PNAS
www.pnas.org

Supplementary Information for

Prior activation state shapes the microglia response to anti-human TREM2 in a mouse model of AD

Daniel C. Ellwanger*, Shoutang Wang*, Simone Brioschi, Zhifei Shao, Lydia Green, Ryan Case, Daniel Yoo, Dawn Weishuhn, Palaniswami Rathanaswami, Jodi Bradley, Sara Rao, Diana Cha, Peng Luan, Shilpa Sambashivan, Susan Gilfillan, Samuel A. Hasson, Ian N. Foltz, Menno van Lookeren Campagne, and Marco Colonna

* D.C.E. and S.W. contributed equally to this work

Menno van Lookeren Campagne
Email: mvanlook@amgen.com

Marco Colonna
Email: mcolonna@wustl.edu

This PDF file includes:

SI Methods
Figures S1 to S5
Table S1
Legends for Dataset S1 and S2
SI References

Other supplementary materials for this manuscript include the following:

Dataset S1 and S2

SI Methods

Animals

Mice were housed in the animal facilities of Washington University in St. Louis. All animal experiments were conducted in compliance with Institutional regulations, under authorized protocols # 20160220 and 19-0981 approved by the Institutional Animal Care and Use Committee of Washington University and Amgen South San Francisco. Only male mice were used for pharmacodynamic analysis, pharmacokinetic analysis and scRNA-seq analysis were performed on both male and female mice in this study.

Anti-hTREM2 antibodies

hTREM2-specific serum titers obtained from immunized mice were monitored by live-cell FACS analysis (Accuri FACS). Lymphocytes from draining lymph nodes of animals with the highest antigen-specific serum native titers directed against hTREM2 were used for hybridoma generation.

Hybridoma supernatants were screened for binding to human TREM2 by ELISA using 384-well plates coated with Neutravidin overnight or coated with control hIgG1 at 2 μg/mL at 37 °C for 1 hour followed by coating with biotinylated-hTREM2 extracellular domain fused to the Fc portion of hIgG1 Fc (hTREM2-Fc, Amgen). After a wash step, the exhausted hybridoma supernatants were diluted with 1% milk/1X PBS (1:5) and added to hTREM2-Fc or control hIgG1 coated 384 well plates and incubated at room temperature for 1 hour. A mixture of goat α-human κ-HRP (2060-05, Southern Biotech) and goat α-human λ-HRP (2070-05, Southern Biotech) were used for detection.

The variable heavy and light chain sequences from a lead candidate identified in the hybridoma screening campaign were cloned and recombinantly expressed with an hIgG1 constant region lacking effector function to generate hT2AB. A murinized version of hT2AB, mT2AB, was generated by grafting the hT2AB variable domains on an effectorless mIgG1 backbone. Preparations of hT2AB, mT2AB, and the effectorless isotype control hIgG1 and control mIgG1 used in animal and cell-based experiments were tested for endotoxin and found to be comparable with < 0.5 EU/mg assuring that responses were not due to TLR signaling.

Antibody binding assay

The purified hT2AB was diluted to 5 μg/ml in assay buffer (10 mM Tris, 0.13 % Triton X-100, 150 mM NaCl, 1 mM CaCl2, 0.1 mg/ml BSA, pH 7.6) and captured on anti-hFc kinetic sensors (18-5090,

ForteBio). Recombinant hTREM1:GSS:Flag:6xHis and recombinant hTREM2:GSS:Flag:6xHis were minimally biotinylated (0.3-0.4 biotin/mol) and immobilized at 70-80 nM onto high precision Streptavidin fiber optic biosensors (SAX, #18-5119) over 2000 seconds to a final loading level of ~2 nm. Each loaded TREM2 protein was then incubated with a dilution series of hT2AB Fab protein (30, 10, 3.3 nM) for 300 seconds and then 500 seconds in buffer alone (dissociation). Raw data was processed with the Octet data analysis software (v10) and processed data were globally fit to a 1:1 binding model and a dissociation constant (K_D) of 50 nM was calculated. An anti-hTREM1 antibody and an isotype-matched hIgG2 control antibody were used as controls.

To test the binding capacity to the hTREM2^{R47H} mutation, bone marrow-derived macrophages (BMMs) from *TREM2*^{CV}, *TREM2*^{R47H} and *Trem2*$^{-/-}$ mice were harvested on day 5 and incubated in FACS buffer (10% FCS in PBS) with hT2AB or control hIgG1 for 30 min, followed by staining with anti-hIgG Fc-PE (9040-09, SouthernBiotech). Dead cells were excluded by DAPI staining.

Generation of differentiated human monocyte-derived macrophages

Large, single donor lots (50-100 million cells per differentiation run) of cryopreserved CD14+ monocytes from healthy human de-identified donors were collected through leukapheresis and negative immunomagnetic selection (Lonza) and used to generate macrophages (hMacs). Suspensions of cryo-recovered monocytes were differentiated in RPMI-1640 medium using plant-derived recombinant M-CSF (50 ng/mL, plant-derived, ultra-low endotoxin 0.05 EU/µg, PromoCell # C-60442A) in a semi-adherent manner with CellGenix VueLife 118-C bio-process bags (Saint-Gobain Performance Plastics). A maximum of 50 million cells were loaded in differentiation medium into each bag (~30 mL of cell suspension initial loading). Differentiation medium was composed of RPMI-1640 + 10% FBS (Gibco PerformancePlus Certified, heat inactivated, ≤5 EU/mL endotoxin, #10082139), 1X GlutaMAX (Gibco #35050061), 1X Pen/Strep (Gibco # 15140122), 1X NEAA (Gibco # 11140050), and 1X Sodium Pyruvate (Gibco # 11360070) in addition to the 50 ng/mL M-CSF. With bags placed on racks in standard tissue culture incubators (humidified, 5% CO2, 37 degrees C) to maximize gas exchange, differentiation was conducted for 9 days total with infusions of fresh differentiation medium on day 3 and day 6. After 9 days of differentiation, macrophages were collected from bio-process bags after agitation to dislodge cells and cryopreserved in BamBanker Serum Free Medium (Wako Chemicals USA # 30214681). Each large-scale production run of human macrophages was qualified for consistent TREM2 expression relative to undifferentiated monocytes by flow cytometry. Throughout the production process, every effort was made to monitor and minimize endotoxin exposure.

Syk phosphorylation assay

Clone G13 cells (8 x 10^5 cell/mL) or hMacs (2 x 10^5 cell/mL) were seeded overnight at 25 µL per well in a 384-well PDL-coated assay plate. On the day of the assay, cells were incubated with a serial dilution of hT2AB for 45-60 min at RT. After removal of the supernatant, the cells were lysed with M-Per+ containing 0.0625 nM anti-pSyk (Tyr525/526, 2710, Cell Signaling Technology) and 0.5 nM biotinylated mouse anti-Syk antibodies, Custom order: BD Bioscienes) at room temperature for 1 hour incubated with 2.5 µg/mL anti-rabbit IgG-Acceptor beads (AL104C, Perkin Elmer) at room temperature for 2 hours followed by addition of 10 µg/mL streptavidin-Donor beads (6760002B, Perkin Elmer) at room temperature for 2 hours. The antibody and bead concentrations stated are final. The AlphaLISA signals (counts) were measured by an EnVision Multilabel Reader.

