



Disease-associated astrocytes in Alzheimer's disease and aging

Naomi Habib^{1,2}✉, Cristin McCabe^{2,8}, Sedi Medina^{3,8}, Miriam Varshavsky^{1,8}, Daniel Kitsberg¹, Raz Dvir-Szternfeld³, Gilad Green¹, Danielle Dionne¹, Lan Nguyen², Jamie L. Marshall¹, Fei Chen², Feng Zhang^{2,4,5}, Tommy Kaplan¹, Aviv Regev^{1,2,7}✉ and Michal Schwartz^{1,2,3}✉

The role of non-neuronal cells in Alzheimer's disease progression has not been fully elucidated. Using single-nucleus RNA sequencing, we identified a population of disease-associated astrocytes in an Alzheimer's disease mouse model. These disease-associated astrocytes appeared at early disease stages and increased in abundance with disease progression. We discovered that similar astrocytes appeared in aged wild-type mice and in aging human brains, suggesting their linkage to genetic and age-related factors.

Alzheimer's disease (AD) is a highly heterogeneous disease, and the most frequent cause of cognitive decline. Recent findings have attributed a major role to nonneuronal cells in disease onset and escalation (for example, see refs. ^{1–3}). However, fully elucidating their function has been challenging, due to the heterogeneity of cells and disease dynamics. Some of this cellular diversity has recently been characterized by single-cell or single-nucleus RNA sequencing (sNuc-seq), highlighting disease-related states of microglia in mouse and human brains^{1,4}. However, astrocytes, which have a wide range of activation states^{5,6}, with variable effects on disease onset and progression⁷, have been less characterized. In particular, massive gliosis has been observed upon AD initiation⁸, to which a negative effect has been attributed⁹. Thus, higher-resolution characterization of astrocytes and other nonneuronal cells could help identify novel cellular components of AD pathogenesis¹⁰.

We used sNuc-seq¹¹ to build a cellular-molecular map of the hippocampi of 7-month-old mice of either WT or a transgenic model of AD (5xFAD)¹² (8 mice, 54,769 high-quality nuclei; Fig. 1a and Supplementary Table 1), using two different protocols to comprehensively capture nuclei across cell types (Extended Data Fig. 1a–c). We partitioned the cells into 23 clusters (Fig. 1b and Extended Data Figs. 1 and 2), revealing largely similar cellular landscapes in WT and AD, with differences in the proportions of several cell populations ($P < 0.01$; Fig. 1c). There were major differences in astrocyte states in AD relative to WT (Fig. 1c), as well as prominent differences in AD microglia, the frequency of which increased along disease progression (relative to all other cells; Fig. 1c), including the disease-associated microglia (DAM) population¹ (Extended Data Fig. 2c). The frequency of pericyte/endothelial cells also increased in AD ($P < 0.01$; Fig. 1c), while that of recently activated pyramidal neurons decreased ($P < 0.01$; Fig. 1c and Extended Data Fig. 2d).

Several rare stromal populations also decreased in frequency ($P < 0.01$; Fig. 1c and Extended Data Fig. 2e).

A continuous range of astrocyte profiles (Extended Data Fig. 3a–c) were aligned on a diffusion map along trajectories between three transcriptional states (Fig. 1d and Extended Data Fig. 3d). The cells were partitioned into six transcriptional sub-clusters (Fig. 1d,e and Extended Data Fig. 3a), and their proportions in the various states differed between AD and WT (Fig. 1f and Extended Data Fig. 3e), with the appearance of a unique state observed only in AD. WT astrocytes spanned a trajectory between a homeostatic Gfap-low state (clusters 1 and 2) and a Gfap-high state (cluster 6), in line with previous observations¹³.

In AD, an additional astrocyte Gfap-high state was found, which we termed disease-associated astrocytes (DAAs; cluster 4; Fig. 1f). DAAs were observed in both male and female mice (Extended Data Fig. 4a–d), as well as in the cortex (in 7- and 10-month-old AD mice; Extended Data Fig. 4e–i), showing that DAAs are not restricted to the hippocampus. The appearance of DAAs was accompanied by a significant decrease in the homeostatic, Gfap-low astrocyte population, relative to WT mice ($P < 0.01$; see Methods). In addition, cluster 5 (in WT and AD) reflected a transitional-like intermediate state between the Gfap-low state and Gfap-high states (cluster 6), while cluster 3, which significantly increased in frequency in AD (Fig. 1f), reflected a transitional-like intermediate state between the Gfap-low state and the DAA (cluster 4; Fig. 1d,e).

The Gfap-high astrocyte population, observed in both WT and AD, and the DAA population, unique in AD, shared a significant number of upregulated genes compared with Gfap-low astrocytes (hypergeometric $P < 10^{-100}$; Fig. 2a,b), with little downregulation of genes (Fig. 2a and Extended Data Fig. 4j); yet, each state had distinct expression features (Supplementary Table 2). Both Gfap-high and DAA populations upregulated genes involved in development and differentiation, metabolic pathways (lipid, cholesterol), response to toxic compounds and inflammatory signaling (false-discovery rate (FDR) < 0.05 ; Fig. 2b, Extended Data Fig. 5 and Supplementary Table 3). Cells in the intermediate states, clusters 3 and 5, upregulated a subset of the genes that were also upregulated in cluster 4 (DAAs) and cluster 6 (Gfap-high), respectively, supporting their definition as intermediate states (Fig. 2b and Extended Data Fig. 5). Notably, the Gfap-high astrocyte population, found in both AD and WT, specifically expressed markers of end-feet

¹Edmond & Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel. ²Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. ³Neurobiology Department, The Weizmann Institute, Rehovot, Israel. ⁴Howard Hughes Medical Institute, McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁵Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁶School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel. ⁷Howard Hughes Medical Institute, Koch Institute of Integrative Cancer Research, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁸These authors contributed equally: Cristin McCabe, Sedi Medina, Miriam Varshavsky.
✉e-mail: naomi.habib@mail.huji.ac.il; regev@broadinstitute.org; michal.schwartz@weizmann.ac.il

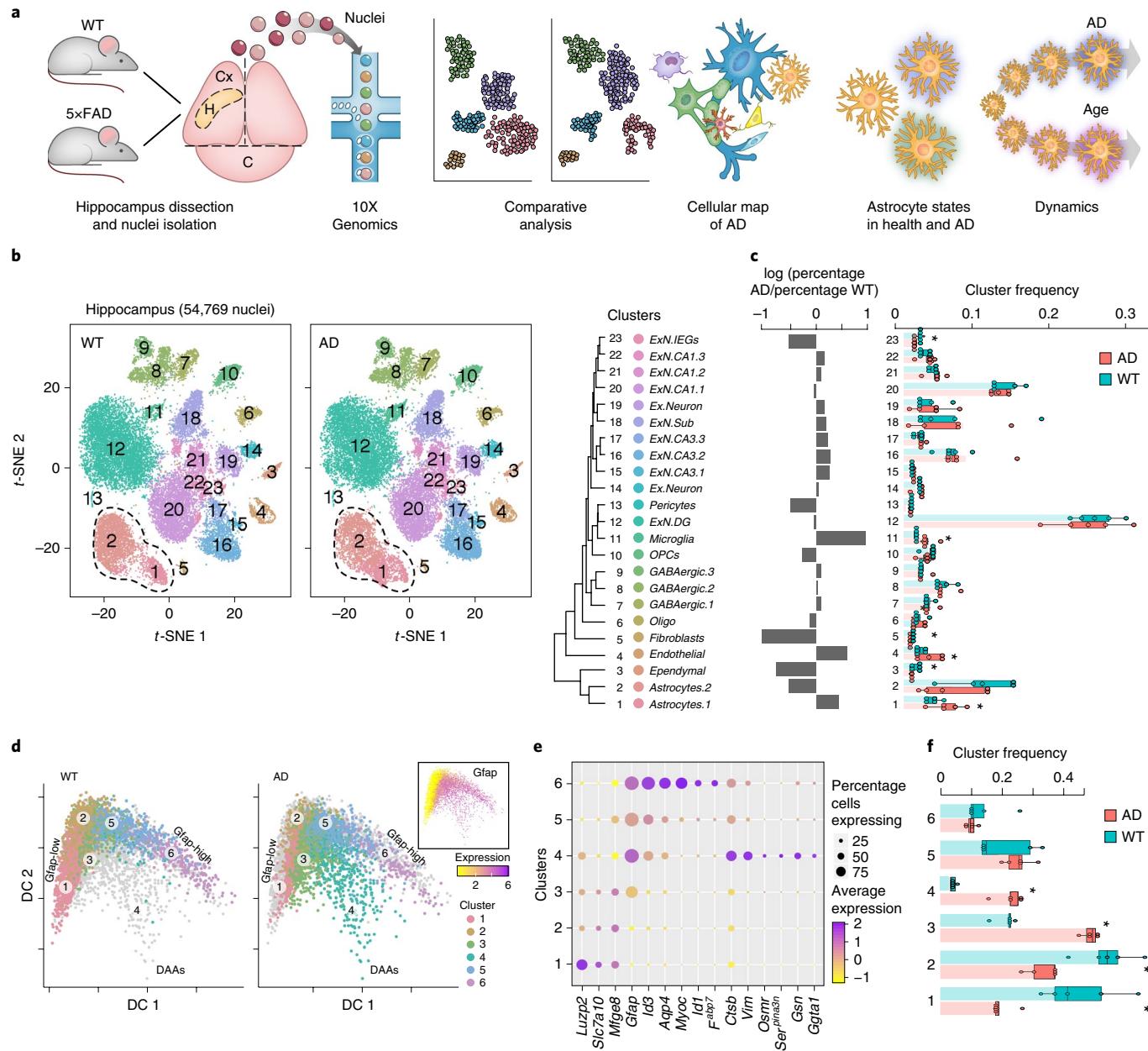


Fig. 1 | A cell atlas of the hippocampus reveals unique DAA states in AD. **a**, Overview of the experimental strategy. **b**, Cell map of mouse hippocampus in WT and AD. t-SNE of 54,769 single-nucleus RNA profiles from hippocampi of 7-month-old male mice, four WT and four 5xFAD (AD); colored by cluster. Right: hierarchical cluster tree. **c**, Changes in frequency of multiple cell types in AD. Right: boxplot showing fraction of nuclei per cluster in WT and AD mice. Box, 75% and 25% quantiles. Line, median. Dots, individual samples. Left: log ratio of average fraction in WT versus AD. Asterisks denote statistically significant differences between AD and WT ($n=8$ animals, two-sided $*P<0.01$, paired t-test or Wilcoxon test for nonnormally distributed samples). **d**, A continuous trajectory across three major astrocyte states in AD and WT brains. Diffusion map embedding of 7,345 WT and AD astrocytes, colored by cluster (gray, cells of other conditions). The three end states are marked: Gfap-low, Gfap-high and DAA. Inset: map colored by Gfap expression level. **e**, Marker genes of astrocyte states. Expression level (color scale) of marker genes across clusters and the percentage of cells expressing them (dot size). **f**, An increase in frequency of DAA states and reduction in frequency of homeostatic Gfap low astrocytes in AD. Boxplot (as in **c**) showing the fraction of nuclei per cluster in WT and AD. Asterisks denote statistically significant differences between AD and WT ($n=8$ animals; statistical test as in **c**). Similar proportions found in cortical brain region and in female mice, see Extended Data Fig. 4a-i.

and blood vessel-associated astrocytes, such as the Aqp4 (ref. ¹⁴). DAA expressed a unique set of genes, including genes involved in endocytosis, complement cascade and aging (FDR < 0.05; see Methods and Fig. 2a,b). Among the DAA-upregulated genes were *Serpina3n*, encoding a serine protease inhibitor linked to increased amyloid accumulation¹⁵, and *Ctsb* (Cathepsin B), encoding a lysosomal cysteine protease involved in proteolytic processing of amyloid precursor

protein¹⁶ (Figs. 1e and 2a and Extended Data Fig. 4j). Interestingly, DAA (cluster 4) and the associated intermediate state cluster 3 cells also expressed *Apoe* and *Clu*¹⁷, along with other genes encoding proteins associated with amyloid metabolism and clearance¹⁷ (FDR < 0.05; Fig. 2b and Extended Data Fig. 5).

To relate DAA to previously described astrocyte profiles, we examined the expression level of RNA signatures from bulk

