Article

Single-cell multiregion dissection of Alzheimer's disease

https://doi.org/10.1038/s41586-024-07606-7

Received: 20 June 2022

Accepted: 24 May 2024

Published online: 24 July 2024

Open access

Check for updates

Hansruedi Mathys[1,2,3,4,13], Carles A. Boix[5,6,7,13], Leyla Anne Akay[1,2,13], Ziting Xia[1,2,8], Jose Davila-Velderrain[9], Ayesha P. Ng[1,2], Xueqiao Jiang[1,2], Ghada Abdelhady[3], Kyriaki Galani[5,6], Julio Mantero[5,6], Neil Band[5,6,12], Benjamin T. James[5,6], Sudhagar Babu[3], Fabiola Galiana-Melendez[1,2], Kate Louderback[1,2], Dmitry Prokopenko[10], Rudolph E. Tanzi[10], David A. Bennett[11], Li-Huei Tsai[1,2,6]✉ & Manolis Kellis[5,6]✉

Alzheimer's disease is the leading cause of dementia worldwide, but the cellular pathways that underlie its pathological progression across brain regions remain poorly understood[1–3]. Here we report a single-cell transcriptomic atlas of six different brain regions in the aged human brain, covering 1.3 million cells from 283 post-mortem human brain samples across 48 individuals with and without Alzheimer's disease. We identify 76 cell types, including region-specific subtypes of astrocytes and excitatory neurons and an inhibitory interneuron population unique to the thalamus and distinct from canonical inhibitory subclasses. We identify vulnerable populations of excitatory and inhibitory neurons that are depleted in specific brain regions in Alzheimer's disease, and provide evidence that the Reelin signalling pathway is involved in modulating the vulnerability of these neurons. We develop a scalable method for discovering gene modules, which we use to identify cell-type-specific and region-specific modules that are altered in Alzheimer's disease and to annotate transcriptomic differences associated with diverse pathological variables. We identify an astrocyte program that is associated with cognitive resilience to Alzheimer's disease pathology, tying choline metabolism and polyamine biosynthesis in astrocytes to preserved cognitive function late in life. Together, our study develops a regional atlas of the ageing human brain and provides insights into cellular vulnerability, response and resilience to Alzheimer's disease pathology.

Alzheimer's disease (AD) is characterized by pathological protein aggregation in a stereotyped pattern across multiple brain regions[1,4]. Post-mortem diagnosis of AD is staged by the severity and distribution of these pathological hallmarks: extracellular amyloid-β deposits and intracellular neurofibrillary tangles (NFTs) in neurons. Tangles are first seen in the entorhinal cortex (EC) (Braak stages I–II), then the hippocampus and thalamus (Braak stages III–IV) and finally the neocortex (Braak stages V–VI), a sequence that is typically synchronous with cognitive decline from mild cognitive impairment to severe dementia[1,2,4–7]. Understanding the cellular architecture of affected brain regions has important implications for early and region-specific therapeutic interventions and may shed light on the molecular mechanisms underlying the regional progression of pathology. Although some brain regions relevant to AD have been studied individually at scale or jointly in samples from a few individuals[8–16], a comprehensive molecular characterization of region-specific differences in AD is currently lacking and could capture differences in regional molecular architecture[17–24] and region-specific neuronal and glial subtype alterations in AD and in cognitive resilience to AD pathology[3,25–27].

Here we present a transcriptomic atlas of the human brain spanning six distinct anatomical regions from persons with and without Alzheimer's dementia as a basis for understanding disease-associated differences. We profile the transcriptomes of over 1.3 million nuclei from the EC, hippocampus (HC), anterior thalamus (TH), angular gyrus (AG), midtemporal cortex (MT) and prefrontal cortex (PFC) from 48 individuals, 26 of whom have a pathologic diagnosis of AD. We annotate region-specific neuronal and glial subtype diversity, present an online resource for navigating this atlas (http://compbio.mit.edu/ad_multiregion) and provide mechanistic insights into cellular vulnerability, response and resilience to AD.

A multiregion atlas of AD

To characterize cellular diversity in the human brain, and the genes, pathways and cell types that underlie AD progression across brain

[1]Picower Institute for Learning and Memory, MIT, Cambridge, MA, USA. [2]Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, USA. [3]University of Pittsburgh Brain Institute, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [4]Department of Neurobiology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [5]Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA. [6]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [7]Computational and Systems Biology Program, MIT, Cambridge, MA, USA. [8]Harvard-MIT Health Sciences and Technology Program, MIT, Cambridge, MA, USA. [9]Human Technopole, Milan, Italy. [10]Genetics and Aging Research Unit, McCance Center for Brain Health, Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [11]Rush Alzheimer's Disease Center, Chicago, IL, USA. [12]Present address: Department of Computer Science, Stanford University, Stanford, CA, USA. [13]These authors contributed equally: Hansruedi Mathys, Carles A. Boix, Leyla Anne Akay. ✉e-mail: lhtsai@mit.edu; manoli@mit.edu