Measurement of sTREM2 and CCL4 levels in hMacs by MSD

hMacs were used for measuring the CCL4 and sTREM2 in conditioned media after treatment with hT2AB or hIgG1 isotype control antibody. Briefly, hMacs (500000 cells/well/ml) were plated in 6-well plates and incubated overnight at 37 °C. Growth media was replaced by culture media (RPMI + GlutaMax + 1% FBS) for 24 hours and the following day an appropriate amount of media was removed and replaced with media that contains hT2AB, hIgG1 isotype control antibody or acetylated LDL at a final concentration of 200 nM. At specified time points (4, 8 or 24 hours) media from each treated well (conditioned media) was removed and saved for analysis until all samples were collected. CCL4 (4, 8 or 24 hours) and sTREM2 (24 hours) levels were measured in conditioned media with an MSD platform-based assay as per manufacturer instructions (Meso Scale Discovery).

Survival assay of BMMs

BMMs from *TREM2*^{CV}, *TREM2*^{R47H} and *Trem2*$^{-/-}$ were harvested at day 5 of culture with CSF1 and transferred to 24-well flat-bottomed plates that were coated with hT2AB or control hIgG1 at 5 × 10^4 cells/well in complete RPMI without CSF1. Survival, measured as % of Propidium Iodide negative cell population, was detected after 48 hours culture by a FACSCalibur.

GFP reporter assay

2B4 NFAT:GFP reporter cells expressing hTREM2^{CV} and hTREM2^{R47H} were used to test the activation of hTREM2 variants. Briefly, stock solutions of hT2AB or control hIgG1 were diluted to the indicated concentrations in a sodium carbonate-bicarbonate buffer, and 50 µl of the resulting solution was added to distinct wells of a 96-well plate for overnight. Each condition was performed in triplicate. The

antibody coated plates were washed three times by cold PBS before transferring 2B4 NFAT:GFP reporter cells expressing hTREM2^{CV} or hTREM2^{R47H} into the corresponding wells at 1000 cells/μl. After stimulating for 16 hours, the cells were transferred to FACS tubes and read on a FACSCalibur for GFP expression.

Pharmacodynamics and Pharmacokinetics of hT2AB

Groups of 10-week old *TREM2*^{CV}, *TREM2*^{R47H} and *Trem2*$^{-/-}$ male mice were injected intravenously (i.v.) with different doses of hT2AB. After 48 hours, mice were sacrificed and brain lysates were used to measure the concentrations of CXCL10, CCL4, CCL2, CXCL2 and CST7 by MSD. In a different treatment group, *TREM2*^{R47H} and *Trem2*$^{-/-}$ male mice were injected i.v. with hT2AB at 30 mg/kg. Mice were sacrificed at 4, 8 and 24 hours after injection. The relative gene expression levels of *Cxcl10*, *Ccl2*, *Ccl4*, *Cst7* and *Tmem119* were measured by qRT-PCR.

For pharmacokinetic analysis, groups of 8-month old *TREM2*^{CV}, *TREM2*^{R47H} and *Trem2*$^{-/-}$ male or female mice were injected intraperitoneally with a single injection of 30 mg/kg hT2AB. Concentrations of hT2AB in mouse serum samples and in homogenate of cold PBS-perfused cerebellum were measured 48 hours later with two different assays. Both assays were sandwich immunoassays, using a recombinant human TREM2 (Amgen, Inc. CA) and a ruthenium conjugated mouse anti-human Fc monoclonal antibody (Amgen, Inc. CA). The Lower Limit of Quantification (LLOQ) for serum and cerebellum homogenate assays were 100 ng/mL and 1 ng/mL, respectively.

Single-cell RNA-seq Analyses

Computational resources

Resource	Source	Version
CellRanger	https://www.10xgenomics.com/	3.0.2
PANTHER	http://www.pantherdb.org/	15.0
R Statistical Software	https://www.r-project.org/	3.6.0
-- *irlba*	https://CRAN.R-project.org/package=irlba	2.3.3
-- *scran*	http://doi.org/10.18129/B9.bioc.scran	1.14.5
-- *batchelor*	http://doi.org/10.18129/B9.bioc.batchelor	1.2.4
-- *dbscan*	https://CRAN.R-project.org/package=dbscan	1.1.5
-- *SingleR*	http://doi.org/10.18129/B9.bioc.SingleR	1.0.5
-- *igraph*	https://CRAN.R-project.org/package=igraph	1.2.4.2
-- *caret*	https://CRAN.R-project.org/package=caret	6.0.86
-- *slingshot*	http://doi.org/10.18129/B9.bioc.slingshot	1.4.0
-- *tradeSeq*	http://doi.org/10.18129/B9.bioc.tradeSeq	1.0.1
Python	https://www.python.org/	3.7.4
-- *umap-learn*	https://pypi.org/project/umap-learn/	0.3.10
Java	http://java.com	1.8.0
-- *Leiden*	https://doi.org/10.5281/zenodo.1466831	1.0.0

Read alignment

We build a reference genome by extending mouse mm10 by human *TREM2* from GRCh38; transcriptome annotations from Ensembl 93 (https://www.ensembl.org/) were used. Sequenced reads from the microfluidic droplet platform were de-multiplexed and aligned using CellRanger version 3.0.2, available from 10x Genomics (www.10xgenomics.com), with default parameters.

Quality control

We performed a multi-step quality assessment. First, each sample was analyzed individually (**Fig. S1A**). Here, all 24 samples passed an initial quality control evaluating the distribution of all mapped reads over the genome and the fraction of sequenced bases with a Phred quality score (Q) > 30. Each sample encompasses thousands of captured events, k, which are either genuine cells or empty droplets with ambient RNA; each event is identified by a unique barcode b. To distinguish single cells, we used a method based on the conventional thresholding on the total UMI count, $\mathbf{u} = (u_b) \in \mathbb{N}^{\boldsymbol{k}}$.

First, each barcode b meeting $u_b > \sum_{b=1}^{k} \frac{u_b}{k}$ was rank transformed in order of decreasing number of total UMI count resulting in a vector $\mathbf{r} = (r_b) \in \mathbb{N}^{\boldsymbol{k}}$; typically, $\sum \frac{u_b}{k} \approx 72$. For barcodes with equal UMI counts, a permutation was used with increasing values at each index set of ties. We then modeled the total UMI count as a function of barcode rank, $\ln(\mathbf{u}) \sim \ln(\mathbf{r})$, by fitting cubic smooth splines with 20 degrees of freedom. Each inflection point of this function may be interpreted as a transition between a subset of barcodes with a larger number of total UMIs, i.e., potentially cell-containing droplets, and the majority of barcodes with ambient RNA. Inflection points were determined by local minima of the first differentiation of the spline basis functions. For each sample, we chose the inflection point φ closest to the expected number of recovered cells and retained all barcodes meeting $u_b > \varphi$. The choice of φ was further guided by the following descriptive metrics of the selected set of barcodes: i) percentage of all reads allotted to the selected barcodes, ii) median number of reads per barcode, iii) median fraction of reads mapped to mitochondrial genes per barcode, iv) median number of UMI per barcode, v) median number of genes with at least one UMI count per barcode, and vi) median fraction of reads originating from an already-observed UMI (saturation). The cell selection was further refined by evaluating the distributions of metrics ii-vi) to remove low quality cells and doublets. Finally, we estimated the cell cycle effect in each sample. We predicted the cell cycle phase per cell using the machine learning based approach proposed by Scialdone *et al.* (1). Briefly, a classifier was trained on pairs of genes that change expression directionality across cell cycle phases. Each cell's cell cycle state can then be projected by examining the sign of the expression difference in the new data set. Cells with a predicted G1 or G2M score above 0.5 were assigned to the G1 or G2M phases, respectively; cells were classified to be in S phase, if the predicted G1 and G2M scores were below 0.5. All calculations were performed using the *cyclone* function in the R package *scran*. Predicted cell cycle scores and phases were not used for cell filtering.

