

# Single-cell transcriptomic analysis of Alzheimer's disease

Hansruedi Mathys<sup>1,2,10</sup>, Jose Davila-Velderrain<sup>3,4,10</sup>, Zhuyu Peng<sup>1,2</sup>, Fan Gao<sup>1,2</sup>, Shahin Mohammadi<sup>3,4</sup>, Jennie Z. Young<sup>1,2</sup>, Madhvi Menon<sup>4,5,6</sup>, Liang He<sup>3,4</sup>, Fatema Abdurrob<sup>1,2</sup>, Xueqiao Jiang<sup>1,2</sup>, Anthony J. Martorell<sup>1,2</sup>, Richard M. Ransohoff<sup>7</sup>, Brian P. Hafler<sup>4,5,6,8</sup>, David A. Bennett<sup>9</sup>, Manolis Kellis<sup>3,4,11\*</sup> & Li-Huei Tsai<sup>1,2,4,11\*</sup>

**Alzheimer's disease is a pervasive neurodegenerative disorder, the molecular complexity of which remains poorly understood. Here, we analysed 80,660 single-nucleus transcriptomes from the prefrontal cortex of 48 individuals with varying degrees of Alzheimer's disease pathology. Across six major brain cell types, we identified transcriptionally distinct subpopulations, including those associated with pathology and characterized by regulators of myelination, inflammation, and neuron survival. The strongest disease-associated changes appeared early in pathological progression and were highly cell-type specific, whereas genes upregulated at late stages were common across cell types and primarily involved in the global stress response. Notably, we found that female cells were overrepresented in disease-associated subpopulations, and that transcriptional responses were substantially different between sexes in several cell types, including oligodendrocytes. Overall, myelination-related processes were recurrently perturbed in multiple cell types, suggesting that myelination has a key role in Alzheimer's disease pathophysiology. Our single-cell transcriptomic resource provides a blueprint for interrogating the molecular and cellular basis of Alzheimer's disease.**

Alzheimer's disease (AD) is a slowly progressing neurodegenerative disorder that starts with mild memory loss and culminates in severe impairment of executive and cognitive functions<sup>1–3</sup>. The pathophysiology of AD involves interactions between neuron and glia; in support of this, transcriptomic and epigenomic analyses reveal downregulation of neuronal functions and upregulation of innate immune responses in AD brains<sup>4–14</sup>. However, bulk-tissue-level resolution may mask the complexity of alterations across cells and within cell groups, especially for less abundant cell types<sup>4</sup>. Potential changes in cell composition during neurodegeneration also confound the distinction between composition and activity changes in a given cell type. Moreover, the complex interplay between protective and damaging molecular processes, both within and across cell types, further contributes to the difficulty in interpreting tissue-resolution signatures of disease.

Single-cell RNA sequencing (scRNA-seq) provides an alternative method to study the cellular heterogeneity of the brain<sup>15–17</sup>, by profiling tens of thousands of individual cells<sup>15,18,19</sup>. With the goal of characterizing the complex cellular changes in AD brain pathology, here we profile 80,660 droplet-based single-nucleus cortical transcriptomes across 48 individuals with varying degrees of AD pathology and including both sexes. The resulting resource—which is the first, to our knowledge, single-cell view of AD pathology—paints a unique cellular-level view of transcriptional alterations associated with AD pathology, and reveals cell-type-specific and shared gene-expression perturbations, disease-associated cellular subpopulations, and sex-biased transcriptional responses.

## Single-nucleus RNA-seq profiling of prefrontal cortex

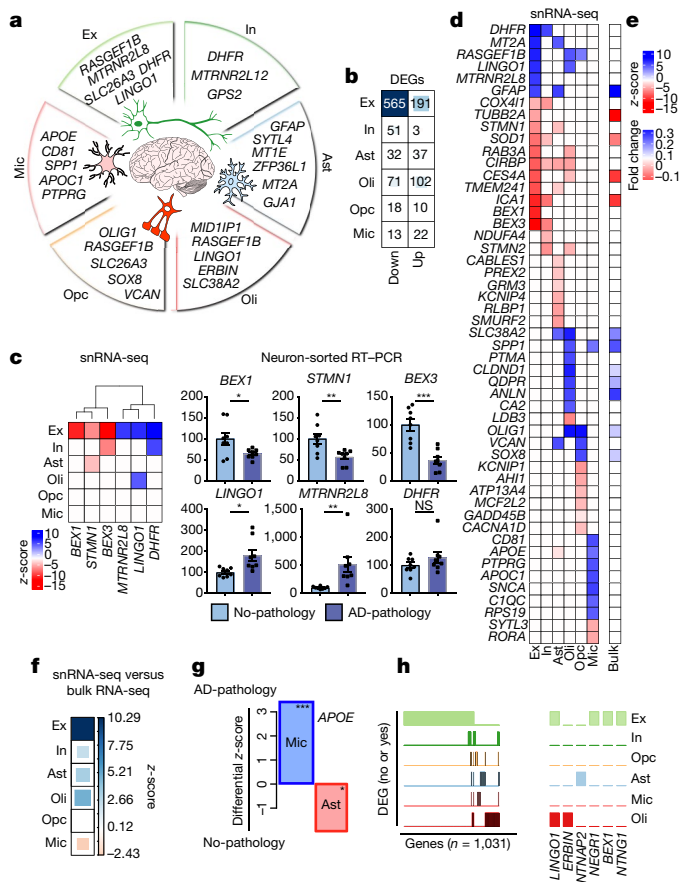
Post-mortem human brain samples came from 48 participants in the Religious Order Study (ROS) or the Rush Memory and Aging Project (MAP), two longitudinal cohort studies of ageing and dementia.

Information collected as part of these studies (which are collectively known as ROSMAP) includes clinical data, detailed post-mortem pathological evaluations, and omics tissue profiling of participants<sup>20</sup>. We selected 24 individuals with high levels of  $\beta$ -amyloid and other pathological hallmarks of AD ('AD-pathology'), and 24 individuals with no or very low  $\beta$ -amyloid burden or other pathologies ('no-pathology'). For each individual, we profiled tissue from the prefrontal cortex (Brodmann area 10)—a region of the brain that has a major role in traits that are affected by AD, including cognition. Immunohistochemistry for  $\beta$ -amyloid confirmed the pathological status of the samples (Extended Data Fig. 1a, b), and bright-field and high-resolution confocal microscopy did not show any apparent physical damage to nuclei isolated from AD-pathology samples relative to no-pathology samples (Extended Data Fig. 1c). We report a total of 80,660 droplet-based single-nucleus RNA-seq (snRNA-seq) profiles (Extended Data Fig. 2a), which are publicly available on the ROSMAP data compendium (see 'Data availability').

## Cellular diversity of the aged prefrontal cortex

To classify the major cell types in the aged cortex, we pre-clustered all cells jointly across the 48 individuals (Extended Data Fig. 2b) to produce 20 transcriptionally distinct pre-clusters with highly consistent expression patterns across individuals (Extended Data Fig. 2c, d). We identified and annotated the major cell types of the human brain by interrogating the expression patterns of known marker genes<sup>18,21</sup>: excitatory neurons (marked by *NRGN*), inhibitory neurons (*GAD1*), astrocytes (*AQP4*), oligodendrocytes (*MBP*), microglia (*CSF1R* and *CD74*), oligodendrocyte progenitor cells (*VCAN*), endothelial cells (*FLT1*), and pericytes (*AMBP*) (Extended Data Fig. 3a, b). The cell types, markers, and proportions of cell types matched previous single-nucleus droplet-based sequencing data from adult human cortex<sup>18</sup>,

<sup>1</sup>Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>2</sup>Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>3</sup>MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA. <sup>4</sup>Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>5</sup>Department of Neurology, Harvard Medical School, Boston, MA, USA. <sup>6</sup>Evergrande Center for Immunologic Diseases, Harvard Medical School, Boston, MA, USA. <sup>7</sup>Third Rock Ventures, Boston, MA, USA. <sup>8</sup>Department of Ophthalmology, Harvard Medical School, Boston, MA, USA. <sup>9</sup>Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA. <sup>10</sup>These authors contributed equally: Hansruedi Mathys, Jose Davila-Velderrain. <sup>11</sup>These authors jointly supervised this work: Manolis Kellis, Li-Huei Tsai. \*e-mail: manoli@mit.edu; lhtsai@mit.edu



**Fig. 1 | Cell-type-specific gene-expression changes in AD pathology.** **a**, Genes most upregulated: excitatory (Ex) and inhibitory (In) neurons, astrocytes (Ast), oligodendrocytes (Oli), oligodendrocyte precursor cells (Opc), and microglia (Mic). **b**, DEG counts for each cell type (two-sided Wilcoxon rank-sum test, FDR < 0.01, log<sub>2</sub>(mean gene expression in AD-pathology/mean gene expression in no-pathology) > 0.25, Poisson mixed-model FDR < 0.05). The intensity of the blue colour and the size of the squares are proportional to entry values. **c**, RT-qPCR validation. snRNA-seq differential scores for excitatory- and inhibitory-neuron DEGs (z-score, Poisson mixed model) (left) and qPCR validation (right). NeuN-positive nuclei were isolated by fluorescence-activated cell sorting (n = 8 AD-pathology individuals and n = 8 no-pathology individuals). Data are mean ± s.e.m.; \*\*\*P < 0.001, \*\*P < 0.01, \*P < 0.05; NS, P > 0.05 (Student's two-tailed t-test). **d, e**, Differential scores (z-scores, Poisson mixed model) for the top DEGs from snRNA-seq (**d**), and the corresponding values from bulk RNA-seq (**e**) (ROSMAP cohorts; n = 484, P < 0.01). **f**, Global consistency analysis of snRNA-seq versus bulk RNA-seq. Agreement or disagreement was estimated by deviation from random expectation (z-score) of single-cell DEG average rank scores in the ranked list of genes from bulk differential analysis. **g**, Differential expression of *APOE* in microglia (n = 955 cells, 24 AD-pathology individuals; n = 965 cells, 24 no-pathology individuals) and astrocytes (n = 1,830 cells, 24 AD-pathology individuals; n = 1,562 cells, 24 no-pathology individuals); z-score, Poisson mixed model; \*\*\*P = 0.001, \*P = 0.05, corresponding P value calculated from two-sided, standard normal distribution). **h**, Binary plots indicating with bars whether a gene (column) is a DEG in a given cell type (rows) or not (n = 1,031 DEGs). Right, six genes associated with myelination and/or axon regeneration.

indicating that our results are robust to the inclusion of pathologically affected brains (Extended Data Fig. 3c–e). We next collapsed the pre-clusters into eight broad cell-type clusters using annotations supported by both direct marker expression and significant (false discovery rate (FDR) < 0.01, hypergeometric test) overlap with previously curated single-cell populations. We used these cell-type categories to characterize the specificity of AD-pathology gene-expression perturbations, to quantify gene–trait associations, and to assess

qualitative differences in cell-type-specific pathological responses between sexes.

### Systematic differential analysis of gene expression

We compared levels of gene expression in cells isolated from AD-pathology versus no-pathology individuals by cell type (Methods), and identified 1,031 unique differentially expressed genes (DEGs) that implicated all major cell types (Fig. 1a, b, Supplementary Table 2). Neurons showed a strong signature of repression—75% of DEGs in excitatory and 95% in inhibitory neurons were downregulated—whereas most DEGs in oligodendrocytes, astrocytes, and microglia were upregulated (53–63%). The numbers of DEGs for non-neuronal populations were substantially smaller, probably owing to reduced power in lower-abundance cell types. These contrasting observations on the number and dominant directionality of DEGs reveal a heterogeneous response to AD pathology between cell types—a recurrent theme that we observed throughout the study. Notably, DEGs were robustly detected at different levels of expression (Extended Data Fig. 4a). Quantitative PCR with reverse transcription (RT-qPCR) in nuclei staining positive for the neuronal marker NeuN (isolated by fluorescence-activated cell sorting) corroborated significant differential expression for five of the six genes tested, including both upregulated and downregulated genes (Fig. 1c).

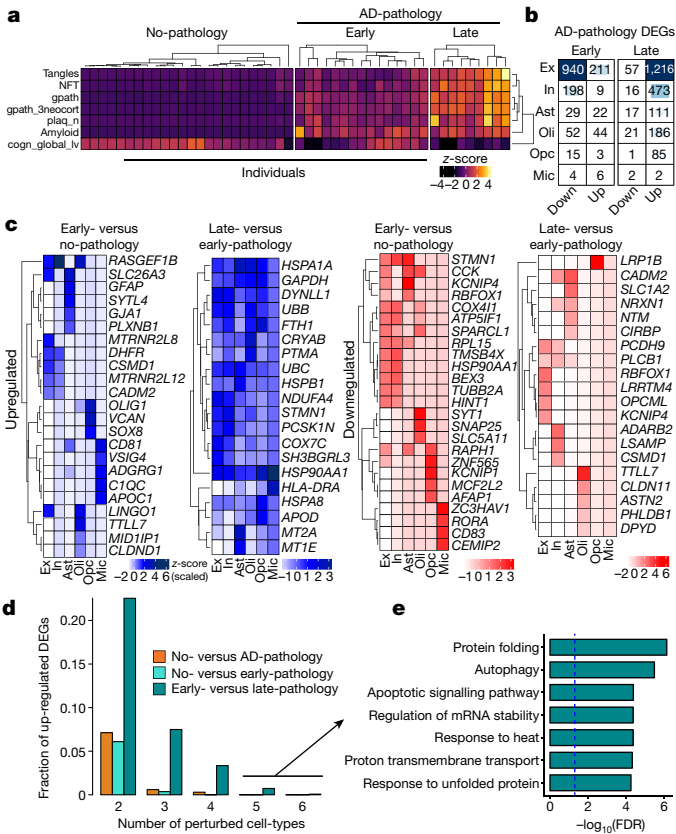
A similar analysis of DEGs using high-quality bulk RNA sequencing (RNA-seq) data from the ROSMAP cohorts (n = 484) validated the DEGs that we identified in our snRNA-seq data (Fig. 1d, e). In addition, a rank permutation test (Methods) revealed that bulk data are dominated by expression changes observed at the single-cell level in excitatory neurons and oligodendrocytes (Fig. 1f). Consequently, changes in other cell types are not well captured, with microglia particularly underrepresented. Bulk RNA-seq data also cannot capture DEGs with opposite directionality in different cell types. With snRNA-seq, we found that *APOE* was strongly upregulated in microglia but downregulated in astrocytes (FDR-corrected P < 0.01, two-sided Wilcoxon rank-sum test) (Fig. 1g), which is consistent with studies of AD in human induced pluripotent stem cells<sup>22</sup> and in mouse models<sup>23–25</sup>.

The vast majority of DEGs (95%) were perturbed only in neurons or in a single glial cell type, which indicates that these perturbations are strongly cell-type specific (Fig. 1h). However, we found that top DEGs are involved in related processes even across cell types. One such example is myelination, axonal outgrowth and regeneration. Top DEGs included *LINGO1*, which was upregulated in excitatory neurons and oligodendrocytes and is a negative regulator of neuronal survival, axonal integrity, and oligodendrocyte differentiation and myelination<sup>26,27</sup>; *ERBIN*, which is required for remyelination of axons<sup>28</sup>; *CNTNAP2*, which mediates axon–myelin interactions<sup>29</sup>; *NEGR1*, which mediates myelin-stimulated outgrowth of axons<sup>30</sup>; *BEX1*, which is involved in regeneration of axons after injury<sup>31</sup>; and *NTNG1*, which promotes neurite outgrowth of axons and dendrites<sup>32</sup> (Fig. 1h). Consistent with the observed downregulation of *NTNG1* in excitatory neurons, RNA in situ hybridization revealed significantly fewer excitatory neurons by expression of *SLC17A7* with detectable *NTNG1* expression in brain sections from AD-pathology individuals compared with no-pathology individuals (Extended Data Fig. 4b).

Overall, these results indicate that all major cell types are affected at the transcriptional level by AD pathology, and that single-cell-level resolution is critical because changes in gene expression—including directionality—can be conditional on cell type.