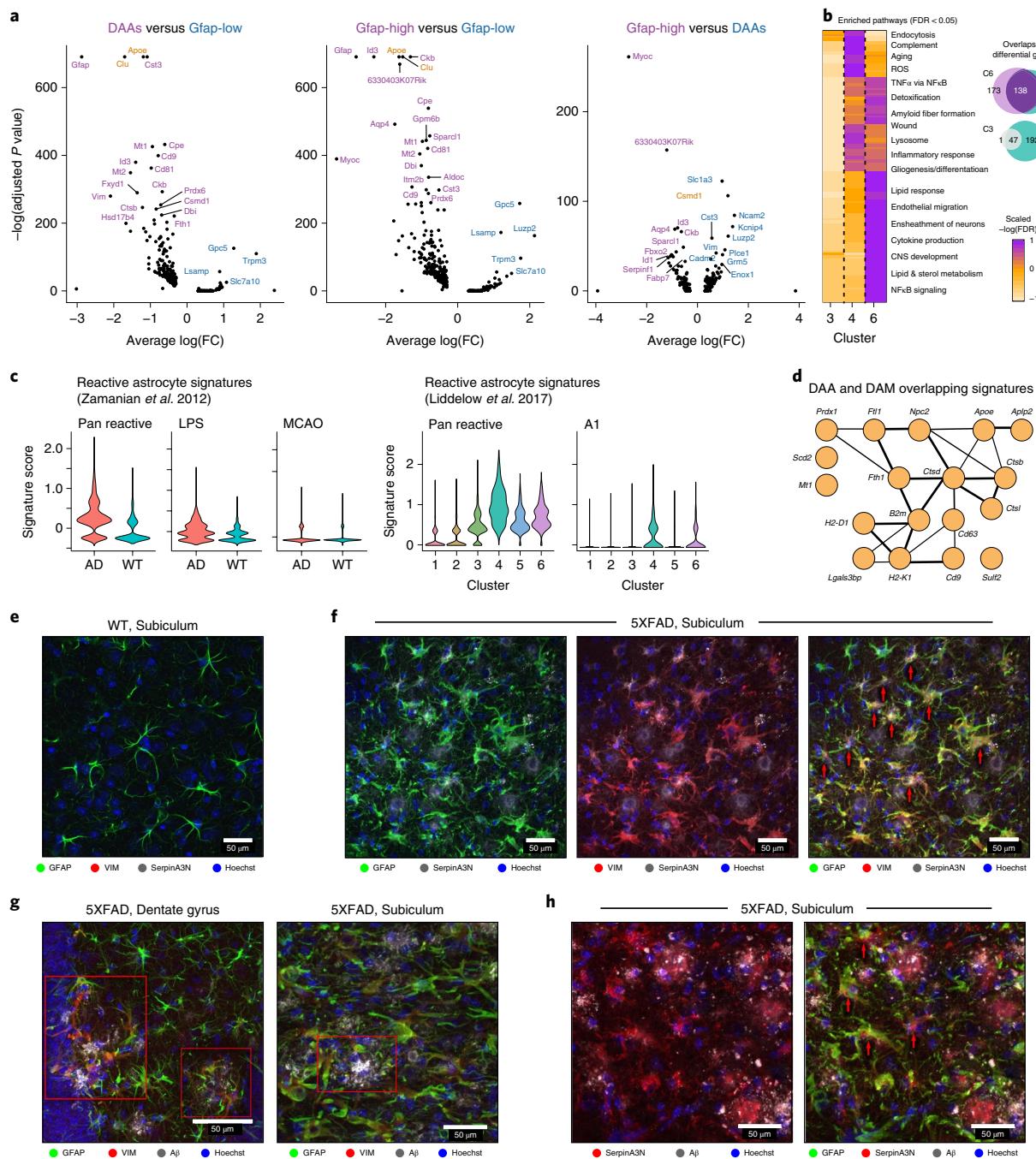


Fig. 2 | DAA markers are associated with diverse molecular pathways and positioned in proximity to A β plaques in AD. **a**, Multiple genes upregulated in DAA and Gfap-high compared with Gfap-low astrocytes (total 7,345 astrocytes across states). Volcano plot showing differential expression of genes in each pair of states ($n=8$ animals; 1,594, 478 and 457 cells in clusters 1, 4 and 6, respectively; negative binomial P value, FDR correction). All cluster pairs are in Extended Data Fig. 5a. AD risk factor genes from genome-wide association studies (GWAS) marked in orange. FC, fold change. **b**, Shared and distinct upregulated genes and pathways in cluster 4 (DAA), 6 (Gfap-high) and 3 (intermediate) astrocytes compared with Gfap $^{\text{low}}$ astrocytes. Left: heatmap showing the significance of upregulated pathways in the three clusters ($n=8$ animals; 2,265 cells in cluster 3, other cell numbers as in **a**; hypergeometric P value (hP), FDR < 0.05 ; -log(FDR) values as color bar) Selected pathway names are highlighted (see full list in Extended Data Fig. 5b and Supplementary Table 3). Right: Venn diagrams of upregulated genes (hP FDR < 0.01). **c**, Signatures of reactive astrocytes found in AD. Violin plots showing the distribution of expression scores for previously defined signature genes for (left) pan-reactive, inflammation-associated (lipopolysaccharides, LPS) and ischemia-associated (middle cerebral artery occlusion, MCAO) reactive astrocytes (from ref. ⁶) across WT (3,831 cells, $n=4$ animals) and AD (3,514 cells, $n=4$ animals), and (right) pan-reactive and A1-reactive astrocytes (from ref. ⁵) across clusters ($n=8$ mice, 10 samples; Extended Data Fig. 6a-c). **d**, Overlapping upregulated genes between DAM¹ and DAA. Eighteen shared genes (nodes), connected by functional links (edges), weighted (edge width) proportionally to the confidence levels (see Methods). **e-h**, Astrocytes expressing DAA markers are found in AD brains, enriched in the subiculum and in proximity to A β plaques. Representative immunofluorescence images in sagittal sections of 7-8-month-old mice. Subiculum, stained for GFAP (green), VIM (red) and serpinA3N (gray), in WT mice (**e**) and 5xFAD mice (**f**). Dentate gyrus (**g**, left) and subiculum (**g**, right), stained for GFAP (green), VIM (red) and A β (gray), in 5xFAD mice. Subiculum, stained for GFAP (green), serpinA3N (red) and A β (gray), in 5xFAD mice (**h**). Cell nuclei are shown in blue (Hoechst). Scale bar, 50 μ m. Red arrows or red boxes, colocalization of three proteins. Experiments were repeated on four brain slices per animal on $n=4$ AD and WT mice.

datasets^{5,6} and found that AD astrocytes overall had an increased expression of pan-reactive and inflammation/A1 astrocyte signatures compared with WT, but did not show an increase in ischemia/A2 astrocyte signatures (Fig. 2c and Extended Data Fig. 6a,b). While most of the inflammation/A1 signature genes were found to be expressed by DAAs, some were found to be expressed by the other astrocyte populations (Fig. 2c and Extended Data Fig. 6c), indicating that previously reported bulk signatures^{5,6} captured a mixed population. Notably, there were 18 shared genes (hypergeometric $P < 3 \times 10^{-4}$; Fig. 2d) between the 239 DAA signature genes and 213 signature genes previously described for DAM¹. These included the AD risk gene *Apoe* and the *Ctsb*, *Ctsd* and *Ctsl* genes, encoding proteins (Cathepsins B, D and L) implicated in AD pathogenesis¹⁶.

We verified the presence of DAAs at the protein level and characterized their spatial distribution by co-staining the hippocampus of WT and 5xFAD animals for the GFAP and DAA markers SERPINA3N and VIM (markers of reactive astrocytes⁶ and NSCs¹⁸, respectively; Fig. 2e,f). GFAP⁺SERPINA3N⁺VIM⁺ astrocytes were observed in 5xFAD but not WT mice, in line with the sNuc-seq profiles, and were most highly concentrated in the subiculum, where disease is known to be most prominent¹². While in WT mice VIM⁺ cells were found in the dentate gyrus, consistent with their known association with the neurogenic niche¹⁸, in AD, VIM⁺GFAP⁺ astrocytes were prevalent throughout the hippocampus and especially in the subiculum. SERPINA3N⁺ and VIM⁺ astrocytes were found adjacent to stained amyloid beta plaques (Fig. 2g,h and Extended Data Fig. 6d). Notably, SERPINA3N, a secreted protein, was also detected embedded in plaques (Fig. 2g,h and Extended Data Fig. 6d), and thus may be expressed by other cells within the diseased brain.

To determine how changes occur along disease progression, we profiled cells by sNuc-seq from AD and WT mice at different ages (1.5–2, 4–5, 7–8, 10, 13–14 and 20 months old, 28 mice, 23,863 astrocyte nuclei; Fig. 3a). As reported¹, microglia frequencies were higher in AD than WT brains at all time points tested, from 4 months, and increased with age (Fig. 3b). Astrocytes mapped along the same continuous trajectories, spanning from Gfap-low to Gfap-high states in WT mice (aged 1.5–10 months) and young AD mice (1.5–2 months old; Fig. 3c and Extended Data Fig. 7). In AD, however, from 4 months of age, before manifestation of cognitive decline¹², there was a consistent reduction in Gfap-low astrocytes and a corresponding increase in DAAs (Fig. 3d). The results suggest

that DAAs arise in the AD mouse model before cognitive decline and increase along disease progression.

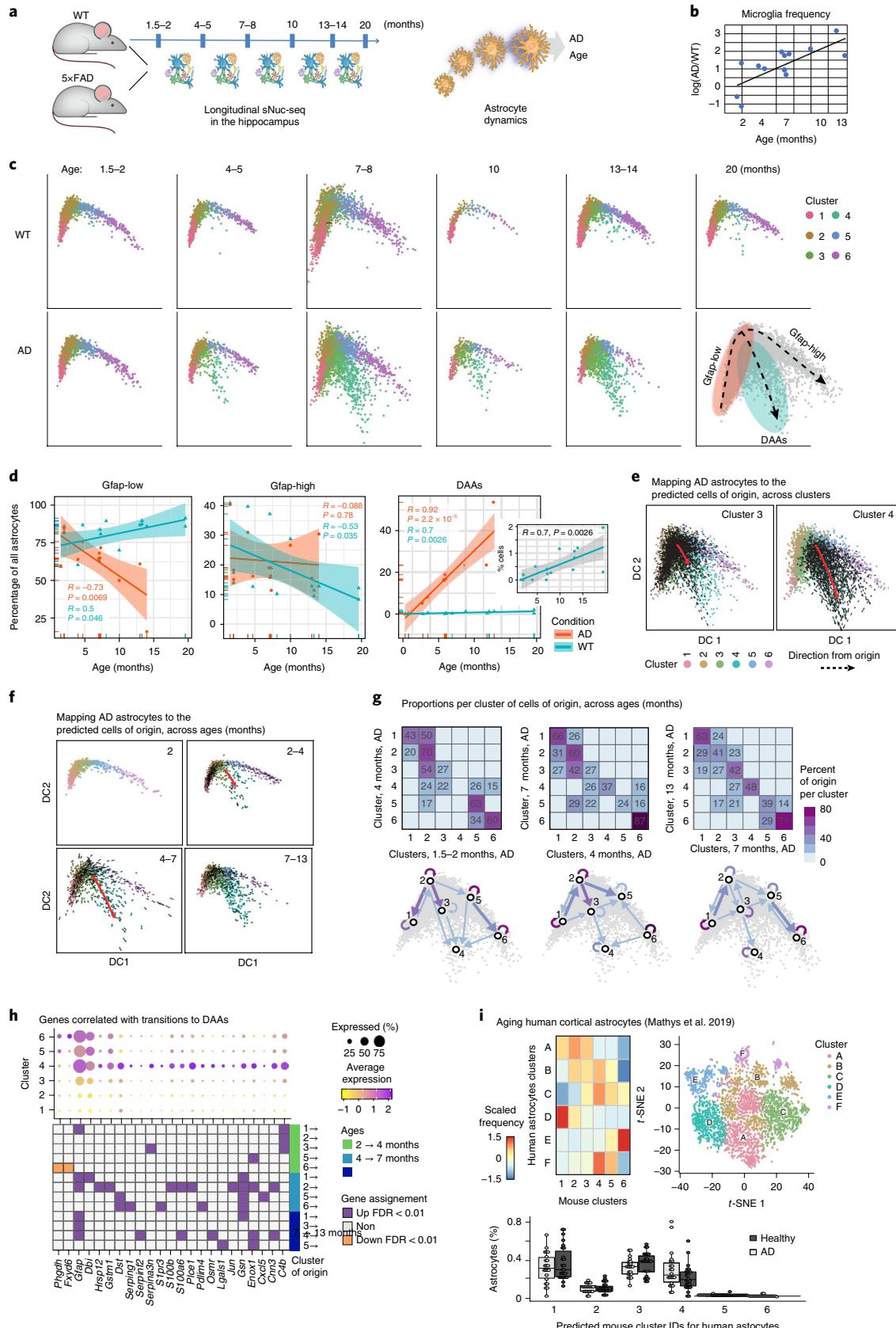
Examining the diffusion map (Figs. 1d and 3c and Extended Data Fig. 7) suggested that Gfap-low astrocytes are the potential source population of both Gfap-high astrocytes and DAAs (Fig. 3c). To explore this possibility, we devised a computational procedure to infer transitions between cell states. We constructed a cell-cell nearest neighbors graph, and found the optimal global assignment to match each AD astrocyte, at a given time point, to a nearest neighbor, termed an origin cell, chosen either among all AD astrocytes outside its cluster (see Methods and Fig. 3e), or among all astrocytes from the preceding measured time point (Fig. 3f and see Methods). Calculating for each cluster the frequency of origin cells within all other clusters revealed that the Gfap-low cells are the main origins for the cells within cluster 3, and that cells in clusters 3 and 5 are the main origins for the DAAs between consecutive time points (cluster 4; Fig. 3g). This is consistent with a model suggesting cluster 3 as an intermediate stage, and Gfap-low astrocytes as the major source of AD-specific cell states. Notably, these results were robust with respect to the source of origin cells (WT versus AD), the measured time point and the algorithm used (Extended Data Fig. 8). Next, we extracted the top genes that are most strongly associated with each transition between clusters across time points, by correlating genes with the average movement from the assigned origin cells to the true positions. These top significant genes reveal a set of unique DAA markers (Fig. 3h), which is consistent with our differential expression analysis (Fig. 2a and Supplementary Table 4), and define a set of genes that are activated or repressed during the transition from one state to the other.

Finally, we searched for evidence of DAA or DAA-like cells in aged brains of both WT mice and humans (Fig. 3c,d,i). In WT mice, DAA-like cells began to emerge at 13–14 months and increased in abundance in 20-month-old mice (up to 2%; Fig. 3d), suggesting that a similar phenotypic state switch also occurs in normal aging, in line with published results¹⁹. Moreover, comparing mouse astrocyte states with aging human astrocytes⁴ (see Methods), we identified astrocyte populations similar to the three mouse states in aging postmortem human brains, including DAA-like cells that appeared at a higher frequency in individuals with AD (Fig. 3i and Extended Data Fig. 9), suggesting that our findings could be relevant to human physiology and disease.

Fig. 3 | DAAs are derived from homeostatic Gfap-low astrocytes and increase with age. **a**, sNuc-seq time-course experiment in WT and 5xFAD mice across six age groups ($n=28$ mice, detailed in Supplementary Table 1). **b**, Relative increase of microglial frequency with age, in AD compared with WT. Log ratio frequency of microglial cells in AD versus WT, by age, in animal pairs matched by batch and age. **c**, Continuous trajectory across three major astrocyte states in AD and WT brains across ages. Embedding of 23,863 astrocytes in diffusion map (as in Fig. 1d), highlighting cells per age and condition (top, WT; bottom, AD), colored by inferred cluster identity (from Fig. 1d, consistent with independent clustering and diffusion map embedding of time-course data in Extended Data Fig. 7). Bottom right: schematic of astrocyte states and predicted transitions. **d**, An increase in the frequency of DAAs and a decrease in homeostatic Gfap-low astrocytes with age. Fraction of Gfap-low, DAA and Gfap-high cells out of all astrocytes across ages, in AD (red) and WT (blue). Lines show linear regression, with confidence intervals. R and P values of the linear fit. Inset: fraction of DAAs in WT ($n=23,863$ cells across ages; $n=28$ mice as detailed in Supplementary Table 1). **e,f**, DAAs are predicted to mainly derive from the population of Gfap-low astrocytes. Diffusion map (as in c) with directions of transition (directed arrows; see Methods) of AD astrocytes ($n=12$ animals, 10,226 cells) from their optimal nearest neighbor (cell of origin) among all other AD astrocytes outside the cluster (e) or among all AD astrocytes from preceding (measured) time points (f). Additional clusters and mappings are in Extended Data Fig. 8. **g**, DAAs are predicted to emerge from Gfap-low astrocytes according to the proportion of cells of origin per cluster. Top: for each pair of consecutive time points, the proportion (color bar) of astrocytes (rows) predicted to be derived from cells in each of the clusters in the preceding time point (columns). Bottom: graph connecting (arrows) pairs of clusters (nodes) between consecutive time points. Edges between clusters are only shown when the proportion of cells of origin is at least 14%. Arrow color, proportion. **h**, Genes correlated with transitions to DAAs across clusters and ages. Top: expression across clusters of genes significantly correlated ($n=12$ animals; Pearson correlation coefficient, $FDR < 10^{-3}$) with the transition to cluster 4 (DAAs), from each cluster in a preceding time point. Dot color, expression level; dot radius, proportion of cells expressing the gene. Bottom: assignment of each gene (column) to the transition from a cluster (row) and time point (color bar): upregulated, purple; downregulated, orange. **i**, DAA-like cells found in aging human cortex. Right: t-SNE of 3,392 sNuc-seq profiles of cortical postmortem human astrocytes from aging brains of healthy individuals and people with AD, taken from Mathys et al.⁴, colored by de novo cluster ID. Left, proportions (color bar, scaled per column) of human astrocyte cluster IDs (rows) mapped to mouse astrocyte cluster IDs (columns). Bottom, proportion of human astrocytes from healthy individuals and people with AD per the predicted mouse astrocyte cluster ID ($n=48$ individuals). Box, 75% and 25% quantiles. Line, median. Dots, individuals.

Overall, using Nuc-seq profiles from 34 WT and 5x FAD mice across ages, we identified a disease-specific state in astrocytes (DAA) that appeared early and increased with disease progression. The DAA-like population also appeared with aging in WT mice and was found in

aging human brains. Consistent with previous reports regarding activation of astrocytes by amyloid plaques^{7,8}, astrocytes expressing DAA markers were found adjacent to amyloid plaques in the hippocampus and in the subiculum, where disease manifestations



are severe¹². The continuous expression spectrum between DAAs and the intermediate cluster 3 astrocytes suggests a dynamic activation process in AD. Moreover, the wide range of activities found in DAAs might reflect changes occurring along disease progression. These dynamics might begin with gliosis as an attempt to contain the damage by demarcating the accumulated misfolded proteins from the still-healthy neurons, and become destructive along disease progression, due to expression of an inflammatory and neurotoxic profile, including SerpinA3N, which might interfere with plaque degradation¹⁵. Notably, the dynamic astrocyte response probably involves environmental factors²⁰ and a cross-talk between various cells present in the microenvironment, including microglia^{5,20}. Intriguingly, our analysis highlighted a shared signature of multiple genes upregulated in both DAAs and DAMs¹ (relative to their respective homeostatic states), suggesting a general transcriptional program in response to the pathological state, shared across cell types under disease conditions. Taken together, the extent and nature of astrocyte alterations that we found, and the early stage at which these changes emerge, support their role, at least in part, at the initial stages of disease pathogenesis. Further studies are required to fully understand their role and to determine whether this astrocyte state is universal or amyloid-associated. Deciphering the different activities of DAAs, along disease progression, may suggest a novel therapeutic target, enhancing their beneficial effects, while dampening the negative properties of these cells, with the potential for disease modification.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41593-020-0624-8>.