Next, we conducted an integrative quality control step to identify unwanted technical artifacts in the data. All filtered 94488 cells from all samples were pooled, and the resulting UMI count matrix was normalized and log-transformed (see *Normalization*). To reveal technical artifacts in the data, we calculated principal components that capture the maximal variance within the data while controlling for the manifest variables sex, age, and treatment. For this purpose, manifest variables were partialled-out from the normalized UMI count matrix. Informative genes were unbiasedly determined by their mean expression-dependent variance in the data (see *Gene expression variance modeling*) and used for calculating principal components of the corrected UMI count matrix (see *Spectral dimensionality reduction*). The resulting 38-dimensional latent space was further modeled with a fuzzy topological structure using Uniform Manifold Approximation and Projection (UMAP; Python package

umap-learn) (2) to unfold the data structure that is either driven by cell type or technical variance. While varying library sizes can be normalized between cells, a large fraction of missing values (i.e., drop-outs) due to poor transcriptome coverage cannot be accurately recovered in the data and will significantly impact downstream analyses. Thus, to score cell quality, we calculated the first principal component of a set of seven quality metrics assessing the transcriptome coverage per cell: i-iv) fraction of reads consumed by the top {500, 200, 100, 50} expressed genes, v) fraction of mitochondrial reads, vi) relative distance to the maximum total number of UMIs (i.e., $1 - \frac{u_b}{\max(\mathbf{u})}$), vii) relative distance to the maximum number of genes with at least one UMI count. The resulting cell quality (CQ) score was then superimposed onto the UMAP revealing a technical confounder in the data (**Fig. S1B**). Using density-based spatial clustering (R package *dbscan*), we extracted a representative group of cells with low CQ score. We determined an optimal CQ score cut-off of -0.1 which removes a maximum fraction of cells within this group (91.4%) and a minimum fraction of cells otherwise (11.1%) by determining the knee of the inverse empirical cumulative CQ score distribution function (**Fig. S1C**). This resulted in a final data set of 71303 high quality single cells (**Fig. S1D**).

Normalization

Library sizes were normalized as proposed by (3). In a nutshell, size factors are computed from pools of similar cells, which are then deconvoluted to cell-based factors and used to scale the counts in each cell. First, for each batch scaling factors for each cell were calculated using the R package *scran*. An initial guess of cell populations contained in the data was calculated by using the function *quickCluster* on a shared nearest neighbor graph; we required a minimum cluster size φ of 10% of the total number of cells, and a minimum average gene expression of 1 for the shared nearest neighbor graph construction. Size factors were then calculated using the function *computeSumFactors* with pool sizes ranging in $[21, \max\{101, \varphi + 1\}] \in \mathbb{N}$. Then, we normalized scaling factors between batches based on their ratio of average UMI counts to provide comparable results to the lowest-coverage batch using the R package *batchelor*. Finally, the scaled UMI count matrix $\mathbf{X}$ was log2 transformed by $\log_2(\mathbf{X} + 1)$.

Supervised CD45+ cell type annotation

Rather than determining cell identities by a biased selection of marker genes, we intended to unbiasedly characterize cell types using the total available transcriptome. For this purpose, we first filtered robustly expressed genes with more than 4 molecules in at least 50 cells, removed ribosomal and mitochondrial genes (obtained from Gene Ontology GO:0005840 and Ensembl, respectively), retained only protein-coding genes, and removed a set of 136 genes highly correlating with

dissociation-induced stress response (4). This resulted in a single-cell expression matrix $\mathbf{X}$ composed of n data vectors $\mathbf{x}_{.j}$ with $j \in 1, \ldots, 71303$ of dimensionality $m = 4453$: $\mathbf{X} = \left(x_{ij}\right) \in \mathbb{R}^{4453 \times 71303}$. We used log-normalized microarray gene expression data of 20 immune cell types measured in 830 pure samples of sorted cells by the Immunological Genome Project (ImmGen) (5); the processed expression matrix $\mathbf{Y} = \left(y_{ij}\right) \in \mathbb{R}^{22134 \times 830}$ was obtained from the *SingleR* R package. We filtered 3697 genes that were highly variable in this dataset by assessing the mean expression-dependent variance (see *Gene expression variance modeling*). Spearman's rank correlation coefficients, $\rho(\mathbf{x}_{.i}, \mathbf{y}_{.j})$, were calculated between all single cells, $\mathbf{x}_{.i} \in \mathbb{R}^{992}$, and all ImmGen samples, $\mathbf{y}_{.j} \in \mathbb{R}^{992}$, using 992 genes overlapping between both datasets. Each single cell was assigned the cell type with the most similar expression profile by $\arg\max_j \rho(\mathbf{x}_{.i}, \mathbf{y}_{.j})$. In this step 15 cell types were detected in the dataset. Since the original expression space of single-cell data is rife with noise and the underlying distributions of scRNA-seq UMI counts and microarray signal intensities differ, we aimed to refine the initial cell type classification. For this purpose, the manifest variables sex, age, and treatment, as well as, three technical confounders (fraction of reads mapped to mitochondrial genes, total number of reads, fraction of reads assigned to dissociation-related stress response genes) were regressed-out from $\mathbf{X}$. Then, a 39-dimensional principal component space was calculated (see *Spectral dimensionality reduction*) on highly variable genes (see *Gene expression variance modeling*) and subjected to UMAP (Python package *umap-learn*) to unfold the data structure driven by cell identity. Next, we generated a weighted cell adjacency matrix; weights were calculated by the Jaccard index between cells using the overlap of their 15-nearest neighborhood. We identified 30 cell communities (segments) in the adjacency matrix using the Leiden algorithm (6) (Java package *Leiden*) with a resolution of 3×10^{-4}. The data was then tabulated by counting for each segment i the number of cells with cell type j resulting in matrix $\mathbf{A} = (a_{ij}) \in \mathbb{N}^{30 \times 15}$. To avoid division by 0 in the subsequent calculation, a pseudocount was added by $\mathbf{A} = \mathbf{A} + 1$. An enrichment score matrix $\mathbf{A}' = \left(a'_{ij}\right) \in \mathbb{R}^{30 \times 15}$ was derived by

$$a'_{ij} = \ln\left(\frac{a_{ij}}{\sum_{j=1}^{15} a_{ij} - a_{ij}}\right) - \ln\left(\frac{\sum_{i=1}^{30} a_{ij} - a_{ij}}{\sum_{i=1}^{30}\sum_{j=1}^{15} a_{ij} - \sum_{j=1}^{15} a_{ij} - \sum_{i=1}^{30} a_{ij} + a_{ij}}\right)$$

Finally, each cell in each segment was assigned the cell type with the highest enrichment score. This revealed 64274 microglia cells in our dataset.

Differential gene expression analysis

To identify differentially expressed genes between two groups of cells, we developed an intuitive and scalable approach based on Bayesian statistics. Here, we calculated the specificity and the detection

rate for each gene in each group. We defined specificity by the posterior probability $P(j = Z \mid x_{ij} > \varphi)$. It defines the probability that cell j is a member of group Z if feature i was observed higher expressed than a threshold φ. Detection level is defined by the posterior probability $P(x_{ij} > \varphi \mid j = Z)$, i.e., the relative fraction of cells expressing a gene i above threshold φ in group Z.The threshold parameter φ was set to 2 (i.e. at least 4 molecules per cell). All conditional probabilities were calculated using Bayes' theorem with adjusted marginal probabilities to avoid sample size bias:

$$P(j = Z) = 0.5$$

$$P(x_{ij} > \varphi\ ) = \frac{P(x_{ij} > \varphi \mid j = Z)\lambda + P(x_{ij} > \varphi \mid j \neq Z)\lambda}{2\lambda}$$

with λ is the maximum cardinality of both groups. In addition, absolute differences between expression intensity means of gene i were estimated using the effect size defined by Cohen's *d*:

$$d(\mathbf{x}_{1.}, \mathbf{x}_{2.}) = \frac{\mu_1 - \mu_2}{\sigma_{1,2}}$$

with μ is the gene expression mean and $\sigma_{1,2}$ is the pooled standard deviation of two samples $\{1,2\}$:

$$\sigma_{1,2} = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}}$$

with σ^2 is the gene expression variance in one group.