### Cell-type-specific changes during AD progression

We next examined whether there are qualitative changes in gene-expression perturbations during early- versus late-stage AD pathology. To define pathology groups, we clustered individuals based on nine clinico-pathological traits (Supplementary Table 3). AD-pathology individuals segregated into two subgroups that correspond to the pathological progression of AD: ‘early-pathology’<sup>1</sup> (amyloid burden, but modest neurofibrillary tangles and modest cognitive impairment)



**Fig. 2 | Gene-expression changes in the progression of AD pathology.** **a**, Phenotypic clustering of 48 individuals (columns) using clinico-pathological variables (rows) measuring neuronal neurofibrillary tangle density (tangles), neurofibrillary tangle burden (NFT), global AD-pathology burden (gpath), global measure of neocortical pathology (gpath\_3neocort), neuritic plaque burden (plaq\_n), overall amyloid level (amyloid), and global cognitive function (last valid score) (cogn\_global\_iv). **b**, Progressive changes in DEG counts for each cell type for early and late AD-pathology individuals (two-sided Wilcoxon rank-sum test,  $FDR < 0.01$ ,  $\log_2(\text{mean gene expression in early-pathology}/\text{mean gene expression in no-pathology}) > 0.25$  or  $\log_2(\text{mean gene expression in late-pathology}/\text{mean gene expression in early-pathology})$  and similarly for the opposite direction, Poisson mixed-model  $FDR < 0.05$ ). The intensity of the blue colour and the size of the squares are proportional to entry values. **c**, Most-significantly altered (based on  $P$  value rank) genes (rows) for each cell type (columns) and comparison. Column-scaled  $z$ -scores computed with a Poisson mixed model are shown ( $FDR < 0.01$ , two-sided Wilcoxon rank-sum test). **d**, Fraction of total upregulated genes ( $y$  axis) as a function of the total number of cell types in which the upregulation occurs. **e**, Gene Ontology terms associated with genes upregulated in the late-pathology group that are common to  $\geq 5$  cell types ( $n = 11$  genes, hypergeometric test,  $FDR$  correction).

and ‘late-pathology’<sup>33</sup> (higher amyloid burden, and also increased neurofibrillary tangles, global pathology, and cognitive impairment) (Fig. 2a). We then quantified gene-expression changes by pairwise comparisons between these groups.

Comparison of the early-pathology and no-pathology subgroups revealed that large-scale transcriptional changes occur before individuals develop severe pathological features (Fig. 2b, c). Both up- and downregulated DEGs were highly cell-type specific, with nearly all genes (96%) perturbed either in neurons (excitatory and inhibitory) or in a single glial cell type. These changes were similar to those found between the no-pathology and AD-pathology groups (Extended Data Fig. 5), suggesting that major transcriptional changes appear early in pathological progression. Comparison of the late-pathology and early-pathology groups revealed common upregulated genes that were shared across cell types (Fig. 2d), in contrast with the cell-type

specificity that we observed at an early stage of pathology. Shared upregulated genes are involved in protein folding (for example, *HSP90AA1* and *HSPA1A*), including molecular chaperones (for example, *HSPB1* and *CRYAB*), and are also associated with autophagy, apoptosis, and the generalized stress response (Fig. 2c, e). These processes are collectively involved in the proteostasis network—the molecular machinery that operates to maintain protein integrity<sup>34–36</sup>. In contrast with upregulated genes, downregulated genes mostly showed cell-type-specific changes.

**Cell-type-specific associations with AD-related traits**

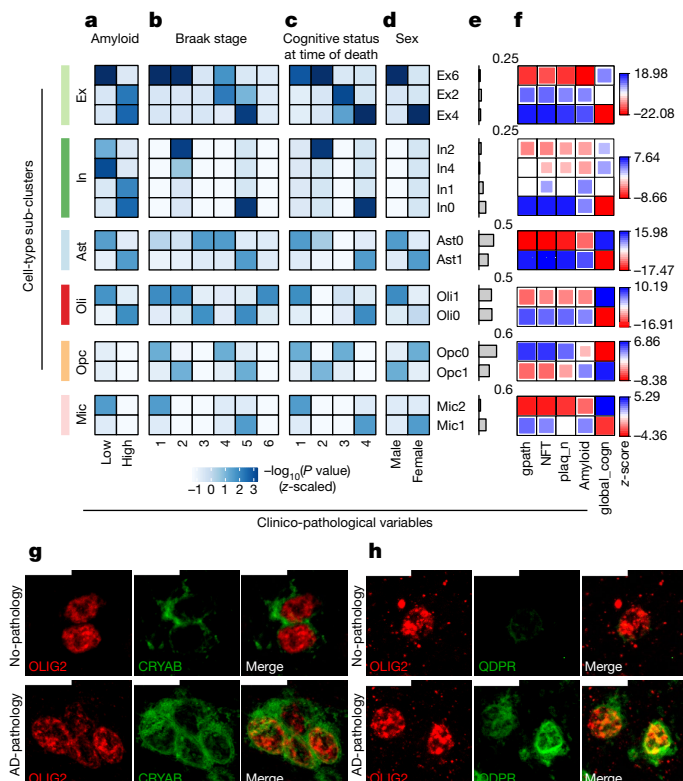
Given the complexity and heterogeneity of the AD phenotype, we next aimed to quantify the association between gene expression in specific cell types and variability of pathological traits. We used the major pathological quantitative traits from the ROS:  $\beta$ -amyloid level, neurofibrillary tangle burden, neuritic plaque count, tangle density, global AD pathology, and global cognition. For each individual, we first computed the correlation between the expression profile of each gene (for a given cell type) and a pathological trait. We then analysed the resultant gene–trait correlation patterns using the self-organizing map (SOM) approach<sup>37</sup> to discover gene sets with similar expression patterns that most strongly correlated with each phenotype (Methods). Genes with similar phenotypic correlations are grouped in the same SOM grid unit, with similar units clustered nearby (Extended Data Fig. 6a–e). We observed that excitatory neurons, inhibitory neurons, astrocytes, microglia, and oligodendrocytes each showed distinct SOM units associated with multiple pathological traits, which indicates that different groups of genes respond to AD pathology in each cell type. Gene sets that correlated with post-mortem interval and age at death were highly distinct from those that correlated with pathological signatures of AD, suggesting that they are orthogonal to each other.

To identify gene groups that showed similar correlations with traits, we used an image segmentation method to identify and manually curate the ten SOM territories that were most strongly associated with AD phenotypes across cell types (Extended Data Fig. 6f). We found that the gene groups defining these territories (gene–trait correlation modules M1–M10) are involved in common functional pathways that are relevant to the pathophysiology of AD (Supplementary Table 4). For example, M7 showed a positive pathological correlation in microglia and was enriched in immune and inflammatory pathways, as well as in pathways associated with the clearance of  $\beta$ -amyloid. M9 was positively correlated with pathology in oligodendrocytes and was enriched in oligodendrocyte differentiation and myelination pathways—possibly reflecting an oligodendrocytic response to myelin loss.

To link genetic risk with coordinated gene activity, we quantified the overrepresentation of genes from genome-wide association studies (GWAS)<sup>38</sup> and identified modules that were linked to genetic risk factors for AD and for general cognitive function. M6 and M7 overlapped with genes that are AD risk factors, including *APOE*, *TREM2*, *MEF2C*, *PICALM*, and the major histocompatibility complex (MHC) class II genes *HLA-DRB1* and *HLA-DRB5*; expression of these genes in microglia is positively correlated with measures of AD pathology. By contrast, M3 overlapped with genes that are associated with cognition<sup>39</sup> and the expression of which is negatively correlated with pathology in neurons (Extended Data Fig. 6g, Supplementary Table 4). These observations provide a link between genetic risk factors and the microglial and neuronal transcriptional responses to pathology, which may partially explain some of the risk conferred by genetic variants.

**AD-associated cellular subpopulations**

To dissect cell-type heterogeneity, we next sub-clustered each major cell type<sup>40</sup>, resulting in 13 excitatory-neuron (Ex), 12 inhibitory-neuron (In), 4 astrocyte (Ast), 5 oligodendrocyte (Oli), 3 oligodendrocyte-precursor-cell (Opc), and 4 microglia (Mic) sub-clusters. The identified subpopulations were not exclusively enriched with cells from any single individual (Extended Data Fig. 7a, b). We examined the cellular composition of each subpopulation in relation to the pathological features



**Fig. 3 | Trait associations in cellular subpopulations.**

**a–d**, Overrepresentation (hypergeometric test) within each sub-cluster of cells isolated from individuals with varying amyloid levels (**a**), Braak stage (braaksc) (**b**), cognitive status at time of death (cogdx, 1 = no impairment; 4 = impairment) (**c**), and sex (**d**). Only significant associations after FDR correction over all variables and subpopulations are displayed. **e**, Fraction of cells in each sub-cluster. **f**, Quantitative enrichment of clinico-pathological features across sub-clusters for: global AD-pathology burden, neurofibrillary tangle burden; neuritic plaque count, overall amyloid level, and global cognitive function (global\_cogn); z-scores were estimated using resampling. The size of the squares is proportional to the numeric entry. **g**, Immunohistochemistry with anti-OLIG2 (red) and anti-CRYAB (green) antibodies in the white matter of Brodmann area 10 of a no-pathology and an AD-pathology individual (scale bars (top left, white), 10  $\mu$ m). The experiment was performed once. **h**, Similar to **g**, for anti-OLIG2 (red) and anti-QDPR (green). The experiment was performed once.

of source brains (Methods), and observed an overrepresentation of cellular subtypes in pathology for most major cell types. For example, the cellular subpopulations Ex4, In0, Ast1, and Oli0 were associated with cells that were isolated from subjects with AD-pathology traits—that is, high amyloid level, high Braak stage (V), low CERAD (Consortium to Establish a Registry for Alzheimer’s Disease) score, low NIA (National Institute on Aging)–Reagan score, and pronounced cognitive decline. By contrast, the subpopulations Ex6, In2, Ast0, and Oli1 were associated with cells from subjects with no pathological traits (FDR < 0.01, hypergeometric test; Fig. 3a–d, Extended Data Fig. 7c, d). These observations were robust to a randomization analysis in which three female and three male subjects were chosen at random from each pathological category for 100 trials (FDR < 0.01, Extended Data Fig. 7d, Supplementary Table 5). Consistent differences were also reflected in the enrichment of quantitative pathological features (Fig. 3f, FDR < 0.01, permutation test). Thus, both categorical assignments and unbiased direct measurements of clinico-pathological variables support a strong association between the cell-type subpopulations and AD pathological status that is robust to individual selection.

To gain further insight into the molecular processes that distinguish these subpopulations, we identified marker genes for each cellular subpopulation (Extended Data Fig. 8 and Supplementary Table 6, FDR < 0.01, two-sided Wilcoxon rank-sum test).

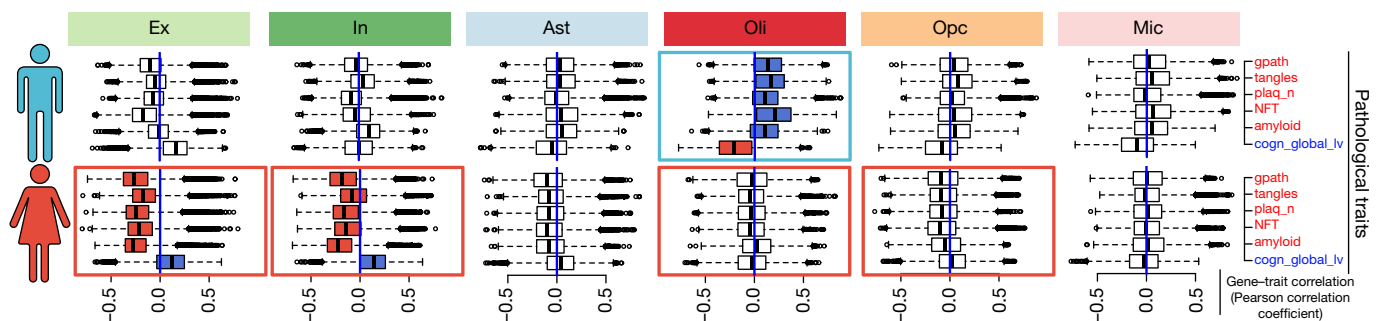
AD-pathology-associated Ex4 neurons were marked by *LINGO1*, *RASGEF1B*, and *SLC26A3*, suggesting that subpopulations preferentially observed in neuropathology may underlie the differences in gene expression that were observed at the cell-type level. AD-pathology-associated Oli0 cells were marked by *CADM2*, *QDPR*, *NLGN1*, and *CRYAB*; the latter is an anti-apoptotic and neuroprotective chaperone, the dysfunction of which could exacerbate inflammation and demyelination<sup>41</sup>. Immunohistochemistry confirmed that cell subsets of the oligodendrocyte lineage (oligodendrocytes and oligodendrocyte precursor cells) express high levels of *CRYAB* or *QDPR* in the white matter of AD-pathology individuals (Fig. 3g, h, Extended Data Fig. 9a, b). The AD-pathology-associated astrocyte subpopulation Ast1 showed preferential expression of *GLUL* and of the AD risk factor *CLU*, which has been shown to be upregulated in reactive astrocytes in response to neurodegeneration<sup>42</sup>. The AD-pathology-associated subpopulations In0, Opc1, and Mic1 were marked by genes that have roles in protein folding and stability, neuronal and necrotic death, and T cell activation and immunity, respectively, suggesting cell-type-specific responses to global cellular stress (Extended Data Fig. 8). Thus, in addition to the previously reported roles of neurons and microglia in AD pathophysiology, the disease-associated signatures of oligodendrocytes and astrocytes reveal additional glial transcriptional responses to pathology.

Previous studies have profiled microglia from mouse models of AD<sup>23,25</sup>. We tested whether the expression signatures associated with AD pathology that we found in the human brain overlapped with the reported states in mice. We found that marker genes of the AD-pathology-associated Mic1 subpopulation, including the MHC-II genes *CD74* and *HLA-DRB1*, significantly overlapped ( $P \leq 0.01$ ; one-sided Fisher’s exact test) with mouse disease-associated<sup>25</sup> and mouse late-response microglia<sup>23</sup> marker genes (Extended Data Fig. 9d, Supplementary Table 7). Immunohistochemistry of samples from AD-pathology individuals confirmed the presence of a subpopulation of microglia that expressed high levels of MHC class II proteins (Extended Data Fig. 9c). The microglial subpopulation identified here in humans revealed AD-associated genes not seen in the animal models, including the complement component *CIQB* and the pattern recognition receptor *CD14*. Next, we tested to what extent Mic1 markers also overlapped with human non-AD, aged microglia<sup>43</sup>. Although we found a significant overlap ( $P \leq 0.01$ ; Fisher’s exact test, Extended Data Fig. 9d), many Mic1 marker genes, including *APOE*, were specific to AD pathology and not identified in aged microglia. Our observations suggest that the Mic1 subpopulation represents a distinct microglial state that shares features with, but is also distinct from, previously reported microglial cell states in mouse models. However, more extensive single-cell references of glial cells in the human brain will be required to contextualize pathological versus normal heterogeneity.

### Sex-specific differential response to AD pathology

We identified robust differences in the association of AD pathology between cells from female versus male individuals (Fig. 3d)—AD-pathology-associated cell subpopulations (Ex4, Ast1, Oli0, and Mic1) were enriched with female cells, whereas no-pathology subpopulations (Ex6, Ast0, and Oli1) were enriched with male cells. This overrepresentation of female cells was not a result of the disproportionate cell contribution of particular individuals (Extended Data Fig. 7a, b), nor was it owing to more severe AD pathology in female individuals (Extended Data Fig. 10a). Differences between sexes were also reflected in the marker genes of AD-pathology-associated subpopulations; these genes showed expression patterns that partially segregated female and male AD individuals, with higher expression in females (Extended Data Fig. 10b, c). We thus hypothesized that the differences might stem from a sex-specific differential transcriptional response to AD pathology.

To discern whether female and male individuals present global differential responses to AD pathology, and whether such variability preferentially involves specific cell types, we recomputed individual-level gene–trait correlations, splitting the dataset by sex (Methods). The resulting data enable correlation distributions to be directly compared



**Fig. 4 | Sex-specific differential response to AD pathology.** Individual-level transcriptome-wide gene–trait correlation analysis. Box plots show the distribution of correlation values (Pearson correlation coefficient) computed between gene-expression profiles ( $n = 17,926$  genes) averaged for cells of each type across each individual, and the corresponding pathological measurements across individuals ( $n = 24$  female individuals and  $n = 24$  male individuals). Box plots are centred around the median, with the interquartile range (IQR) defining the box. The upper whisker

by cell type and pathological feature (Fig. 4). Notably, we found contrasting, qualitatively distinct responses between sexes, involving multiple cell types (Supplementary Table 8). The most extreme differences were observed in neurons and oligodendrocytes. In males, increased pathology correlated with a global transcriptional activation in oligodendrocytes, as reflected in positive correlation shifts (median correlations, 0.203 (neurofibrillary tangle burden (NFT)); 0.165 (tangles); 0.108 (amyloid); and 0.101 (plaques)) (Fig. 4, top). By contrast, global shifts in transcription in oligodendrocytes were not observed in females; instead, correlations with pathological traits and with cognition remained centred around zero (median correlations,  $-0.04$  (NFT);  $-0.042$  (tangles); 0.028 (amyloid); and  $-0.031$  (plaques)) (Fig. 4). Extreme responses were also observed in neuronal cells, in particular in response to amyloid. In females, increased pathology correlated with a global downregulation (median negative correlation) of gene activity in both excitatory neurons (median correlation,  $-0.272$  (amyloid)) and inhibitory neurons (median correlation,  $-0.225$  (amyloid)). In males, excitatory neurons showed a qualitatively similar but much less pronounced response (median correlation,  $-0.01$ , amyloid). Conversely, inhibitory neurons did not, in males, show a clear overall shift in response to pathology or a correlation with cognition, except for a slight increase in response to amyloid (median correlation, 0.1 (amyloid)). We also observed differences, albeit more subtle, in other cell types. In particular, oligodendrocyte precursor cells in females presented a shift towards downregulation in response to pathology—as reflected by a shift in the directionality of correlation for all pathological variables—which was not observed in males. Estimations of median sex differences (bootstrap-estimated) that support these observations are provided in Supplementary Table 8.