Received: 24 April 2019; Accepted: 18 March 2020;

Published online: 27 April 2020

References

1. Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**, 1276–1290.e17 (2017).
2. Strooper, B. & Karan, E. The cellular phase of Alzheimer's disease. *Cell* **164**, 603–615 (2016).
3. Baruch, K. et al. PD-1 immune checkpoint blockade reduces pathology and improves memory in mouse models of Alzheimer's disease. *Nat. Med.* **22**, 135–137 (2016).
4. Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
5. Liddelow, S. A. et al. Neurotoxic reactive astrocytes are induced by activated microglia. *Nature* **541**, 481–487 (2017).
6. Zamanian, J. L. et al. Genomic analysis of reactive astrogliosis. *J. Neurosci.* **32**, 6391–6410 (2012).
7. Matias, I., Morgado, J. & Gomes, F. Astrocyte heterogeneity: impact to brain aging and disease. *Front. Aging Neurosci.* **11**, 59 (2019).
8. Rodríguez-Arellano, J., Parpura, V., Zorec, R. & Verkhratsky, A. Astrocytes in physiological aging and Alzheimer's disease. *Neuroscience* **323**, 170–182 (2016).
9. Bates, K., Fonte, J., Robertson, T., Martins, R. & Harvey, A. Chronic gliosis triggers Alzheimer's disease-like processing of amyloid precursor protein. *Neuroscience* **113**, 785–796 (2002).
10. Liao, M.-C. et al. Single-cell detection of secreted A β and sAPP α from human iPSC-derived neurons and astrocytes. *J. Neurosci.* **36**, 1730–1746 (2016).
11. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
12. Oakley, H. et al. Intraneuronal β -amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *J. Neurosci.* **26**, 10129–10140 (2006).
13. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014.e22 (2018).
14. Leitão, R. et al. Aquaporin-4 as a new target against methamphetamine-induced brain alterations: focus on the neurogliovascular unit and motivational behavior. *Mol. Neurobiol.* **55**, 2056–2069 (2018).
15. Mucke, L. et al. Astroglial expression of human $\alpha 1$ -antichymotrypsin enhances Alzheimer-like pathology in amyloid protein precursor transgenic mice. *Am. J. Pathol.* **157**, 2003–2010 (2000).
16. Wang, C., Sun, B., Zhou, Y., Grubb, A. & Gan, L. Cathepsin B degrades amyloid- β in mice expressing wild-type human amyloid precursor protein. *J. Biol. Chem.* **287**, 39834–39841 (2012).
17. Roussotte, F. F. et al. Combined effects of Alzheimer risk variants in the CLU and ApoE genes on ventricular expansion patterns in the elderly. *J. Neurosci.* **34**, 6537–6545 (2014).
18. von und Halbach, O. Immunohistological markers for proliferative events, gliogenesis, and neurogenesis within the adult hippocampus. *Cell Tissue Res.* **345**, 1–19 (2011).
19. Clarke, L. E. et al. Normal aging induces A1-like astrocyte reactivity. *Proc. Natl Acad. Sci. USA* **115**, E1896–E1905 (2018).
20. Rothhammer, V. et al. Microglial control of astrocytes in response to microbial metabolites. *Nature* **557**, 724–728 (2018).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Experimental design. Heterozygous 5xFAD transgenic mice¹² (Tg6799; on a C57BL6-SJL background) co-overexpress mutant forms of human amyloid precursor protein associated with familial AD, the Swedish mutation (K670N/M671L), the Florida mutation (I716V) and the London mutation (V717I). The study included sNuc-seq profiles of samples from WT and 5xFAD male mouse hippocampi across six different age groups: 1.5–2 ($n=6$ mice), 4–5 ($n=4$), 7–8 ($n=10$, 12 samples), 10 ($n=2$), 13–14 ($n=6$) and 20 ($n=2$, WT only) months, as well as 5xFAD and WT 7-month-old female mouse hippocampi ($n=2$), and WT and 5xFAD male mouse prefrontal cortex (7 months, $n=2$; 10 months, $n=2$; same mice used for profiling the hippocampus) (see Supplementary Table 1). Throughout the study, WT controls in each experiment were nontransgenic littermates from the same mouse colony. Additional 5xFAD and WT 7-month-old mice were used for immunohistochemistry validations.

Animal care and tissue dissection. Animals were bred and maintained by the Animal Breeding Center of the Weizmann Institute of Science. Animal handling complied with the regulations formulated by the Institutional Animal Care and Use Committee of the Weizmann Institute of Science. Mice were anaesthetized by intraperitoneal injection of ketamine (100 mg kg⁻¹) and xylazine (10 mg kg⁻¹), and tissue dissection was performed immediately. Tissue was frozen after dissociation and kept at -80°C until further processing.

Nucleus isolation and sNuc-seq library preparation. Working on ice throughout, hippocampus tissue was transferred into a Dounce homogenizer (Sigma Cat No: D8938) with 2 ml of EZ Lysis Buffer (as in ref.¹¹, Sigma-Aldrich: NUC101-1KT) or NP40 Lysis Buffer (as in ref.²¹, 0.1% NP40, 10 mM Tris, 146 mM NaCl, 1 mM CaCl₂, 21 mM MgCl₂, 40 U ml⁻¹ RNase inhibitor). Tissue was carefully Dounced while on ice 22 times with pestle A followed by 22 times with pestle B, then transferred to a 15-ml conical tube. Next, 3 ml of lysis buffer was added to the Dounce homogenizer, to rinse residual nuclei, and the suspension was transferred to a 15-ml tube at a final volume of 5 ml. When using EZ, the lysis homogenate was incubated on ice for 5 min; for NP40 buffer no incubation was done. The samples were centrifuged with a swinging bucket rotor at 500g for 5 min at 4°C . Supernatant was removed and the pellet was resuspended in 100 μl of ice-cold PBS (for EZ lysis) or Tris buffer (for NP40 lysis, 10 mM Tris, 146 mM NaCl, 1 mM CaCl₂, 21 mM MgCl₂), plus 0.04% BSA (NEB B9000S) and 40 U ml⁻¹ RNase inhibitor. Then, 40- μm FlowMi cell strainers were pre-wetted with 200 μl of ice-cold PBS and the resuspended nuclei were gently filtered through the FlowMi cell strainers into 1.5-ml Eppendorf tubes. Nuclei were counted using Nextelom Cellometer Vision and a DAPI stain. DAPI was diluted to 2.5 μg μl^{-1} in PBS and 20 μl of the DAPI was pipette-mixed with 20 μl of the filtered nucleus suspension, then 20 μl of the stained nuclei were pipetted into the Cellometer cell-counting chamber (Nextelom CHT4-SD100-002). Nuclei were counted using a custom program with dilution factor set to 2. Finally, 10,000 nuclei were used as input to 10X Genomics single-cell 3' Gene Expression v2 assay (22 samples) or v3 assay (10 samples) (Supplementary Table 1).

Libraries were prepared following the manufacturer's protocol. Briefly, single nuclei were partitioned into Gel Beads in Emulsion in the GemCode/Chromium instrument with cell lysis and barcoded reverse transcription of RNA, followed by amplification, shearing and 5' adaptor and sample index attachment. Complementary DNA was amplified for 12 cycles, and the resulting whole-transcriptome amplification measured by Qubit HS DNA (Thermo Fisher Scientific: Q32851) and quality assessed by BioAnalyzer (Agilent: 5067-4626). The whole-transcriptome amplification material was diluted to <8 ng μl^{-1} and processed through v.2 or v.3 library construction, according the manufacturer's protocol. The resulting libraries were quantified again by Qubit and BioAnalyzer. Libraries were pooled and sequenced on two lanes of Illumina HiSeqX by the Broad Institute's Genomics Platform (22 samples) or on two lanes of NextSeq 500 by the Hebrew University Genomics Platform (10 samples).

Preprocessing of sNuc-seq data. De-multiplexing, alignment to the mm10 transcriptome and unique molecular identifier (UMI)-collapsing were performed using the Cellranger toolkit (v.2.1.1, chemistry V2, or v.3.0.2, chemistry V3) provided by 10X Genomics (with the Single Cell 3' chemistry parameter), and run using cloud computing on the Terra platform (<https://Terra.bio>). Since nuclear RNA is expected to include roughly equal proportions of intronic and exonic reads, we built and aligned reads to genome references with pre-messenger RNA annotations, which account for both exons and introns. For every nucleus, we quantified the number of genes for which at least one read was mapped, and then excluded all nuclei with fewer than 400 detected genes. Genes that were detected in fewer than ten nuclei were excluded. Expression values E_{ij} for gene i in cell j were calculated by dividing UMI counts for gene i by the sum of the UMI counts in nucleus j , to normalize for differences in coverage, and then multiplying by 10,000 to create TP10K (transcript per 10,000) values, and finally computing $\log_2(\text{TP10K} + 1)$ (using the NormalizeData function from the Seurat²² package v.2.3.4).

Identifying variable genes. Selection of variable genes was performed as previously described by Haber et al.²³. Briefly, we fit a logistic regression to the nuclear detection fraction, using the total number of UMIs per nucleus as a

predictor. Outlier genes expressed in a lower fraction of nuclei than expected were chosen as the variable genes, using a threshold of P value of 0.05. To minimize batch effect, variable genes were found for each batch independently, and their intersection was then used as the variable genes for downstream analysis. We restricted the expression matrix to this subset of variable genes, which was used for batch correction and scaling. For the integrated analysis of the time-course data, brain regions and sex comparisons, we used the integrated workflow of Seurat²² v.3, including finding variable genes by using the function FindVariableFeatures (setting the selection method to vst and the number of features to 2,000).

Batch correction and scaling data matrix. For the 7-month-old mouse dataset, batch correction was done using ComBat (ComBat²⁴ function from the sva package in R) on the normalized dataset. The batch-corrected data were scaled using the ScaleData function from Seurat²² with default parameters (v.2.3.4), yielding the relative expression of each gene by scaling and centering. The scaled data matrix was then used for dimensionality reduction and clustering. To rule out the possibility that the resulting clusters were driven by batch or other technical effects, we examined the distribution of samples within each cluster and the distribution of the number of genes detected across clusters (as a measure of nucleus quality). Overall, the nuclei separated into distinct point clouds in t -distributed stochastic neighbor embedding (t -SNE) space that were not driven by batch; each cluster/cloud was an admixture of nuclei from all technical and biological replicates, with variable numbers of genes. Related to the number of genes, we note that there was a distinct biological difference in the number of transcripts (and expected RNA content) between neuronal and glial cells in the brain.

Dimensionality reduction, clustering and visualization. We used the scaled expression matrix restricted to the variable genes for principal component analysis (PCA), using the RunPCA method in Seurat (a wrapper for the irlba function), computing the top 50 principal components. After PCA, significant principal components were identified using the elbow method, plotting the distribution of standard deviation of each principal component (PCElbowPlot function in ref.²²), choosing 25 principal components for analysis of all cells and 15 principal components for astrocytes (across all analyses, 7-month-old mice, time-course astrocytes, female/male, brain regions, human data). Scores from only significant principal components were used as the input to downstream clustering, visualization by t -SNE or Uniform Manifold Approximation and Projection (UMAP)²⁵, and diffusion map modeling.

Clustering the nuclei into transcriptionally similar clusters was done using a k -nearest neighbor (k -NN) graph, followed by the Louvain community detection algorithm²⁶, using the top significant principal components as input to a graph-based clustering algorithm, as previously described¹¹, with small modifications and improvements. Briefly, in the first step, we computed a k -NN graph, connected each nucleus to its k -NNs (based on Euclidean distance), then used the k -NN graph as an input to the Louvain algorithm, which decomposes an input graph into communities. We used $k=40$ for clustering of each full dataset, with varying resolutions from 0.9 to 1.2. For the full dataset across all cell types, at the final stage, cells were hierarchically clustered and re-ordered (using BuildClusterTree method from the Seurat package, given the same set of variable genes as input to the PCA), providing the cluster ordering described throughout this study.

For visualization, the dimensionality of the datasets was further reduced by t -SNE, using the same top principal components as input to the algorithm (using the runTSNE method in the Seurat package, which runs the Barnes-Hut approximation of t -SNE implemented in the Rtsne function), or the UMAP²⁵ dimensionality reduction technique (using the RunUMAP method which runs the umap-learn Python package). Cell populations were matched to cell types based on the expression of known marker genes and previously identified expression signatures^{11,27}. Recently activated neuronal populations were identified based on expression of immediate early genes, such as Egr4.

Subclustering of astrocytes and diffusion map embedding. To identify subtypes of cells within the astrocyte populations, the same analyses described in the “Dimensionality reduction, clustering and visualization” section were performed, but limited to nuclei classified by their cluster identity as astrocytes. Astrocytes were identified by marker gene expression, including high expression of Slc1a3, Gfap, Aldoc and Glul, and low/no expression of Rarres2 and Slc6a13. As shown in Extended Data Fig. 2, the expression of Slc1a3, Gfap, Rarres2 and Slc6a13 is sufficient to uniquely identify the clusters of astrocyte cells (excluding clusters of doublet cells). Specifically, for the 7-month hippocampus male mouse dataset, astrocyte clusters 1 and 2 from Fig. 1b were combined and used in all of the downstream analysis.

To model continuous transitions in astrocyte states, we calculated the diffusion components for all astrocytes (applied on the 7-month-old dataset or on the combined time-course datasets). The diffusion components were calculated using the cell embedding values in the top 15 principal components (generated either on the scaled expression matrix restricted to the variable genes in the 7-month-old mouse dataset or on the aligned canonical correlation analysis (CCA) subspace for the entire time-course data), using the DiffusionMap function from the destiny²⁸ package in R (with $k=30$ and a local sigma). We then chose the top two diffusion components for data visualization (other combinations of components revealed similar continuous trajectories, as shown in Extended Data Fig. 3d).

Doublet detection and removal. For doublet detection and elimination, we used the Scrublet²⁹ method over a random subset of 30,000 nuclei to identify clusters of doublets. Scrublet identifies heterotypic doublets comprising two different cell types by simulating doublets and building a nearest neighbor classifier. Scrublet was run with the following parameters: $k=30$ for the k -NN graph, doublet simulation rate of 0.3 and expected doublet rate of 0.1. Nuclei receiving Scrublet scores over 0.25 were classified as doublets. We then clustered our data at high resolution (resolution parameter = 2.5), generating 39 clusters, and found multiple small clusters that had over 70% of nuclei classified as doublets, which we then excluded from downstream analysis. We validated that these clusters were ‘doublet clusters’ based on a traditional inspection of expression patterns of cell type marker genes, showing that they coexpressed markers of at least two different cell types (the main markers we use in these inspections are shown in Extended Data Fig. 2). Notably, some clusters, such as the fibroblast cells, could have been misclassified as a doublet cluster based on the marker gene approach alone, since they express markers of several other cell types, demonstrating the importance of a computational approach that does not depend entirely on previous knowledge. In the downstream analysis of astrocytes, a second inspection for doublets was done using the traditional inspection of expression patterns of cell type marker genes. Specifically, this was done in the cortex and the female datasets before integration, removing small clusters highly expressing neuronal cell markers (such as Meg3).