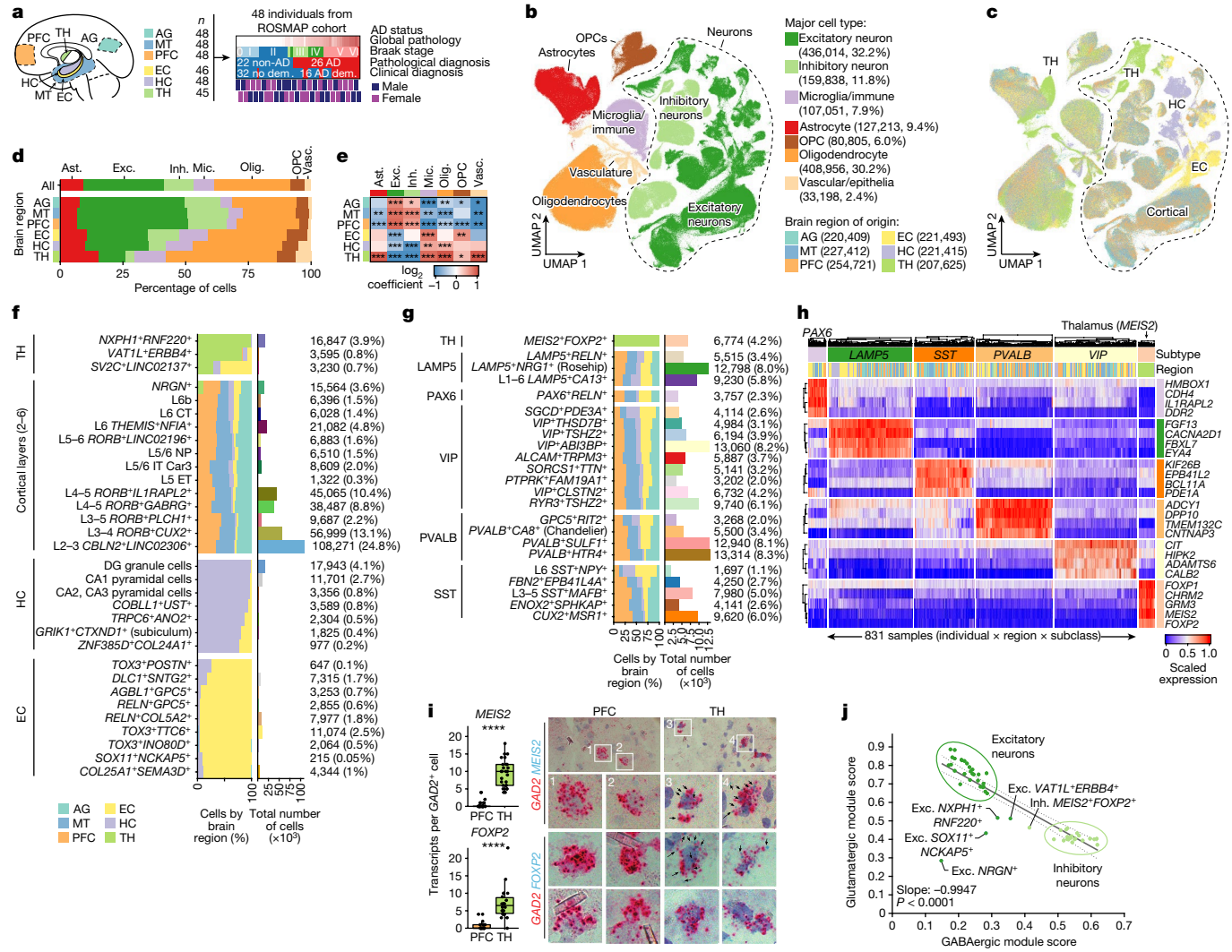


Fig. 1 | snRNA-seq analysis of six distinct regions of the aged human brain. **a**, snRNA-seq profiling summary, covering 283 samples across 6 brain regions from 48 participants from ROSMAP, showing global pathology, Braak stage and pathological (26 AD and 22 non-AD) or clinical diagnosis of AD (16 AD dementia (dem.) and 32 no dementia). **b**,**c**, Joint uniform manifold approximation and projection (UMAP), coloured by major cell type (**b**) and region of origin (**c**). **d**, The regional composition of major cell types. **e**, Relative enrichment of major cell types across regions by quasi-binomial regression. False discovery rate (FDR)-corrected *P* values are indicated by asterisks; ***$P < 0.001$, **$P < 0.01$, *$P < 0.05$. **f**,**g**, Global breakdown, region composition, enrichment and number of nuclei for excitatory (**f**) and inhibitory (**g**) neuronal subtypes. **h**, Gene expression analysis of the top four markers per inhibitory subclass, averaged at the sample by subclass level (columns). **i**, RNAscope validation of *FOXP2* and *MEIS2* as markers of the unique thalamus subtype, with quantification (left) performed using Student's *t*-tests and representative images (right). The blue puncta represent *MEIS2* (top) or *FOXP2* (bottom) transcripts and red puncta represent *GAD2* transcripts. *FOXP2*: $n = 19$ (PFC) and $n = 22$ (TH) cells; *MEIS2*: $n = 35$ (PFC) and $n = 26$ (TH) cells; each dot represents an individual cell, pooled from eight samples (four individuals; each had one PFC and one thalamus sample). **j**, Glutamatergic versus GABAergic scores for all neuron subtypes. The dotted lines represent the 95% confidence interval around the linear fit. *P* values were calculated using two-sided *F* tests. Ast., astrocytes; exc., excitatory neurons; inh., inhibitory neurons; mic., microglia/immune cells; olig., oligodendrocytes; vasc., vascular/epithelial cells.