Microglia baseline annotation

For subsequent downstream analyses, we filtered the normalized and processed gene expression matrix $\mathbf{X}$ by robustly expressed genes with more than 2 molecules in at least 50 cells, removed ribosomal and mitochondrial genes (obtained from Gene Ontology GO:0005840 and Ensembl, respectively), retained only protein-coding genes, and removed a set of genes highly correlating with dissociation-induced stress response (4). To annotate the microglia baseline cell heterogeneity, we extracted 5694 cells from male and female control hIgG1 treated *TREM2*^{CV}-5XFAD mice. This resulted in the filtered single-cell expression matrix $\mathbf{X}' = (x'_{ij}) \in \mathbb{R}^{11361 \times 5694}$. We embedded the data vectors of highly variable genes (see *Gene expression variance modeling*) proximal to a non-linear lower dimensional manifold using Diffusion Maps (see *Spectral dimensionality reduction*). This revealed a temporal axis in the data, presumably a microglia activation trajectory. A weighted cell adjacency matrix was calculated from the diffusion components by using the Jaccard index of the overlap of each cell's nearest neighborhoods. The resulting graph was clustered using Louvain's community detection method available in the R package *igraph*. To better understand the underlying temporal topology of the data, we fitted an unconstrained maximum parsimony tree between all 11

clusters using the *getLinages* function from the *slingshot* R package. It revealed a branching trajectory with 5 terminal ends. The cluster with the highest expression of common microglia marker was determined to be the resting microglia population.

We further compared our clusters to a reference expression profile of DAMs from a previous study (7). First, we calculated vectors with gene expression log2 fold-changes between all clusters and the resting microglia population in our data. Then, we computed a single vector of log2 fold-changes between stage 2 DAMs and the resting population in the reference study. Overall, 10124 genes overlapped between both studies, resulting in a log2 fold-change vector $\mathbf{z}_i \in \mathbb{R}^{10124}$ for each cluster and the reference. Only genes with an absolute log2 fold-change of > 0.5 were considered differentially expressed and retained for the subsequent analysis. Each log2-fold change vector was binarized by using the signum function $\mathrm{sign}(\mathbf{z}_i)$. We tabulated the data by counting the agreements and disagreements in expression directionality between our data and the reference study resulting in a contingency table, $\mathbf{A} = (a_{ij}) \in \mathbb{N}^{2\times 2}$. We interpreted this table as confusion matrix, i.e., how well our data predicts the expression trend of the refence DAM population. Here, we calculated a similarity score by subtracting the sum of the off-diagonal values (false positives + false negatives) from the matrix trace (true positives + true negatives). We further tested the null hypothesis that the overall agreement between both datasets defined by $\mathrm{trace}(\mathbf{A}) \,/\, \sum_i \sum_j a_{ij}$ is less or equal than the no-information rate defined by the fraction of the largest class in the data; test statistics were calculated using the R package *caret*.

The predicted cell cycle phases of each cell were used to identify a cluster of cycling microglia cells; cell cycle phase prediction and scoring were performed in the quality control step (see *Quality control*). Gene Ontology term enrichment analyses were performed using the PANTHER classification system (8) to functionally characterize the expression profile of selected clusters. Marker genes were determined by contrasting one cluster against a cell pool of all other clusters (see *Differential gene expression analysis*); genes were selected by meeting minimum specificity/detection rate/effect size thresholds (IFN-R: 75.0%/10.0%/1.5, MHC-II: 50.0%/10.0%/1.0). We used the 11361 genes contained in our dataset as reference list for each statistical overrepresentation test. False discovery rate was used to correct Fisher's exact test *P*-values for multiple testing.

Sample harmonization

Given that hT2AB and control hIgG1 injections were performed in mice of different sex and carrying distinct TREM2 variants, we next developed a supervised sample harmonization approach to define

the impact of sex, genotype, and treatment on the four microglial trajectories (**Fig. S3A**). We used our results from the baseline analysis of microglia control hIgG1-treated mice as a reference for classifying cells of samples from different conditions. To minimize classification errors, diffusion components of the reference dataset and each query dataset were calculated (see *Spectral dimensionality reduction*). Machine learning was applied on the lower-dimensional manifold to learn cell types from the reference set and to classify the query dataset using Xtreme gradient boosting. Hyperparameters of the machine learning model were optimize using grid search provided by the R package *caret*. Since the manifold is describing a developmental continuum with overlapping cell cluster boundaries, we accepted a reasonable average training accuracy of 87.5% to avoid overfitting. To assess overall classification accuracy, we used reverse projection. We trained a classifier on the predicted classes of the query dataset and projected the cell types for the reference dataset. An average training accuracy of 92.9% indicates that the predicted classes of the query dataset are highly coherent. Our model achieved a high overall average prediction accuracy of 86.3% (**Fig. S3B**).

Microglial cell type trajectory reconstruction

To analyze distinct trajectories for each microglia cell type, we extracted the branching cluster, all terminal clusters, and the intermediate cluster *t6* from the expression matrix $\mathbf{X}$ containing all samples. The data was then split by cell type: DAM trajectory = {*t5*, *t6*, *DAM*}, Cyc-M trajectory = {*t5*, *Cyc-M*}, IFN-R trajectory = {*t5*, *IFN-R*}, MHC-II trajectory = {*t5*, *MHC-II*}. Cells of each cell type trajectory were embedded onto a lower-dimensional manifold using diffusion maps (see *Spectral dimensionality reduction*) on the highest variable genes (see *Gene expression variance modeling*) and projected onto 2-dimensions with UMAP to unfold the latent temporal axis. To further order cells chronologically, we fitted smooth trajectory curves for each linage using principal curves and projected datapoints orthogonally using the R package *slingshot*; here, arc lengths were interpreted as pseudotime values (9).

Cell type compositional analysis

We aimed to delineate differences in microglial cell-type proportions due to sex, genotype and treatment. Cell type fractions in the single-cell data are not linearly corresponding to true proportions *in vivo* due to technical artifacts, including cell damage induced by sample preparation and handling or by pressure changes during cell sorting, amongst others. Further, small sample sizes may not properly represent the actual population. This may cause large variances in cell type proportions between biological replicates. By using stratified bootstrap resampling we can estimate the population mean and correct the sampling bias of individual replicates, as well as, measure estimate uncertainty.