Cell fraction estimations. The change in fraction of the different cell populations (clusters) was separately computed for each mouse/sample across all clusters, as the fraction of nuclei in each cluster, out of the total number of nuclei. To assess statistically significant changes in a fraction of a specific population, we first performed the Shapiro–Wilk test for normality (using the shapiro.test function in R), followed by paired *t*-test (for normally distributed populations) and an additional paired nonparametric Wilcoxon test (using t.test and wilcox.test functions in R). Results with $P < 0.01$ were reported as significant.

Differential expression, pathway analysis and scoring gene signatures.

Differentially expressed signatures were calculated using a negative binomial test and controlled false-discovery rates (FDRs) using the Benjamini–Hochberg procedure, to find genes that are upregulated within each cluster compared with the rest of the nuclei in the dataset or between pairs of clusters, including genes with less than 1% FDR. Genes were required to be expressed in at least 10% of nuclei in the given cluster, and at least 0.3-fold less in all other cells.

Gene set enrichment. The differential expression signatures were tested for enriched pathways and gene sets, using a hypergeometric test (function *enrichment* in the bc3net package in R), and corrected for multiple hypotheses by FDR. Results with $FDR < 0.05$ were reported as significantly enriched pathways. Gene sets and pathways were taken from the MSigDB/GSEA resource (combining data from Hallmark pathways, KEGG, Reactome, Pathway Interaction Database (PID), Canonical Pathways and Gene Ontology (GO) biological processes)³⁰. Gene signatures of reactive astrocytes were obtained from Zamani et al.⁶ and Liddelow et al.⁵.

Scoring of gene signatures across cells was done by following a procedure adapted from Smillie et al.³¹. Briefly, for every cell the log-transformed geometric average of expression values of all genes in the signature was computed over the scaled log-normalized data matrix of UMI counts. To correct for highly expressed genes, the log-transformed geometric mean of a random set of genes with similar expression levels was then subtracted from the expression average and used as the signature score.

To compare intermediate (cluster 3), Gfap-high and DAA states, we identified genes upregulated in each of the three states versus the Gfap-low homeostatic astrocytes (cluster 1) and performed pathway analysis as described as described in the “Gene set enrichment” section. We then combined the FDR Q values across the union of all enriched pathways, using a threshold of $FDR < 0.05$. Enriched pathways with higher FDR values were marked as nonsignificant.

DAM. The list of DAM-upregulated genes compared with the homeostatic microglia was taken from the literature (supplementary table 2 in Keren-Shaul et al.), and the DAM signature was defined as all genes with fold change greater than 1.2 and $-\log FDR > 3.3$, resulting in 213 genes. Genes were filtered to include only genes found to be detected in the sNuc-seq data (retaining a total of 152 genes). All cells in our dataset from WT and AD mice were scored for the DAM signature (as described for scoring gene signatures in the “Gene set enrichment” section), and the distribution of scores was compared. To identify overlaps in the DAA and DAM signatures, we used the differentially expressed genes in DAA (cluster 4) compared with the homeostatic state (cluster 1, following the procedure described in the “Differential expression” section). Display of gene–gene functional links within the shared DAM–DAA signature genes was done using the STRING database³², setting edges to reflect the confidence of the prediction, and based on the following categories: coexpression, databases, text mining and experiments.

Comparison across brain regions and sex by CCA integrated analysis.

Comparison of astrocyte profiles across two groups (that is, sex, in Extended Data Fig. 4a–d, or brain regions, in Extended Data Fig. 4e–i) was done by integrated

clustering of cells from the two groups (that is, males and females or hippocampus and cortex samples), and annotating the de novo clusters by projecting the astrocyte clusters defined for the hippocampus of 7-month-old male mice (from Fig. 1d). These annotated clusters were used to identify DAAs, homeostatic Gfap-low and Gfap-high astrocytes in female mice and in the cortex brain region, and to compare the relative abundances of astrocyte states between AD and WT mice across sex and brain regions. This was done by the following steps: we first identified the astrocyte cells in each of the additional datasets: sNuc-seq profiles of cortex brain region or of female hippocampus region (Supplementary Table 1) were analyzed following the procedures for normalization, scaling and clustering described in the section “Dimensionality reduction, clustering and visualization” (using Seurat v.3). Astrocyte clusters were identified by marker gene expression (as described in the section “Subclustering of astrocytes”) and selected for further analysis.

To compare the astrocyte cell states across regions and sex, we combined the astrocyte cells (from female mice or male cortex) with sNuc-seq profiles from four 7-month-old mouse hippocampi (Supplementary Table 1). Datasets were combined and batch corrected by anchor-based joint CCA^{22,33}. Using Seurat v.3, we first normalized each dataset separately (log-transform) and found variable genes (using the function FindVariableFeatures, and setting the parameters: selection.method to vst and nfeatures to 2,000). We then searched for integration anchors, which are pairwise correspondences between individual cells, with the underlying assumption that cells originating from the same biological state will be matched (using the function FindIntegrationAnchors, over the first 20 dimensions). We integrated the data using this list of anchors (using the function IntegrateData). Following the data integration, the integrated expression matrix was analyzed by repeating the standard analysis steps on the integrated expression matrix: data scaling, PCA, clustering and two-dimensional t-SNE for visualization. The top 15 principal components were used for the clustering and t-SNE.

To annotate the clusters, we used a similar analysis, based on the CCA^{22,33} approach (using Seurat v.3), to project the cluster identities (IDs) from the 7-month-old mouse dataset (IDs as in Fig. 1d) to the integrated set. We set the 7-month-old male mouse clusters as a reference, and predicted the cluster identity of all other cells (of different sex or brain regions), using the functions FindTransferAnchors and TransferData (in Seurat v.3). Each cluster in the integrated dataset was matched to one (or two) of the astrocyte states (homeostatic Gfap-low, Gfap-high, DAA state, intermediate-state DAA and Gfap-high intermediate state) by a majority vote within all cells in the cluster, while ambiguous results were not matched with any state. Finally, the fraction of astrocytes across clusters/states was calculated for each mouse, comparing the relative frequencies of the DAA, homeostatic Gfap-low and Gfap-high clusters in matching pairs of WT and 5xFAD mice.

The CCA integrated analysis was used to enable comparison of astrocyte states, while excluding technical batch effects (differences between 10X versions), sex-specific expression patterns or regional biases (such as regional specific ambient RNA). It is of note that not all cells in the comparison matched the original dataset of 7-month-old male mouse hippocampi. For example, the homeostatic astrocytes split to separate clusters between the cortex and hippocampus brain regions, while the DAAs and Gfap-high cells clustered together across brain regions. Thus, when using the CCA approach, we were able to identify distinct expression profiles in astrocytes (across brain regions), and also to identify shared expression profiles, such as DAAs in female 5xFAD mice and in the cortex of 5xFAD mice.

For comparison of expression profiles of clusters across regions or sex, we used the top marker genes, identified for the 7-month male hippocampus mice (as in Fig. 1e), of the three major astrocyte states, DAAs, homeostatic Gfap-low and Gfap-high.

Immunohistochemistry. Immunohistochemistry was performed on mouse brain sections. Mice were perfused with PBS before brain tissue fixation, followed by tissue processing to form 30-μm-thick floating sections. The following primary antibodies were used: rabbit anti-GFAP (1:150; Dako no. Z0334, lot 20056262), chicken anti-VIM (1:150; Abcam no. 24525, lot GR3216660-15), goat anti-SerpinA3N (1:200; R&D Systems no. AF4709-SP, lot CBKW0318051) and mouse anti-Aβ 1–16 (1:150; Biologend no. 803001, lot B247600). Secondary antibodies were Cy2/Cy3/Cy5 donkey anti-rabbit/chicken/goat/mouse antibodies, respectively (1:150; Jackson ImmunoResearch). Staining with secondary antibody alone was used to rule out nonspecific staining. The following secondary antibodies were used: Cy2/Cy3/Cy5 donkey anti-rabbit/chicken/goat/mouse antibodies, respectively (1:150; Jackson ImmunoResearch; as reported in the Reporting Summary: 711-225-152, Cy2 donkey anti-rabbit, lot 142845; 715-225-150, Cy2-AffiniPure donkey anti-mouse, lot 142843; 705-225-147, Cy2-AffiniPure donkey anti-goat IgG, lot 122781; 711-165-152, Cy3 donkey anti-rabbit, lot 143202; 715-165-151, Cy3-AffiniPure donkey anti-mouse IgG, lot 143017; 705-165-147, Cy3 donkey anti-goat, lot 143201; 711-175-151, Cy5-AffiniPure donkey anti-mouse IgG, lot 144119; 711-175-152, Cy5-AffiniPure donkey anti-rabbit IgG, lot 144221; 705-175-147, Cy5-AffiniPure donkey anti-goat, lot 134531). For nuclear staining, Hoechst was used (1:4,000; Invitrogen Probes) for 30 s before sealing of the slides. Confocal microscopy was used for analysis (Zeiss, LSM880). Representative images were merged and optimized using ImageJ (Java v. 1.8.0_66).

Time-course dataset preprocessing and integration. The time-course dataset consisted of four datasets (or batches), a 7-month-old mice dataset (including four 5xFAD and WT mice from the dataset in Figs. 1 and 2 of samples prepared with EZ lysis buffer to match the conditions at the other time points) and three additional datasets from 5xFAD and WT mice ranging in age from 1.5 to 20-months (Supplementary Table 1). Each dataset in the time course was analyzed separately, repeating the analysis steps described for the 7-month-old mice. Clusters of astrocytes or microglia were selected from each batch for further analysis.

For microglia, the fraction of microglial cells was calculated for each sample out of the total number of cells, and compared across ages and conditions, matching pairs of littermate AD and WT mice that were processed in the same sequencing batch.

For astrocytes, data across all four batches were combined and batch corrected by anchor-based joint CCA^{2,3,5}, using Seurat v.3, as described in the section “Subclustering of astrocytes and diffusion map embedding”. Following the data integration, the integrated expression matrix was analyzed by repeating the standard analysis steps: data scaling, PCA, clustering and two-dimensional t-SNE embedding for visualization, as well as diffusion map modeling (using the top 15 principal components for the clustering, t-SNE and diffusion map steps). The diffusion map embedding of the time-course dataset aligned cells along two continuous trajectories, a Gfap-low to Gfap-high trajectory and a homeostatic-DAA trajectory (shown in Extended Data Fig. 7), which matched the trajectories found for the 7-month-old mice (Fig. 1d). Notably, for consistency and clarity in Fig. 3 we show the embedding of the time-course datasets in the diffusion map of the 7-month-old mice (as in Fig. 1d, which was done as described in the next section, “Dynamic modeling, k-NN embedding and gene expression dynamics”).

Dynamic modeling, k-NN embedding and gene expression dynamics. We devised a computational approach to computationally model dynamic transitions between astrocyte cell states along disease progression by assigning a unique ‘origin cell’ for each astrocyte in AD. This was done using the Hungarian algorithm³⁴ (munkres function in MATLAB) by globally optimizing the pairwise distances, in the PCA or CCA space, between all AD astrocyte cells and their matched neighbors, selected from all AD cells at the preceding time point or among all other AD cells outside its cluster. We predicted origin cells for both reference sets to capture the dynamics in astrocyte states along different stages of the disease as well as the continuous range of astrocytes within a given time point. The predictions of origin cells were used to: (1) map the trajectory of AD cells along disease stages in the diffusion space using force fields; (2) predict the source population of the DAAAs by quantifying the transitions between clusters; and (3) find putative genes underlying the transitions in astrocyte states. For assigning the ‘origin’ cells, we searched for the optimal assignment of AD astrocytes, by globally minimizing the sum of pairwise Euclidean distances in the PCA/CCA space between all pairs of cells and their assigned origin cell. For each AD time point (or AD cluster), we searched for the globally optimal assignment between all AD cells at a given time point (or a given cluster) and all other AD cells at a preceding time point (or all other AD cells outside the cluster), which we termed ‘origin’ cells.

We devised a computational approach to identify the genes most strongly associated with the transitions in astrocyte states. To this end, we first computed the average trajectory (in the PCA/CCA space) of cells from each AD cluster to every other AD cluster (between time points or within a given time point). We then correlated the average trajectory with the PCA/CCA loading values of each gene to identify genes that are up- or downregulated in the transition from one state to the other. Genes with significant Pearson correlation ($FDR < 1 \times 10^{-2}$) were marked as putative marker genes for each transition.

Importantly, diffusion map analysis of the time-course data aligned the astrocytes along trajectories similar to the ones identified in the 7-month dataset, in which astrocytes traverse a homeostatic-to-Gfap-high trajectory and a homeostatic-to-DAA trajectory (as shown in Extended Data Fig. 7). However, for visualization and annotation purposes, we embedded the time-course data into the diffusion map of the 7-month dataset (Fig. 1d) using a k-NN approach. To embed each cell in the 7-month diffusion space, we first identified the k nearest nuclei within the CCA space ($K=10$), and calculated their relative Euclidean distances, d_i . We then calculated the relative normalized weights for each of the k neighbors, as: $w_i \propto \frac{1}{d_i^2}$, such that $\sum w_i = 1$, and calculated the estimated position \hat{Y}_i of that nucleus within the 7-month diffusion space as a weighted average $\hat{Y}_i = \sum w_i X_i$ of the positions of the k neighboring nuclei (X_i). Similarly, we predicted the cluster IDs (as in Fig. 1d) of embedded cells using a weighted majority vote based on the cluster IDs of their k neighbors, using the same relative normalized weights as above.

Finally, we computed the relative trajectory of each nucleus from its ‘origin’ cell (or embedded position \hat{Y}_i) to its position Y_i in the diffusion space, and visualized this trajectory as a directed arrow using force fields (quiver function in MATLAB). To validate robustness, the same analysis was done comparing AD cells with all WT cells in the 7-month dataset across all cells, or with all WT and AD cells outside the cluster, per cluster.

Comparison with aging human astrocytes. The count matrix of sNuc-seq data from human astrocytes was obtained from Mathys et al.⁴, using the published cluster annotations to select all astrocyte cells. To compare the human and

the mouse astrocytes, mouse genes were first mapped to human genes, using annotations from the UCSC genome browser (for mouse mm10 genome). The human data were re-analyzed following all steps described for the mouse data, including normalization, variable gene detection, scaling, clustering and visualization (all using Seurat v.3). Finally, cluster identity was assigned to the human clusters using the CCA approach (Seurat v.3), using the data and cluster IDs of 7-month-old male mice as reference (as in Fig. 1d and Supplementary Table 1; following the steps detailed in the section “Comparison across brain regions and sex by CCA integrated analysis”). The cluster of human astrocytes with the highest fraction of cells mapped to the mouse DAA cluster was termed DAA-like. The fraction of astrocytes assigned to each of the mouse clusters (out of the total number of human astrocytes) was calculated for each individual, and split to show the distributions in healthy individuals and individuals with AD. For comparison of human and mouse expression signatures, we chose top marker genes from the mouse dataset that are expressed in human astrocytes. Notably, while a single human cluster showed the highest similarity to the mouse DAAAs, more than one human cluster could be associated with the mouse DAAAs by the predicted cluster IDs and by the expression of marker genes (Extended Data Fig. 9 and Fig. 3i), suggesting high diversity of astrocyte states in the human brain.