regions, we performed single-nucleus RNA-sequencing (snRNA-seq) analysis of nuclei isolated from 283 post-mortem brain samples across six brain regions from 48 participants in the Religious Order Study (ROS) or the Rush Memory and Aging Project (MAP)[28] (together, ROSMAP; Fig. 1a). We selected 48 participants on the basis of pathologic diagnosis of AD (stratified by NIA-Reagan score of 26 (with AD) and 22 (without AD; labelled non-AD)) and on the basis of clinical diagnosis of Alzheimer's dementia ($n = 16$) versus non-dementia ($n = 32$)[29,30] (Fig. 1a, Extended Data Fig. 1a and Supplementary Table 1). From these 48 individuals, we profiled six brain regions: the EC (221,493 cells), which is affected in early AD (stages I–II); the HC (221,415) and TH (207,625), which are affected in mid-AD (stages III–IV); and the AG (220,409), MT (227,412) and PFC (254,721), which are affected in late AD (stages V–VI), for a total of 1.35 million transcriptomes of independent nuclei after removing doublets, low-quality cells and highly sample-specific clusters. We annotated 76 high-resolution cell types in 14 major cell type groups, including 32 excitatory neuron subtypes (436,014 nuclei, 32.2% of total) and 23 inhibitory subtypes (159,838 nuclei, 11.8% of total) (Extended Data Fig. 1b–d, Supplementary Figs. 1 and 2 and Supplementary Table 2). We characterized these cell types in terms of their transcriptome size and proliferative status, compared our atlas with previously published data across species[31–33] (Extended Data Fig. 1e,f and Supplementary Figs. 3–5) and identified broad cell type identity programs using non-negative matrix factorization (NMF)[34] and transcriptional regulons using SCENIC[35,36] (Extended Data Figs. 2 and 3 and Supplementary Tables 3 and 4).

To gain insights into the cellular architecture of the human brain, we investigated differences in the composition of major cell types between

the six brain regions. The fraction of neurons increased significantly from the TH (14.4% neurons) to the three-layer allocortical HC (32.2%), the entorhinal periallocortex (36.6%) and the six-layered neocortical regions (AG, MT and PFC, 58.9%) (Fig. 1b–e and Supplementary Fig. 6). Glia, including astrocytes, oligodendrocytes, oligodendrocyte precursor cells (OPCs) and microglia/immune cells, tended to be less abundant in neocortical samples (Fig. 1b–e), in agreement with previous studies in humans[37,38] and mice[39,40] (Supplementary Fig. 7a–d). Differences in the composition of major cell types between regions were reproducibly observed across study participants, irrespective of the individual's disease status (Supplementary Fig. 7e–h), suggesting that variability in the major cell type composition between regions is a fundamental characteristic of the human brain and is not affected by AD pathology.

Neuronal diversity across brain regions

We first characterized the regional diversity of excitatory neuron subtypes, which were consistent across individuals and were either highly region-specific to the HC, EC and TH (7, 9 and 2 subtypes, respectively) or were predominantly shared across neocortical regions (12 subtypes) (Fig. 1f and Supplementary Fig. 8–12). Hippocampal subtypes included neurons from the highly structured CA1 and CA2/CA3 subfields and dentate gyrus and the more entorhinal-proximal subiculum and para/presubiculum areas[9]. EC-specific subtypes that clustered separately from neocortical subtypes for the same layers were often marked by expression of *RELN*, *TOX3* and *GPC5*, and contained subtypes from both the lateral (L2 *RELN*$^+$*GPC5*$^+$) and medial (L2 *TOX3*$^+$*POSTN*$^+$) EC[41–43] (Supplementary Fig. 10). Excitatory neurons in the TH were predominantly composed (74%) of a subtype (*NXPH1*$^+$*RNF220*$^+$) that was not observed in the neocortex and is predicted to be regulated by *LHX9*, *SOX2*, *SHOX2* and *TCF7L2*[34,36] (Extended Data Fig. 2a,b and Supplementary Figs. 9e–i and 11n,o). We found that the thalamic–neocortex separation is conserved in mice and recapitulated both this divide and thalamic marker genes in independent single-cell, bulk and microarray data in both mice and humans[8,39,40,43] (Supplementary Fig. 12).

In contrast to excitatory neuron subtypes, the majority of inhibitory neuron subtypes (22 out of 23 subtypes) were observed in all five cortical regions (Fig. 1g and Supplementary Figs. 13–17), although some inhibitory subtypes had regional biases, including *PVALB*$^+$*HTR4*$^+$ and *CUX2*$^+$*MSR1*$^+$ (enriched in neocortex), layer 6 *SST*$^+$*NPY*$^+$ (EC and HC) and *GPC5*$^+$*RIT2*$^+$ (EC), suggesting that there are significant differences in inhibitory neuron composition between the neocortex and allocortex (Fig. 1g and Supplementary Fig. 14). Moreover, in the HC, EC and MT, caudal ganglionic eminence-derived GABAergic neurons (*VIP*$^+$*LAMP5*$^+$) were significantly more abundant than medial ganglionic eminence-derived neurons (*SST*$^+$*PVALB*$^+$), but these two major clades were not significantly different in the PFC (Extended Data Fig. 1g). By contrast, the TH contained a single unique, thalamus-specific inhibitory subtype (*MEIS2*$^+$*FOXP2*$^+$) marked by genes that are involved in neurite outgrowth, such as the semaphorins *SEMA3C* and *SEMA3E*, *DISC1* and *SPON1*, and receptors for serotonin (*HTR2A*), acetylcholine (*CHRM2*, *CHRNA3*) and glutamate (*GRM3*) (Fig. 1g,h and Supplementary Figs. 14 and 15). These genes were in a single inhibitory program (Inh-22, from NMF) that included the SCENIC-predicted subtype regulators *FOXP2* and *LEF1*[34,36] (Extended Data Figs. 2c,d and 3b). We recapitulated this thalamic difference and program genes in the mouse thalamus and human lateral geniculate nucleus (dLGN) using previously published single-cell data (Supplementary Figs. 16 and 17). To validate the localization and specificity of markers of the thalamic inhibitory neuron subtype, we performed in situ hybridization for both *FOXP2* and *MEIS2* with *GAD2* on TH and PFC post-mortem brain samples from four individuals, and found significant thalamus-specific co-localization of both marker genes with *GAD2* (Fig. 1i).