Let $\mathbf{W} = (w_{ij}) \in \mathbb{N}^{m \times 2}$ a label matrix composed of 2-dimensional vectors with a label indicator (e.g., cell type or time interval) and a biological replicate indicator for m cells, l label levels, and d biological replicates. For example, $\mathbf{w}_{1.} = \{10, 2\}$ is the label vector for the first cell listed in $\mathbf{W}$, which was sampled from time interval 10 and from the second biological replicate. We constructed stratified resamples of size k, where k was set to the maximum replicate size. Each resample was then tabulated resulting in a matrix $\mathbf{V} = (v_{ij}) \in \mathbb{R}^{d,l}$ which contained absolute numbers of cells per cell label and replicate. To conservatively correct biases in single replicates, resamples were aggregated by

$$\mathbf{b} = \frac{\min_i v_{ij}}{k}, j = 1, \dots l$$

The resulting vector $\mathbf{b} \in \mathbb{R}^l$ represented a sample of cell type proportions. We performed $N = 500$ bootstrapping iterations resulting in a matrix $\mathbf{B} = \left(\mathrm{b}_{ij}\right) \in \mathbb{R}^{N \times l}$. Estimates of cell-type proportions $\boldsymbol{\mu} \in \mathbb{R}^l$ were derived by

$$\boldsymbol{\mu} = \frac{\sum_{i=1}^{N} b_{ij}}{N}, j = 1, \dots l$$

The standard error for the 95% confidence interval for an estimated cell type proportion of label j were derived by:

$$\mathrm{SE}_j = 1.96 \frac{\sigma_j}{\sqrt{N}} \text{ with } \sigma_j = \sqrt{\frac{\sum_{i=1}^{N} b_{ij} - \mu_j}{N - 1}}$$

Inference of expression dynamics

We fitted gene expression as a function of pseudotime for each microglia cell type trajectory on a subset of cells from a specified genotype, sex and treatment using negative binomial generalized additive models (NB-GAM) (10). Wald tests were performed to test the hypothesis if the beginning or the end of a gene's dynamic curve differs between conditions and corresponding log fold-changes were calculated using parameters of the NB-GAM smoothers. These computations were performed with the R package *tradeSeq*. To statistically classify gene dynamics, Wald test *P*-values were corrected for multiple testing via false discovery rate. Each corrected *P*-value p was weighted by the sign of the log fold-change S by p^S and -log10 transformed. Resulting negative values denoted downregulation and positive values indicated upregulation, respectively.

Gene expression variance modeling

By assuming that biological heterogeneity is driven by a subset of genes having high variance between cells, we aim to improve resolution by removing genes driven by technical noise. Total

variance of each gene is decomposed into its biological and technical components by fitting the variance as a function of mean expression (11). Significance is inferred by modeling the residuals of this fit with an *F*-distribution. To retain condition-specific variance, the union of all highly variable genes of all batches is always used. Variance decomposition and *F*-distribution statistics were calculated using the R package *scran*.

Spectral dimensionality reduction

This step aims to reduce redundancy and to improve the signal-to-noise ratio in the data, which eventually will reveal latent biological factors in the data. For this purpose, we employed spectral dimensionality reduction methods. Linear spectral embedding is obtained by a Principal Component Analysis (PCA). This method captures the maximal variance in the data. It may miss substructures in the data but is sufficient for data with low intrinsic complexity or may be used to get a first insight into the data structure. PCA was calculated with the R package *irlba*. Non-linear embedding was performed by using Diffusion Maps (12). This method resolves non-linear and linear substructures in the data based on a cell dissimilarity matrix calculated from a distance function D on each cell's expression profile. However, instead of calculating the diffusion components using an estimated global sigma in the diffusion kernel K with $\mathrm{K}(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{-\mathrm{D}(\mathbf{x}_i,\mathbf{x}_j)^2}{2\sigma}\right)$, we used multiple local sigmas as proposed by Haghverdi *et al.* (13):

$$K(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\frac{2\sigma_i\sigma_j}{\sigma_i^2 + \sigma_j^2}} \exp\left(\frac{-\mathrm{D}(\mathbf{x}_i, \mathbf{x}_j)^2}{\sigma_i^2 + \sigma_j^2}\right)$$

We account for batch effects in the data as follows. For PCA, we center the input expression matrix by the mean of the center vectors of each batch and scale the covariance matrix by the total cell count of each batch. This ensures that each batch contributes equally to the identification of the loading vectors (i.e., the PCA won't be dominated by samples with high cell count). For Diffusion Maps, we use the robust pairwise cosine correlation distance. For both techniques, the resulting lower-dimensional matrix was corrected for batch effects using mutual nearest neighbors' batch correction (14) (R package *batchelor*). The number of selected components was guided by a scree plot analysis.

Immunostaining for Aβ and Image Analyses

Free-floating brain sections were initially blocked and permeabilized with PBS + 3% BSA solution containing 0.25% Triton X-100. Primary antibodies were added at a dilution of 1:1,000 for $A\beta_{1-16}$ (6E10, conjuncted Alexa Fluor 488, 803013, BioLegend), and 1:500 for $A\beta_{42}$ (rabbit recombinant monoclonal, Thermo Fisher Scientific) at 4°C for overnight. Secondary antibodies, anti-rabbit IgG Alexa Fluor 647 (goat recombinant polyclonal,1:1,000; Invitrogen) and methoxy-X04 (3 μg/ml; Tocris) were added for 1.5 hours at room temperature. The confocal pictures were taken on a Nikon A1Rsi+ confocal laser-scanning microscope using a 20 × 0.95-NA objective. z-Stacks with 1.1-mm steps in the z direction, 1,024 × 1,024-pixel resolution, were recorded. The percentage of Aβ area coverage was calculated automatically by batch processing in ImageJ.

Aβ and Chemokine/Cytokines Quantification

Human $A\beta_{1-40}$ and $A\beta_{1-42}$ were measured by MSD V-plex 6E10 kit (Meso Scale Discovery) and mouse IL-1b, CXCL10 (IP-10) and CCL4 (MIP-1b) were measured by MSD U-plex biomarker group 1 Assay (Meso Scale Discovery). Briefly, cortical tissues from antibody or control hIgG1 treated mice were homogenized with Tissue Lyser II (Qiagen) in 12x v/w of PBS containing 0.5% of Triton x-100 and 1x Halt protease inhibitor cocktail. The insoluble fraction was pelleted by ultracentrifugation at 100,000g for 1 hour. Supernatant was collected as soluble PBS fraction. The pellet was resuspended in 250 μl of 6 M guanidine and 50 mM Tris, pH 8.0, buffer, and was further homogenized by sonication, followed by ultracentrifugation at 75,000 rpm to clarify the denatured pellet. The supernatant was collected as the insoluble guanidine fraction.