Reproducibility and blindness. Sample size, data exclusion and randomization: the main dataset consisted of four WT and four transgenic 5xFAD mice, based on published results by us and other groups, showing high consistency in sNuc-seq libraries between mice; due to technical variability it is recommended to have at least three animals per group. The time-course data had an additional 12 WT and 10 5xFAD animals across time points. Immunohistochemistry was repeated, for every set of antibodies, over $n=4$ AD and WT mice, with 4 brain slices per animal. No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to those reported in previous publications^{3,11,21,27,35}. Number of nuclei per animal was determined to enable detection of rare populations around 2% of all cells (calculated based on our previous data in the mouse brain^{11,21}). No samples or animals were excluded from the analysis. As commonly done, nucleus libraries were filtered using previously established methods: nuclei with fewer than 400 detected genes were excluded from the analysis. Animals were assigned randomly to the various experimental batches according to their age, while matching transgenic Alzheimer’s model 5xFAD animals with WT nontransgenic littermates from the same mouse colony and of the same age. Notably, the clustering analysis, and specifically the identification of the different astrocyte populations, was done blinded to the animal strain, sex, brain region and age.

Statistics. Several statistical tests were used throughout this work, as mentioned in each relevant section. These tests can be divided into the following tasks: (1) Finding statistically significant changes in fractions of cellular populations between WT and 5xFAD mice: we first performed the Shapiro-Wilk test for normality. On normally distributed cellular populations we used the two-sided paired *t*-test, and for nonnormally distributed samples we used the paired, nonparametric, two-sided Wilcoxon test (matching mice by experimental batch and littermate). In both cases, tests were applied to $n=8$ animals (10 samples), and cellular populations with $P < 0.01$ were reported. The number of cells per population varied (as shown in Fig. 1 and Extended Data Fig. 1). (2) Identifying differentially expressed genes between clusters: negative binomial test and controlled FDRs using the Benjamini-Hochberg procedure, between pairs of clusters. For Fig. 2a and Supplementary Table 2: $n=8$ animals, $FDR < 1\%$ and genes were required to be expressed in at least 10% of nuclei in the given cluster, and at least 0.3-fold less in all other cells, as routinely done. Data distribution was estimated to be negative binomial, although this assumption might not fit every gene. (3) Finding enriched pathways: done by calculating the hypergeometric *P* value and controlled FDRs using the Benjamini-Hochberg procedure (all gene sets used are reported in the section “Differential expression, pathway analysis and scoring gene signatures”). For Fig. 2b and Supplementary Table 3: $n=8$ animals and a threshold of $FDR < 5\%$. (4) Testing for consistent changes in cell frequency across ages we used linear regression, computing a confidence interval per time point. Calculated with: $n=12$ 5xFAD animals or $n=16$ WT animals across ages, for each astrocyte state (reporting the *R* and *P* value). (5) Identifying the genes associated with the transitions in astrocyte states: we used our computational approach described in section “Dynamic modeling, k-NN embedding and gene expression dynamics.” Briefly, we find genes that are statistically significantly (*P* value controlled for FDRs using the Benjamini-Hochberg procedure <0.001) correlated with the direction of average trajectory by computing the Pearson correlation coefficient. This was done using $n=12$ animals, and 10,226 cells from 5xFAD mice across ages.

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

Data availability

Raw and processed mouse sequencing data that support the findings of this study have been deposited in the Gene Expression Omnibus database under accession number GSE143758 and are also available at https://singlecell.broadinstitute.org/single_cell/study/SCP302/mouse-alzheimers-and-disease-astrocytes. Source data

for Fig. 1 are presented with the paper. Code is available at: <https://github.com/naomihabiblab/5xFAD-sNucSeq>.

References

21. Gaublomme, J. T. et al. Nuclei multiplexing with barcoded antibodies for single-nucleus genomics. *Nat. Commun.* **10**, 2907–2908 (2019).
22. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
23. Haber, L. A. et al. A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).
24. Johnson, E. W., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
25. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. Preprint available at *arXiv* <https://arxiv.org/abs/1802.03426> (2018).
26. Lancichinetti, A. & Fortunato, S. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Phys. Rev. E* **80**, 016118 (2009).
27. Habib, N. et al. Div-Seq: single-nucleus RNA-seq reveals dynamics of rare adult newborn neurons. *Science* **353**, 925–928 (2016).
28. Angerer, P. et al. destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**, 1241–1243 (2016).
29. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291.e9 (2019).
30. Liberzon, A. et al. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
31. Smillie, C. S. et al. Intra- and inter-cellular rewiring of the human colon during ulcerative colitis. *Cell* **178**, 714–730.e22 (2019).
32. Szklarczyk, D. et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
33. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
34. Munkres, J. Algorithms for the assignment and transportation problems. *J. Soc. Ind. Appl. Math.* **5**, 32–38 (1957).
35. Baruch, K. et al. Breaking immune tolerance by targeting Foxp3⁺ regulatory T cells mitigates Alzheimer’s disease pathology. *Nat. Commun.* **6**, 7967–12 (2015).

Acknowledgements

We thank R. Herbst, I. Avraham-David and members of the Habib laboratory for discussions and helpful suggestions, and L. Gaffney for graphics and J. Rood for editing. We thank I. Shif and M. Bronstein for help with library preparation and sequencing. This work was supported by the NIH BRAIN Initiative grant (grant no. U19MH114821 for A.R.), the Klarman Incubator and Klarman Cell Observatory (A.R.), the HHMI (A.R.) and a MOST-IL flagship grant (for N.H.). M.S. holds the Maurice and Ilse Katz Professorial Chair in Neuroimmunology, and N.H. holds the Goren-Khazzam Chair in Neurobiology. The work was supported in part by Advanced European Research Council grant no. 232835 (M.S.); Israel Science Foundation (ISF) research grants no. 991/16 (M.S.), no. 913/15 (T.K.) and no. 1250/18 (T.K.); ISF-Legacy Heritage Biomedical Science Partnership research grant no. 1354/15 (M.S.); the Adelis Foundation and the Thompson Foundation (M.S.); the MOST-IL-China research grant no. 3-15687 (N.H.); and the Alon Fellowship (N.H.).

Author contributions

N.H., M.S., A.R. and F.Z. conceived the study. M.S., N.H. and S.M. designed the experiments. C.M. and D.K. isolated nuclei and generated libraries with assistance from D.D. and L.N. S.M., J.L.M. and F.C. conducted the imaging experiments and S.M. analyzed the imaging data. S.M. and R.D.-S. conducted the animal work. N.H., A.R. and T.K. devised the data analysis. N.H., M.V., G.G. and T.K. conducted the data analysis. T.K. and N.H. developed a new algorithmic approach. M.S., N.H., T.K. and A.R. wrote the paper with input from all of the authors.

Competing interests

A.R. is a founder and equity holder of Celsius Therapeutics, an equity holder in Immunitas Therapeutics and an SAB member of Syros Pharmaceuticals, Thermo Fisher Scientific, Neogene Therapeutics and Asimov. The other authors declare no competing interests.

Additional information

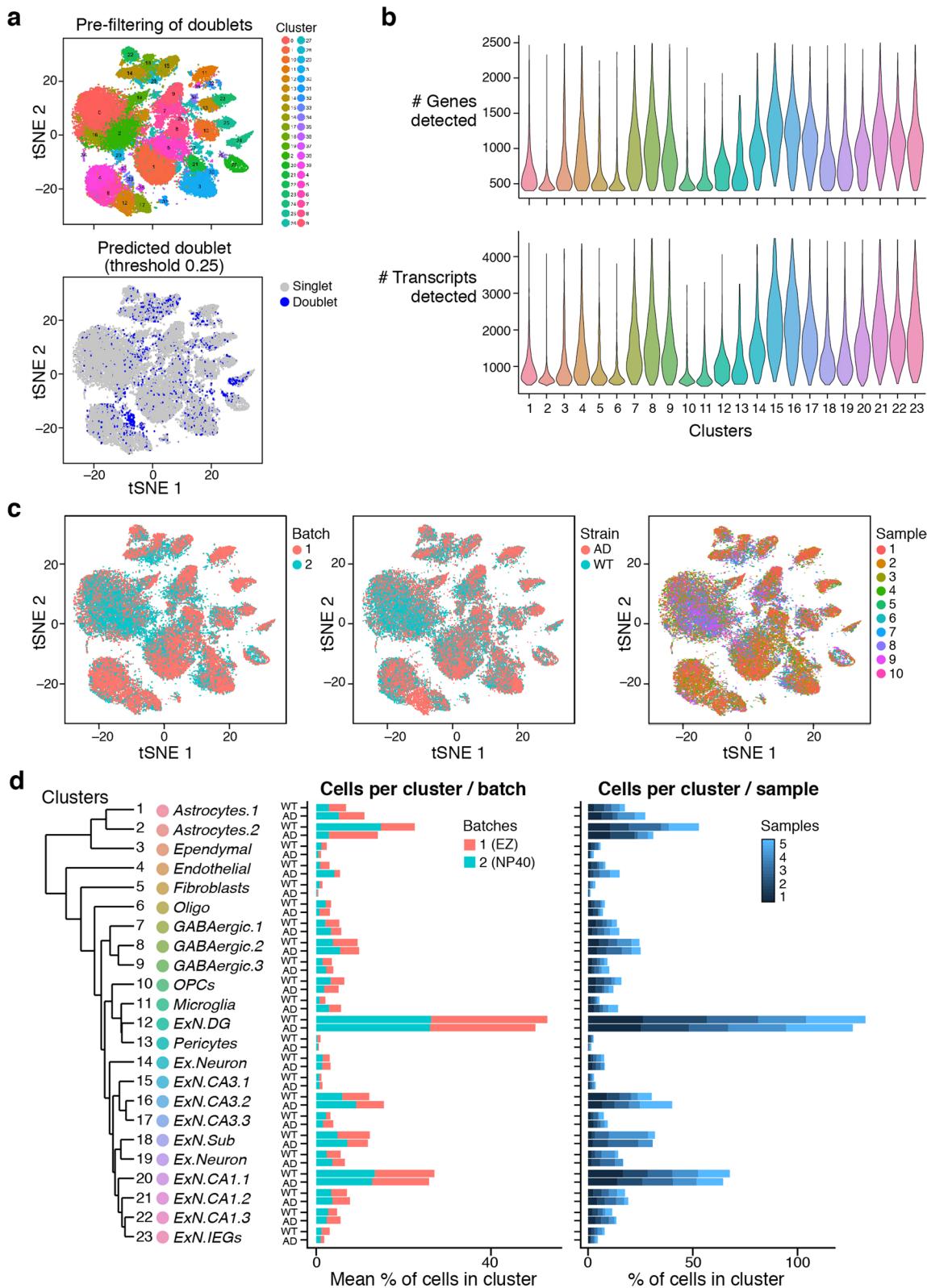
Extended data is available for this paper at <https://doi.org/10.1038/s41593-020-0624-8>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41593-020-0624-8>.

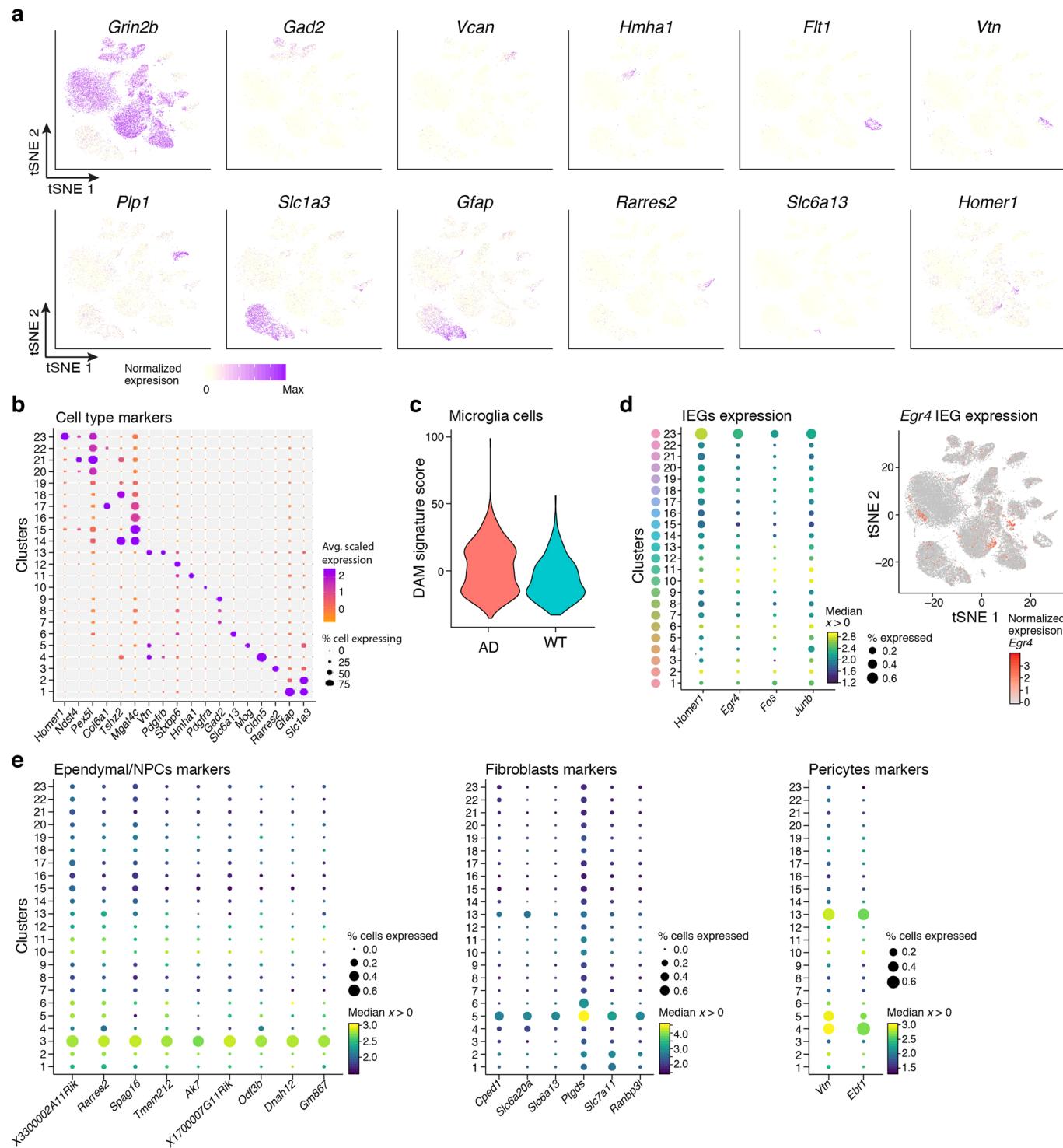
Correspondence and requests for materials should be addressed to N.H., A.R. or M.S.