To examine a biological role of possible sex bias in relation to white-matter pathology, we looked at the total volume of white-matter hyperintensities (WMH) from MRI data from 505 individuals in the ROSMAP cohorts. We found a significant association between the volume of white-matter lesions and lower cognition (*cogn\_global\_lv*) in female subjects, but not in males (Extended Data Fig. 10d, e). These observations are consistent with a scenario of reduced transcriptional response, particularly in oligodendrocytes, in females with AD pathology. Although these analyses support an underlying sex bias in molecular processes linked to white-matter changes in AD pathophysiology, further single-cell-level analyses of larger sample size and additional experimental follow-up studies will be needed to fully understand the relationship between transcriptional and pathological differences between sexes.

## Discussion

We report here 80,660 single-cell transcriptomes across 48 men and women with a range of AD-associated pathology. Whereas previous

extends to the largest value no further than  $1.5 \times$  IQR from the end of the box. The lower whisker extends to the smallest value at most  $1.5 \times$  IQR from the end of the box. Pathological traits are represented in red font and cognition in blue. Rectangles around plots highlight the most contrasting differences observed between male and female transcriptional responses; blue indicates a dominant positive correlation and red a dominant negative correlation.

studies on AD pathophysiology focused primarily on neurons and microglia (and bulk RNA is dominated by expression signals of neurons and oligodendrocytes), here we provide pathology-responsive transcriptional signatures across six major cell types—excitatory neurons, inhibitory neurons, astrocytes, oligodendrocytes, oligodendrocyte precursor cells, and microglia—as well as 40 transcriptionally distinct subpopulations of cells, some of which are preferentially overrepresented in AD pathology and differentially represented between sexes.

Although most genes presented distinct cell-type-specific perturbations, many of the top DEGs were involved in related processes across cell types. Myelination-related genes were recurrently perturbed not only in oligodendrocytes and oligodendrocyte precursor cells, but also in cells of most major cell types—possibly indicating a major regulatory response to maintain myelin integrity. Although white-matter pathologies have been documented in AD<sup>44</sup>, single-cell resolution enabled us to identify regulators of myelination—such as *LINGO1*—that are perturbed across neuronal and glial cells, as well as other myelin-related genes that are perturbed only in neurons (*PRNP*, *CNTNAP2*, *ERBIN*, *NEGR1*, and *BEX1*) or only in glial cells (*CRYAB*). Cell-type-specific regulatory complexity may thus need to be taken into consideration in the design of therapeutic interventions in AD, and genes with a more homogeneous response, such as *LINGO1*, might have more potential for intervention<sup>26</sup>.

Whereas perturbations in gene expression were largely cell-type specific at an early stage of pathology, genes upregulated in late-stage pathology tended to be common across cell types and were associated with a global stress response. Previous studies at the bulk-tissue level have implicated the downregulation of aggregation-prone proteins as a mechanism to mitigate compromised protein homeostasis in AD<sup>45</sup>. The global cell-agnostic upregulation of proteostasis pathways observed herein may similarly constitute an intrinsic adaptive response in an attempt to balance impaired protein homeostasis in late-stage pathology. Alternatively, it might reflect a disruption of the base-level functioning of the proteostasis network—a factor known to contribute to disease progression<sup>34–36</sup>.

We found that transcriptional alterations seemed to stem from changes in cell state, with certain cell-type subpopulations more readily captured in AD pathology. The observed alterations are consistent with a scenario in which existing subtypes of the normal brain are preferentially responsive to pathology and overexpress a set of responsive genes in addition to constitutive markers of cell subtypes. However, changes in the relative abundance of subtypes as a consequence of differential vulnerability to pathology is an alternative explanation that we currently cannot rule out. We observed that cells isolated from female individuals were overrepresented in many of the AD-pathology-associated cell subpopulations; responses between sexes were contrasting and qualitatively distinct, particularly in neurons and

oligodendrocytes. Our data highlight myelination-related processes in AD pathogenesis, and provide support for a sexual dimorphism in AD that manifests at the transcriptional level, even for individuals matched for age and pathology. The transcriptional sex differences may be interpreted as a greater transcriptional disease burden in female individuals. Alternatively, our findings may be viewed as indicating greater resilience in the female individuals in our study; that is, despite strong alterations at the transcriptional level, the degree of pathophysiological and cognitive decline remains similar in females and males.

Overall, our single-cell-resolution analysis highlights the complexity of glial-neuronal interactions in response to AD pathology. Building on our data, future research could investigate how to distinguish between neuroprotection versus pathogenicity, and the responsive versus the driving nature of the transcriptional alterations observed. The mechanistic basis of the observed changes during the course of AD progression remains unknown.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41586-019-1195-2>.

Received: 23 July 2018; Accepted: 24 April 2019;

Published online 1 May 2019.

- Masters, C. L. et al. Alzheimer's disease. *Nat. Rev. Dis. Primers* **1**, 15056 (2015).
- Hardy, J. & Selkoe, D. J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353–356 (2002).
- Braak, H. & Braak, E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* **82**, 239–259 (1991).
- De Strooper, B. & Karran, E. The cellular phase of Alzheimer's disease. *Cell* **164**, 603–615 (2016).
- Canter, R. G., Penney, J. & Tsai, L.-H. The road to restoring neural circuits for the treatment of Alzheimer's disease. *Nature* **539**, 187–196 (2016).
- Heneka, M. T. et al. Neuroinflammation in Alzheimer's disease. *Lancet Neurol.* **14**, 388–405 (2015).
- Bishop, N. A., Lu, T. & Yankner, B. A. Neural mechanisms of ageing and cognitive decline. *Nature* **464**, 529–535 (2010).
- Pimenova, A. A., Raj, T. & Goate, A. M. Untangling genetic risk for Alzheimer's disease. *Biol. Psychiatry* **83**, 300–310 (2018).
- Fisher, D. W., Bennett, D. A. & Dong, H. Sexual dimorphism in predisposition to Alzheimer's disease. *Neurobiol. Aging* **70**, 308–324 (2018).
- Nativio, R. et al. Dysregulation of the epigenetic landscape of normal aging in Alzheimer's disease. *Nat. Neurosci.* **21**, 497–505 (2018).
- Bossers, K. et al. Concerted changes in transcripts in the prefrontal cortex precede neuropathology in Alzheimer's disease. *Brain* **133**, 3699–3723 (2010).
- Zhang, B. et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**, 707–720 (2013).
- Miller, J. A., Woltjer, R. L., Goodenbour, J. M., Horvath, S. & Geschwind, D. H. Genes and pathways underlying regional and cell type changes in Alzheimer's disease. *Genome Med.* **5**, 48 (2013).
- Gjoneska, E. et al. Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature* **518**, 365–369 (2015).
- Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
- Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
- Zhong, S. et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).
- Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
- Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
- Bennett, D. A. et al. Religious Orders Study and Rush Memory and Aging Project. *J. Alzheimers Dis.* **64**, S161–S189 (2018).
- He, Z. et al. Comprehensive transcriptome analysis of neocortical layers in humans, chimpanzees and macaques. *Nat. Neurosci.* **20**, 886–895 (2017).
- Lin, Y.-T. et al. APOE4 causes widespread molecular and cellular alterations associated with Alzheimer's disease phenotypes in human iPSC-derived brain cell types. *Neuron* **98**, 1141–1154 (2018).
- Mathys, H. et al. Temporal tracking of microglia activation in neurodegeneration at single-cell resolution. *Cell Reports* **21**, 366–380 (2017).
- Krasemann, S. et al. The TREM2-APOE pathway drives the transcriptional phenotype of dysfunctional microglia in neurodegenerative diseases. *Immunity* **47**, 566–581 (2017).
- Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**, 1276–1290 (2017).
- Fernandez-Enright, F. & Andrews, J. L. Lingo-1: a novel target in therapy for Alzheimer's disease? *Neural Regen. Res.* **11**, 88–89 (2016).
- Mi, S. et al. LINGO-1 negatively regulates myelination by oligodendrocytes. *Nat. Neurosci.* **8**, 745–751 (2005).
- Liang, C. et al. Erbin is required for myelination in regenerated axons after injury. *J. Neurosci.* **32**, 15169–15180 (2012).
- Scott, R. et al. Loss of *Cntnap2* causes axonal excitability deficits, developmental delay in cortical myelination, and abnormal stereotyped motor behavior. *Cereb. Cortex* **29**, 586–597 (2019).
- Poplawski, G. H. D. et al. Adult rat myelin enhances axonal outgrowth from neural stem cells. *Sci. Transl. Med.* **10**, eaa12563 (2018).
- Khazaei, M. R. et al. Bex1 is involved in the regeneration of axons after injury. *J. Neurochem.* **115**, 910–920 (2010).
- Seiradake, E. et al. Structural basis for cell surface patterning through NetrinG-NGL interactions. *EMBO J.* **30**, 4479–4488 (2011).
- Holtzman, D. M., Morris, J. C. & Goate, A. M. Alzheimer's disease: the challenge of the second century. *Sci. Transl. Med.* **3**, 77sr1 (2011).
- Andrade, W. A. et al. Early endosome localization and activity of RasGEF1b, a toll-like receptor-inducible Ras guanine-nucleotide exchange factor. *Genes Immun.* **11**, 447–457 (2010).
- Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. Adapting proteostasis for disease intervention. *Science* **319**, 916–919 (2008).
- Labbadia, J. & Morimoto, R. I. The biology of proteostasis in aging and disease. *Annu. Rev. Biochem.* **84**, 435–464 (2015).
- Kohonen, T. The self-organizing map. *Proc. IEEE* **78**, 1464–1480 (1990).
- Karch, C. M. & Goate, A. M. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol. Psychiatry* **77**, 43–51 (2015).
- Davies, G. et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* **9**, 2098 (2018).
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
- Ousman, S. S. et al. Protective and therapeutic role for  $\alpha$ B-crystallin in autoimmune demyelination. *Nature* **448**, 474–479 (2007).
- Shin, Y.-J. et al. Clusterin enhances proliferation of primary astrocytes through extracellular signal-regulated kinase activation. *Neuroreport* **17**, 1871–1875 (2006).
- Olah, M. et al. A transcriptomic atlas of aged human microglia. *Nat. Commun.* **9**, 539 (2018).
- Caso, F. et al. White matter degeneration in atypical Alzheimer disease. *Radiology* **277**, 162–172 (2015).
- Ciryam, P. et al. A transcriptional signature of Alzheimer's disease is associated with a metastable subproteome at risk for aggregation. *Proc. Natl Acad. Sci. USA* **113**, 4753–4758 (2016).

**Acknowledgements** We thank the study participants and staff of the Rush Alzheimer's Disease Center; T. F. Andreassen for technical assistance; and S. J. Barker for discussions and comments. This work was supported in part by the Cure Alzheimer's Fund (CAF), the JBP Foundation and by NIH grants RF1AG054321, RF1AG062377, RF1AG054012, U01NS110453, R01AG062335, and R01AG058002 (L.-H.T.); P30AG10161, R01AG15819, R01AG17917, U01AG46152, and R01AG57473 (D.A.B.); and U01NS110453, R01AG062335, R01AG058002, R01MH109978, R01HG008155, RF1AG054012, RF1AG062377, and U01MH119509 (M.K.). H.M. was supported by an Early Postdoc Mobility fellowship from the Swiss National Science Foundation (P2BSP3\_151885).

**Reviewer information** Nature thanks Hongjun Song and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Author contributions** This study was designed by H.M., J.D.-V., D.A.B., M.K., and L.-H.T., and directed and coordinated by M.K. and L.-H.T. H.M., Z.P., M.M., F.A., X.J., and A.J.M. performed the experiments. M.M. performed the RNAscope experiment under the supervision of B.P.H. H.M. and J.D.-V. performed the bioinformatics analysis with help from F.G., S.M., and L.H. H.M., J.D.-V., J.Z.Y., R.M.R., D.A.B., M.K., and L.-H.T. wrote the manuscript.

**Competing interests** The authors declare no competing interests.

### Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41586-019-1195-2>.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-019-1195-2>.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

**Correspondence and requests for materials** should be addressed to M.K. or L.-H.T.

**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 Limited 2019

## METHODS

**Data reporting.** No statistical methods were used to predetermine sample size.

**Selection of individuals from ROSMAP.** We selected a total of 48 individuals from the ROS, a longitudinal cohort study of ageing and dementia in elderly nuns, priests, and brothers. The study includes clinical data collected annually, detailed post-mortem pathological evaluations, and extensive genetic, epigenomic, transcriptomic, proteomic, and metabolomic bulk-tissue profiling<sup>20</sup>. For analyses in this paper, we took into account a total of 12 clinical, cognitive, and pathological hallmarks of AD identified through the ROS (Supplementary Table 3; ROS clinico-pathological variables can be found in the Supplementary Information). We selected 24 control individuals with no or very little pathology (no-pathology) and 24 age-matched individuals with a spectrum of mild to severe  $\beta$ -amyloid and other pathologies (AD-pathology) (Supplementary Table 1). The no-pathology group included some individuals with clinical cognitive impairment, as we selected study participants solely based on pathology and let all other variables freely associate. Details of the clinical and pathological data collection methods have been previously reported<sup>46–50</sup>. Individuals were balanced between sexes (12 each), and matched for age (medians 86.7 (AD-pathology) and 87.1 (no-pathology)) and years of education (medians 19.5 (AD-pathology) and 18 (no-pathology)). Informed consent was obtained from each subject, and the Religious Orders Study and Rush Memory and Aging Project were approved by an Institutional Review Board (IRB) of Rush University Medical Center.

**Isolation of nuclei from frozen post-mortem brain tissue.** The protocol for the isolation of nuclei from frozen post-mortem brain tissue was adapted from a previous study<sup>51</sup>. All procedures were carried out on ice or at 4 °C. In brief, post-mortem brain tissue was homogenized in 2 ml homogenization buffer (320 mM sucrose, 5 mM CaCl<sub>2</sub>, 3 mM Mg(CH<sub>3</sub>COO)<sub>2</sub>, 10 mM Tris HCl pH 7.8, 0.1 mM EDTA pH 8.0, 0.1% IGEPAL CA-630, 1 mM  $\beta$ -mercaptoethanol, and 0.4 U  $\mu$ l<sup>-1</sup> recombinant RNase inhibitor (Clontech)) using a Wheaton Dounce tissue grinder (10 strokes with the loose pestle). Homogenization buffer (3 ml) was added (to a final volume of 5 ml) and the homogenized tissue was incubated on ice for 5 min. Then the homogenized tissue was filtered through a 40- $\mu$ m cell strainer, mixed with an equal volume of working solution (50% OptiPrep density gradient medium (Sigma-Aldrich), 5 mM CaCl<sub>2</sub>, 3 mM Mg(CH<sub>3</sub>COO)<sub>2</sub>, 10 mM Tris HCl pH 7.8, 0.1 mM EDTA pH 8.0, 0.1 mM  $\beta$ -mercaptoethanol) and loaded on top of an OptiPrep density gradient (10 ml 29% OptiPrep solution (29% OptiPrep density gradient medium, 134 mM sucrose, 5 mM CaCl<sub>2</sub>, 3 mM Mg(CH<sub>3</sub>COO)<sub>2</sub>, 10 mM Tris HCl pH 7.8, 0.1 mM EDTA pH 8.0, 1 mM  $\beta$ -mercaptoethanol, 0.04% IGEPAL CA-630, and 0.17 U  $\mu$ l<sup>-1</sup> recombinant RNase inhibitor) on top of 5 ml 35% OptiPrep solution (35% OptiPrep density gradient medium, 96 mM sucrose, 5 mM CaCl<sub>2</sub>, 3 mM Mg(CH<sub>3</sub>COO)<sub>2</sub>, 10 mM Tris HCl pH 7.8, 0.1 mM EDTA pH 8.0, 1 mM  $\beta$ -mercaptoethanol, 0.03% IGEPAL CA-630, and 0.12 U  $\mu$ l<sup>-1</sup> recombinant RNase inhibitor). The nuclei were separated by ultracentrifugation using an SW32 rotor (20 min, 9,000 r.p.m., 4 °C). A total of 3 ml of nuclei was collected from the 29%/35% interphase and washed with 30 ml of PBS containing 0.04% BSA. The nuclei were centrifuged at 300g for 3 min (4 °C) and washed with 20 ml of PBS containing 0.04% BSA. Then the nuclei were centrifuged at 300g for 3 min (4 °C) and resuspended in 500  $\mu$ l PBS containing 0.04% BSA. The nuclei were counted and diluted to a concentration of 1,000 nuclei per microlitre in PBS containing 0.04% BSA.

