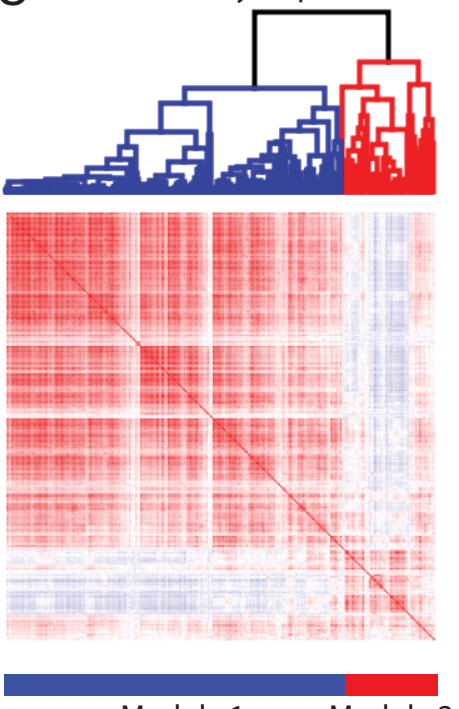
Enriched Gene Ontology terms (Biological Processes) associated with genes of module 1

Term	Bonferroni-corrected P-Value
translation	2.75E-48
immune response	6.14E-06
protein folding	0.000882
generation of precursor metabolites and energy	0.000908
cellular carbohydrate catabolic process	0.005164

Enriched Gene Ontology terms (Biological Processes) associated with genes of module 2

Term	Bonferroni-corrected P-Value
cell cycle	2.43E-48
cell cycle process	4.01E-34
cell cycle phase	6.62E-33
DNA metabolic process	2E-32
cell division	8.84E-32

C Cluster 7 (early response)



D Cluster 7 (early response)

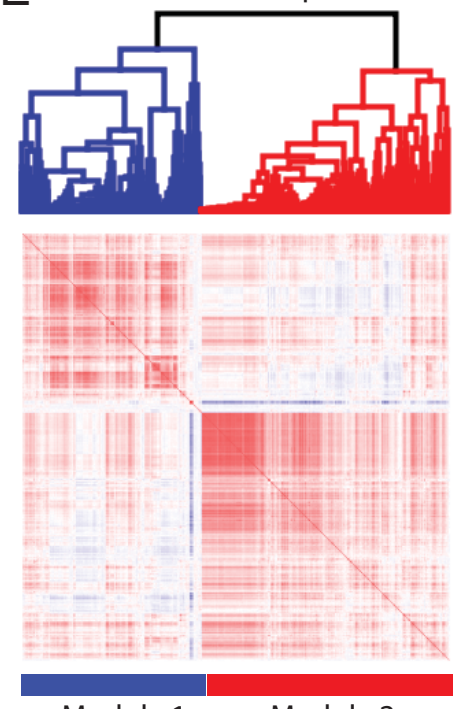
Enriched Gene Ontology terms (Biological Processes) associated with genes of module 1

Term	Bonferroni-corrected P-Value
cell cycle	8.56E-65
M phase	1.18E-64
cell cycle process	9.44E-64
cell cycle phase	9.64E-62
M phase of mitotic cell cycle	3.61E-57

Enriched Gene Ontology terms (Biological Processes) associated with genes of module 2

Term	Bonferroni-corrected P-Value
inflammatory response	0.003856
immune response	0.008748
glucose metabolic process	0.045337

E Cluster 6 (late response)



F Cluster 6 (late response)

Enriched Gene Ontology terms (Biological Processes) associated with genes of module 1

Term	Bonferroni-corrected P-Value
immune response	3.19E-42
defense response	2.23E-17
antigen processing and presentation	8.47E-11
innate immune response	1.23E-10
response to virus	4.61E-09

Enriched Gene Ontology terms (Biological Processes) associated with genes of module 2

Term	Bonferroni-corrected P-Value
translation	1.54E-62
monosaccharide metabolic process	0.016953
hexose metabolic process	0.018147
antigen processing and presentation of exogenous antigen	0.031755
glycolysis	0.049229

Figure S5

Figure S5. Functionally distinct modules of co-regulated genes in microglia cells. Related to Figure 3

A) Heat map showing the correlation matrix after hierarchical clustering between genes up-regulated in Cluster 3 (Distance metric: one minus pearson correlation; linkage method: complete linkage). We first calculated all pairwise correlations between significantly up-regulated genes across all single cells. Then, to enhance the resulting correlation matrix, we calculated the correlation of the initially determined correlation values between each pair of up-regulated genes across all up-regulated genes. B) Table showing the top 5 GO terms (Biological Processes) associated with the gene modules up-regulated in Cluster 3. C) Heat map showing the pairwise correlation between genes up-regulated in Cluster 7. D) Table showing the top 5 GO terms (Biological Processes) gene modules up-regulated in Cluster 7. E) Heat map showing the pairwise correlation between genes up-regulated in Cluster 6. F) Table showing the top 5 GO terms (Biological Processes) gene modules up-regulated in Cluster 6.