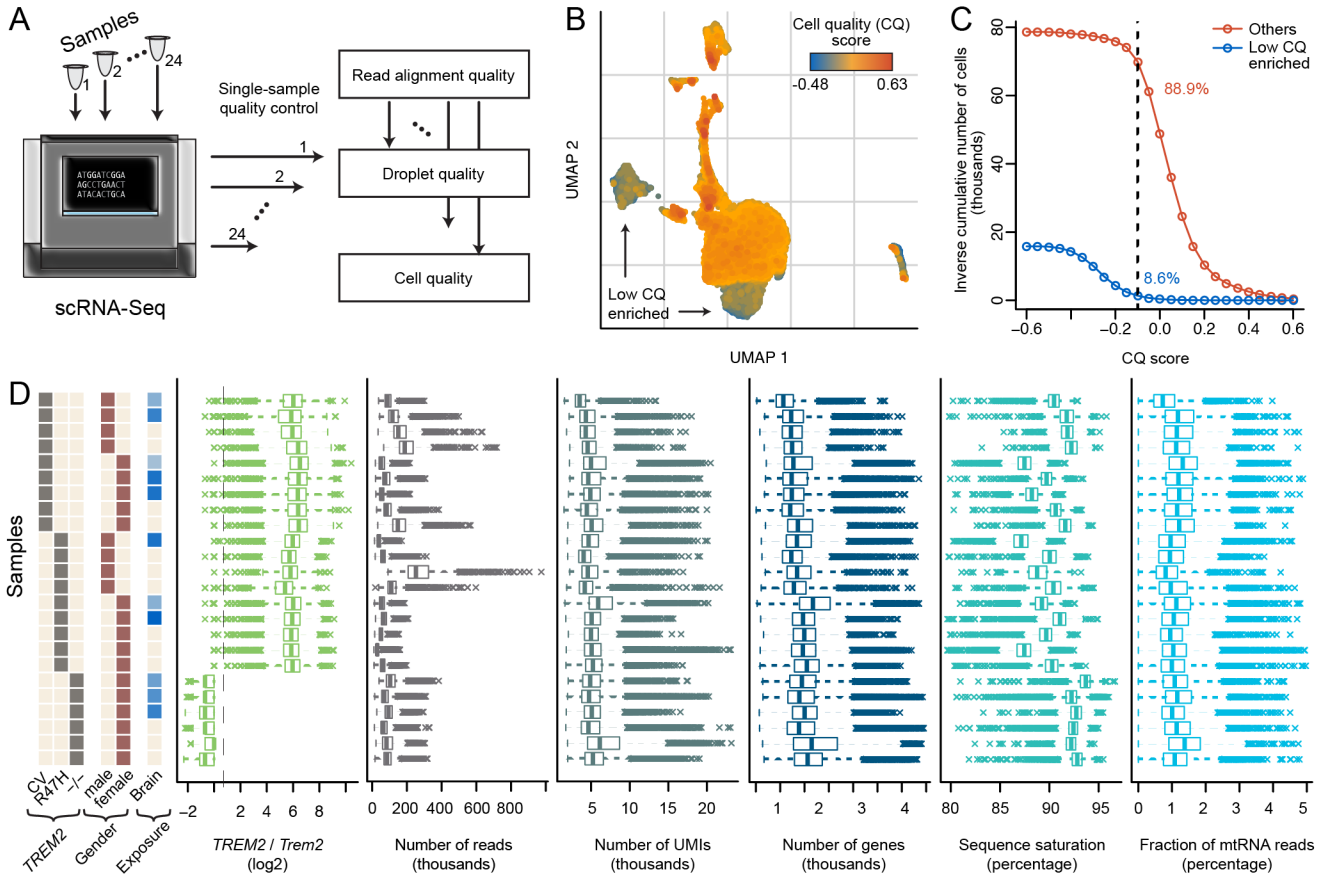

Fig. S1. Single-cell RNA-seq quality control. (A) 24 samples were collected and subjected to individual read alignment, droplet, and cell quality control. (B) Integrative data quality assessment revealed technical artifacts in the data, likely driven by low cellular transcriptome coverage. Cells are colorized by the cell quality (CQ) score derived from a principal component analysis of selected quality metrics. Groups enriched in cells with low CQ scores are indicated. (C) Determination of the cut-off for cell filtering using the inverse empirical cumulative CQ score distribution function. The determined threshold of –0.1 retained 8.6% of cells within the low CQ enriched group and 88.9% of all other cells. (D) Distribution of common quality metrics for each cell per sample. Samples are characterized by human TREM2 variant, sex, and measured hT2AB brain exposure. Shown is also the ratio between the raw expression of the endogenous *Trem2* locus and the human *TREM2* transgenic locus in all collected cells.

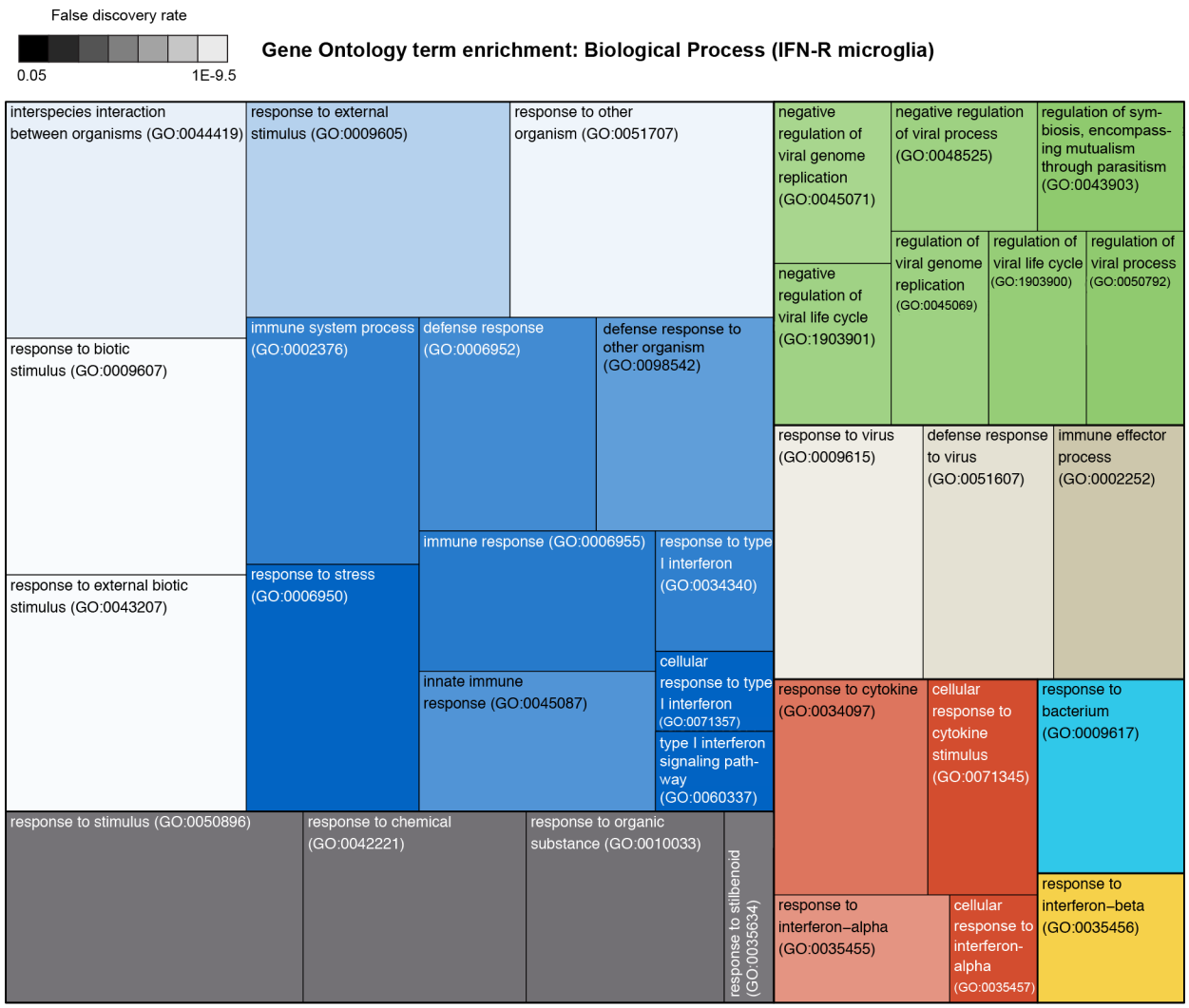

Fig. S2. GO Term Enrichment of IFN-R microglia marker genes. IFN-R marker genes meeting a specificity threshold of 0.75 and an effect size of 1.5 were subjected to a Gene Ontology term enrichment analysis. Shown are biological process terms with a false-discovery rate (*FDR*) corrected Fisher's exact test *P*-value < 0.05. Terms are grouped and colored by their ontological relation, square sizes correspond to the number of IFN-R marker genes found in this category, and color depth is scaled by *FDR*.