As thalamic *MEIS2* neurons expressed several typically glutamatergic neuron genes, we determined glutamatergic and GABAergic module scores for every neuronal cell to further examine the chimeric nature of this subtype (Supplementary Fig. 15g–k and Supplementary Table 5). These scores matched the cortical excitatory and inhibitory split and were negatively correlated both across and within broad neuronal classes (Fig. 1j and Supplementary Fig. 15g,h). Both thalamic *MEIS2*$^+$ inhibitory and *NXPH1*$^+$ excitatory neurons had intermediate scores, placing them between the cortical excitatory and inhibitory clusters, suggesting that they are less polarized with regard to the expression of cortical glutamatergic versus GABAergic programs (Fig. 1j and Supplementary Fig. 15i–k). Predicted cell–cell communication interactions were mostly shared across multiple regions, but the thalamus had multiple differential interactions (Supplementary Figs. 18 and 19). The top thalamus-specific interactions were between excitatory *NXPH1* and neuronal *NRXN1* or *NRXN3*, whereas inhibitory neurons expressed *NXPH1* in the other regions, suggesting that neurexophilin signalling swaps from excitatory neurons in the thalamus to inhibitory neurons in cortical brain regions (Extended Data Fig. 4).

Glial diversity annotated by gene modules

We next tested whether glial cells also had transcriptional differences between brain regions. We identified multiple transcriptionally distinct subsets for each major glial cell type and determined their characteristic marker genes (Fig. 2a and Supplementary Fig. 20–25). Among glial cell types, astrocytes had the highest regional heterogeneity, containing both highly neocortex-enriched (*GRM3*$^+$*DPP10*$^+$) and thalamus-enriched (*LUZP2*$^+$*DCLK1*$^+$) subtypes (Fig. 2a–c and Supplementary Fig. 20). Region-specific astrocyte subtypes were experimentally validated using RNA in situ hybridization (Fig. 2d) and confirmed by analysing a separate snRNA-seq dataset[14] (Supplementary Fig. 23m). Cortical astrocytes were enriched for markers involved in glutamate processing and transport, whereas hippocampus- and anterior thalamus-enriched *DCLK1* astrocytes had lower glutamate transporter activity and were enriched instead for focal-adhesion-related genes (Fig. 2c and Supplementary Fig. 25a). Thalamic astrocytes (*LUZP2*$^+$) expressed GABA-uptake genes *SLC6A1* and *SLC6A11* at much higher levels compared with other subtypes, even though the proportion of inhibitory neurons was not markedly higher in the thalamus (Fig. 2c). Notably, the thalamic *MEIS2*$^+$*FOXP2*$^+$ interneurons shared multiple markers with neocortical *GRM3* astrocytes, including *GRM3*, *MEIS2* and *VAV3*[39,40,43] (Supplementary Fig. 23n), suggesting that astrocytes in evolutionarily newer regions may share some functions with inhibitory neurons in older regions.

We developed a method, single-cell decorrelated module networks (scdemon), to identify gene expression modules from highly correlated sets of genes in atlas-scale snRNA-seq datasets (Fig. 2e). Highly imbalanced cell type composition in single-cell datasets, in which rare cellular states are outnumbered by common cell types, can lead to under-recovery of gene–gene interactions, especially for genes that are expressed at low levels. To account for these issues, our method estimates a sample-decorrelated gene–gene correlation matrix, thresholds gene–gene pairs based on their sparsity and uses the adjusted matrix to identify modules of highly correlated genes (Methods). We used our method to identify modules both across all cells in the atlas and for each major cell type independently, and recovered modules expressed to varying degrees, ranging from identity modules for each glial cell type to a cell cycle module found in just 0.7% of microglia (Fig. 2f, Extended Data Fig. 5, Supplementary Figs. 26–36 and Supplementary Table 6). Cells expressing these modules were enriched for diverse aspects of our dataset, including cellular subtype identity (205 modules), brain region (156, with 77 thalamus specific and 34 EC specific), AD status (73), *APOE* genotype (78) and sex (24) (Fig. 2f and Supplementary Table 7). We hierarchically clustered modules across the cell types and found that many cell types expressed gene programs for cholesterol biosynthesis (C10), chaperones (C5), ribosomes (C1 and C2), ER protein processing