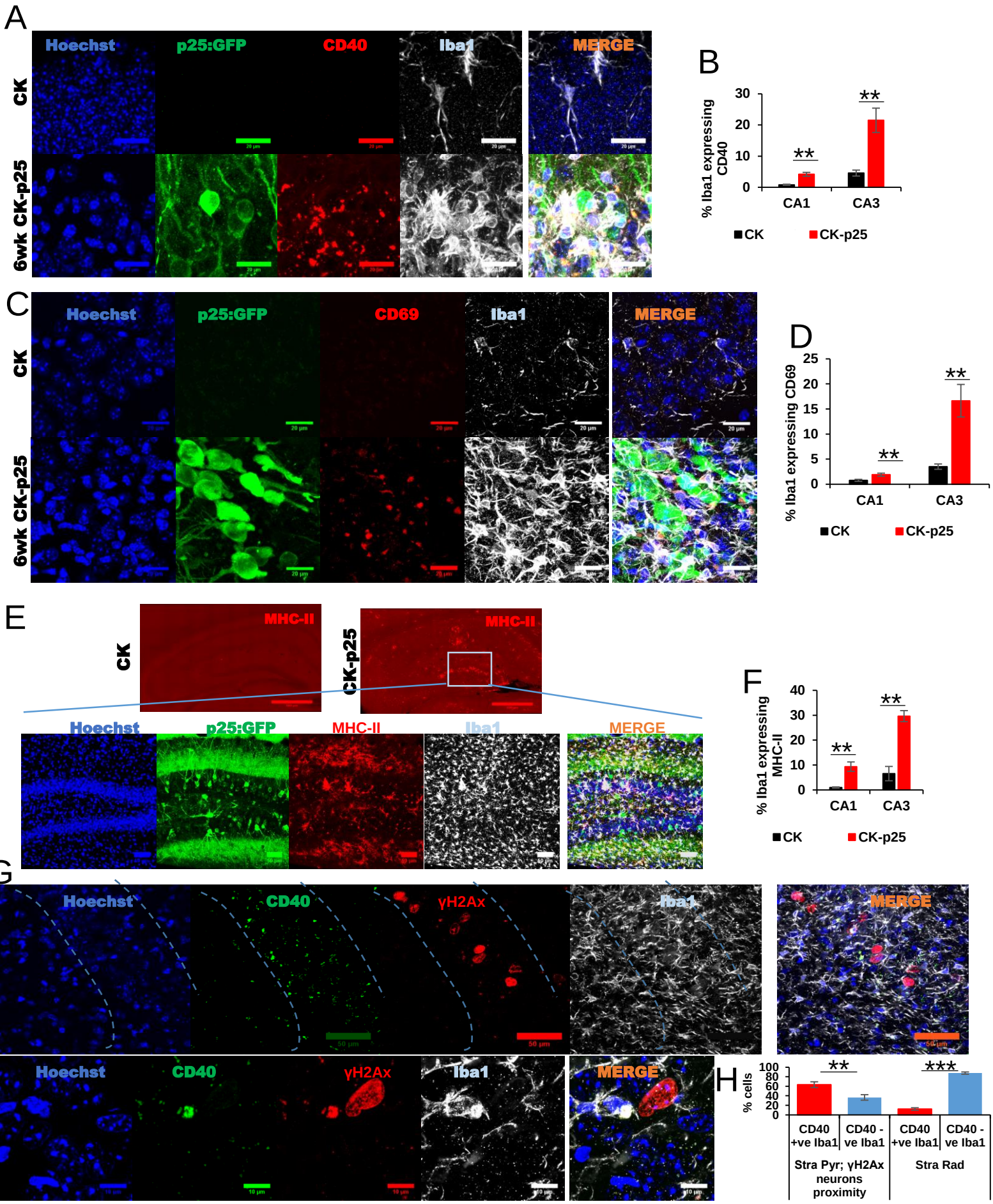

A
Hoechst
p25:GFP
CD40
Iba1
MERGE
CK
6wk CK-p25
B
% Iba1 expressing CD40
30
20
10
0
**
**
CA1
CA3
CK
CK-p25
C
Hoechst
p25:GFP
CD69
Iba1
MERGE
CK
6wk CK-p25
D
% Iba1 expressing CD69
25
20
15
10
5
0
**
**
CA1
CA3
CK
CK-p25
E
CK
CK-p25
MHC-II
MHC-II
Hoechst
p25:GFP
MHC-II
Iba1
MERGE
F
% Iba1 expressing MHC-II
40
30
20
10
0
**
**
CA1
CA3
CK
CK-p25
G
Hoechst
CD40
γH2Ax
Iba1
MERGE
Hoechst
CD40
γH2Ax
Iba1
MERGE
H
% cells
100
80
60
40
20
0
**
***
CD40 +ve Iba1
CD40 - ve Iba1
CD40 +ve Iba1
CD40 - ve Iba1
Stra Pyr; γH2Ax neurons proximity
Stra Rad