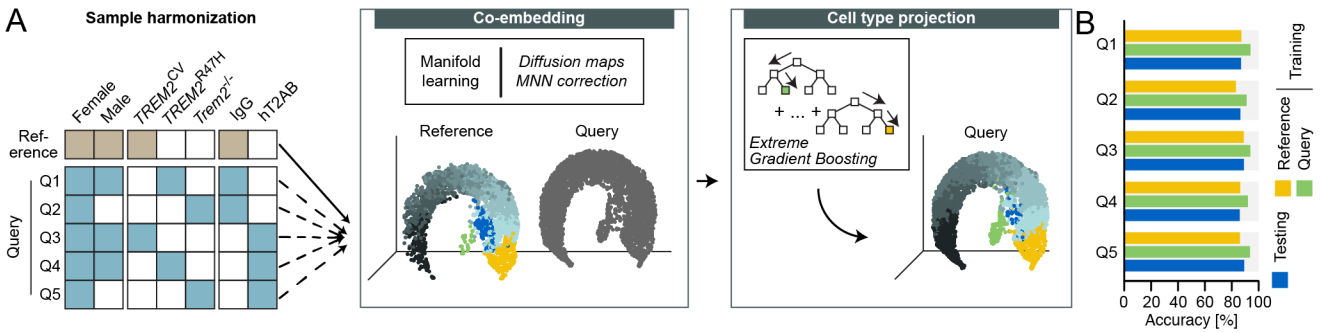

Fig. S3. Sample harmonization. Annotated microglia cells from control hIgG1-treated female and male *TREM2*^{CV}-5XFAD mice were used as reference to classify cells from the remaining 5 conditions (query Q1-Q5). First, each query set was co-embedded with each reference set using diffusion maps and subsequently corrected for batch effects using mutual nearest neighbor correction. Then cell types were projected using an Extreme Gradient Boosting classifier trained on the diffusion components of the query cells. (B) Model accuracy assessments. Training accuracy on the reference or query dataset, as well as prediction accuracy of projecting cell types on the reference dataset using a classifier trained on the query dataset.

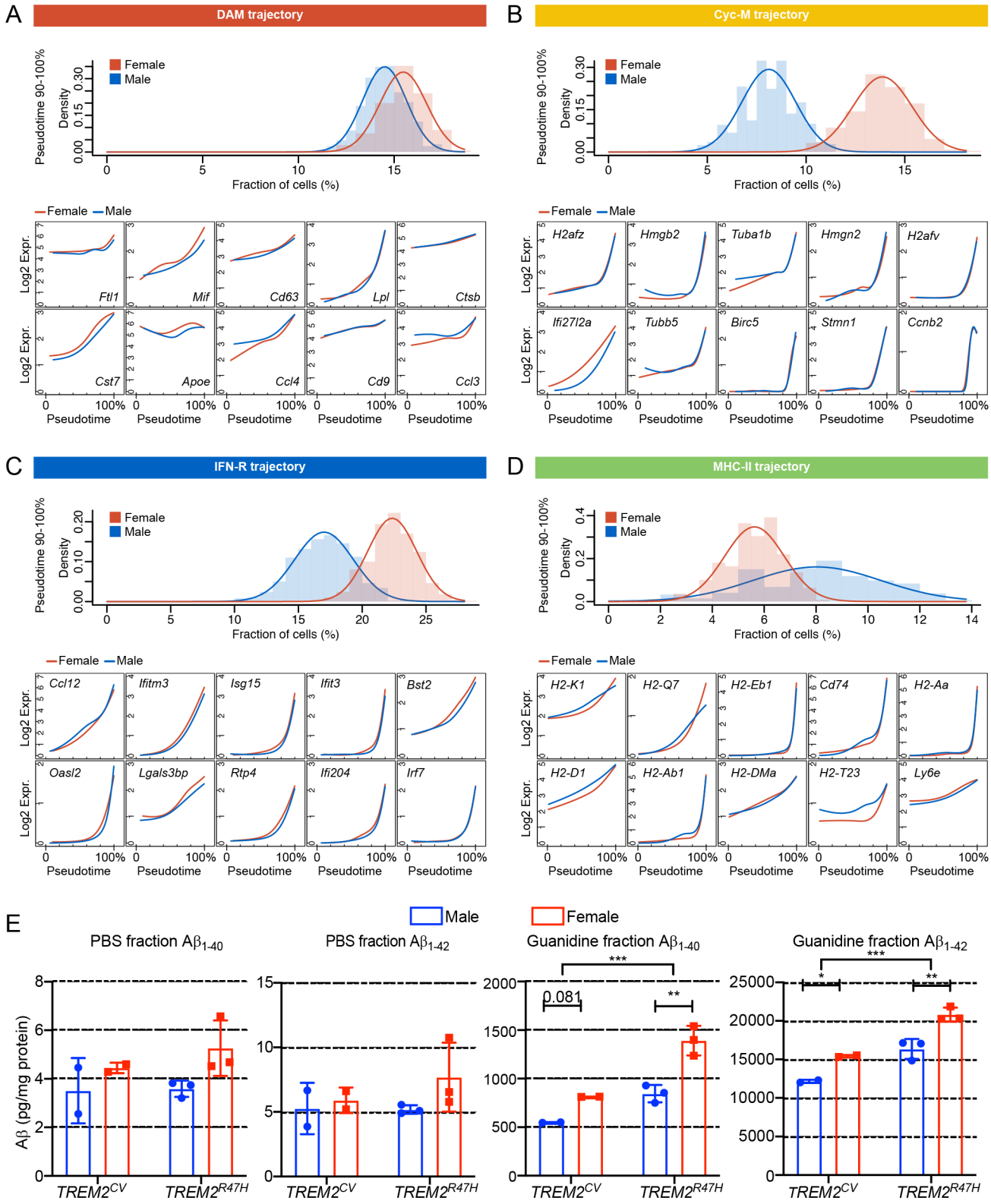

Fig. S4. Mouse sex-associated differences in microglia fates in control hIgG1-treated *TREM2*^{CV}-5XFAD mice. Shown are the distributions of the estimated fractions of cells in the terminal

90-100% pseudotime interval using Bootstrapping (upper panel) and the fitted expression dynamics for the top 10 marker genes (lower panel) for the DAM (A), Cyc-M (B), IFN-R (C), and MHC-II (D) clusters. (E) Quantification of soluble and insoluble $A\beta_{1-40}$ and $A\beta_{1-42}$ in the hippocampus lysates of control hIgG1-treated *TREM2*^{CV}-5XFAD mice. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ by two-way ANOVA with Sidak's multiple comparisons test. Data are shown as mean ± SD.