**Droplet-based snRNA-seq.** For droplet-based snRNA-seq, libraries were prepared using the Chromium Single Cell 3' Reagent Kits v.2 according to the manufacturer's protocol (10x Genomics). The generated scRNA-seq libraries were sequenced using NextSeq 500/550 High Output v2 kits (150 cycles).

**Analysis of droplet-based snRNA-seq data.** Gene counts were obtained by aligning reads to the hg38 genome (GRCh38.p5 (NCBI:GCA\_000001405.20) using Cell Ranger software (v.2.0.0) (10x Genomics). To account for unspliced nuclear transcripts, reads mapping to pre-mRNA were counted. After quantification of pre-mRNA using the Cell Ranger count pipeline on each of the 48 individual libraries, the Cell Ranger aggr pipeline was used to aggregate all libraries and equalize the read depth between libraries before data merging (with the default parameters) to generate a gene-count matrix. Then, a cut-off value of 200 unique molecular identifiers (UMIs) was used to select single cells for further analysis. From our pilot sample analysis, we realized that the default cell-detection method used by 10x Genomics (assuming UMI values—a reflection of the RNA content—vary by roughly an order of magnitude among cells) failed to detect a large fraction of the microglia population. Therefore, to determine a more appropriate UMI cut-off value, we plotted a histogram showing cell density as a function of UMI values. On the basis of this analysis we determined 200 UMIs as the lower cut-off for cell filtering. This resulted in an initial dataset that was then further examined to exclude low-quality libraries (see 'Quality control for cell inclusion', below).

**External data sources.** Markers of human cortical layers and marker genes of cell-type clusters were obtained from ref. <sup>21</sup> and ref. <sup>18</sup>, respectively, and microglia cell-state signatures were obtained from previously published studies<sup>23,25</sup>.

**Quality control for cell inclusion.** The initial dataset contained 80,660 cells, with a median value of 1,496 total read counts over protein-coding genes. As initial reference, the entire dataset was projected onto the two-dimensional space using *t*-distributed stochastic neighbour embedding (*t*-SNE) on the top 10 principal components. The *t*-SNE coordinates were used to visualize potential biases in apparent cell similarity caused by differential cell quality. For each cell, the following quality measures were quantified: (1) the number of genes for which at least one read was mapped (which is indicative of library complexity); (2) the total number of counts; (3) the percentage of counts mapping to the top 50 genes; and (4) the percentage of reads mapped to mitochondrial genes (which may be used to approximate the relative amount of endogenous RNA, and is commonly used as a measure of cell quality). Cells with a high ratio of mitochondrial relative to endogenous RNAs had low starting amounts of RNA, which might indicate that source cells were dead or stressed and thus result in RNA degradation. Outlier cells in these quality metrics were found to cluster together in the *t*-SNE two-dimensional space. On the basis of these observations and subsequent scatter-plot analyses, cells with fewer than 200 detected genes and cells with an abnormally high ratio of counts mapping to mitochondrial genes (relative to the total number of detected genes) were removed. Specifically, given a highly skewed empirical distribution of the mitochondrial ratio values (that is, having an elbow shape clearly separating high and low scores), outlier cells were classified in two groups using the *k*-means clustering algorithm (*k* = 2) on the mitochondrial ratio, and subsequently removed. Only counts associated with protein-coding genes were considered; mitochondrially encoded genes and genes detected in fewer than 2 cells were excluded. After applying these filtering steps, the dataset included 17,926 genes profiled in 75,060 nuclei.

**Cell clustering.** All 75,060 cells were combined into a single dataset. Normalization and clustering were done with the SCANPY package<sup>52</sup>. In brief, counts for all nuclei were scaled by the total library size multiplied by 10,000, and transformed to log space. A total of 3,188 highly variable genes were identified based on dispersion and mean, the technical influence of the total number of counts was regressed out, and the values were rescaled. These preprocessing steps were performed by sequentially using the functions `normalize_per_cell`, `filter_genes_dispersion`, `log1p.regress_out`, and `scale` in SCANPY. Principal component analysis (PCA) was performed on the variable genes, and *t*-SNE was run on the top 10 principal components (PCs) using the Multicore-TSNE package (<https://github.com/DmitryUlyanov/Multicore-TSNE>). The top 50 PCs were used to build a *k*-nearest-neighbours cell-cell graph with *k* = 30 neighbours; subsequently spectral decomposition over the graph was performed with 15 components, and the Louvain graph-clustering algorithm was applied to identify cell clusters. These analyses were performed using the functions `pca`, `neighbours`, and `louvain` in SCANPY. We confirmed that a number of PCs greater than 30 captures 100% of the variance of the data. The initial pre-clustering analysis resulted in 20 pre-clusters with a median number of 2,990 cells, ranging from 413 to 15,900 cells, after excluding 2 pre-clusters of 360 and 791 cells that reflected low-quality cells (that is, cells that showed mixed cell-type markers; extreme complexity with many more genes expressed than other cells; either too many or too few reads; and, in one case, cells isolated almost exclusively from one individual). For each pre-cluster, differentially expressed genes were detected using the variance-adjusted *t*-test as implemented in the function `rank_genes_groups` in SCANPY. The top 500 ranking genes were extracted for each cluster, and used to test for overlap with markers as previously reported<sup>18</sup>. The same clustering protocol was used for both pre- and sub-clustering analyses. During sub-clustering, additional potentially spurious clusters representing low-quality or doublet cells were detected on the basis of extreme separation from the rest of the sub-clusters from the same cell type. Of these, those having a distinctly high number of total counts and mixed expression of markers from different cell types were tagged as potential doublets and not considered for downstream analyses, resulting in a total of 70,634 cells.

**Cell type annotation and sub-clustering.** For each pre-cluster, we assigned a cell-type label using statistical enrichment for sets of marker genes<sup>18,21</sup>, and manual evaluation of gene expression for small sets of known marker genes. Enrichment was statistically assessed using the hypergeometric distribution (Fisher's exact test) and FDR correction over all gene sets and pre-clusters. Broad cell-type clusters were defined by grouping together all pre-clusters corresponding to the same cell type. Sub-clustering analysis was performed independently over each broad cell-type cluster.

**Sub-cluster-trait associations.** Cell sub-clusters were annotated for phenotypic traits and AD-pathology status by assessing, within each sub-cluster, the overrepresentation of cells isolated from individuals annotated with the corresponding trait. For categorical phenotypic variables, enrichment was evaluated using the hypergeometric distribution (Fisher's exact test) and FDR correction over all gene sets and sub-clusters. Enrichment or depletion of quantitative pathological variables was assessed individually by contrasting the average observed value across the cells of a given sub-cluster with a corresponding null distribution estimated by randomly resampling sub-cluster label assignments and computing an average

score 10,000 times. The deviation of the observed value from the random expected distribution was quantified using a  $z$ -score statistic. Statistical overrepresentation analysis for cells isolated from each individual across the sub-clusters (Fisher's exact test) verified that individual sub-clusters were not exclusively composed of cells from one or a few subjects (Extended Data Fig. 7b). Robustness of sub-cluster–trait associations was assessed using a randomization study in which cells from 3 female and 3 male subjects randomly chosen from each pathological category (AD-pathology and no-pathology) were subjected to a computing enrichment analysis for 100 trials. Aggregate  $P$  value calculations were conducted by computing a meta- $p$ -value using the R package *metap* (Extended Data Fig. 7d).

**Phenotypic analysis.** Post-mortem scores on AD-related pathologic indices for amyloid, Braak staging, CERAD score, cognitive diagnosis, global cognitive function, global AD-pathology burden, neurofibrillary tangle burden, NIA–Reagan score, diffuse and neuritic plaque burden, and neuronal neurofibrillary tangle density were used to quantitatively describe the pathological phenotype of each individual (Supplementary Information). The quantitative values tangles, NFT, gpath (aggregate of neuritic plaque, diffuse plaque, and neurofibrillary tangle scores), plaq\_n, amyloid, and cogn\_global\_lv were selected as the primary markers for quantifying AD progression. To identify AD pathological groups, distance-based  $k$ -means clustering was applied to the quantitative pathological features using  $k = 3$ . The number of clusters was selected by testing multiple increasing values of  $k$ , and taking the value at which the within-cluster sum-of-square distance dropped.

**Marker identification.** For sub-clusters, a set of markers (specifically over-expressed) genes was defined by a differential expression analysis of the cells grouped in each sub-cluster against the remaining cells within the corresponding broad cell-type cluster. This analysis was applied to all cell types independently. Significantly overexpressed genes were defined based on the Wilcoxon rank-sum test with a FDR-corrected  $P$  value  $\leq 0.01$  and a  $\log_2$ (mean gene expression across cells in sub-cluster/mean gene expression across cells in other sub-clusters) of 0.5. Only genes detected in at least 25% of the cells within the given sub-cluster were considered.

Gene Ontology enrichment analyses were performed using the R package *gProfileR* (<https://cran.r-project.org/web/packages/gProfileR/index.html>) and using the  $P$  value-ranked gene lists as input. For all cases we used the set of 17,926 protein-coding genes included in the quality control data as background.

**Differential gene-expression analysis.** snRNA-seq-based differential expression analysis was assessed using two tests. First, a cell-level analysis was performed using the Wilcoxon rank-sum test and FDR multiple-testing correction. Second, a Poisson mixed model accounting for the individual of origin for nuclei and for unwanted sources of variability was performed using the R packages *lme4* and *RUV-seq*, respectively. The consistency of DEGs detected using the cell-level analysis model with those obtained with the Poisson mixed model was assessed by comparing the directionality and rank of DEGs in the two models. Consistency in directionality for all cell types was measured by counting the fraction of the top 1,000 DEGs (ranked by FDR scores) detected in cell-level analysis that showed consistent direction in the mixed model. High consistency was found, with a median fraction of 0.99 (Supplementary Table 2). Global consistency between the two models was assessed statistically using a resampling test. We tested whether the differential  $P$  value and  $z$ -score ranks corresponding to genes detected as upregulated or downregulated in the cell-level analysis were significantly higher or lower than those that were expected by chance when computed using the mixed model. Expected scores were estimated by randomly sampling same-sized gene sets ( $n = 1,000$  replicates). Both significance rank and direction deviated significantly from expectation, with directionality consistent for up- or downregulated genes. Results from the consistency tests and from both the cell-level and mixed-model differential tests are included in Supplementary Table 2. For analyses involving DEG counts, only genes that were significantly supported by both models using the criteria FDR-corrected  $P < 0.01$  in a two sided Wilcoxon-rank sum test, absolute  $\log_2$ (mean gene expression in AD category  $x$ /mean gene expression in AD category  $y$ )  $> 0.25$ , and FDR-corrected  $P < 0.05$  in a Poisson mixed model were considered. Owing to their low cell counts, pericyte and endothelial cell populations were excluded from differential analyses.

Bulk RNA-seq differential analysis was performed by fitting a linear model using the R package *limma*<sup>53</sup>, accounting for the covariates age, RNA integrity number, post-mortem interval, and plate batches. The pathological definition of AD groups provided by ROSMAP was used. This classification defines AD or non-AD groups based exclusively on the overall pathological burden, without considering clinical diagnosis. We used the labels AD-pathology and no-pathology for consistency.

Consistency of gene-expression perturbations in the different cell types observed in snRNA-seq data with those detected in tissue-level bulk RNA-seq was assessed using a resampling approach. To test whether the genes identified as DEGs in single-cell data were also detected as high ranking in the differential analysis in bulk data, a  $z$ -score statistic was computed to quantify the deviation of the observed differential ( $P$  value) rank scores obtained in the bulk analysis for the

genes detected as DEGs in single cells, relative to those observed in 1,000 randomly chosen gene sets (Supplementary Table 2). This analysis was performed for each cell type independently.

**Correlation analysis of gene expression and AD-related neuropathological traits using SOMs.** First, for each major cell type, a gene-wise correlation coefficient (Spearman's rank correlation coefficient) was computed using gene expression and AD-related neuropathological trait values across all the annotated cells as variables. The AD-related neuropathological traits included in this analysis were cogn\_global\_lv, age\_death (age at death), educ (years of education), msex (self-reported sex), parksc\_lv (global Parkinsonian summary score (last valid score)), gpath, gpath\_3neocort (global measure of neocortical pathology), pmi (post-mortem interval), amyloid, plaq\_d (diffuse plaque burden), plaq\_n, NFT, and tangles. Only significant correlations after Bonferroni correction at  $P < 0.01$  were considered. The resulting correlation matrices for each major cell type were concatenated and analysed using a computational algorithm (SOM)<sup>37</sup>. All SOMs were created using the kohonen R package<sup>54</sup>. To identify the territories of the SOM most strongly correlated with AD-related neuropathological traits, we used an image segmentation method and further manual curation to identify territories (gene–trait correlation modules) based on all the individual cell-type-specific SOM plots for each neuropathological trait. Enrichment analysis for Gene Ontology terms among the genes of a gene–trait correlation module was performed using Metascape<sup>55</sup>. The robustness of gene–trait associations to single-cell heterogeneity and noise was confirmed examining individual-level correlations for the genes in a gene–trait module. Individual-level correlations were computed by first averaging for each individual the normalized gene-expression profiles across cells of the same cell. This resulted in cell-type-specific averaged gene-expression profiles across the 48 individuals. Average profiles were subsequently mean-centred and scaled to finally compute gene-wise correlation coefficients versus corresponding pathological values. Individual-level gene–trait correlations were computed independently for all 48 individuals, for only the 24 male individuals, and for only the 24 female individuals. The robustness of gene–trait associations to potential confounding variables was corroborated by confirming the cell-type-specific recovery of identified gene–trait modules using partial correlation. In brief, the partial Pearson's correlation coefficient between average gene expression and each pathological trait, after correcting for the effects of post-mortem interval, age, gender, and education level of each individual, was computed by first orthogonalizing the normalized expression with respect to the normalized covariates and then computing the correlation in the residual subspace.

**Immunohistochemistry.** Fixed human brain tissue (prefrontal cortex, BA10) was sectioned at 50  $\mu$ m using a vibratome (Leica). To retrieve the antigens, sections were incubated at 95 °C in immunohistochemistry (IHC) antigen retrieval solution (ThermoFisher Scientific; 00-4955-58) containing 0.05% Tween-20 for 45 min and then placed in PBS for 20 min at room temperature. After washing with PBS (3  $\times$  10 min), the brain sections were incubated in quenching solution (50 mM ammonium acetate, 100 mM CuSO<sub>4</sub>) at room temperature overnight. After washing with double-distilled water (1  $\times$  15 min) and PBS (3  $\times$  15 min), the sections were permeabilized in PBS containing 0.3% Triton X-100 for 10 min and blocked in PBS containing 0.3% Triton X-100 and 5% normal donkey serum at room temperature for 1 h. The sections were incubated overnight (anti- $\beta$ -amyloid (D54D2) antibody) or for 48 h (anti-IBA1 and anti-human HLA-DP, HLA-DQ, and HLA-DR antibodies) at 4 °C in primary antibody in PBS with 0.3% Triton X-100 and 5% normal donkey serum. Primary antibodies were anti-IBA1 (1:500; Synaptic Systems; 234 004, polyclonal guinea pig anti-serum), anti-human HLA-DP, HLA-DQ, and HLA-DR antigens, clone CR3/43 (1:100, Agilent; M077501-2), and anti- $\beta$ -amyloid (D54D2) (1:500, Cell Signaling Technology; 8243). The sections were washed with PBS containing 0.1% Triton X-100 at room temperature (4  $\times$  15 min), and then incubated with secondary antibodies (dilution 1:2,000) overnight at 4 °C. Primary antibodies were visualized with Alexa-Fluor 488, Alexa-Fluor 594, and Alexa-Fluor 647 antibodies (Molecular Probes), and cell nuclei were visualized with Hoechst 33342 (Sigma-Aldrich; 94403). The sections were washed with PBS containing 0.1% Triton X-100 at room temperature (4  $\times$  15 min) and then mounted on Fisherbrand Superfrost Plus microscope slides in ProLong Gold Antifade Mountant. Images were acquired using a confocal microscope (LSM 880; Zeiss) with a 20 $\times$  or 63 $\times$  objective.