Figure S6

Figure S6. Immunostaining of hippocampal sections from 6-week induced CK-p25 and CK control littermates. Related to Figure 7

A) Immunohistochemistry with anti-GFP (green), anti-CD40 (red), and anti-IBA1 (white) antibodies in the dentate gyrus of CK and CK-p25 animals. B) Quantification of CD40 and Iba1 immunostaining in hippocampal CA1 and CA3 of CK and CK-p25 animals. Values are percent of Iba1-positive cells expressing CD40. Quantification is based on 7 CK animals and 4 CK-p25 animals, with 2 sections per animal. C) Immunohistochemistry with anti-GFP (green), anti-CD69 (red), and anti-IBA1 (white) antibodies in the dentate gyrus of CK and CK-p25 animals. D) Quantification of CD69 and Iba1 immunostaining in hippocampal CA1 and CA3 of CK and CK-p25 animals. Values are percent of Iba1-positive cells expressing CD69. Quantification is based on 7 CK animals and 4 CK-p25 animals, with 2 sections per animal. E) Immunohistochemistry with anti-GFP (green), anti-CD74 (MHC2, red), and anti-IBA1 (white) antibodies of hippocampal sections from CK and CK-p25 animals. F) Quantification of MHC2 and Iba1 immunostaining in the CA1 and CA3 region of CK and CK-p25 animals. Values are percent of Iba1-positive cells expressing CD74. The quantification is based on 7 CK animals and 4 CK-p25 animals, with 2 sections per animal. G) Immunohistochemistry with anti-γH2Ax (red), anti-CD40 (green), and anti-IBA1 (white) antibodies in the CA3 region of the hippocampus of a CK-p25 animal 6 weeks after p25 induction. H) Quantification of the immunostaining shown in Figure S6G. The bar graph shows the percentage of CD40 positive- and negative microglia in an area of high γH2Ax-positive neurons and an area that contained almost no γH2Ax positive cells. The quantification is based on 4 CK-p25 animals, with 2 sections per animal. For all graphs, error bars show the standard error of the mean, asterisks denote statistical significance (**, P-Value < 0.01; ***, P-Value < 0.001).