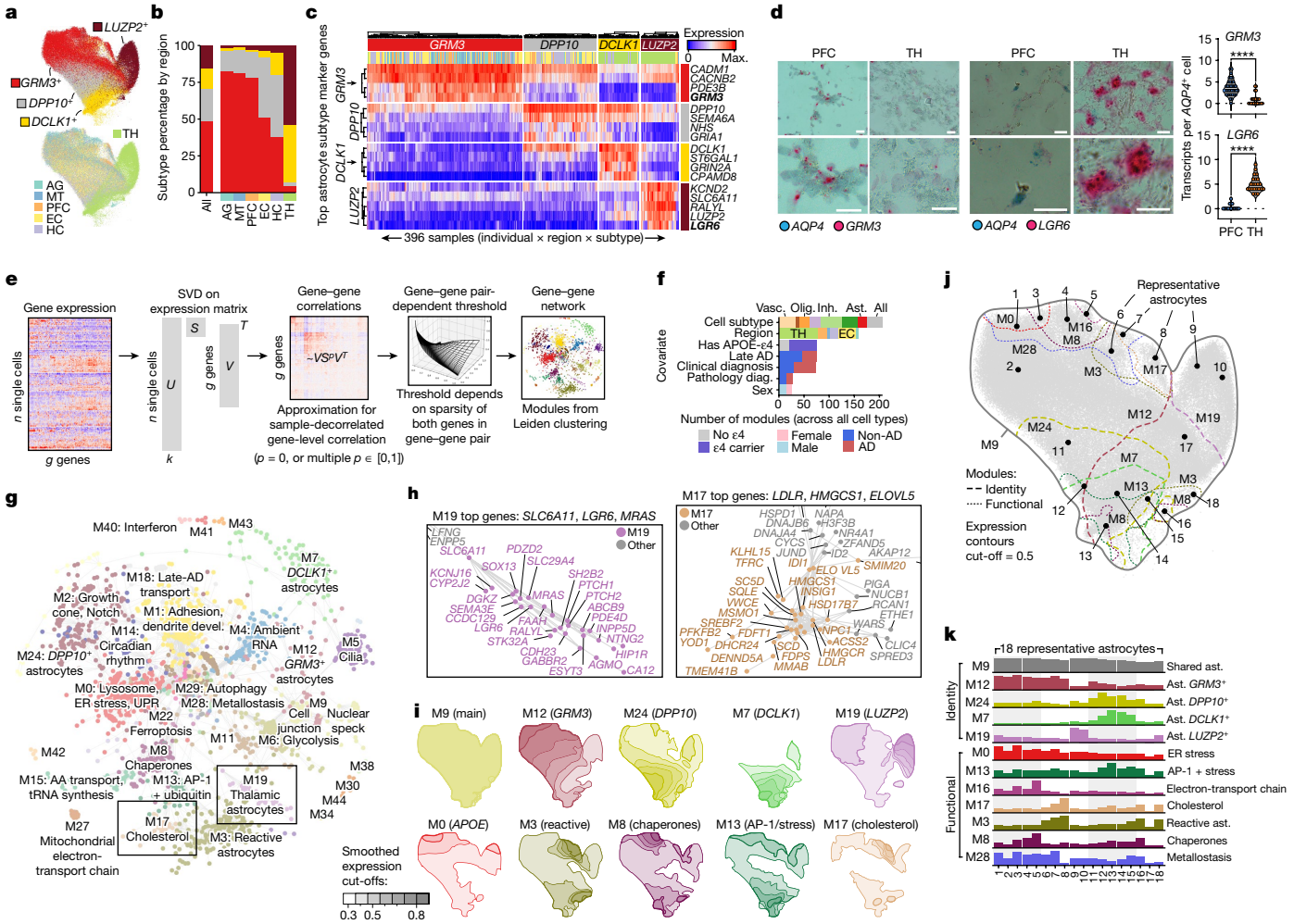


Fig. 2 | Astrocyte diversity across regions annotated by gene expression modules. a, UMAP plot for astrocyte nuclei, coloured by astrocyte subtype or brain region of origin. **b**, Global breakdown and regional composition of astrocyte subtypes. **c**, Gene expression heat map for the top markers of each astrocyte subclass, averaged to sample by subtype and scaled to the row maximum (max.). **d**, RNAscope validation of *GRM3* and *LGR6* as markers of *AQP4*$^+$ neocortical and TH astrocytes, respectively (bold markers in **c**). Representative images (left) showing *AQP4* transcripts (blue puncta) and *GRM3* or *LGR6* transcripts (red puncta). Scale bars, 20 µm. Quantification (right) was performed using two-tailed unpaired Student's *t*-tests; ****$P < 0.0001$. Each dot represents an individual cell, pooled from eight samples (four individuals; each with one PFC and one thalamus sample). *GRM3*: $n = 37$ (PFC) and $n = 23$ (TH) cells; *LGR6*: $n = 17$ (PFC) and $n = 23$ (TH) cells. **e**, The framework for detecting gene expression modules using scdemon. **f**, The number of modules enriched for each covariate across all module sets (hypergeometric test, $P < 0.001$). Bar plots are coloured by the covariate level for which the modules are enriched (or by the major cell type used for module discovery for cell subtype). Diag., diagnosis. **g**,**h**, Gene–gene network (**g**) and magnification of the indicated regions (**h**) for astrocyte modules, with insets for M19, a subtype identity module for LUZP2 astrocytes (**h**, left) and M17, a functional program involved in cholesterol biosynthesis (**h**, right). AA, amino acid. **i**, Contour plots on the astrocyte UMAP for module expression of five identity (top row) and five functional (bottom row) programs. Expression was smoothed on a 500 × 500 grid with a 2D Gaussian kernel (size = 25 × 25; $\sigma = 1$). **j**,**k**, Module contours showing regions of top expression on the astrocyte UMAP for selected identity modules (**j**) and corresponding module scores (**k**) for the 18 labelled representative cells across the astrocyte UMAP for selected identity and expression models, scaled to the maximal expression of each module. ER, endoplasmic reticulum; SVD, singular value decomposition.

(C7), oxidative phosphorylation (C18), synapse interaction (C16), and glycolysis and response to hypoxia (C20) (Extended Data Fig. 6a,b).

Using this approach, we identified 32 modules in astrocytes, including an astrocyte-wide program (M9, expressed in >99% of astrocytes) marked by *GPM6A* and *GPC5* and enriched for cell junction assembly, and subtype- and region-specific identity programs such as thalamus-associated M19 (*SLC6A11*, *LGR6*, *MRAS*), which were enriched for sonic hedgehog signalling, M12 (*GRM3*: forebrain neuron development) and M7 (*DCLK1*: synaptic membrane) (Fig. 2g,h and Supplementary Fig. 31). Other modules spanned a diverse set of functions, including metallostasis, RNA splicing, glycolysis, oxidative phosphorylation and cholesterol biosynthesis and were shared by multiple subtypes (Fig. 2i–k and Extended Data Figs. 5b and 6a–c). For example, chaperone-enriched and APOE-ε4-associated M8 was expressed in multiple different astrocyte subtypes and regions, and expression of AD-associated M28 (metallostasis) overlapped with expression of both *APOE*$^+$ (M0) and reactive (M3) astrocytes (Fig. 2i–k). Module–module correlations across samples revealed co-expressed programs, such as reactive astrocytes (M3, marked by *TPST1*, *CLIC4* and *EMP1*) with cholesterol biosynthesis module M17 ($r = 0.60$), and glycolysis (M6) with AP-1 module M13 ($r = 0.39$, including *FOS*/*JUN* and ubiquitin), a pair that is potentially co-expressed in astrocytes under metabolic stress (Extended Data Fig. 6d,e).