For tyramide signal amplification (TSA) labelling, sections were incubated at 95 °C in IHC antigen retrieval solution (ThermoFisher Scientific; 00-4955-58) containing 0.05% Tween-20 for 45 min and then placed in PBS for 20 min at room temperature. After washing with PBS (3  $\times$  10 min), the brain sections were incubated in quenching solution (50 mM ammonium acetate, 100 mM CuSO<sub>4</sub>) at room temperature overnight. After washing with double-distilled water (1  $\times$  15 min) and PBS (3  $\times$  15 min), endogenous peroxidases were quenched with 0.3% H<sub>2</sub>O<sub>2</sub> for 20 min at room temperature followed by washing with PBS (3  $\times$  10 min). Then the sections were blocked in TNB blocking buffer (0.1 M Tris HCl pH 7.5, 0.15 M NaCl, and 0.5% TSA blocking reagent (PerkinElmer; FP1020)) for 30 min at room



temperature. Primary antibody incubation was performed for 48 h at 4°C in TNB blocking buffer. The sections were washed with PBS containing 0.3% Triton X-100 at room temperature (4 × 15 min) and then incubated with horseradish peroxidase (HRP)-labelled secondary antibody (1:1,500) in TNB blocking buffer for 30 min at room temperature. Then the sections were washed with PBS containing 0.3% Triton X-100 at room temperature (4 × 15 min) and the fluorophore reaction was performed using the TSA Plus Cyanine 5 and Fluorescein System (PerkinElmer; NEL754001KT) for 3 min at room temperature. Then the sections were washed with PBS containing 0.3% Triton X-100 at room temperature (4 × 15 min) and incubated at 95°C in IHC antigen retrieval solution (ThermoFisher Scientific; 00-4955-58) containing 0.05% Tween-20 for 45 min. After washing with PBS containing 0.3% Triton X-100 at room temperature (4 × 15 min) the sections were either mounted or blocked in TNB blocking buffer for another round of TSA labelling. The following primary antibodies were used: anti-OLIG2 (1:1,000; Atlas; HPA003254; rabbit polyclonal), anti-CRYAB (1:200; LSBio; LS-B3696; rabbit polyclonal), and anti-QDPR (1:2,500; Atlas; HPA065649; rabbit polyclonal). The following secondary antibody was used: HRP-labelled anti-rabbit IgG (goat) (PerkinElmer; NEF812001EA). Images were acquired using a confocal microscope (LSM 880; Zeiss) with a 20× or 63× objective.

**RNA in situ hybridization.** Frozen brain tissue was embedded in Tissue-Tek OCT compound (VWR; 25608-930), cryo-sectioned to 16-μm thickness, and placed onto Fisherbrand Superfrost Plus microscope slides (ThermoFisher Scientific; 12-550-15). The RNAscope 2.5 HD Duplex Detection Kit (Chromogenic) was used according to the manufacturer's instructions, with minor modifications to tissue preparation. In brief, sections frozen at -80°C were washed with 1 × PBS, baked at 60°C for 15 min, and fixed in 10% neutral buffered formalin for 90 min at room temperature. Following fixation, the tissue was dehydrated using 50%, 70% and 100% ethanol and then baked at 60°C for 30 min to avoid detachment. Tissues were then treated with H<sub>2</sub>O<sub>2</sub> for 10 min, followed by protease IV for 30 min. Brain sections were hybridized with two mRNA probes: *SLC17A7* (RNAscope probe Hs-SLC17A7-C2; ACD; 415611-C2) and *NTNG1* (RNAscope probe Hs-NTNG1; ACD; 446101). Probes in channel 1 (*NTNG1*) were labelled with HRP enzyme and visualized with a green substrate; and probes in channel 2 (*SLC17A7*) were labelled with alkaline phosphatase enzyme and a red substrate. Haematoxylin was used to mark cell nuclei in blue. Slides were then imaged using a Leica Dmi8 microscope to take 40× bright-field images. For the quantification, a blinded researcher manually counted the cells labelled with red substrate (*SLC17A7*) and determined how many of these cells were co-labelled with green substrate (*NTNG1*). The quantification was based on the analysis of BA10 tissue sections from four low-amyloid individuals and four high-amyloid individuals, with five to six images per individual.

**Isolation of neuronal nuclei, cDNA synthesis and RT-qPCR analysis.** Post-mortem brain tissue was homogenized in 2 ml NF1 buffer (0.5% Triton X-100, 0.1 M sucrose, 5 mM MgCl<sub>2</sub>, 1 mM EDTA, 10 mM Tris HCl pH 8.0, 1 mM β-mercaptoethanol, 0.4 U μl<sup>-1</sup> recombinant RNase inhibitor (Clontech)) using a Wheaton Dounce tissue grinder (15 strokes with the loose pestle). After homogenization, an additional 8 ml of NF1 buffer was added and the samples were inverted 10 times to mix. The homogenate was passed through a 40-μm cell strainer (VWR; 21008-949) and centrifuged at 300g for 3 min at 4°C. Then the supernatant was removed and the nuclei were incubated with Alexa-Fluor-488-conjugated anti-NeuN antibody (Millipore Sigma; MAB377X) in PBS containing 1% BSA for 15 min at 4°C on a shaker. The nuclei were washed by adding 35 ml PBS containing 1% BSA, spun down at 300g for 3 min at 4°C, resuspended in 0.5 ml PBS containing 0.04% BSA and 0.4 U μl<sup>-1</sup> recombinant RNase inhibitor (Clontech), and stained with NucBlue Live ReadyProbes Reagent (ThermoFisher Scientific; R37605). NeuN-positive nuclei were then directly sorted into RNA lysis buffer (Qiagen, 74134) using fluorescence-activated cell sorting. Total RNA was extracted using the RNeasy Plus Mini Kit (Qiagen; 74134) according to the manufacturer's instructions, and reverse transcribed using the iScript cDNA Synthesis Kit (Bio-Rad; 170-8891). For gene-expression analysis, cDNA was quantitatively amplified on a thermal cycler (BioRad) using SsoFast EvaGreen Supermix (BioRad; 1725202) and gene-specific primers (*DHFR* forward, GCTGGAGTATTGATCCCGCC, *DHFR* reverse, CTGGACTATGTTCCGCCAC; *STMN1* forward, ATACACT GCCTGTCGCTTGT, *STMN1* reverse, CTTTGTACCGAGGGCTGAGA; *MTRNR2L8* forward, ATTGACCTGCCCCGTGAAGAG, *MTRNR2L8* reverse, AGGGCCCTGTGGACTTGTTAAG; *LINGO1* forward, ACCGCATC AAAACGCTCAAC, *LINGO1* reverse, CTAGCGGGATGAGCTTCAGG; *BEX3* forward, ATGGGCCATACCAATAGGC, *BEX3* reverse, AGAGAGCTCC CCCATAAGGA; *BEX1* forward, CGCTACTCGTGTCTCGCTAC, *BEX1* reverse, CCATTACTCCTGGCCCTATCC; *RPL13* forward, CCTTCCGCTCGG CTGTTTT, *RPL13* reverse, GGCCTTACGTCTGCGGATCT). The comparative C<sub>t</sub> method was used to examine differences in gene expression and values were normalized to expression levels of *RPL13*.

### Statistical analysis of white-matter data from ROSMAP cohort studies.

Measures of WMH were obtained through in vivo brain MRI data collected through the ROSMAP cohort studies. Participants undergo MRI in the Rush Alzheimer's Disease Center (RADC) cohort studies and substudies biennially on 3T scanners including the T1-weighted 3D Magnetization Prepared Rapid Acquisition Gradient Echo (MPRAGE) and T2-weighted 2D Fluid-Attenuated Inversion Recovery (FLAIR) sequences. Raw MPRAGE images were processed to generate total volumes including grey matter, white matter, cerebrospinal fluid and intracranial volumes using the statistical parametric mapping (SPM) package<sup>56</sup>. White-matter lesions that appeared hyperintense in T2-weighted images were segmented based on FLAIR and MPRAGE data using BIANCA<sup>57</sup>. A mask of WMH was generated and the total volume of hyperintensities calculated. Cognitive scores were taken from the recorded variable *cogn\_global\_lv* (Supplementary Information), which measures averaged *z*-scored values from a battery of 19 cognitive tests, yielding a global cognitive function summary. A negative *z*-score simply means that someone has an overall score that is lower than the average of the entire cohort at baseline. High- and low-cognition groups were defined based on whether a subject has an overall score lower (low-cognition, *z*-score <0) or higher (high-cognition, *z*-score >0) than the average. WMH values were compared between low- and high-cognition groups in male and female subjects, using the Wilcoxon rank-sum test (Extended Data Fig. 10d). Statistical estimation of significant difference in WMH between low-cognition and high-cognition groups in females, and between low-cognition and high-cognition groups in males, was assessed by bootstrap point and 95% confidence interval estimation of the effect size (mean difference) between groups, correcting for sample size (resampling *n* = 40 observations per group 1,000 times) (Extended Data Fig. 10e).

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

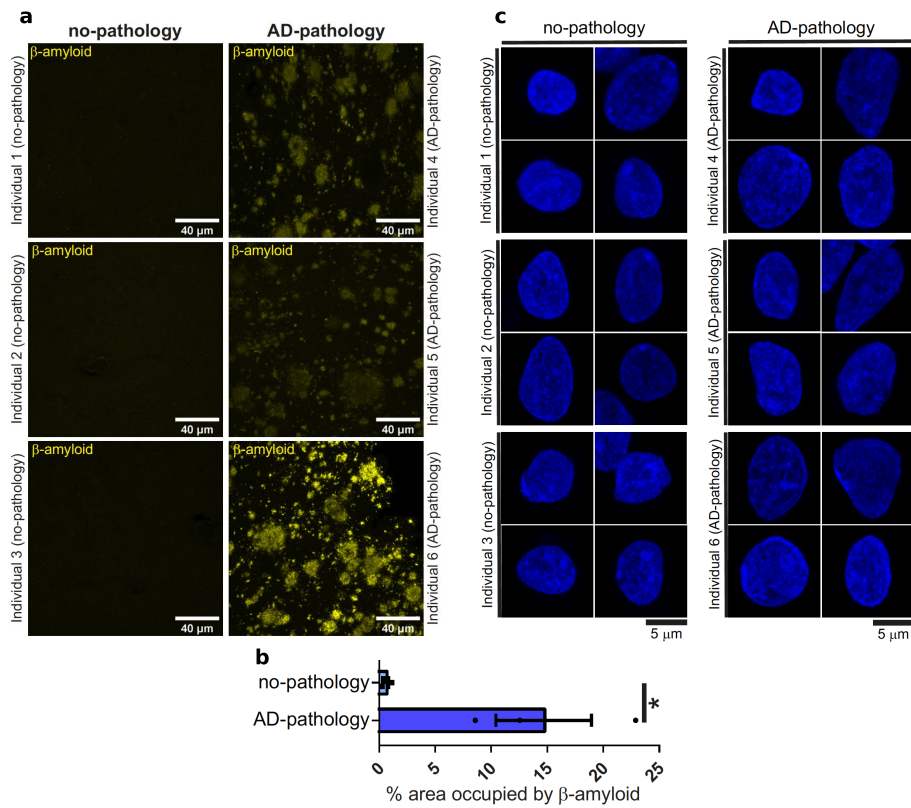
### Data availability

The snRNA-seq data are available on The Rush Alzheimer's Disease Center (RADC) Research Resource Sharing Hub at <https://www.radc.rush.edu/docs/omics.htm> (snRNA-seq PFC) or at Synapse (<https://www.synapse.org/#!Synapse:syn18485175>) under the doi 10.7303/syn18485175. The ROSMAP metadata can be accessed at <https://www.synapse.org/#!Synapse:syn3157322>. The data are available under controlled use conditions set by human privacy regulations. To access the data, a data use agreement is needed. This registration is in place solely to ensure anonymity of the ROSMAP study participants. A data use agreement can be agreed with either Rush University Medical Center (RUMC) or with SAGE, who maintains Synapse, and can be downloaded from their websites.

### Code availability

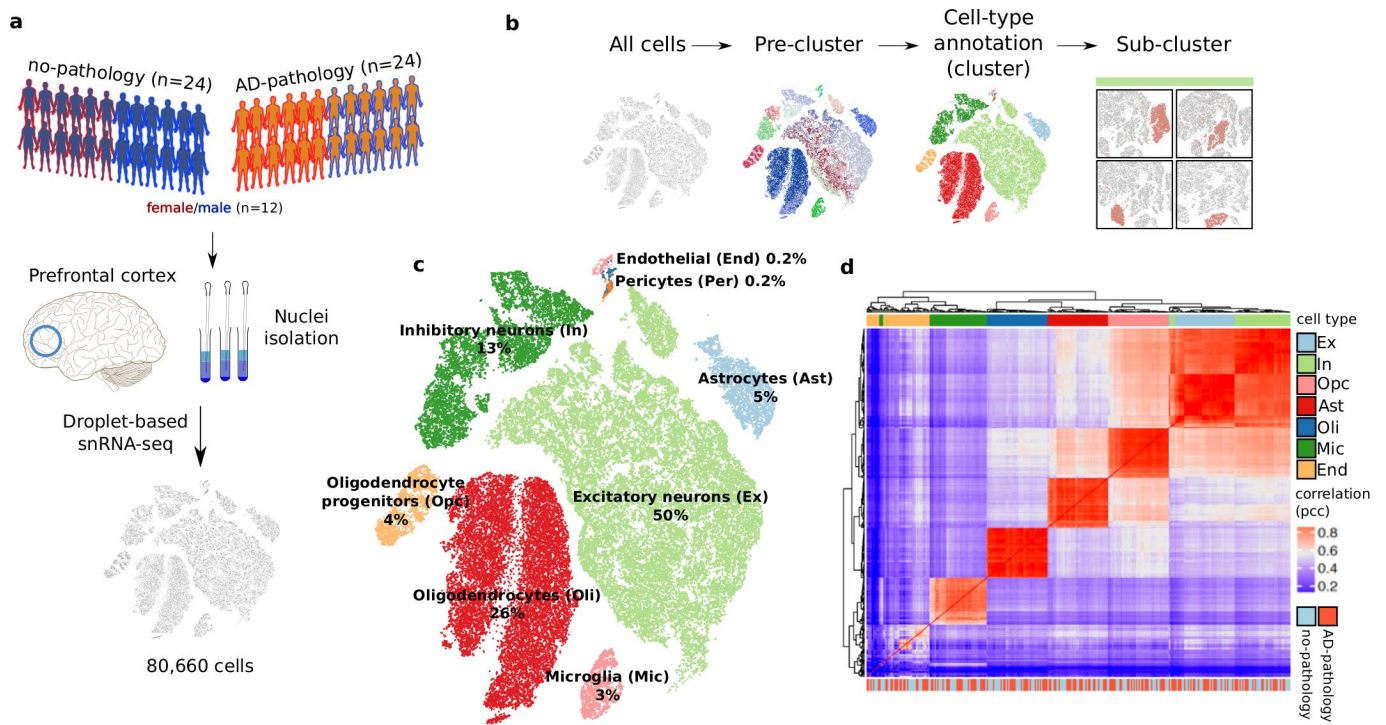
Code used throughout this study is available upon reasonable request from the corresponding authors.

- Bennett, D. A. et al. Natural history of mild cognitive impairment in older persons. *Neurology* **59**, 198–205 (2002).
- Bennett, D. A. et al. Apolipoprotein E ε4 allele, AD pathology, and the clinical expression of Alzheimer's disease. *Neurology* **60**, 246–252 (2003).
- Bennett, D. A., Schneider, J. A., Wilson, R. S., Bienias, J. L. & Arnold, S. E. Neurofibrillary tangles mediate the association of amyloid load with clinical Alzheimer disease and level of cognitive function. *Arch. Neurol.* **61**, 378–384 (2004).
- Bennett, D. A. et al. Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology* **66**, 1837–1844 (2006).
- Bennett, D. A. et al. Decision rules guiding the clinical diagnosis of Alzheimer's disease in two community-based cohort studies compared to standard practice in a clinic-based cohort study. *Neuroepidemiology* **27**, 169–176 (2006).
- Swiech, L. et al. In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nat. Biotechnol.* **33**, 102–106 (2015).
- Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
- Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
- Wehrens, R. & Buydens, L. M. C. Self- and super-organizing maps in R: the kohonen package. *J. Stat. Softw.* **21**, 1–19 (2007).
- Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
- Friston, K. J. et al. Spatial registration and normalization of images. *Hum. Brain Mapp.* **3**, 165–189 (1995).
- Griffanti, L. et al. BIANCA (Brain Intensity AbNormality Classification Algorithm): a new tool for automated segmentation of white matter hyperintensities. *Neuroimage* **141**, 191–205 (2016).