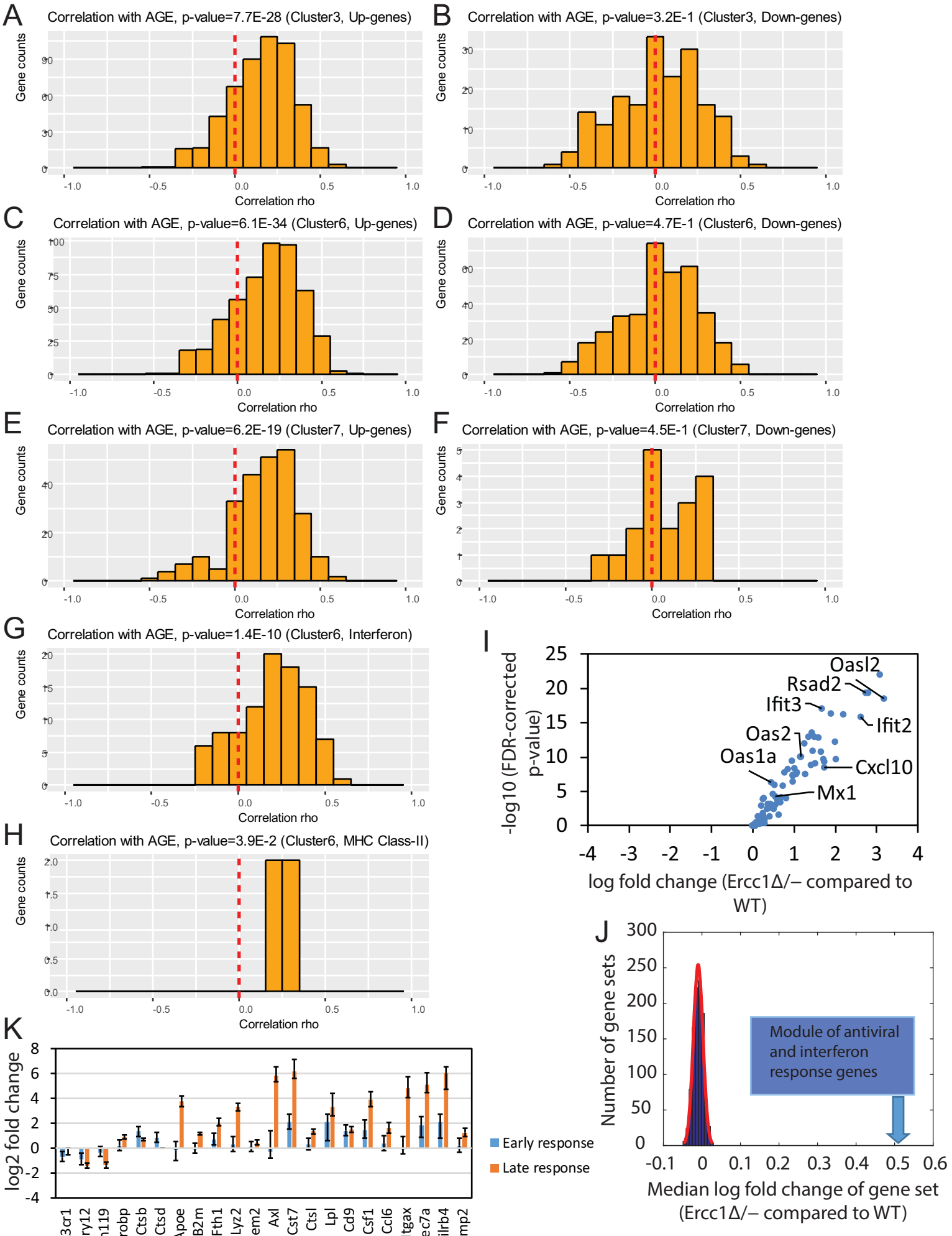

A
Correlation with AGE, p-value=7.7E-28 (Cluster3, Up-genes)
B
Correlation with AGE, p-value=3.2E-1 (Cluster3, Down-genes)
C
Correlation with AGE, p-value=6.1E-34 (Cluster6, Up-genes)
D
Correlation with AGE, p-value=4.7E-1 (Cluster6, Down-genes)
E
Correlation with AGE, p-value=6.2E-19 (Cluster7, Up-genes)
F
Correlation with AGE, p-value=4.5E-1 (Cluster7, Down-genes)
G
Correlation with AGE, p-value=1.4E-10 (Cluster6, Interferon)
H
Correlation with AGE, p-value=3.9E-2 (Cluster6, MHC Class-II)
Gene counts
Correlation rho
I
-log10 (FDR-corrected p-value)
log fold change (Ercc1Δ/– compared to WT)
Oasl2
Rsad2
Ifit3
Ifit2
Oas2
Oas1a
Cxcl10
Mx1
J
Number of gene sets
Median log fold change of gene set (Ercc1Δ/– compared to WT)
Module of antiviral and interferon response genes
K
log2 fold change
Early response
Late response
Cx3cr1
P2ry12
Tmem119
Tyrobp
Ctsb
Ctsd
Apoe
B2m
Fth1
Lyz2
Trem2
Axl
Cst7
Ctsl
Lpl
Cd9
Csf1
Ccl6
Itgax
Clec7a
Lilrb4
Timp2

Figure S7

Figure S7. Gene sets induced in the CK-p25 mouse model of neurodegeneration also tend to be upregulated in microglia of aged human individuals. Related to Figure 6

A to H) Histograms showing the distribution of the correlation between gene expression and age in human microglia for the following gene sets: A) genes up-regulated in Cluster 3 (compared to Cluster 2), B) genes down-regulated in Cluster 3 (compared to Cluster 2), C) genes up-regulated in Cluster 6 (compared to Cluster 2), D) genes down-regulated in Cluster 6 (compared to Cluster 2), E) genes up-regulated in Cluster 7 (compared to Cluster 2), F) genes down-regulated in Cluster 7 (compared to Cluster 2), G) module of 132 co-regulated antiviral and interferon response genes, H) four genes encoding components of the MHC class II, I) Analysis of the expression of the module of co-regulated antiviral and interferon response genes (identified in this study) in microglia of the Ercc1Δ/- DNA-repair deficient mouse model compared to wild type control animals (analysis of data previously published by Holtman et al. Acta Neuropathol Commun. 2015 May 23;3:31). 84 genes of the module of 132 antiviral and interferon response genes were detected in the list of genes reported by Holtman et al. The Volcano plot shows the fold change of the expression of these antiviral and interferon response genes in microglia isolated from the DNA-repair deficient model compared to WT animals. Data show the log fold change and the minus log10 of the FDR-adjusted P-Value as reported by Holtman et al. The median log fold change of the 84 genes of the module of antiviral and interferon response genes was 0.54. The data reported by Holtman et al. are based on the primary study by Raj et al. Neurobiol Aging. 2014 Sep;35(9):2147-60. J) Among the genes reported by Holtman et al we selected 1000 random sets of 84 genes and computed the median log fold change (Ercc1Δ/− compared to WT) of each set. The histogram shows the distribution of the median log fold change of 1000 random sets of 84 genes (number of gene sets per bin) and the normal density function fit to the data (red). The mean and standard deviation of the distribution were determined and the normcdf function in Matlab (Normal cumulative distribution function) was used to determine the probability p that a single observation from a normal distribution with parameters μ and σ will fall in the interval (-∞ 0.54]. The resulting probability was p=1. K) Bar graph showing the fold change in gene expression in early and late response microglia. Key genes characterizing stage 1 disease-associated microglia (DAM) (Cx3cr1, P2ry12, Tmem119, Tyrobp, Ctsb, Ctsd, Apoe, B2m, Fth1, Lyz2) and stage 2 DAM (Trem2, Axl, Cst7, Ctsl, Lpl, Cd9, Csf1, Ccl6, Itgax, Clec7a, Lilrb4, Timp2) as defined by Keren-Shaul et al. (Keren-Shaul et al., 2017) are shown. Definition of early response microglia: cells of Cluster 3 that were isolated from CKp25 animals 1 week after p25 induction. Definition of late response microglia: cells of Cluster 6.