In contrast to astrocytes, immune cells showed little regional specificity and oligodendrocyte-lineage cells had thalamus-enriched subtypes with minor transcriptomic differences to neocortex-enriched subtypes (Supplementary Figs. 21–25). Immune modules included identity programs, such as for T cells (M6), macrophages (M7) and cycling microglia

(*MKI67*$^+$, M5) as well as modules found across immune cells and enriched for genes involved in NF-κB signalling (M18), interferon (M20), p53 and DNA damage response (M22) and TGFβ signalling (M14) (Extended Data Fig. 6f and Supplementary Fig. 32). Oligodendrocyte-lineage modules showed high regional specificity, and two OPC modules—thalamus-enriched M11 and EC-enriched M25—were marked by synapse-associated genes such as neural adhesion-related *SEMA3D*, *SEMA6D* and *CNTN5*, and glutamate receptor *GRIA4*, suggesting a role for OPC sensation and response to neuronal activity in specific brain regions (Extended Data Fig. 5c,e and Supplementary Figs. 33, 34 and 36).

Vulnerable neuronal subtypes in AD

After constructing our atlas across AD-affected brain regions, we examined how AD affects the cellular composition. At the level of major cell types, we observed slight, non-significant decreases in the number of both excitatory neurons (odds ratio (OR) = 0.94, individuals stratified by pathologic diagnosis of AD), inhibitory neurons (OR = 0.93) and OPCs (OR = 0.85), as well as an increase in the number of oligodendrocytes (OR = 1.14, adjusted P (P_{adj}) = 0.01) and vascular cells (OR = 1.24), mostly driven by differences in the EC, HC and PFC regions, especially in late AD (Extended Data Fig. 7a,b). We next tested whether the fractions of region-specific neuronal subtypes were significantly altered relative to both individual-level pathologic and clinical diagnoses of AD and region-level NFT and plaque accumulation (Fig. 3a and Extended Data Fig. 7c,d). Among excitatory neurons, we identified one HC-specific (CA1 pyramidal neurons) and four EC-specific (L2 *RELN*$^+$ lateral EC, L3 *RELN*$^+$, L5 and L2/3 *TOX3*$^+$*TTC6*$^+$ neurons) subtypes that were significantly less abundant (OR = 0.38–0.66) in individuals with a pathologic diagnosis of AD (Fig. 3a and Supplementary Fig. 37a–c). Neocortical L2–3 neurons were also significantly less abundant in samples with high NFT levels and in individuals with neocortical NFT involvement (Fig. 3a). Individuals with lower percentages of these vulnerable excitatory neuron subtypes performed significantly worse on cognitive tasks, with the strongest observed impacts on episodic memory and global cognitive function for subtypes marked by *GPC5* (EC L5 and L2 *RELN*$^+$)[2] (Extended Data Fig. 7e,f). Notably, while the overall excitatory fraction was not associated with cognition, lower OPC fraction across regions and, in particular, in non-neocortex regions was significantly associated with impaired cognition (Supplementary Fig. 37d).

Given that these neuronal subtypes lie in highly interconnected regions, we next examined whether neuronal subtypes connected across regions were coordinately depleted. We found that vulnerable neuronal subtypes were co-depleted specifically in individuals with AD, with some of the strongest effects observed in established connections between the CA1, subiculum, EC–L3 and EC–L5 (Fig. 3b,c and Extended Data Fig. 7g). These included co-depletion for entorhinal L5 versus L5-projecting subiculum (Kendall's τ = 0.37 (AD); −0.1 (non-AD)) or CA1 (τ = 0.42 (AD) and −0.16 (non-AD)); and for CA1 versus L2-lateral EC (LEC, τ = 0.26 (AD) and −0.07 (non-AD)) and L3 *RELN*$^+$ (τ = 0.24 (AD) and −0.13 (non-AD)) EC neurons, both of which project in part to the CA1 subfield[44,45] (Fig. 3b,c and Extended Data Fig. 7g).

We next investigate whether vulnerable subtypes share marker genes that might mediate their vulnerability, and identified 391 genes with significantly higher baseline (non-AD) expression in vulnerable subtypes (Fig. 3d and Supplementary Table 8). These included Reelin signalling pathway genes *RELN* and *DAB1*; kinase-associated genes *MAP2K5*, *PRKCA* and *SPHKAP*; and multiple genes associated with heparan sulfate proteoglycan biosynthesis (including *HS6ST3*, *XYLT1* and *NDST3*) (Fig. 3d and Extended Data Fig. 7h,i). Notably, while *RELN* expression, which is typically restricted to inhibitory neurons, was highly specific to two EC excitatory subtypes, its downstream partner *DAB1* was present across subtypes (Extended Data Fig. 7h,i and Supplementary Fig. 37e,f).

We next examined whether vulnerable inhibitory neuron subtypes in the PFC share characteristics with vulnerable excitatory neuron subtypes across our brain regions using single-cell transcriptomes from 621 ROSMAP study participants[27,46]. We identified specific inhibitory neuron subtypes that are depleted in individuals with a high tangle density burden, consistent with our previous findings[27] (Extended Data Fig. 7j). Vulnerable and non-vulnerable inhibitory neuron subtypes differed in the expression of genes involved in neuron projection morphogenesis (*ROBO2*, *SEMA6A* and *EPHB6*), enzyme-linked receptor protein signalling pathways (*FGFR2*, *TGFBR1* and *PLCE1*) and heparan sulfate proteoglycan biosynthesis (Extended Data Fig. 7k and Supplementary Table 8). Notably, vulnerable inhibitory neuron subtypes expressed significantly higher levels of the Reelin signalling pathway components *RELN* and *DAB1*, mirroring the observed higher expression of these two genes in vulnerable excitatory neuron subtypes (Fig. 3e). Furthermore, the Reelin receptors LRP8 (also known as ApoER2) and NRP1 exhibited significantly different baseline expression in vulnerable compared with non-vulnerable inhibitory neuron subtypes (Fig. 3f).