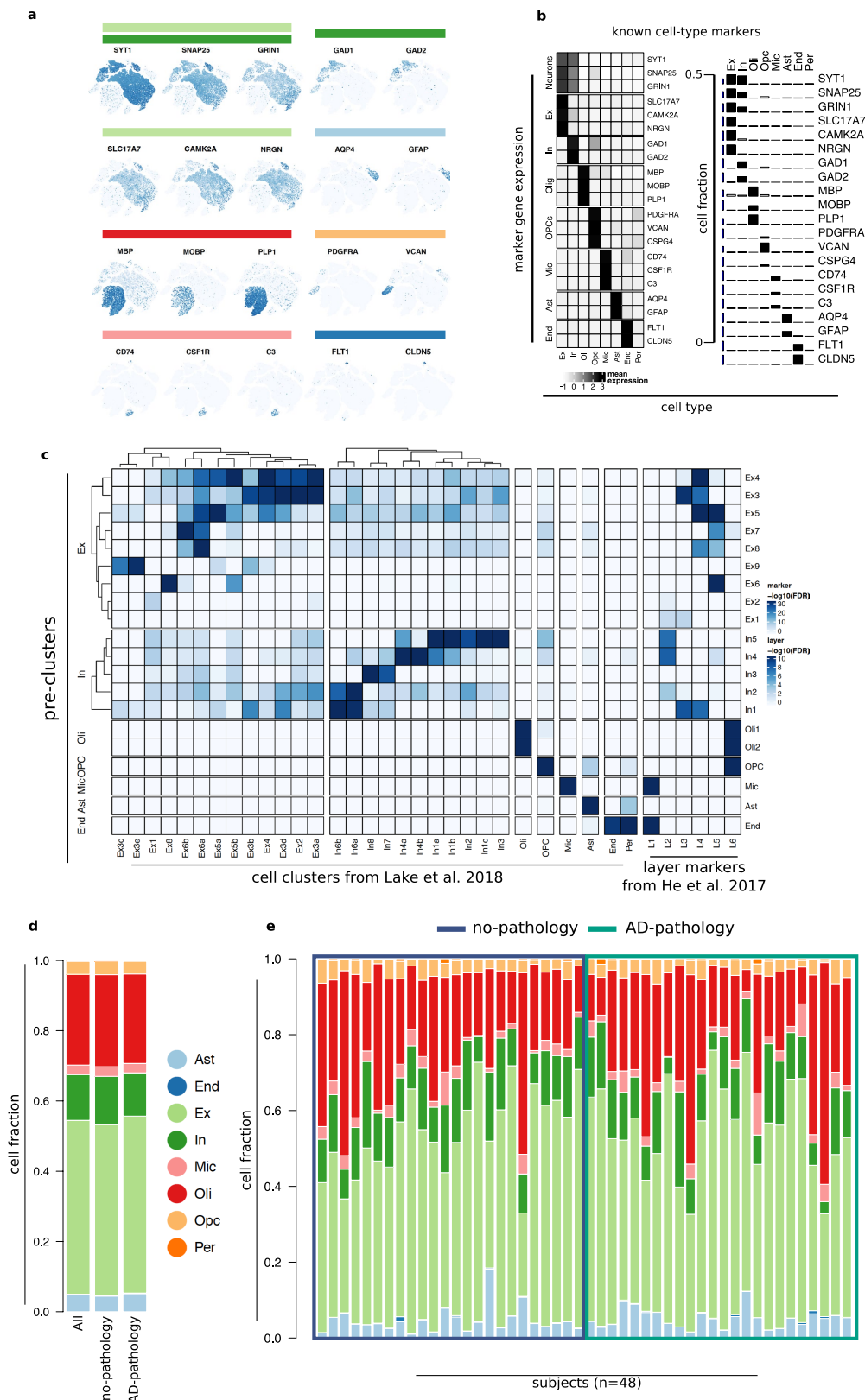
**Extended Data Fig. 1 | Pathological status verification and physical integrity of isolated neuronal nuclei. a**, Immunohistochemistry with anti- $\beta$ -amyloid antibody (D54D2, yellow) in the grey matter of Brodmann area 10 of no-pathology and AD-pathology individuals. **b**, Quantification of the  $\beta$ -amyloid immunostaining in **a**. Data are mean  $\pm$  s.e.m.;

\* $P = 0.030$  (Student's two-tailed  $t$ -test). **c**, High-resolution confocal microscopy images of neuronal nuclei isolated from no-pathology and AD-pathology individuals and stained with Hoechst. The experiment was performed once.



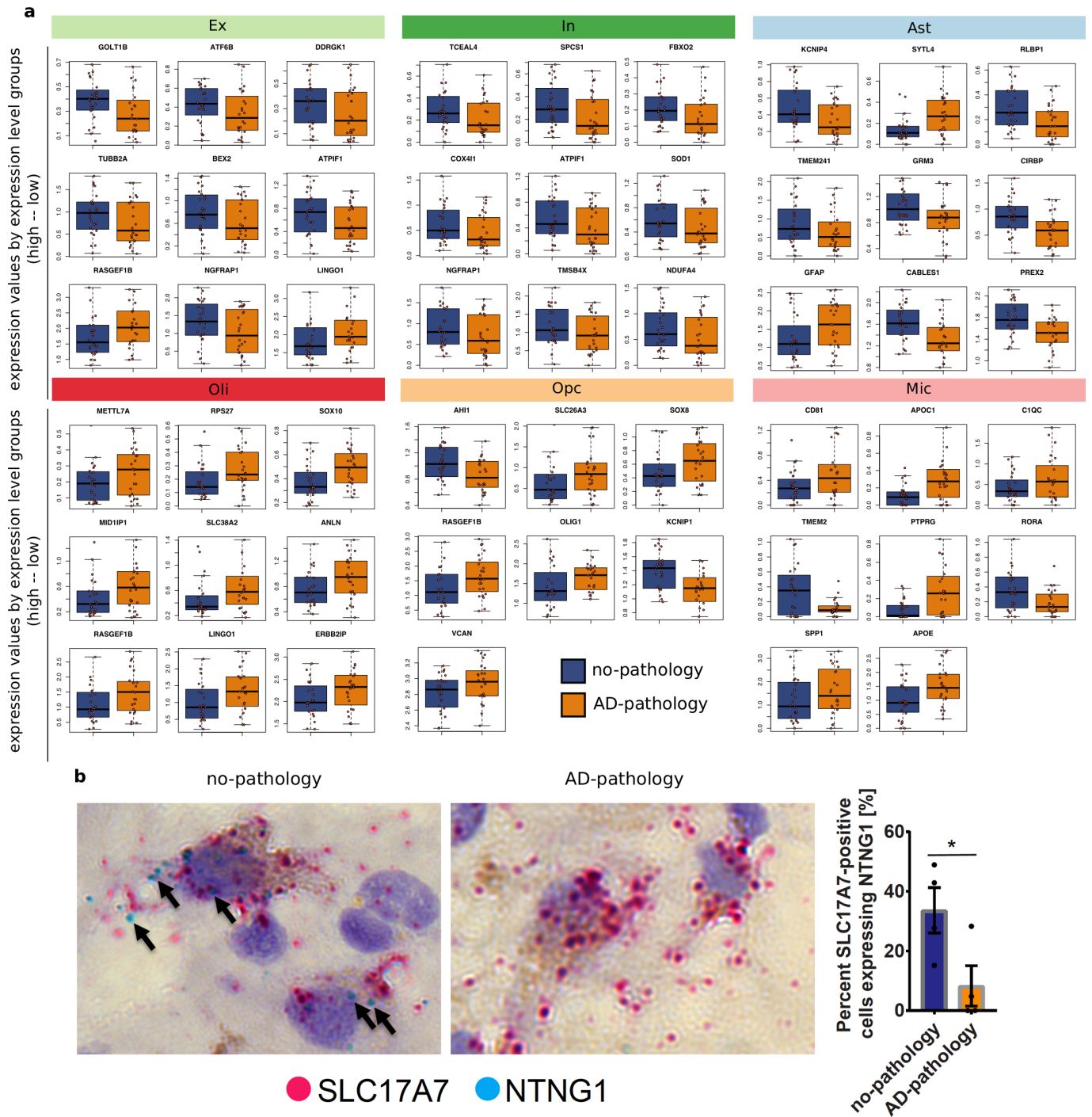
**Extended Data Fig. 2 | snRNA-seq profiling and cell-type characterization.** **a**, Study cohort and sample preparation. **b**, Clustering analysis workflow. **c**, Two-dimensional *t*-SNE projection of all annotated

cells ( $n = 75,060$  from 24 pathology and 24 no-pathology individuals). **d**, Correlation matrix (Pearson correlation coefficient; pcc) of the average expression profiles by cell type for each individual.



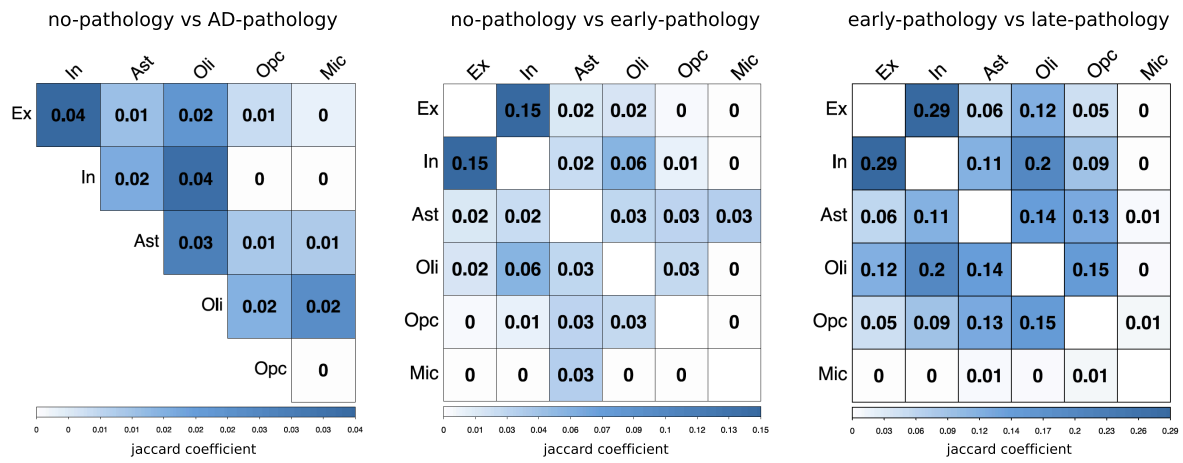
**Extended Data Fig. 3 | Consistency of cells of the same type across individuals.** **a**, Expression of known cell-type marker genes for each cell type. **b**, Left, expression of known cell-type marker genes in each cluster; right, fraction of cells in each cluster that express each marker gene. Vertical dashed blue line represents a scale bar referencing 0.5. **c**, Overrepresentation analysis (hypergeometric test) within each of the

pre-cluster marker sets (rows) of genes previously identified as markers<sup>18</sup> (columns, left;  $n = 1,729$  total genes) and genes previously identified as markers of cortical layers<sup>21</sup> (columns, right;  $n = 3,400$  total genes). **d**, Fraction of cells of each type isolated across all ( $n = 48$ ), no-pathology ( $n = 24$ ), and AD-pathology ( $n = 24$ ) individuals. **e**, Fraction of cells of each type isolated from each individual (columns;  $n = 48$ ).



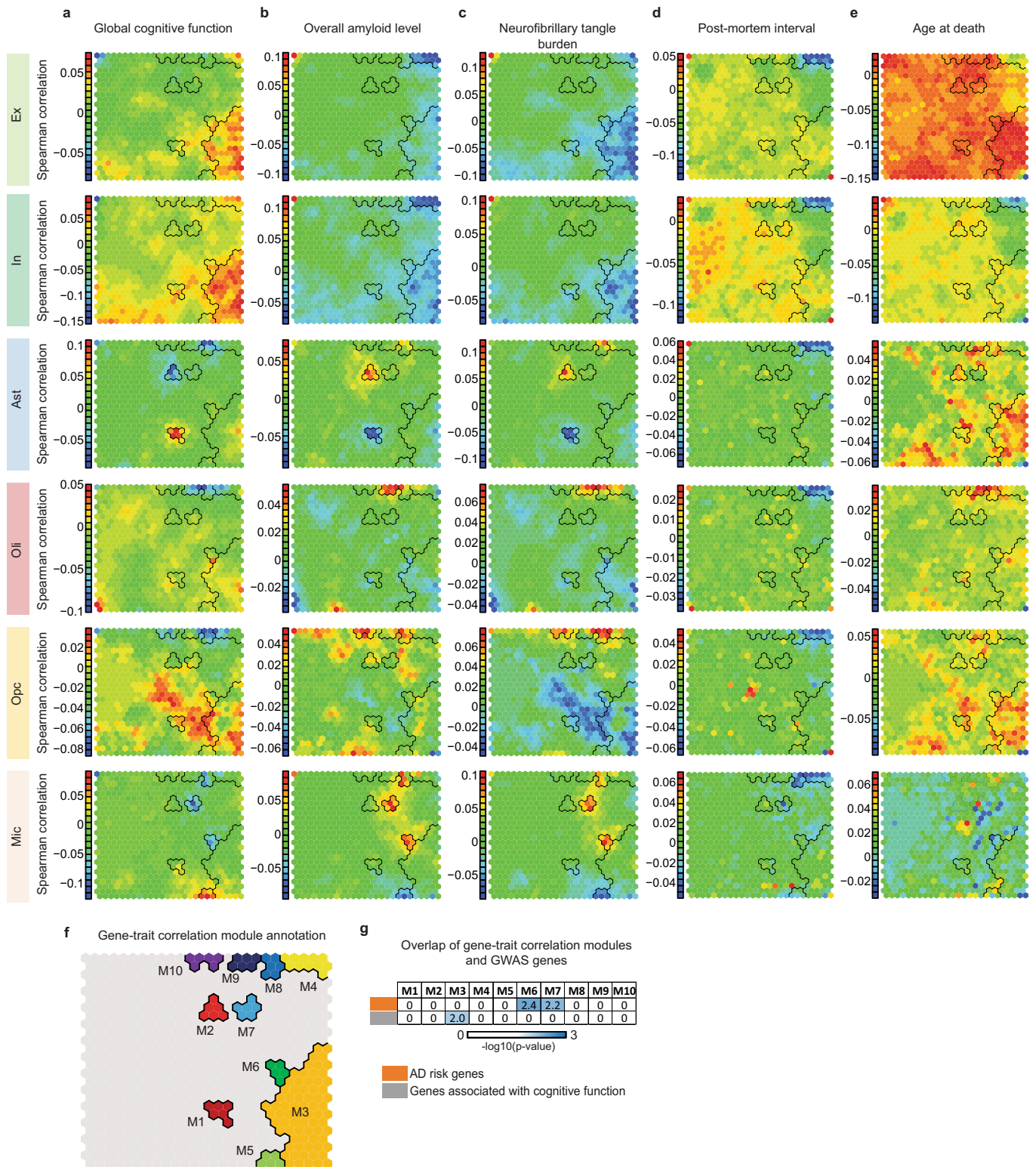
**Extended Data Fig. 4 | Expression values and validation of top DEGs.**  
**a**, Mean expression values of genes across the nuclei isolated from each individual. Each point represents one individual. DEGs were classified as low, mid, or high in expression, based on their median expression level across the cells of the corresponding cell type. Groups were defined based on *k*-means clustering (*k* = 3). The top three genes for each group (low-, mid-, and high-expression levels) and for each cell type are shown. For oligodendrocyte precursor cells and microglia, only one and two genes, respectively, were classified within the high-expression group. *ATPIF1* is also known as *ATP5IF1*, *NGFRAP1* is also known as *BEX3*,

*TMEM2* is also known as *CEMP2*, and *ERBB2IP* is also known as *ERBIN*.  
**b**, Left, RNA in situ hybridization (RNAscope) with probes that detect the excitatory neuron marker *SLC17A7* (red) and *NTNG1* (blue) in the grey matter of Brodmann area 10 of a no-pathology and an AD-pathology individual. The tissue was counterstained with haematoxylin. Right, quantification of RNA in situ hybridization on Brodmann area 10 tissue sections. Data are mean  $\pm$  s.e.m.; \**P* = 0.047 (Student's two-tailed *t*-test). *n* = 4 no-pathology and *n* = 4 AD-pathology individuals; *n* = 5 or 6 images per individual.