Supplemental experimental procedures

Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Animals

All animal work was approved by the Committee for Animal Care of the Division of Comparative Medicine at the Massachusetts Institute of Technology. Adult (3 months old) male double transgenic CK-p25 mice (Jonathan C. Cruz, Tseng, Goldman, Shih, & Tsai, 2003)(J. C. Cruz et al., 2006)(Fischer, Sananbenesi, Pang, Lu, & Tsai, 2005) and their CK control littermates were used for the experiments. Mice were housed in groups of three to five on a standard 12 h light/12 h dark cycle, and all experiments were performed during the light cycle. Food and water were provided *ad libitum*. No animals were excluded from the analysis.

Isolation of microglia from the hippocampus

The hippocampus was rapidly dissected and placed in ice cold Hanks' balanced salt solution (HBSS) (Gibco by Life Technologies, catalogue number 14175-095). The tissue was then enzymatically digested using the Neural Tissue Dissociation Kit (P) (Miltenyi Biotec, catalogue number 130-092-628) according to the manufacturer's protocol, with minor modifications. Specifically, the tissue was enzymatically digested at 37 °C for 15 min instead of 35 min and the resulting cell suspension was passed through a 40 µm cell strainer (Falcon Cell Strainers, Sterile, Corning, product 352340) instead of a MACS SmartStrainer, 70 µm. The resulting cell suspension was then stained using allophycocyanin (APC)-conjugated CD11b mouse clone M1/70.15.11.5 (Miltenyi Biotec, 130-098-088) and phycoerythrin (PE)-conjugated CD45 antibody (BD Pharmingen, 553081) according to the manufacturer's (Miltenyi Biotec) recommendations. FACS was then used to purify CD11b and CD45 positive microglial cells. Standard, strict side scatter width versus area and forward scatter width versus area criteria were used to discriminate doublets and gate only singlets. Viable cells were identified by staining with propidium iodide (PI) and gating only PI-negative cells. Single cells were sorted into 96-well plates (Eppendorf twin.tec PCR Plate 96, catalogue number 951020401) containing 5 µl Buffer TCL (QIAGEN, catalogue number 1031576) per well containing 1% β-mercaptoethanol (Sigma-Aldrich, catalogue number M6250), snap frozen on dry ice, and then stored at −80 °C before whole transcriptome amplification, library preparation and sequencing.

Reverse transcription, Whole-transcriptome amplification, library construction, sequencing, and processing

Single-cell RNA sequencing libraries were generated based on the SMART-Seq2 protocol (Picelli et al., 2014) with the following modifications. RNA from single cells was first purified with Agencourt RNAClean XP beads (Beckman Coulter, catalogue number A63987) before oligo-dT primed reverse transcription with Maxima H Minus Reverse Transcriptase (Thermo Fisher, catalogue number EP0753), which was followed by 21 cycle PCR amplification using KAPA HiFi HotStart ReadyMix (KAPA Biosystems, catalogue number KK2602) with subsequent Agencourt AMPure XP bead (Beckman Coulter, catalogue number A63881) purification. Libraries were tagmented using the Nextera XT DNA Library Preparation Kit (Illumina, catalogue number FC-131-1096) and the Nextera XT Index Kit v2 Sets A,B,C, and D according to the manufacturer's instructions with minor modifications. Specifically, reactions were run at one fourth the recommended volume, the tagmentation step was extended to 10 minutes, and the extension time during the PCR step was increased from 30 seconds to 60 seconds. Libraries from 384 cells with unique barcodes were combined and sequenced on the Illumina HiSeq 2000 platform at the MIT BioMicro Center.

The raw fastq data of 40-bp paired-end sequencing reads were aligned to the mouse mm9 reference genome using TopHat2.0. The mapped reads were processed by Cufflinks 2.2 with UCSC mm9 reference gene annotation to estimate transcript abundances. A median of 251353 mapped sequencing reads was obtained for 2183 single-cell libraries. Relative abundance of transcripts was measured by fragments per kilobase of exon per million fragments mapped (FPKM). Alternatively, expression values were calculated using RSEM.