Peer review information *Nature Neuroscience* thanks Valentina Fossati and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

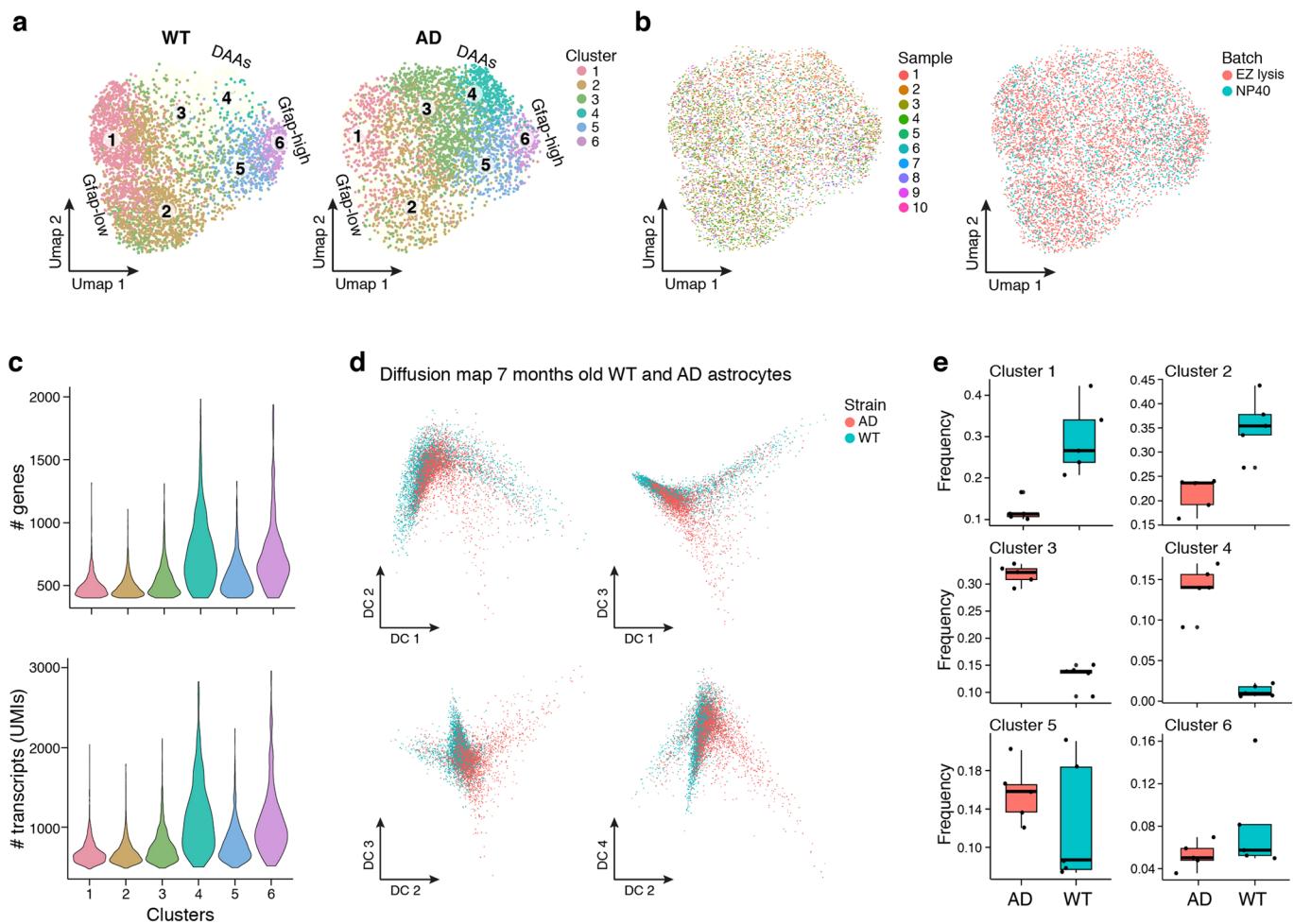
Reprints and permissions information is available at www.nature.com/reprints.



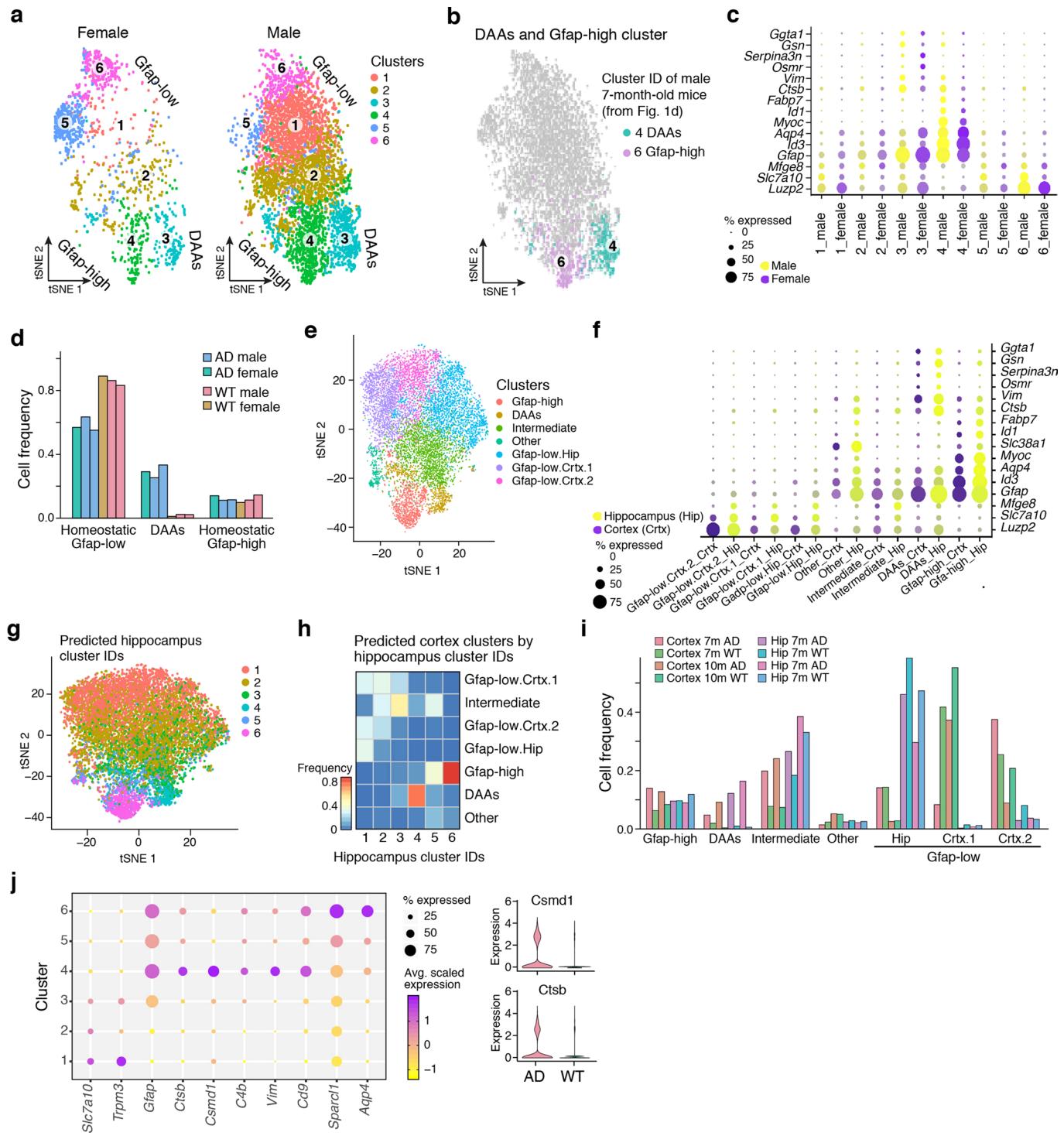
Extended Data Fig. 1 | A cellular map of the mouse hippocampus of WT and 5xFAD mice and quality controls. **a**, Doublet detection and elimination. 2-D tSNE embedding of 60,818 single nuclei RNA profiles from hippocampus of four WT and four 5xFAD 7-month old mice, before filtration. Top: Color coded by cluster assignment. Bottom: Color coded by doublet score assigned per cell by the Scrublet²⁹ software, used to infer doublet cells and clusters to exclude from the analysis. **b**, Number of genes and transcripts across clusters. Violin plots showing the distribution of number of genes (top) and transcripts (Unique UMIs, bottom) detected in each cluster ($n=8$ mice, 10 samples). Cluster numbers as in Fig. 1b. **c**, 2-D tSNE embedding of single nuclei RNA profiles from hippocampus of WT and 5xFAD mice (as in Fig. 1b), colored by (from left to right): batch, mouse strain (WT or AD) and sample. **d**, Similar distribution of samples and batches across clusters. The percent of cells per cluster, in WT and 5xFAD mice. Middle: Colored by batch/lysis buffer (red = EZ lysis. Blue = NP40 lysis, Methods). Right: Colored by sample (blue color scale, 4 animals and 5 samples per mouse strain, AD or WT). Left: The hierarchical cluster tree and annotations of clusters, as in Fig. 1b.



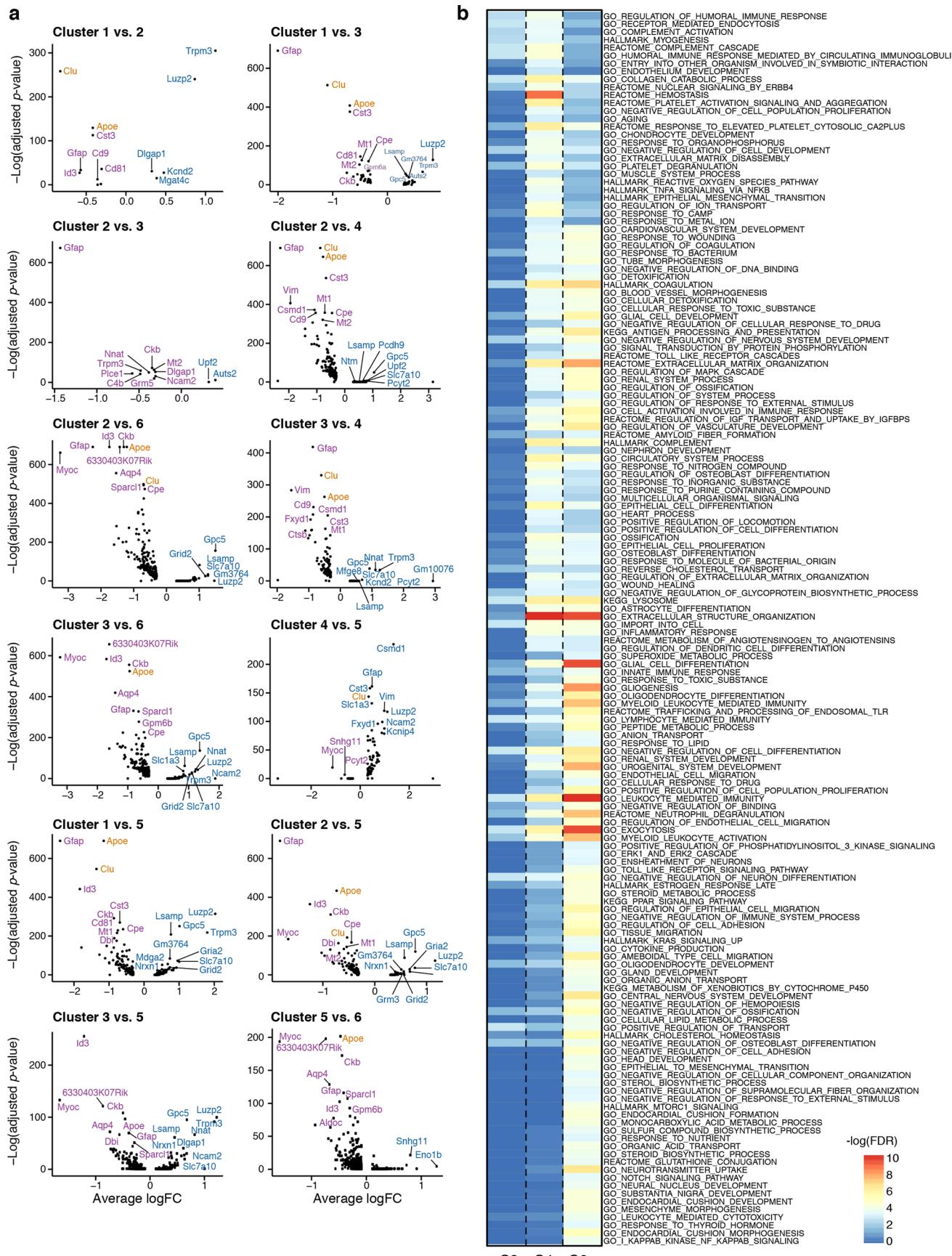
Extended Data Fig. 2 | Cell type marker genes and assignments. **a**, Expression of marker genes across clusters. 2-D tSNE embedding of single nuclei RNA profiles from hippocampi of WT and 5xFAD mice (as in Fig. 1b), colored by expression levels of marker genes: *Grin2b* (neurons), *Gad2* (GABAergic neurons), *Vcan* (Oligodendrocytes precursor cells, OPCs), *Hmha1* (microglia), *Flt1* (endothelial), *Vtn* (pericytes), *Plp1* (oligodendrocytes), *Slc1a3* (astrocytes), *Gfap* (astrocytes), *Rarres2* (ependymal/NPCs), *Slc6a13* (fibroblasts), *Homer1* (immediate early gene, IEGs). **b**, Clusters and marker genes. Dot plot showing the expression level (color scale) and the percent of cells expressing (dot size) marker genes across all clusters (rows). Cluster numbers as in Fig. 1b. **c**, Disease associated microglia (DAM) signature enriched in AD. Violin plots showing the distribution in WT (n=8 animals, 896 cells) and AD (n=8 animals, 1,540 cells) of microglia expression scores for signatures of genes up-regulated in DAM compared to homeostatic microglia (from Keren-Shaul et al.¹, Methods). Expression score per cell is the geometric mean normalized expression level (TPMs) across all signature genes, corrected by subtraction of the geometric mean expression of a random set of genes of similar expression levels (Methods). **d**, Recently activated pyramidal neurons. Left: Dot plot as in (b) showing the expression of immediate early genes (IEGs) across all clusters (as in Fig. 1b), showing cluster 23, capturing pyramidal neurons expressing IEGs. Right: tSNE plot of all cells, color coded by the expression level of the *Egr4* gene in CA3/CA1/Subiculum (cluster 23), and DG excitatory neurons (part of cluster 12). **e**, Cell type specific markers. Dot plots as in (b) showing the expression level, across all clusters (rows, as in Fig. 1b), of markers found to be specific to cells classified as (from left to right): ependymal/NPCs, fibroblasts and pericytes.



Extended Data Fig. 3 | Diversity of astrocyte states in WT and 5xFAD mice. **a**, 2-D Umap²⁵ embedding of 7,345 single nuclei RNA profiles of astrocytes (as in Fig. 1d) from hippocampus of 4 WT (left) and 4 5xFAD (AD, right) 7-month old mice. Colored by cluster, all other cells in light yellow in the background. **b**, 2-D Umap²⁵ embedding as in (a), colored by sample (left), or batch (right). **c**, Number of genes and transcripts (unique UMIs, bottom) detected per cluster (n=8 animals, total of 7,345 cells). Cluster numbers as in (a). **d**, Diffusion maps²⁸ of 7,345 single nuclei RNA profiles of astrocytes in the hippocampus of WT and 5xFAD mice, showing 2-D embedding of cells in combinations of the top four diffusion components (DC), colored by mouse strain, WT (blue) and 5xFAD (AD, red). **e**, Distribution of astrocyte states in WT and AD brains. Box plots showing the fraction of each astrocyte cluster (compared to total number of astrocyte cells, clusters as in Fig. 1d, n=8 animals, 10 samples), in WT and 5xFAD mice. Displaying the median (thick lines), 25% and 75% quantiles (box), and individual samples (dots).