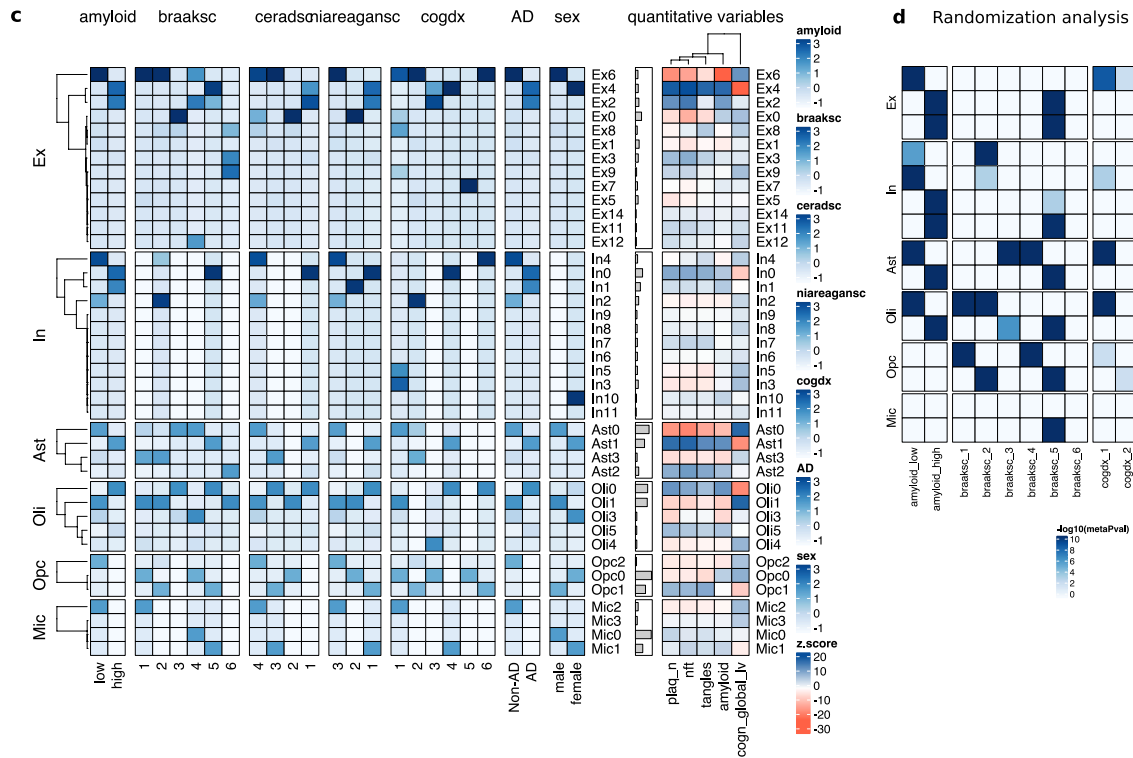
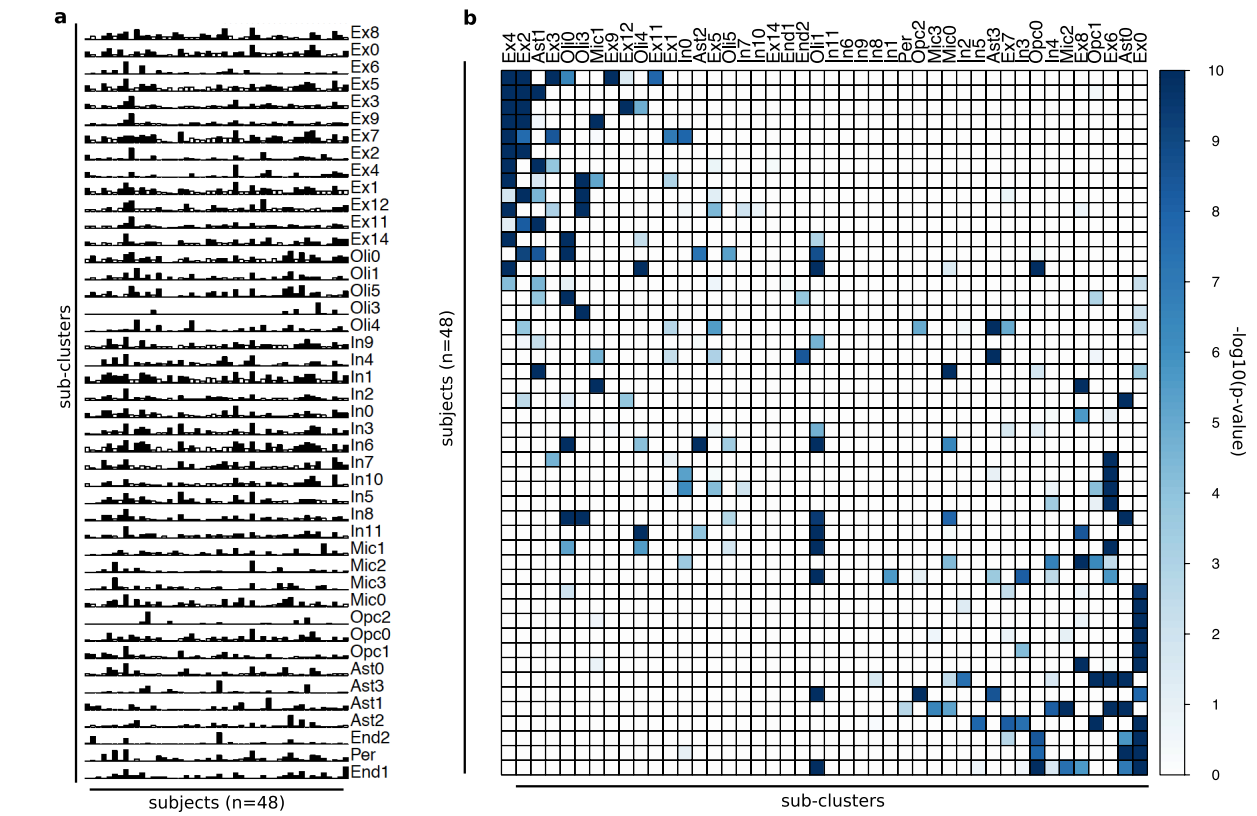
**Extended Data Fig. 5 | Overlap of genes that are altered in the progression of AD pathology.** Quantification of the overlap (Jaccard coefficient) between pairs of gene sets identified as differentially expressed

in each of the major cell types when comparing cells isolated from AD-pathology individuals with cells isolated from no-pathology individuals, and combinations of early- and late-pathology individuals.



**Extended Data Fig. 6 | Cell-type-specific and phenotype-specific gene-trait correlation analysis.** a–e, SOM generated from transcriptome-wide gene-expression correlation of each gene with neuropathological signatures of AD. Genes with similar correlation patterns are mapped to the same SOM unit and similar units group close together. SOM grid layout is common and built jointly across all phenotypes and all cell types. Colour indicates the average Spearman’s rank correlation for genes in each

unit. f, Selected SOM territories (M1–M10). g, Overlap (one-sided Fisher’s exact test) between gene-trait correlation-module genes ( $n = 1,472$  genes (M3),  $n = 70$  (M6),  $n = 80$  (M7)) and AD GWAS-risk genes (top;  $n = 28$  genes), as well as genes associated with general cognitive function (bottom;  $n = 709$  genes). The  $P$  values have been adjusted for multiple hypothesis testing;  $-\log_{10}(\text{Bonferroni-corrected } P \text{ values})$  are shown.

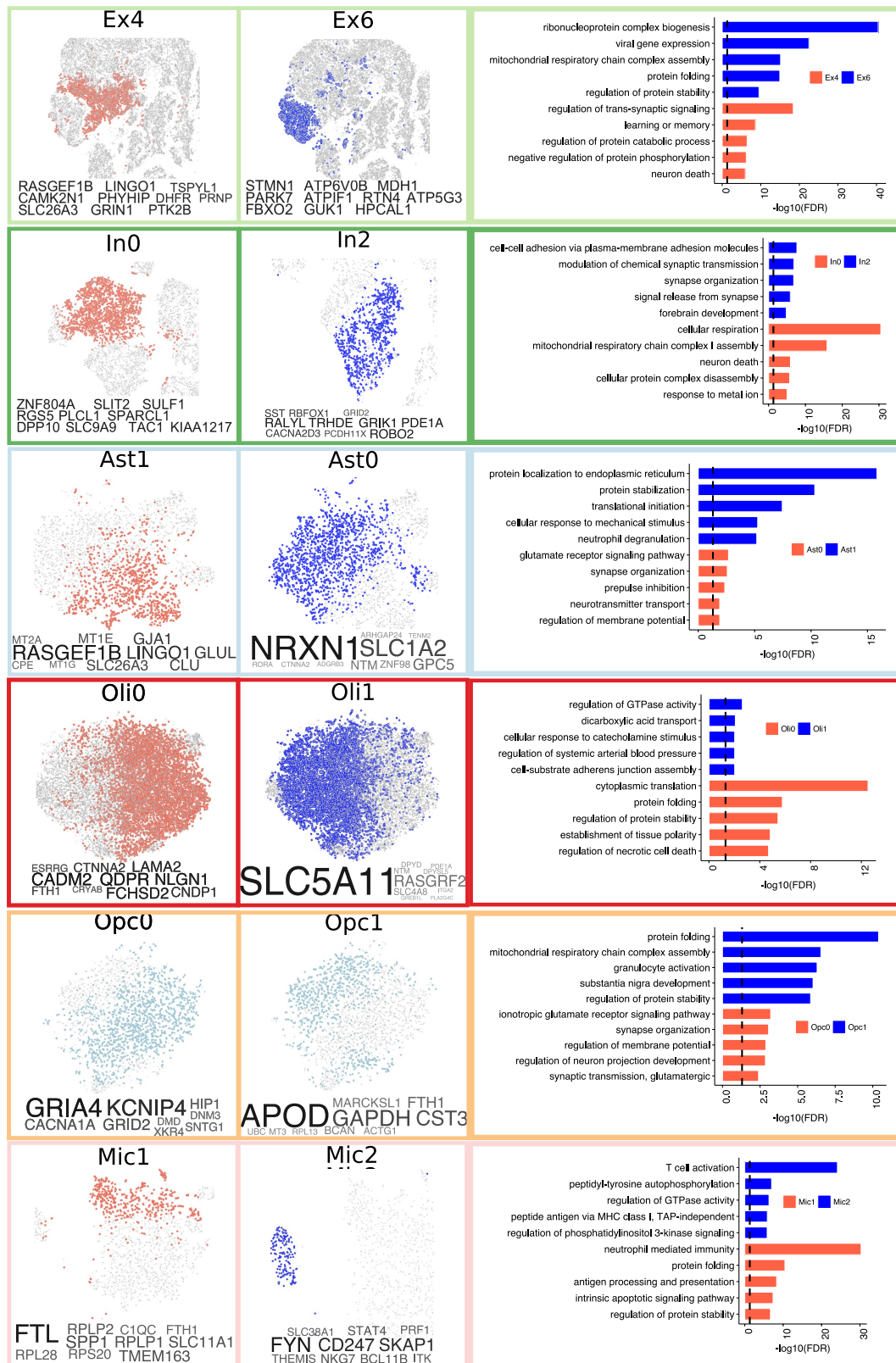


Extended Data Fig. 7 | See next page for caption.



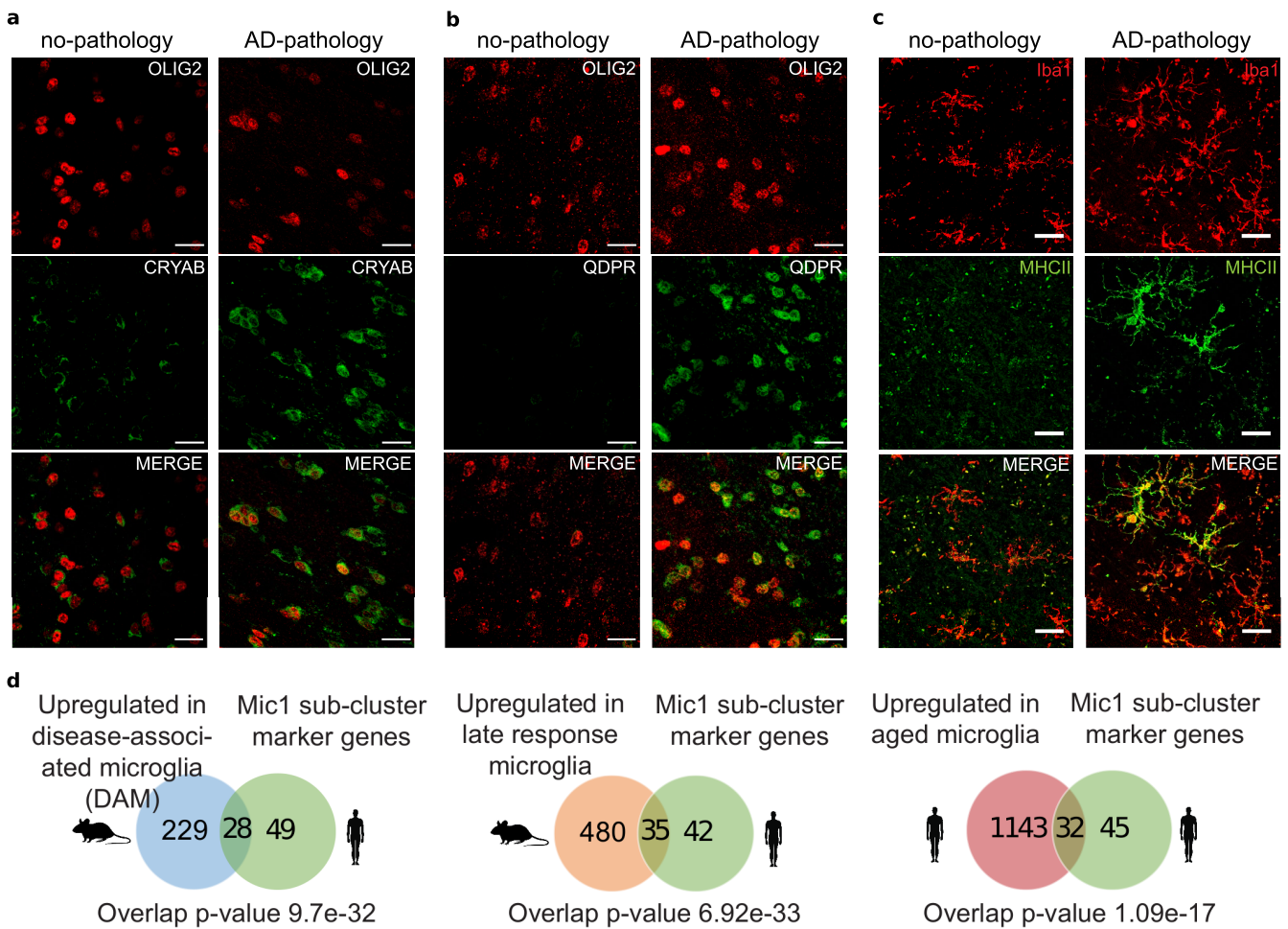
**Extended Data Fig. 7 | Overrepresentation analyses for cells in sub-clusters.** **a**, Cell composition of each identified sub-cluster (rows) across individuals (columns). Bars represent the fraction of cells corresponding to each individual. Bar colour indicates whether the corresponding value exceeds (black) or does not exceed (white) the average value measured across all the entries in the row. **b**, Overrepresentation analysis (hypergeometric test) within each pre-cluster (columns) of cells isolated from each individual (rows). **c**, Overrepresentation analysis within each sub-cluster of cells isolated from individuals with different values of discrete clinico-pathological variables (overall amyloid level, Braak stage, CERAD score (ceradsc), NIA-Reagan score (niareagansc), clinical consensus diagnosis of cognitive status at the time of death (cogdx), and sex). The scale bars on the right indicate the significance of the overrepresentation (hypergeometric test,  $-\log_{10}(P \text{ value})$ ,  $z$ -scaled, FDR

multiple-testing correction). For quantitative variables, enrichment was computed based on an estimated  $z$ -score quantifying the deviation from random expected values using resampling (Methods). The quantitative variables considered were neuritic plaque count, neurofibrillary tangle burden, tangle density, overall amyloid level, and global cognitive function. For a detailed description of clinico-pathological variables, see Supplementary Information. **d**, Overrepresentation analysis (hypergeometric test) similar to that in **a**, but computed only across cells isolated from randomly chosen female and male individuals for AD-pathology and no-pathology groups (Methods). Scores represent aggregated  $P$  values (meta-p values, meanp method, metap R package) computed across 100 random realizations. Only scores with a FDR  $< 0.01$  (correction across traits  $\times$  subpopulations) are plotted.