Sample filtering

For each library, we computed the following quality scores/parameters: (1) number of read pairs, (2) total number of reads, (3) number of mapped reads, (4) mapping rate, (5) percentage of transcripts identified (compared with the

overall number of transcripts identified by at least one cell), (6) mapped locations, (7) number of coding reads, (8) number of CpG island reads, (9) number of exonic reads, (10) number of intergenic reads, (11) number of intronic reads, (12) number of miRNA reads, (13) number of ncRNA reads, (14) number of reads mapping to promoter regions, (15) number of protein coding reads, (16) number of reads mapping to pseudogenes, (17) number of rRNA reads, (18) number of snoRNA reads, (19) number of snRNA reads, (20) number of reads mapping to transcription termination sites, (21) number of 3' UTR reads, (22) number of 5' UTR reads, (23) mean insert size, (24) standard deviation of the insert size.

Following an approach described by Gaublomme et al. (Gaublomme et al., 2015), we excluded libraries with poor values in either the number of mapped reads, the percentage of mapped reads, or the percentage of identified transcripts. To this end for a given performance measure x, we set a minimum cutoff value cx by taking the maximum over: {AVG(x) – 1.645 * STD(x), MED(x) – 1.645 * MAD(x)} (AVG, MED, MAD stand for average, standard deviation, median, median absolute deviation, respectively). In addition, we introduced hard lower bounds (hlb) for the cutoff values (number of mapped reads > 25000; percentage of mapped reads > 20%; percentage of identified transcripts > 5%). We then re-set the cutoff to be max{cx,hlb}. We only retained cells that scored above the cutoff (number of mapped reads > 25000 [hlb]; percentage of mapped reads > 20% [hlb]; percentage of identified transcripts > 6.873% [cx = Median-1.645*MAD = 6.873%]) in all three cases. Using this approach, we excluded 498 libraries retaining 1685 libraries.

Normalization technique to reduce the effects of confounders

To reduce the effects of confounders we followed an approach described by Gaublomme et al (Gaublomme et al., 2015). We performed PCA over the quality score matrix (a matrix with columns corresponding to libraries/cells and rows corresponding to all 24 quality scores). We then used a global-scaling normalization approach to remove the effects of the top PCs, until we covered >90% of the variance in the quality matrix (PCs 1 to 5). For a given PC, the cells are divided into 10 equally sized bins based on their projected values. The normalized expression measures are defined as:

$$E'(i,j) = E(i,j) - \text{Median}(\{E(i,j'), \text{s.t. } j' \in k(j)\}) + \text{Median}(\{E(i,:)\})$$

Where E(i,j) is the original expression value of gene i in cell j; k(j) denotes the PC-value bin to which cell j belongs; and E(i,:) denotes the values for gene i across all cells.

The results of the t-SNE analysis after applying this normalization technique are shown in Figures S3A and S3B.

Computation of module scores (weighted mean of fold induction) taking into account false negatives

In order to quantitate the activation of sets of co-regulated genes, we computed a module score for these sets of genes following an approach described by (Shalek et al., 2014). Single-cell expression values were transformed to "fold induction values" by dividing by the weighted average expression of the gene in microglia cells isolated from the CK animals. Then we applied a probabilistic weighting strategy to down weight the influence of false negative expression values. Specifically, genes were placed into 25 bins based on their mean expression level. For each cell, we calculated, for each of the 25 bins, the fraction of genes within the bin that was detected in that particular cell. Finally, we fitted a sigmoid function to the estimated values. Then, for each gene G and cell C we computed the following weight:

$$w(C,G) = \frac{1}{1 + e^{-(B_{C,0} + B_{C,1} * \mu G)}}$$

where $B_{C,0}$ and $B_{C,1}$ are the fitted parameters for the sigmoid function and μG is the mean expression of gene G in cell C. The module score is then computed as the weighted mean of "fold induction values" over all module genes within an individual cell. Module scores were calculated separately for cells isolated from CK animals, CKp25 animals 1 week after p25 induction, CKp25 animals 2 weeks after p25 induction, and CKp25 animals 6 weeks after p25 induction.

Principal component analysis and t-distributed stochastic neighbor embedding (t-SNE)

Principal component analysis and t-distributed stochastic neighbor embedding (t-SNE) were performed using Seurat, an R package designed for the analysis of single-cell RNA-sequencing data (Satija, Farrell, Gennert, Schier, &

Regev, 2015). Clusters were determined by density clustering (DBSCAN) using the DBClust_dimension function of Seurat (reduction.use = "tsne", G.use = 6). The t-SNE perplexity was set to 30. Marker genes of the Clusters 4 and 5 were identified using the find.markers function of Seurat.