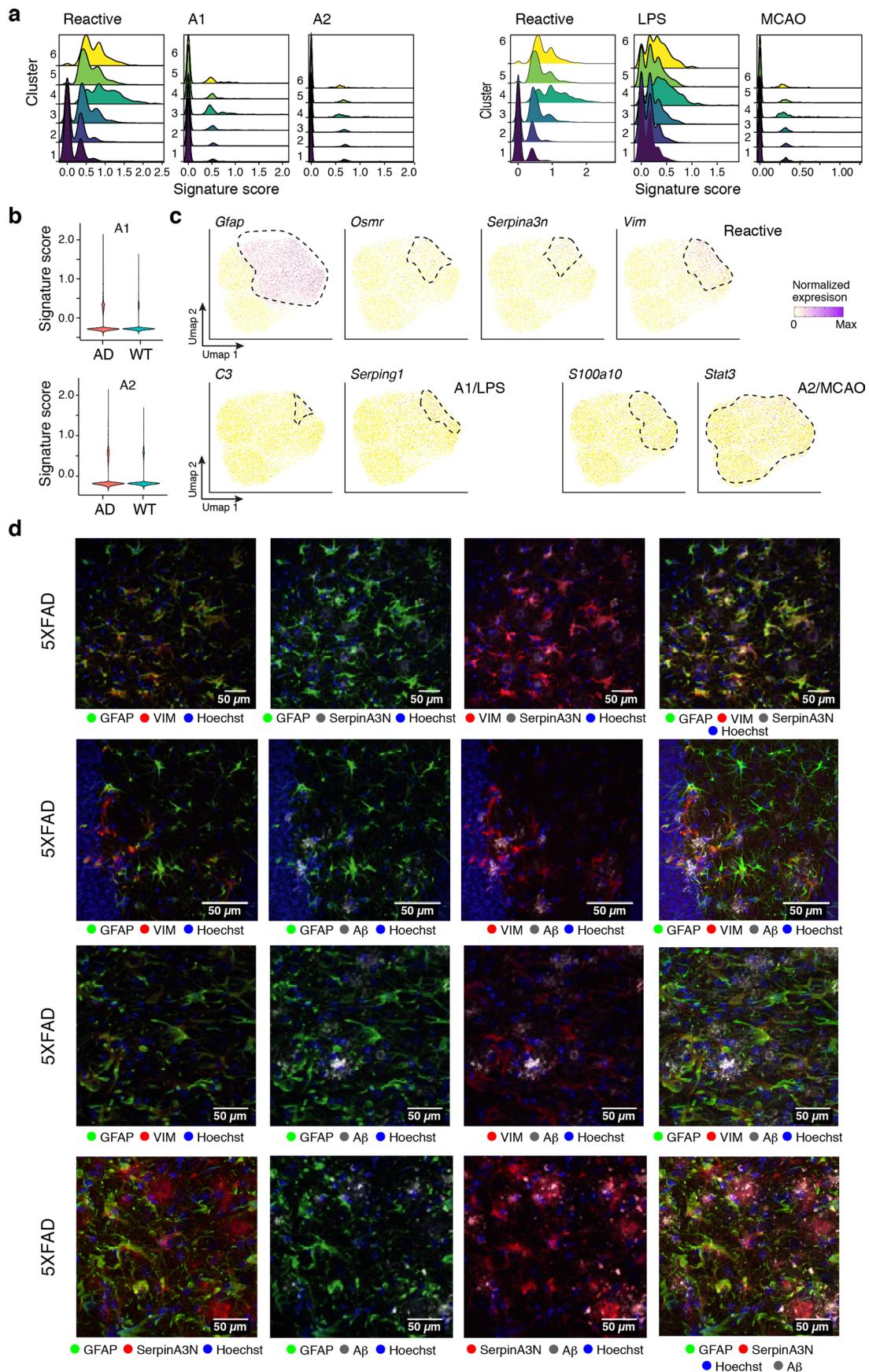


Extended Data Fig. 4 | DAs found in female 5xFAD mice and in the cortex. **a**, tSNE embedding of single nuclei RNA profiles of astrocytes from hippocampus of 7-month old female (1,500 nuclei, 2 mice, left) and male (5,183 nuclei, 4 mice, right) of three WT and three 5xFAD (AD) mice. Colored by cluster assignment. The three astrocyte end states are marked on the graph (annotated by projection from Fig. 1d, Methods). **b**, tSNE embedding as in (a), colored by cluster 4 (DAA) and cluster 6 (Gfap-low) cluster IDs of male mice as in Fig. 1d (other cells in grey). **c**, Expression levels (color scale) and the percent of cells expressing (dot size) of marker genes across clusters, split by sex: for DAA (*Ggt1*, *Gsn*, *Osmr*, *Vim*, *Serpina3n*, *Ctsb*, *Gfap*), Gfap-low (*Fabp7*, *Slc38a1*, *Myoc*, *Aqp4*, *Id1*, *Id3*, *Gfap*) and Gfap-low (*Mfge8*, *Slc7a10*, *Luzp2*). **d**, The proportion of astrocytes classified as Gfap-low (clusters 1,2,5,6 in (a)), DAs (cluster 3 in (a)) and Gfap-high (cluster 4 in (a)) in male and female mice. Bar: individual mice, colored by strain and sex. **e**, DAs in 5xFAD mice cortex at age 7 and 10 months. tSNE embedding of single nuclei RNA profiles of astrocytes from WT and AD mice, from the cortex of 7 and 10 month old mice (6,062 nuclei, 4 mice) and hippocampus of 7-month old mice (5,344 nuclei, 4 mice). Colored by cluster. **f**, Similar astrocytes marker genes in the hippocampus and cortex. Dot plot showing the expression level (as color scale) and the percent of cells expressing (as dot size) marker genes for DAs, Gfap-high and Gfap-low across clusters (as in (e)), split by brain region to hippocampus and cortex. **g**, tSNE embedding as in (e), colored by predicted cluster ID of hippocampal astrocytes from 7-month old male mice (inferred by CCA^{22,33} projections, as in Fig. 1d, Methods). **h**, Cortical astrocyte populations match astrocyte states identified in the hippocampus. Heat map showing the correspondence between the *de novo* cluster IDs (rows, from (e)) of the cross regional dataset, and the predicted cluster IDs (columns) using the hippocampal astrocytes cluster IDs as reference (as from Fig. 1d). Color scale based on the proportion of predicted IDs per *de novo* cross-regional cluster. **i**, DAs appear in the cortex of 5xFAD (AD) mice. The proportion of astrocytes, per sample, across clusters, including clusters of Gfap-low, DAs, and Gfap-high astrocytes. Bars: Individual mice, color annotated by region: cortex or hippocampus (Hip), age: 7 or 10 months (m), and strain: AD or WT. **j**, Astrocyte marker genes. Left: Average expression level (color scale) and the percent of cells expressing (dot size) marker genes for: Gfap-low (*Slc7a10*; *Trpm3*), DAA (*Ctsb*¹⁶; *Csmd1*, associated with cognitive functions; *C4b*, encoding complement factor 4; *Vim*, a marker of adult neurogenesis/NSCs¹⁶), common to DAA and Gfap-high (*Cd9*, expressed by neural stem cells like astrocytes), and Gfap-high (*Sparcl1/Hevin*, encoding astrocytes pro-synaptic protein; *Aqp4*, an endfeet marker¹⁴) (cells clusters as in Fig. 1e). Right: Violin plots (n=8 animals, 10 samples, 7,345 cells), showing the expression level distributions of *Csmd1* and *Ctsb* in WT and AD astrocytes.



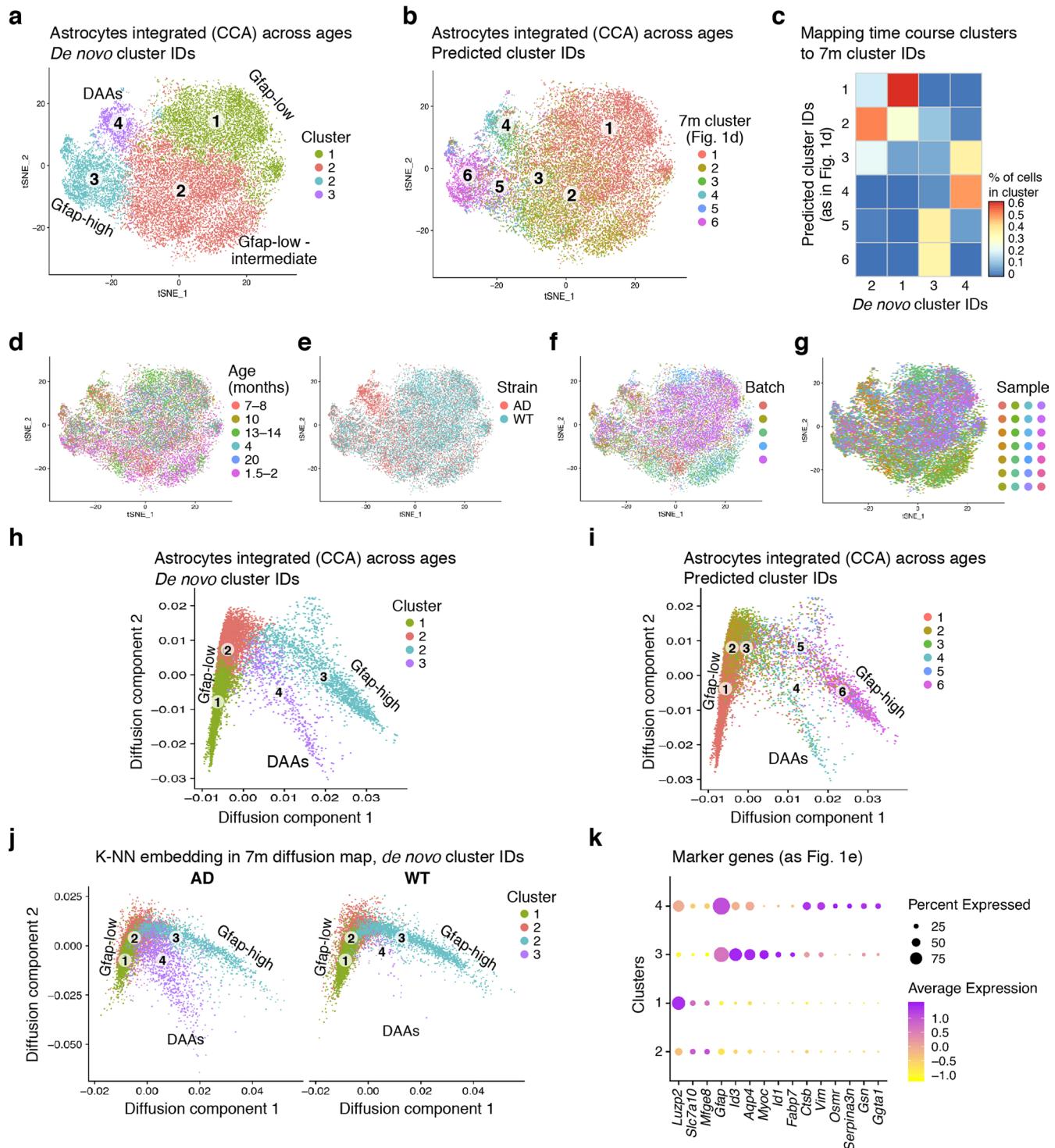
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Shared and distinct transcriptional programs of DAAs and physiological Gfap-high astrocytes. **a**, Differential expression across astrocyte states. Volcano plots showing differentially expressed genes in each pair of states (n=8 animals, 10 samples. y-axis: -log adjusted hypergeometric p-value, following FDR multiple hypothesis correction, x-axis: average log fold change). AD risk factor genes from GWAS marked in orange (as in Fig. 2a). **b**, DAAs, Gfap-high astrocytes, and cluster 3 astrocytes share multiple upregulated genes and pathways compared to the homeostatic Gfap-low astrocyte population, but also have distinct expression programs. Pathway (rows) enrichment for upregulated genes in cluster 4 (C4, DAAs, n=478 cells), cluster 6 (C6, Gfap-high, n=457), or cluster 3 (C3, intermediate state, n=1,666 cells), compared to Gfap-low astrocytes (n=1,594 cells). Enriched pathways³⁰ (hypergeometric p-value with FDR<0.05. n=8 animals, 10 samples), colored by -log FDR values (as in Fig. 2b, with full list of pathway annotations and no scaling).

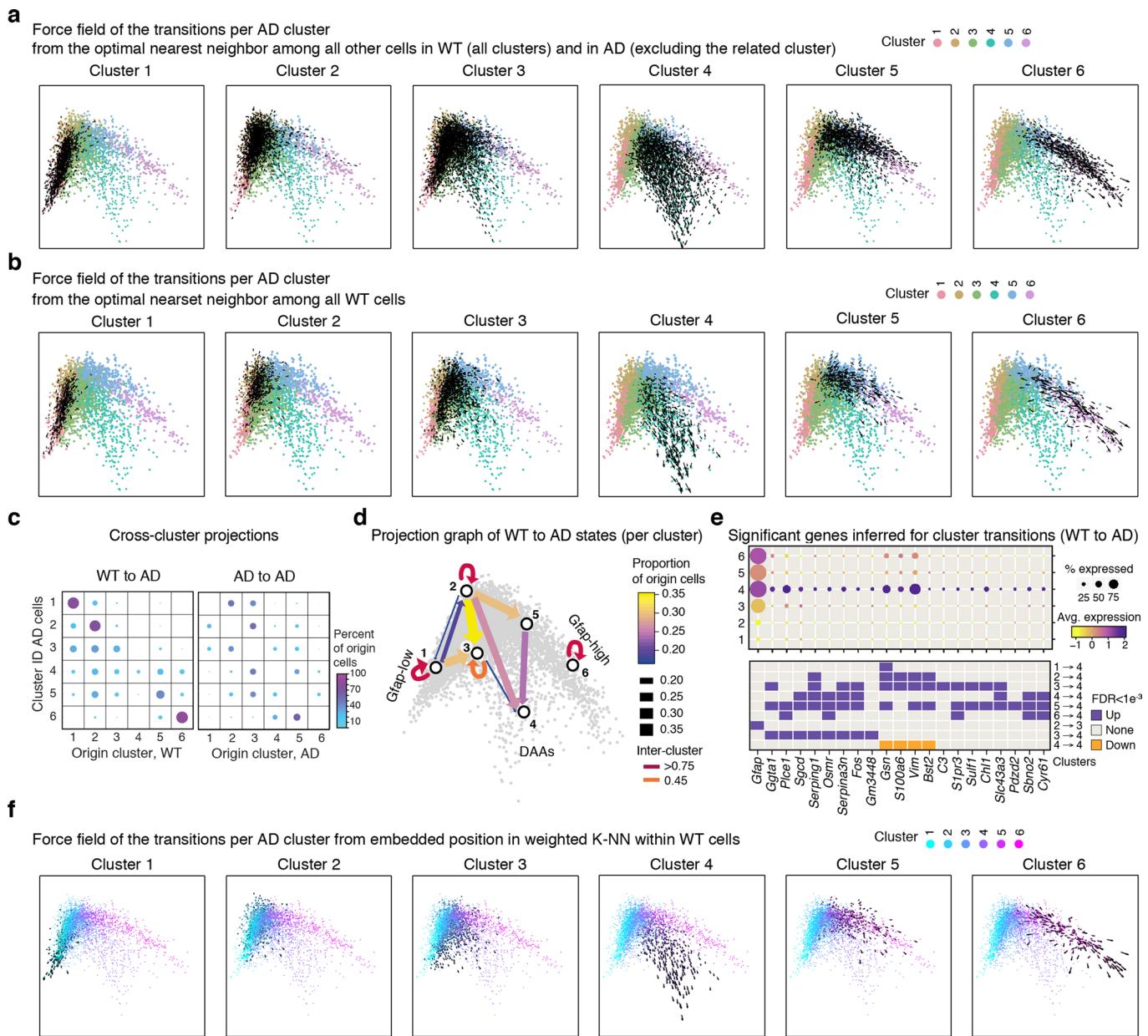


Extended Data Fig. 6 | See next page for caption.