**Extended Data Fig. 8 | Cell-type subpopulations.** Cells from sub-clusters enriched (red) or depleted (blue) with cells for individuals with AD pathology and cognitive decline shown using *t*-SNE for major cell types (Ast1 *n* = 1,134, Ast0 *n* = 1,728, Oli0 *n* = 8,310, Oli1 *n* = 8,032, Ex4 *n* = 3,198, Ex6 *n* = 2,757, In0 *n* = 2,368, In2 *n* = 984, Opc0 *n* = 1,589, Opc1 *n* = 976, Mic1 *n* = 509, and Mic2 *n* = 169 cells). Left, corresponding marker genes (font proportional to enrichment level); right, enriched Gene

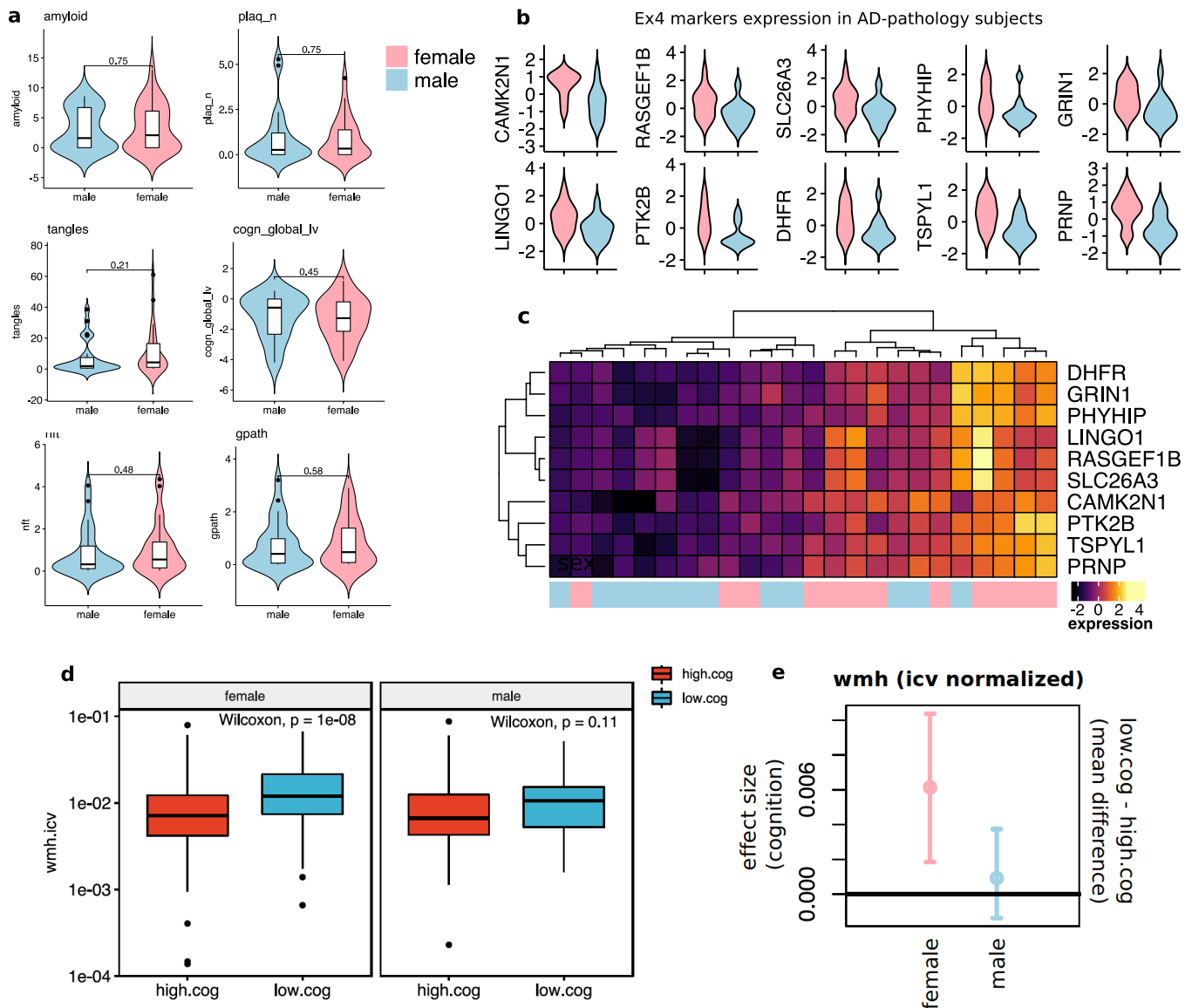
Ontology terms. Gene Ontology enrichment was based on FDR-corrected cumulative hypergeometric *P* values, with *P* value-ranked gene-marker lists (FDR < 0.01,  $\log_2(\text{mean gene expression across cells in sub-cluster} / \text{mean gene expression across cells in other sub-clusters}) > 0.5$ , two-sided Wilcoxon rank-sum test) used as input (Ex4 *n* = 783, Ex6 *n* = 2,438, In0 *n* = 1,702, In2 *n* = 350, Ast1 *n* = 574, Ast0 *n* = 73, Oli0 *n* = 227, Oli1 *n* = 73, Opc0 *n* = 19, Opc1 *n* = 536, Mic1 *n* = 487, and Mic2 *n* = 646).



### Extended Data Fig. 9 | Immunohistochemistry of subpopulation markers in oligodendrocyte lineage cells and microglia.

**a**, Oligodendrocyte lineage cell subpopulation marked by alpha B-crystallin (CRYAB). Immunohistochemistry with anti-OLIG2 (red) and anti-CRYAB (green) antibodies in the white matter of Brodmann area 10 of a no-pathology and an AD-pathology individual (scale bars, 20  $\mu\text{m}$ ). A selected area of these images is shown in Fig. 3g. The experiment was performed once. **b**, Oligodendrocyte lineage cell subpopulation marked by quinoid dihydropteridine reductase (QDPR). Immunohistochemistry with anti-OLIG2 (red) and anti-QDPR (green) antibodies in the white matter

of Brodmann area 10 of a no-pathology and an AD-pathology individual (scale bars, 20  $\mu\text{m}$ ). A selected area of these images is shown in Fig. 3h. The experiment was performed once. **c**, Immunohistochemistry with anti-IBA1 (red) and anti-MHC class II (green) antibodies in the white matter of Brodmann area 10 of a no-pathology and an AD-pathology individual (scale bars, 20  $\mu\text{m}$ ). The experiment was performed once. **d**, Overlap (one-sided Fisher's exact test) between Mic1 marker genes and genes upregulated in mouse disease-associated microglia (left), in mouse late-response microglia (middle), and in aged human microglia (right).



**Extended Data Fig. 10 | Sex comparisons in pathology, gene expression, and white matter.** **a**, Quantitative clinico-pathological measurement comparison between male and female individuals ( $n = 24$  female and  $n = 24$  male individuals; two-sided Wilcoxon rank-sum test). Violin plots are centred around the median with interquartile ranges, and the shape represents individual distribution. The quantitative clinico-pathological variables considered were overall amyloid level, neuritic plaque burden, neurofibrillary tangle burden, tangle density, global cognitive function, and global AD pathology burden. **b**, Violin plots showing aggregate expression levels ( $z$ -scaled) across excitatory neurons in female (red) versus male (blue) individuals ( $n = 12$  each) of the top 10 marker genes of the AD-associated Ex4 subpopulation of excitatory neurons. **c**, Hierarchical clustering of pathology-affected individuals (columns) based on average expression level (colour) of the top 10 marker genes (rows) of the AD-enriched Ex4 subpopulation of excitatory neurons for female versus male individuals. **d**, Statistical comparison of in vivo brain MRI imaging from ROSMAP cohorts. **d**, Intracranial volume-normalized WMH (wmh.icv) measures for female ( $n = 399$ ) and male

( $n = 106$ ) individuals and high-cognition ( $n = 252$  female and  $n = 63$  male individuals) and low-cognition ( $n = 147$  female and  $n = 43$  male individuals) groups. Groups were defined based on whether subjects had an overall cognition score lower (low.cog,  $z$ -score  $< 0$ ) or higher (high.cog,  $z$ -score  $> 0$ ) than the average. Mean rank-difference values between cognition groups were compared using the two-sided Wilcoxon rank-sum test. **e**, Statistical estimation of significant difference in WMH between low-cognition and high-cognition groups in females, and between low-cognition and high-cognition groups in males, assessed by bootstrap point and 95% confidence interval estimation of the effect size (mean difference) between groups. Bootstrap resampling was performed by resampling  $n = 40$  observations per group 1,000 times. Horizontal line highlights zero difference. The positive effect-size points and confidence interval estimates do not overlap the zero line in the female group, which provides statistical evidence of an increment in WMH (wmh.icv) in the low-cognition group relative to the high-cognition group in females but not in males.

## 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

no software was used

Data analysis

Cell Ranger software (2.0.0 version) (10x Genomics) was used to align clean reads to the hg38 human genome (GRCh38.p5 (NCBI:GCA\_000001405.20)). The single-cell analysis package scanpy (version 0.4.4) implemented in python was used for clustering analyses. The R packages scran ((version 1.8.2) and scater ( (version 1.8.1) were used for single-cell data manipulations and QC analyses. The R package Seurat (2.3.4) was used to compared reproducibility of markers. Bulk RNA-seq differential analyses we performed by fitting a linear model using the R package limma (version 3.38.3). The R package lme4 (1.1-21) was used to fit Poisson mixed models, the R package RUVseq (1.16.1) to remove unwanted variation from RNA data, and the R package metap (1.1) to compute aggregate p-values. All statistical analyses and visualizations were implemented in R (version 3.4.3). GSEA was applied to identify priori defined gene sets that show statistically significant differences between two given clusters. Gene Ontology (GO) enrichment analyses were performed using Metascape (version 3.0). Raw MPRAGE images were processed to generate total volumes including gray matter, white matter, CSF and intracranial volumes using SPM (version 12). White matter lesions appearing hyperintense in T2-weighted images were segmented based on FLAIR and MPRAGE data using BIANCA. All self-organizing maps were created using the Kohonen R package (version 3.0.8).

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

The single-nucleus RNA-Sequencing data is available at Synapse (<https://www.synapse.org/#!Synapse:syn18485175>). The DOI for this dataset is: 10.7303/syn18485175. The data is available under controlled use conditions set by human privacy regulations.

## 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](https://www.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	No explicit calculations were performed to determine sample size. Rather, we aimed to analyze brain tissue from an equal number of men and women and at least 12 individuals per group. Therefore we analyzed brain tissue from 24 individuals with low amyloid burden and 24 individuals with high amyloid burden. These numbers of samples was sufficient to perform a confident data analysis.
Data exclusions	Low quality snRNA-seq libraries were excluded and the exclusion criteria are described in the manuscript as follows. Cells with a high ratio of Mitochondrial (MT) relative to endogenous RNAs had low starting amounts of RNA, which might indicate that source cells were dead or stressed, resulting RNA degradation. Outlier cells in these quality metrics were found to cluster together in the tSNE 2D space. Based on these observations and subsequent scatter plot analyses, cells with less than 200 detected genes, and cells with abnormally high ratio of counts mapping to MT genes, relative to the total number of detected genes were removed. Specifically, given a highly skewed empirical distribution of the MT ratio values (i.e., having an elbow shape clearly separating high and low scores) outlier cells were classified in two groups using the k-means clustering algorithm (k=2) on the MT ratio, and subsequently removed.
Replication	Verification of the single-nucleus RNA-seq data was performed through validation using RNA in situ hybridization, quantitative RT-PCR, and immunohistochemistry on tissue derived from a subset of the individuals analyzed using snRNA-seq. These experiments validated the findings derived from snRNA-seq.
Randomization	The study participants were allocated into groups based on the overall amyloid level.
Blinding	Investigators were blinded to group allocation for the quantification of the RNAscope data.

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

### Methods

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

## Antibodies

Antibodies used

anti-Iba1 (Synaptic Systems; Cat. No. 234 004, Polyclonal Guinea pig antiserum, Lot number: 2-13, 1:500);  
 anti-Human HLA-DP, DQ, DR antigen (Agilent, M077501-2, clone CR3/43, Lot number: 20047190, 1:100);  
 anti-β-Amyloid (Cell Signaling Technology, #8243, D54D2, Lot number: 4, 1:500);  
 anti-OLIG2 (Atlas, HPA003254, rabbit polyclonal, Lot number: C114365, 1:1000);

anti-CRYAB (LSBio, LS-B3696, rabbit polyclonal, Lot number: 124639, 1:200);  
 anti-QDPR (Atlas Antibodies, HPA065649, rabbit polyclonal, Lot number: R92923, 1:2500);  
 anti-Rabbit IgG (goat), HRP-labeled (PerkinElmer, NEF812001EA, from goat serum, Lot number: 10311573, Dilution);  
 Alexa Fluor®488 conjugated anti-NeuN antibody (MilliporeSigma, catalog number MAB377X, clone A60, Lot number: 3101114, 1:500)

## Validation

anti-Iba1: reactivity validated by the company for Human. Validated by the company for IHC.  
 anti-Human HLA-DP, DQ, DR antigen: reactivity validated by the company for Human. The antibody was included in the First International Workshop and Conference on Monoclonal Antibodies to Human MHC Class II Antigens (1983) and its specificity and other characteristics were ascertained by a variety of techniques, including reactivity with isolated antigen, immunoblotting, and labelling of transfected cells.  
 anti-β-Amyloid: reactivity validated by the company for Human. Validated by the company for IHC on paraffin-embedded human Alzheimer's brain tissue sections.  
 anti-OLIG2: reactivity validated by the company for Human. Validated by the company for IHC. Has been validated by the Human Protein Atlas in 44 human control brain samples.  
 anti-CRYAB: reactivity validated by the company for Human. Validated by the company for IHC.  
 anti-QDPR: reactivity validated by the company for Human. Validated by the company for IHC. Has been validated by the Human Protein Atlas in 44 human control brain samples.  
 anti-Rabbit IgG (goat), HRP-labeled: Tested by the company to react with rabbit IgG and may recognize other immunoglobulin types that have light chains in common with IgG.  
 Alexa Fluor®488 conjugated anti-NeuN: reactivity validated by the company for Human.

## Human research participants

Policy information about [studies involving human research participants](#)

## Population characteristics

We selected 24 individuals with elevated β-amyloid pathology and 24 age-matched control individuals with no or very low β-amyloid burden. Individuals were balanced between male and female subjects (12 in each group), and matched for both age (median 86.7 high-amyloid, 87.1 low-amyloid) and years of education (median 19.5 high-amyloid, 18 low-amyloid).

## Recruitment

No donors were recruited, the tissue has been obtained from participants in the Religious Order Study.

## Ethics oversight

The Religious Orders Study and Rush Memory and Aging Project were approved by an IRB of Rush University Medical Center.

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

## Magnetic resonance imaging

### Experimental design

## Design type

N/A

## Design specifications

N/A

## Behavioral performance measures

N/A

### Acquisition

## Imaging type(s)

Structural

## Field strength

3T

## Sequence &amp; imaging parameters

3D magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence: echo-time (TE)=2.98 ms, repetition time (TR)=2.3 s, inversion time (TI)=900 ms, flip angle=8 degrees, 176 sagittal slices, slice thickness=1 mm, field of view (FOV)=25.6 cm x 25.6 cm, 256x256 acquisition matrix. T2-weighted fluid-attenuated inversion recovery (FLAIR): TE=150 ms, TR=9 s, TI=2.49 s, 35 axial slices, slice thickness=4 mm, FOV=22 cm x 22 cm, 256x256 acquisition matrix.

## Area of acquisition

Whole brain

## Diffusion MRI

Used

Not used

### Preprocessing

## Preprocessing software

FLIRT (affine registration of T1-weighted MPRAGE to T2-weighted FLAIR), BET (brain extraction), WMLS (automated white matter hyperintensity segmentation based on both T1-weighted MPRAGE and T2-weighted FLAIR), Freesurfer (intracranial volume calculation).

## Normalization

T1-weighted MPRAGE data were spatially registered to the T2-weighted FLAIR data using affine registration.

## Normalization template

The data were not normalized to a standardized template.

Noise and artifact removal

N/A

Volume censoring

N/A

### Statistical modeling & inference

Model type and settings

The total volume of white matter hyperintensities was measured for each participant and then normalized by the corresponding intracranial volume.

Effect(s) tested

N/A

Specify type of analysis:  Whole brain  ROI-based  Both

Statistic type for inference  
(See [Eklund et al. 2016](#))

N/A

Correction

N/A

### Models & analysis

n/a | Involved in the study

Functional and/or effective connectivity

Graph analysis

Multivariate modeling or predictive analysis