We found cells from only one of three biological replicates in Cluster 8 (Figure 2A). This indicates that Cluster 8 likely resulted from technical confounders and was therefore not compared to other clusters. Because the expression profile of other cells from the affected biological replicates could have been altered by the technical confounders, we repeated the single-cell differential gene expression analysis excluding all cells from the two affected animals. When we repeated the differential gene expression analysis excluding all of the cells from the affected biological replicates, we found that the GO terms enriched among the genes up-regulated in the early response and late response microglia were very similar as in the original analysis. The top GO terms enriched among genes upregulated in the early response Cluster 3 were cell cycle, DNA replication, and DNA replication initiation, all of which were also found in the top 10 GO terms in the original analysis. Similarly, the top GO terms enriched among genes upregulated in the late response Cluster 6 were translation, immune system process, and response to virus, all of which were also found in the top 10 GO terms in the original analysis. The top GO terms enriched among genes upregulated in the early response Cluster 7 were cell cycle, cell division, and mitotic nuclear division. In the original analysis, although these specific terms were not in the top 10, we found other GO terms associated with mitosis. All the genes highlighted in the manuscript were also found to be significantly differentially expressed when all of the cells from the affected biological replicates are excluded from the analysis.

Single-cell differential gene expression analysis

The SCDE (single-cell differential expression) software package (Kharchenko, Silberstein, & Scadden, 2014) was used to determine expression differences between groups of single cells. Read counts observed for each gene determined using HTSeq were used as input.

Click-iT Plus EdU cell proliferation assay

Three CK-p25 and three CK control animals were intraperitoneally injected with 50 mg/kg 5-ethynyl-2'-deoxyuridine (EdU) every second day during the first two weeks after p25 induction. Two weeks after p25 induction the animals were transcardially perfused with ice cold PBS and microglia were isolated using the Neural Tissue Dissociation Kit (P) as described above. The resulting cell suspension was then stained using allophycocyanin (APC)-conjugated CD11b mouse clone M1/70.15.11.5 (Miltenyi Biotec, 130-098-088) and phycoerythrin (PE)-conjugated CD45 antibody (BD Pharmingen, 553081) according to the manufacturer's (Miltenyi Biotec) recommendations. EdU incorporation was then detected using the Click-iT Plus EdU Pacific Blue Flow Cytometry Assay Kit (ThermoFisher Scientfic, catalogue number: C10636) according to the manufacturer's instructions. Finally, the cells were analyzed by flow cytometry.

Cell cycle analysis

We determined the average expression of a core set of 43 G1/S phase and 55 G2/M phase genes for each cell following an approach described by Tirosh et al. (Tirosh et al., 2016). The G1/S and G2/M scores were calculated as the natural logarithm of the average expression level (expression measure FPKM) of the set of genes plus 1 (to account for cells in which only very low levels of cell cycle genes were detected).

Single-cell qRT-PCR

Microglia were isolated from the hippocampus of a wild type mouse and a mouse of the genotype $CX3CR1^{CreERT2\text{-}IRES\text{-}EYFP/CreERT2\text{-}IRES\text{-}EYFP}$; $R26^{TdTomato/+}$ as described above. The mouse of the genotype $CX3CR1^{CreERT2\text{-}IRES\text{-}EYFP/CreERT2\text{-}IRES\text{-}EYFP}$; $R26^{TdTomato/+}$ was intraperitoneally injected with 50 mg/kg 4-Hydroxytamoxifen once a day for five consecutive days starting 8 days before the animal was sacrificed and the microglia cells were isolated. The resulting cell suspensions was stained using allophycocyanin (APC)-conjugated CD11b mouse clone M1/70.15.11.5 (Miltenyi Biotec, 130-098-088). and CD11b-positive cells were sorted into 96-well plates containing 5 µl of a 1:500 dilution of 5X Phusion HF Buffer (NEW ENGLAND BioLabs, catalogue number B0518S). To verify that our FACS sorting protocol resulted in individual cells in each well, an aliquot of the cell suspensions from the two genetically distinct animals was mixed 1:1 before FACS sorting. After sorting the cells, Proteinase K (Thermo Fisher Scientific, catalogue number AM2548) was added to a final concentration of 0.16 mg/ml and the 96-well plate was incubated for 15 minutes at 50 °C. To dry the samples, the seal was removed and the 96-well plate was

incubated for 10 minutes at 95 °C. Then the RNA was reverse transcribed in a volume of 5 µl using the iScript Reverse Transcription Supermix for RT-qPCR (BIORAD, catalogue number 170-8891) (5 minutes at 25 °C, 30 minutes at 42 °C, 5 minutes at 85 °C). The cDNA was diluted 1:7 and 9 µl were used for RT-qPCR amplification in a 20 µL reaction using TaqMan Fast Advanced Master Mix (ThermoFisher Scientific) and TaqMan Gene Expression Assays targeting Actb (ThermoFisher Scientific, catalogue number Mm00607939_s1), Cx3cr1 (ThermoFisher Scientific, catalogue number Mm02620111_s1), or TdTomato (ThermoFisher Scientific, Custom TaqMan Gene Expression Assay). The PCR cycling conditions used were: 2 minutes at 50 °C, 20 seconds at 95 °C, and then 45 cycles of 3 seconds at 95 °C and 30 seconds at 60 °C.