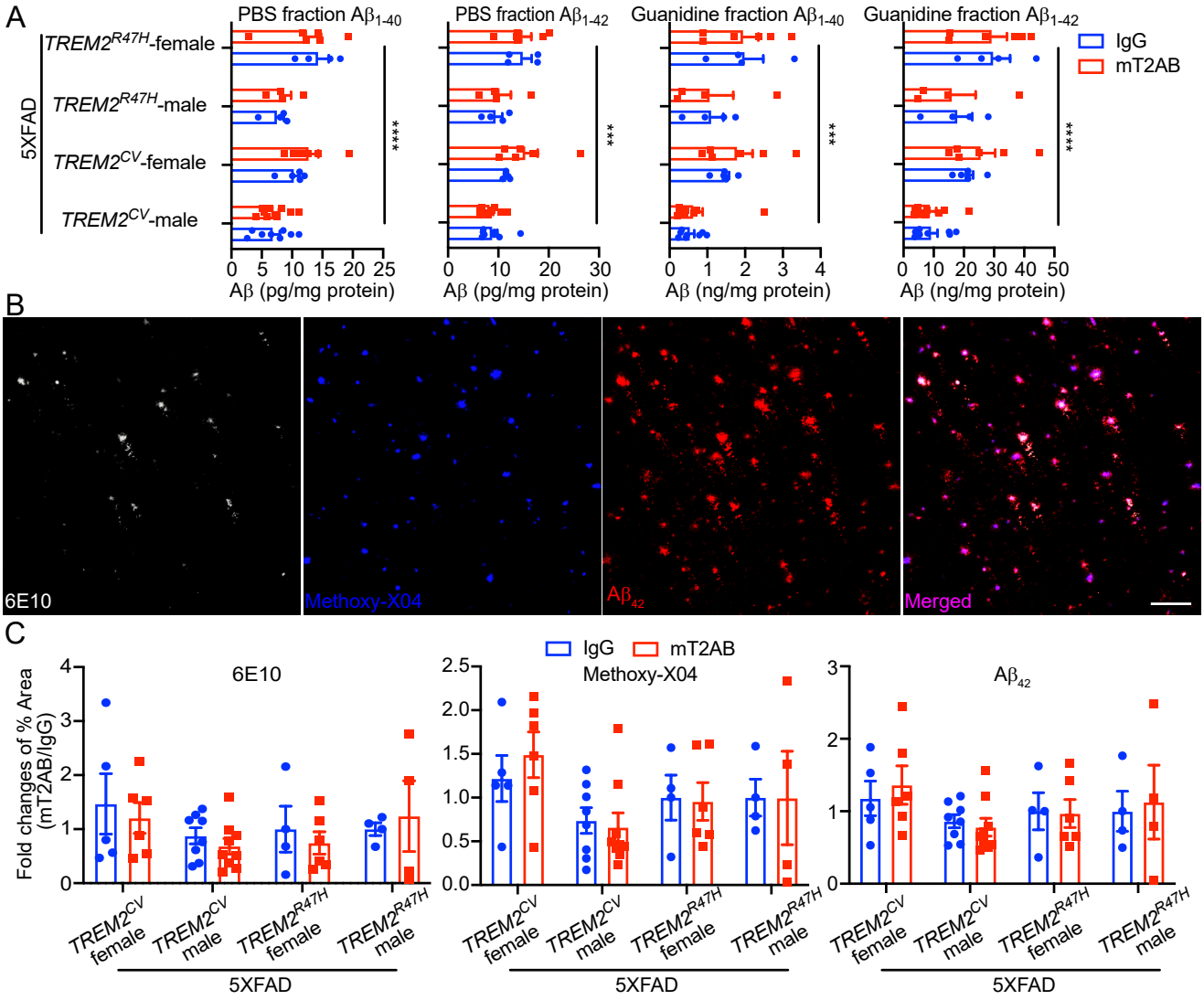

Fig. S5. Acute treatment with mT2AB does not change Aβ load. (A) Quantification of soluble and insoluble $A\beta_{1\text{-}40}$ and $A\beta_{1\text{-}42}$ in the cortex lysates among different treatment groups. (B) Representative confocal images from *TREM2*^{CV}-5XFAD mice treated with mT2AB stained with 6E10 (white), methoxy-X04 (blue), and $A\beta_{42}$ (red), displaying the distribution of Aβ in the cortex. Bar = 100 μm. (C) Fold changes of 6E10^{+}, methoxy-X04^{+} or $A\beta_{42}^{+}$ area coverage in the cortex between mT2AB and control mIgG1 treated groups. Data are shown as mean ± SEM. *TREM2*^{CV}-5XFAD, male, mIgG1, n = 8; *TREM2*^{CV}-5XFAD, male, mT2AB, n = 9; *TREM2*^{CV}-5XFAD, female, mIgG1, n = 5; *TREM2*^{CV}-5XFAD, female, mT2AB, n = 6; *TREM2*^{R47H}-5XFAD, male, mIgG1, n = 4; *TREM2*^{R47H}-5XFAD, male, mT2AB, n = 4; *TREM2*^{R47H}-5XFAD, female, mIgG1, n = 4; *TREM2*^{R47H}-5XFAD, female, mT2AB, n = 6. ***, P <0.001; ****, $P < 0.0001$ by two-way ANOVA with Sidak's multiple comparisons test. Data are shown as mean ± SEM.

Table S1. Cerebellar and serum concentration of hT2AB after a single i.p. injection. Analysis of hT2AB dosed samples in serum/brain showed exposure to hT2AB and the ratio of BBB passing for hT2AB was around 0.19% - 0.76%.

			hT2AB (nM)		
Genotype	Gender	Treatment (48 h)	Serum	Brain	% Ratio (Concentration Br/Sr)
TREM2 CV -5XFAD	Male	hT2AB	2153.33	6.56	0.30
TREM2 CV -5XFAD	Male	hT2AB	2020.00	3.75	0.19
TREM2 CV -5XFAD	Female	hT2AB	394.67	2.98	0.76
TREM2 CV -5XFAD	Female	hT2AB	2460.00	7.27	0.30
TREM2 CV -5XFAD	Female	hT2AB	2793.33	7.53	0.27
TREM2 R47H -5XFAD	Male	hT2AB	1773.33	8.00	0.45
TREM2 R47H -5XFAD	Female	hT2AB	1733.33	4.18	0.24
TREM2 R47H -5XFAD	Female	hT2AB	1673.33	9.07	0.54
Trem2-/- -5XFAD	Female	hT2AB	2100.00	7.20	0.34
Trem2-/- -5XFAD	Female	hT2AB	1993.33	5.42	0.27
Trem2-/- -5XFAD	Female	hT2AB	2280.00	5.69	0.25

Dataset S1. Differential Gene Expression Analysis. Listed are contrasted gene expression profiles of (A) each CD45-positive, (B) each microglia cell group against all other cells, as well as (C) IFN-R microglia against MHC-II microglia, respectively. Genes are characterized by their Ensembl (https://www.ensembl.org/) identifiers, symbols, and names. Absolute differences in expression are quantified by effect size and log2 fold-change. Gene expression specificity and gene detection rate were determined using conditional probabilities. Specificity is defined by the posterior probability of a cell being a member of a cell type; the detection level is defined by the relative fraction of cells expressing a given gene in a group. A gene was classified as expressed in a cell, if it had at least four RNA molecules detected. Posterior probabilities are calculated using uniform priors for cell types to avoid a sample size bias.

Dataset S2. Expression Dynamics Analysis. For each trajectory gene expression was fitted as a function of pseudotime. Genes that are expressed in at least 10% of cells with an average expression of at least 2 mRNA molecule counts in any condition were used. Early and late terminal timepoints of the resulting expression dynamic curves were compared between sex and genotype matched hT2AB and control hIgG1 treated samples. *P*-values were calculated using Wald Statistics and corrected for multiple testing via false discovery rate (FDR). The FDR was weighted by the sign of the log fold-change S by FDR^S and -log10 transformed. Negative values denote hT2AB -induced downregulation, positive values indicate upregulation. The resulting -log10(FDR^S) values for early and late terminal timepoints, as well as, the type of expression change are listed for the (A) DAM, (B) Cyc-M, (C) IFN-R, and (D) the MHC-II trajectory. Genes are characterized by their Ensembl (https://www.ensembl.org/) identifiers, symbols, and names.

SI References

1. A. Scialdone, *et al.*, Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods* **85**, 54–61 (2015).

2. E. Becht, *et al.*, Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–47 (2019).

3. A. T. L. Lun, K. Bach, J. C. Marioni, Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 1–14 (2016).

4. S. C. Van Den Brink, *et al.*, Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* **14**, 935–936 (2017).

5. T. S. P. Heng, *et al.*, The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).

6. V. A. Traag, L. Waltman, N. J. van Eck, From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 1–12 (2019).

7. H. Keren-Shaul, *et al.*, A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276-1290.e17 (2017).

8. H. Mi, *et al.*, Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.* **14**, 703–721 (2019).

9. K. Street, *et al.*, Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).

10. K. Van den Berge, *et al.*, Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.* **11**, 1–13 (2020).

11. A. T. L. Lun, D. J. Mccarthy, J. C. Marioni, A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor [version 2 ; referees : 3 approved , 2 approved with reservations]. *F1000Research* (2016).

12. R. R. Coifman, S. Lafon, Diffusion maps. *Appl. Comput. Harmon. Anal.* **21**, 5–30 (2006).

13. L. Haghverdi, F. Buettner, F. J. Theis, Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989–2998 (2015).

14. L. Haghverdi, A. T. L. Lun, M. D. Morgan, J. C. Marioni, Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).