To test the selective vulnerability of Reelin-expressing excitatory neurons in AD, we performed in situ hybridization (RNAscope) analysis of Reelin and vGlut (excitatory neuron marker) in EC tissue samples from both patients with AD and healthy individuals without AD. We found a significant decrease in the percentage of Reelin-expressing excitatory neurons in the EC of individuals with AD (Fig. 3g). To determine whether this finding was conserved in animal models of AD, we used immunohistochemistry to assess the expression of Reelin in the EC of both 12-month-old *App* knock-in (KI) mice and 9-month-old Tau(P301S) transgenic mice. We found that, relative to wild-type littermate controls, *App*-KI mice and Tau(P301S) mice had a significantly decreased percentage of Reelin-positive neurons in the EC (Fig. 3h,i), in agreement with our human transcriptomic data suggesting a selective vulnerability of Reelin-expressing neurons (Fig. 3d–f).

To understand how vulnerable subtypes are altered in AD, we computed differentially expressed genes (DEGs) for each excitatory neuron subtype (Methods and Supplementary Fig. 38a–c). We partitioned DEGs into sets associated with either vulnerable or non-vulnerable subtypes according to their expression levels in individuals with late AD (Extended Data Fig. 7l,m). DEGs linked to non-vulnerable subtypes were enriched for a diverse set of functions, including ubiquitin-ligase binding, heat-shock-family chaperones, ER protein processing and mediators of neuronal death, whereas vulnerability-associated DEGs were highly enriched only for mitochondrial oxidative phosphorylation but included *CRK* and *NEUROD2*, which are both associated with Reelin signalling[17,18] (Extended Data Fig. 7l,m and Supplementary Fig. 38d–f). Some DEGs associated with non-vulnerable subtypes had higher differential effect sizes in the vulnerable subtypes, and showed additional enrichment for aerobic glycolysis (including *PGK1*, *LDHB* and *SLC2A3*) and clathrin-mediated endocytosis (including *AP2M1*/*AP2S1*, *OCRL* and *COPS8*) (Extended Data Fig. 7m).

Regional expression differences in AD

To identify regional differences in cellular expression and function specific to individuals with pathologic AD, we computed DEGs for each major cell type in every region alone and across regions using a negative binomial linear mixed model framework, adjusting for both known covariates and potential unknown batch effects (Methods) (Extended Data Fig. 8a and Supplementary Table 9). Astrocytes and inhibitory and excitatory neurons showed the highest number of DEGs over all of the regions, with the largest number of changes in the EC (Extended Data Fig. 8a). Notably, neuronal DEGs showed little overlap across regions, indicating that neuronal differences in AD are primarily determined by subtype or region of origin (Extended Data Fig. 8b). By contrast, microglia and OPC DEGs overlapped within the non-neocortex regions, and astrocyte and oligodendrocyte changes were more consistent across all regions (Extended Data Fig. 8b). AD DEGs were consistent with published results both for region-specific DEGs and for