Additional computational analysis and bioinformatics

Enrichment analysis for Gene Ontology (GO) terms was performed in DAVID (Huang, Sherman, & Lempicki, 2008). In DAVID, gene lists were uploaded, and the GOTERM_BP_DIRECT annotation category was selected. Hierarchical clustering analyses and heat map visualizations were performed using the GENE-E (Broad Institute, https://software.broadinstitute.org/GENE-E/) matrix analysis platform. Violin plots were created in R using the ggplot2 package (Wickham, 2009). Singe-cell consensus clustering was performed in R using the SC3 package (Kiselev et al., 2016). The optimal clustering solution (number of clusters k) was determined based on the average silhouette width.

Transcription factor (TF) motif analysis was done using the HOMER software tool. Mouse TF motif database was collected from HOCOMOCO, and the enriched motifs at core promoters (±500bp from TSS) of selected gene sets were ranked by -log10(P-value).

Sequence fragments of P25 transcripts were detected from tissue and microglia RNA-Seq data using BWA aligner. Read counts were normalized as CPM (counts per million reads) for cross-sample comparison.

To identify modules of co-regulated genes, we first calculated all the pairwise correlations between significantly up-regulated genes across all single cells. Then, to enhance the resulting correlation matrix, we calculated the correlation of the initially determined correlation values between each pair of up-regulated genes across all up-regulated genes. Hierarchical clustering of the resulting correlation matrix revealed at least two major modules of co-regulated genes (Figures S5A, S5C, and S5E).

Human aging microglia RNA-Seq FASTQ files were download from GSE99074 and processed using STAR/CUFFLINKS pipeline for gene quantification. As noted, 39 normal human samples in this study were processed using either Illumina TruSeq (23 samples, each with 60ng of total RNA) or Clontech SMARTer (16 samples, each with 500pg of total RNA) kit for RNA-Seq library preparation. Our PCA analysis of gene quantification tables (FPKM values) showed a cluster of 22 samples with high unique mapping rates (average 83.7%) compared to the other samples (average 54.2%). Thus the gene quantification result from these 22 human microglia samples were selected as high quality RNA-Seq result for downstream analysis. Gene expression FPKM values were correlated with human age information using spearman's correlation model. Human orthologs of mouse gene sets were converted using the Human and Mouse Homology Classes information from JAX Mouse Genome Informatics database. Distribution of correlation rho values of a particular gene set was tested against distribution of rho for all expressed genes to calculate statistical p-values (Wilcoxon Signed-Rank test).

Immunohistochemistry

Immunohistochemistry experiments were performed as described previously (Iaccarino et al., 2016). Briefly, mice were perfused with 4% paraformaldehyde under deep anaesthesia, and the brains were post-fixed overnight in 4% paraformaldehyde. Brains were sectioned at 40 µm using a vibratome (Leica). Sections were permeabilized and blocked in PBS containing 0.3% Triton X-100 and 5% normal donkey serum at room temperature for 2 h. Sections were incubated overnight at 4 °C in primary antibody in PBS with 0.3% Triton X-100 and 5% normal donkey serum. Primary antibodies were anti-CD40 Monoclonal Antibody (3/23) (1:200; ThermoFisher Scientific; MA5-17852), anti-CD69 (1:200; R&D Systems; MAB23861), anti-Iba1 (1:500; Novus Biologicals; NB100-1028), anti-MHC class II (1:200; I-A/I-E), clone M5/114 (EMD Millipore; MABF33), anti-γH2Ax (1:500; Millipore, #05-636). Primary antibodies were visualized with Alexa-Fluor 488, Alexa-Fluor 594, and Alex-Fluor 647 secondary antibodies (Molecular Probes), and cell nuclei visualized with Hoechst 33342 (Sigma-Aldrich; 94403). Images were acquired using a confocal microscope (LSM 710 or LSM 880; Zeiss) with a 40x or 63X objective at identical

settings for all conditions. Images were quantified using ImageJ 1.42q. Three to seven CK-control and 3-4 CK-p25 animals were used for staining and two coronal sections from each animal were used for quantification.

The number of cells positive for Iba1 was manually quantified. MHC class II immunostaining was very clear and we used Image J to identify the number of Iba1 cells expressing MHC-II. Punctate like signals were observed from CD69 and CD40 immunohistochemical staining. Therefore, signal threshold was applied to extract CD69 and CD40 immunostaining signals and the size of particles that are more than 5 μM^2 was quantified.