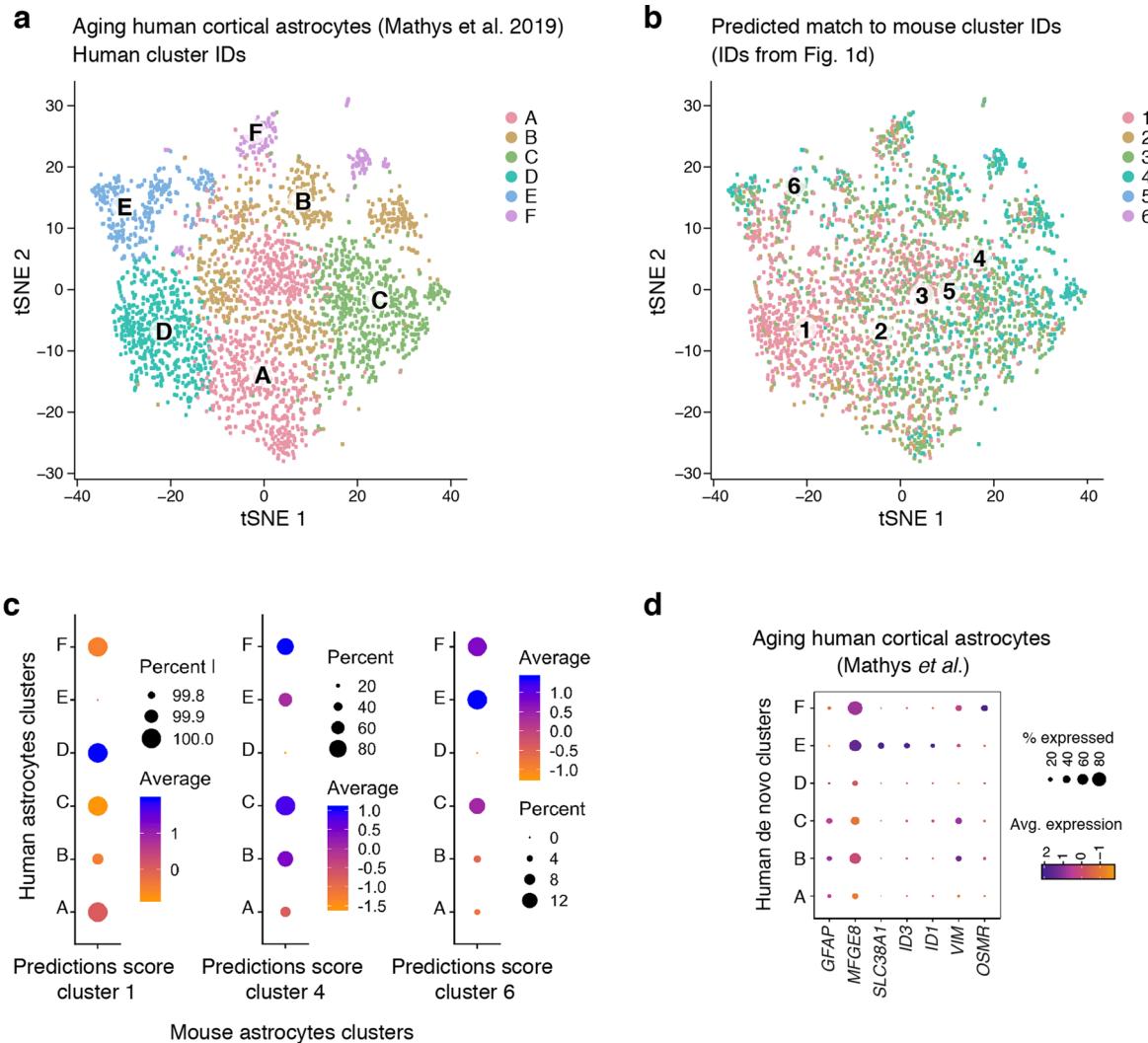
Extended Data Fig. 6 | DAA_s express signatures of reactive astrocytes and are found across brain regions. **a-b**, Signatures of inflammatory reactive astrocytes found in AD. Ridge plots showing the distribution of expression scores across each astrocyte cluster, for signature genes of: pan-reactive, A1, and A2 (from Liddelow *et al.*⁵, left), or of: pan-reactive, inflammation (LPS-induced), and ischemia (MCAO-induced) (from Zamanian *et al.*⁶, right). Expression score per cell is the geometric mean normalized expression level (TPMs) across all signature genes, corrected by subtraction of the geometric mean expression of a random set of genes of similar expression levels. **b**, Violin plots showing the distribution of expression scores across WT (3,831 cells, n=4 animals) and AD (3,514 cells, n=4 animals), for signature genes for: A1 and A2 astrocytes (as defined in Liddelow *et al.*⁵). Scores computed as in (a). **c**, Expression of genes from reactive, A1, and A2 signatures, showing diversity of gene patterns across astrocyte clusters. 2-D umap²⁵ embedding of 7,345 single nuclei RNA profiles of astrocytes in the hippocampus of WT and 5xFAD mice (as in Supplementary Fig. 3a), colored by the gene expression level. Dotted black lines: areas of highest expression. Gene name: on top of each graph. The associated gene (pan reactive, A1 or A2) on the right side of each panel. **d**, Astrocytes expressing DAA_s markers are present in AD brains, enriched in the subiculum and in proximity to A β plaques. Representative immunofluorescence images (staining repeated over n=4 AD and WT mice, with 4 brain slices per animal) in sagittal sections of 7-8-month old 5XFAD mice (with Fig. 2f-h). From top to Bottom: Subiculum, stained for GFAP (green), VIM (red), and serpinA3N (gray), (as in Fig. 2f). Dentate gyrus (in Fig. 2g, left) and subiculum (in Fig. 2g, right), stained for GFAP (green), VIM (red), and A β (gray). Subiculum, stained for GFAP (green), serpinA3N (red), and A β (gray) (as in Fig. 2h). Cell nuclei are shown in blue (Hoechst). Scale bar, 50 μ m.



Extended Data Fig. 7 | Clustering of astrocyte cells from WT and 5xFAD mice across ages. **a-b**, tSNE embedding of sNuc-Seq profiles of astrocytes in WT and 5xFAD (AD) mice across 6 different age groups (in months): 1.5-2 (n=6 mice), 4-5 (n=4), 7-8 (n=8), 10 (n=2), 13-14 (n=6), and 20 (n=2, WT only). (23,863 cells, integrated across four batches^{22,33}, Methods). Colored by: cluster IDs (in **a**), or predicted cluster IDs (using reference cluster IDs of astrocytes of 7-month old mice as in Fig. 1d, in **b**). **c**, The correspondence between the *de novo* cluster IDs (columns) of the time course data, and the predicted cluster IDs (rows) from astrocytes of 7-month old mice (as in Fig. 1d). Color scale = proportion of predicted IDs per *de novo* cluster. **d-g**, 2-D tSNE embedding of astrocytes across ages as in (a), colored by: age (**d**), mouse strain, AD/WT (**e**), batch (**f**), or sample (**g**). **h**, Continuous trajectory of astrocyte states across ages in WT and AD. Diffusion map embedding of astrocytes across ages. Colored by *de novo* clusters: Gfap-low (1), Gfap-low/intermediate (2), Gfap-high (3), and DAAs (4). 15,113 astrocyte cells, down sampled randomly from 23,863 cells to capture 2,500 cells of each age group (or the maximum number of cells available if less than 2,500). **i**, Diffusion map embedding (as in **h**), colored by predicted cluster IDs (as in **b**). **j**, 2-D embedding of the 15,113 astrocytes of AD (left) and WT (right) cells across ages, projected onto the diffusion map of 7-month old mice (by the weighted average position of the K-NN from Fig. 1d, Methods). Colored by the *de novo* clusters (as in **a**). **k**, The expression level (as color scale) and the percent of cells expressing (as dot size) of marker genes for DAAs (Ggt1, Gsn, Osmr, Vim, Serpina3n, Ctsb, and Gfap), Gfap-high (Fabp7, Slc38a1, Myoc, Aqp4, Id1, Id3, and Gfap) and Gfap-low (Mfge8, Slc7a10, Luzp2), across *de novo* clusters of astrocytes across ages.



Extended Data Fig. 8 | DAA are largely derived from Gfap-low astrocytes. **a-b**, Direction of transition of AD astrocytes from their global optimal nearest neighbor (origin) cell (predicted by the Hungarian algorithm³⁴, Methods) to their given position in the diffusion map. For each cluster, force field (black arrows) marking the directionality of transitions along the diffusion map²⁸ (as in Fig. 1d), cells colored by cluster ID. Showing transitions from the predicted cell of origin among all WT cells or AD cells outside the cluster (a), or among all WT cells only (b) (transitions among AD cells only are shown in Fig. 3). **c**, Proportion of cells of origin per cluster (dot size and color) for each AD cluster (rows), from all WT cells (left) or from AD cells from all other clusters (right). **d**, Scheme of predicted transitions, shown as a graph with arrows between pairs of clusters with high proportion of origin cells (>15%) when mapping AD to all WT cells. Color and width of edge reflects the proportion. Diffusion map as in (a) in the background. **e**, Genes correlated with predicted transitions from WT cells to DAA in AD. The expression level across clusters (dot color) and the percent of cells expressing (dot size) significant (Pearson Correlation coefficient, FDR q-value<0.001, n=28 mice, 25,076 cells, Supplementary Table 1) genes that correlated with the transition to cluster 3 (intermediate) or cluster 4 (DAAs) from WT cells. Bottom: Assignment of each gene to a transition between pairs of clusters. **f**, Direction of transition on the diffusion space of AD astrocytes in each cluster from their weighted K-nearest neighbor position among all other WT cells. Force field showing the directionality of transitions between the expected position in the diffusion map (weighted average position of K-NN among all WT cells, k=10) to the true position along the diffusion map (as in Fig. 1d) for each cell. Colored by cluster IDs. Cluster numbers on top of each graph.



Extended Data Fig. 9 | DAA-like astrocytes found in aging human cortex. **a-b**, Diversity of human astrocytes. De novo tSNE embedding of 3,392 sNuc-Seq profiles of cortical human astrocytes from post-mortem aging brains of AD and non-AD individuals ($n=48$, from Mathys et al.⁴), colored by de novo human cluster ID (**a**) or by the predicted mouse cluster IDs (**b**) (predicted by CCA^{22,33} projections from the mouse clusters as in Fig. 1d). **c**, The average prediction scores (as color scale) and the percent of cells with score above 0 (as dot size) for the three major end-states mouse clusters (from Fig. 1d), across the human astrocyte clusters (as in a, Mathys et al.⁴). From left to right: Gfap-low (mouse cluster 1), DAA (mouse cluster 4), and Gfap-high (mouse cluster 6). **d**, Dot plot showing the expression level (as color scale) and the percent of cells expressing (as dot size) marker genes for mouse astrocyte states: DAA (OSMR, VIM, GFAP), Gfap-high (ID1, ID3, SLC38A1, GFAP) and homeostatic Gfap-low (MFGE8), across clusters of the human cortical astrocytes (from Mathys et al.⁴).

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

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

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Nuclei were counted and imaged using the Nexcelom Cellometer Vision with provided software (version 2.1.7.6 Running in Standard Mode). Single nucleus RNA-seq libraries were made using 10X Genomics single-cell 3' Gene Expression assay (v2 or v3, as indicated per sample in supplementary table 1), sequenced on Illumina HiSeqX using the provided software (Control Software - Version 3.5.0.7, Real Time Analysis - version 2.7.7, Chemistry - v2.0420) or NextSeq500 (control software version 2.2.0, chemistry v2), and quality controlled using Agilent Technologies 2100 BioAnalyzer with the provided software (2100 Expert version B.02.08.SI648(SR3)).

Data analysis

Data analysis was done using the commercial software Cellranger (version 2.1.1, chemistry V2, 10x Genomics) and downstream analysis was done in R using code written by us (available on github), publicly available R packages (version 3.6.1) for statistics, machine learning, visualizations, and vector manipulations, including packages for analysis of scRNA-seq data: Seurat (V2.3.4), ComBat (Version 0.0.4) and Destiny (through BioConductor Release 3.10). Doublet detection was done using the Scrublet software (version 0.2).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data And Software Availability. Raw and processed mouse sequencing data is available at https://singlecell.broadinstitute.org/single_cell/study/SCP302/mouse-alzheimers-and-disease-astrocytes and at the Gene Expression Omnibus (GEO) database accession GSE143758, and the code at: <https://github.com/naomihabiblab/5xFAD-sNucSeq>

Field-specific reporting

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

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

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

Life sciences study design

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

Sample size

The main dataset consisted of 4 WT and 4 transgenic 5xFAD mice, based on published results by us and other groups, showing high consistency in single nucleus RNA-seq libraries between mice, but due to technical variability it is recommended to have at least 3 animals per group. The time course data had additional 6 WT and 6 5xFAD animals across time points. Number of nuclei per animal was determined to enable detection of rare populations around 2% of all cells (calculated based on our previous data in the mouse brain).

Data exclusions

No animal or sample was excluded from the analysis. As routinely done, specific single nucleus RNA-seq libraries were excluded based on low technical quality (low number of transcripts and genes captured) or after identification of these libraries as doublets.

Replication

Biological replicates (different mice) and technical replicates (left and right brain hemispheres from the same mouse) of the single nuclei RNA-seq libraries were made, and used as validations. Replication were successful and we report differences between WT and 5xFAD mice which are confirmed by all samples (animals and technical replicates). For the time course data, since not all time points had technical replicates, we report differences between WT and 5xFAD mice that consistently increase or decrease across all time points.

Randomization

In each experimental batch, animals were chosen randomly based on their age. However, we matched transgenic Alzheimer's model 5xFAD animals with WT non-transgenic litter-mate from the same mouse colony and age.

Blinding

Multiple analysis steps of the single cell RNA-seq data was done blindly to the mouse strain, including: clustering, detection of differentially expressed genes and marker genes, cell type identification and frequency assessments.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	<input type="checkbox"/> Involved in the study <input checked="" type="checkbox"/> Antibodies <input checked="" type="checkbox"/> Eukaryotic cell lines <input checked="" type="checkbox"/> Palaeontology <input type="checkbox"/> Animals and other organisms <input checked="" type="checkbox"/> Human research participants <input checked="" type="checkbox"/> Clinical data
-----	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Methods

n/a	<input type="checkbox"/> Involved in the study <input checked="" type="checkbox"/> ChIP-seq <input checked="" type="checkbox"/> Flow cytometry <input checked="" type="checkbox"/> MRI-based neuroimaging
-----	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Antibodies

Antibodies used

The following primary antibodies were used: rabbit anti-GFAP (1:150; Dako #Z0334 #LOT 20056262), chicken anti-VIM (1:150; Abcam #24525, # LOT GR3216660-15), goat anti-serpinA3N (1:200; R&D systems #AF4709-SP, # LOT CBKW0318051), mouse anti-AB 1-16 (1:150; Biolegend #803001 #LOT B247600). Secondary antibodies were Cy2/Cy3/Cy5 donkey anti-rabbit/chicken/goat/mouse antibodies, respectively (1:150; Jackson ImmunoResearch, 711-225-152, cy2 Donkey anti Rabbit, lot 142845; 715-225-150, Cy2-AffiniPure Donkey Anti-Mouse; lot 142843; 705-225-147, Cy2-AffiniPure Donkey Anti-Goat IgG; lot 122781 711-165-152, Cy3 Donkey anti Rabbit, lot 143202; 715-165-151, Cy3-AffiniPure Donkey Anti-Mouse IgG, lot 143017; 705-165-147, Cy3 Donkey Anti-Goat, lot 143201; 715-175-151, Cy5-AffiniPure Donkey Anti-Mouse IgG, lot 144119; 711-175-152, Cy5-AffiniPure Donkey Anti-Rabbit IgG, lot 144221; 705-175-147, Cy5-AffiniPure Donkey Anti-Goat, lot 134531; 703-485-155, DyLight 488 AffiniPure Donkey Anti-Chicken IgG, lot 153002).

Validation

Full validations were done and staining with secondary antibody alone was used as a negative control, to rule out nonspecific staining. The Rabbit anti-GFAP Polyclonal antibody directed to bovine GFAP, was shown to be specific and functional in mouse brain in the following publications: Rosenzweig et al., 2019.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals

In this study, the following mouse strains and ages were used: male heterozygous 5XFAD transgenic mice (Tg6799 strain on a C57/BL6-SJL strain background) and WT strain C57/BL6-SJL mice, at ages: 1.5, 2, 4, 7, 10, and 13 months old. The study included multiple male mice and two female mice. All information regarding the age, sex and strain is provided per sample in Supplementary Table 1.

Wild animals

The study did not include any wild animals

Field-collected samples

The study did not include any samples collected at the field

Ethics oversight

All experiments detailed herein complied with the regulations formulated by the Institutional Animal Care and Use Committee (IACUC) of the Weizmann Institute of Science.

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