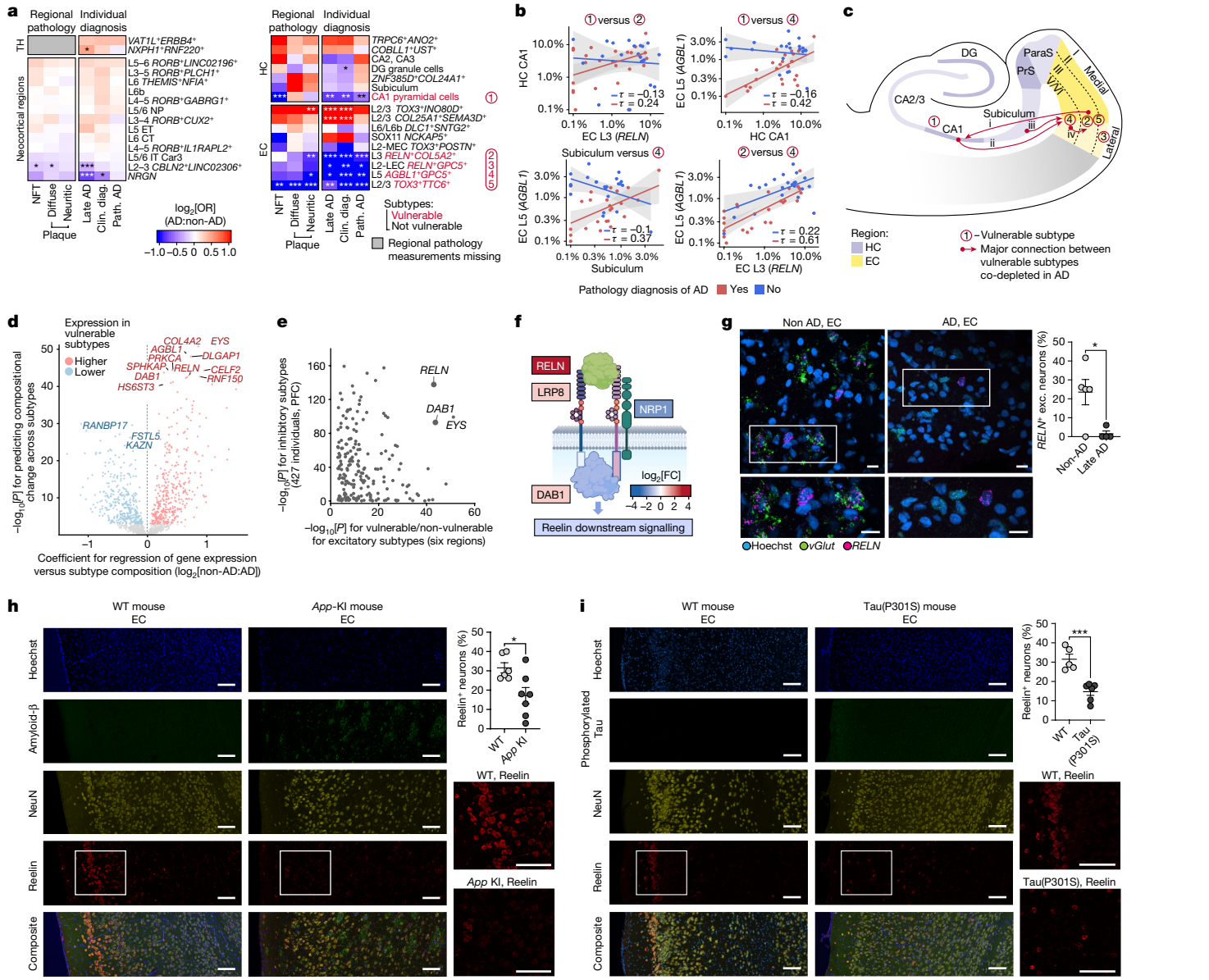


Fig. 3 | Subtype-specific neuronal vulnerability in AD. a, Compositional differences in excitatory neuron subtype enrichment and depletion in AD by quasi-binomial regression with FDR correction. Clin. diag., clinical diagnosis; path. AD, pathologic AD. **b**, Scatter plot and correlations (Kendall's τ) of the subtype fraction between four pairs of neuronal subtypes in the HC and EC (linear fit with 95% confidence intervals). **c**, Schematic of the HC and EC, highlighting the locations of vulnerable excitatory subtypes and co-depleted connections. **d**, Genes associated with excitatory neuron subtype vulnerability across all brain regions. Linear regression between normalized sample + subtype-level gene expression and $\log_2$[OR] for late-AD, with FDR-corrected *P* values. **e**, Genes associated with excitatory and inhibitory subtype vulnerability (FDR-corrected *P* values, only genes significantly and positively associated with excitatory subtype vulnerability). **f**, Schematic of Reelin signalling pathway genes that are differentially expressed in vulnerable inhibitory subtypes (colour indicates the $\log_2$-transformed fold change in expression between vulnerable and non-vulnerable subtypes). The diagram was created using BioRender. **g**, In situ hybridization (RNAscope) validation of depletion of *RELN*$^+$ excitatory neurons in the EC of individuals with AD relative to individuals without AD. Representative images (left) include Hoechst (blue), *vGlut* transcripts (green puncta) and *RELN* transcripts (magenta puncta). Scale bars, 20 µm. Quantification (right) was performed using unpaired two-tailed Student's *t*-tests ($P = 0.0242$). Data are mean ± s.e.m. $n = 5$ (non-AD) and $n = 4$ (AD) individuals. **h**,**i**, Immunohistochemistry analysis of Reelin, NeuN and amyloid-β (**h**) or phosphorylated tau (**i**) in 12-month-old *App-KI* mice (**h**) or 9-month-old Tau(P301S) transgenic mice (**i**), showing depletion of Reelin-positive neurons in the ECs of the KI and transgenic mice compared with those of the wild-type controls. Representative images (left) show Hoechst (blue); amyloid-β (**h**; D54D2) or phosphorylated-tau (**i**) (green); NeuN (yellow); and Reelin (red). Scale bars, 100 µm (**h** and **i**). Quantification (right) was performed using unpaired two-tailed Student's *t*-tests; $P = 0.0181$ (*App*-KI, **h**; $n = 7$ (*App*-KI) and $n = 6$ (wild type) mice) and $P = 0.0005$ (Tau(P301S), **i**; $n = 6$ mice (Tau(P301S)) and $n = 5$ (wild type) mice). Data are mean ± s.e.m. ParaS, parasubiculum; PrS, presubiculum.

DEGs computed jointly over all regions for multiple AD variables, and were further corroborated by comparisons with various independent studies[11,12,15,19,47–53] (Supplementary Fig. 39).

Excitatory DEGs were strongly enriched for electron-transport functional terms across regions and showed weak region-specific enrichments for protein-folding-, ubiquitination- and synapse-associated terms (Extended Data Fig. 8c). Inhibitory DEGs were also broadly enriched for protein-folding- and synapse-associated terms and for oxidative phosphorylation uniquely in the thalamus (Extended Data Fig. 8c). While microglia DEGs were broadly enriched for clathrin-coated endocytosis (up) and viral response (down), they also had diverse region-specific enrichments, including upregulation of

major histocompatibility complex type II (MHC-II) binding in the EC and HC, RNA processing in thalamus and glycolysis in the PFC and EC; and HC-driven downregulation of phagocytosis, phospholipase signalling and protein kinase activity (Extended Data Fig. 8c).

The majority of region-specific DEGs was either broadly shared (on average, 11% of genes were differentially expressed in 3+ cell types in a region) or were in cell-type-specific programs (40% of DEGs were in 3+ regions for a cell type) (Extended Data Fig. 8d,e). Such genes included *SLC38A2* and *EIF4G2* (broadly shared across regions) and *PRDX5*, *HLA-DRA* or *CD44*, upregulated DEGs in excitatory neurons, microglia and astrocytes, respectively (Extended Data Fig. 8f and Supplementary Fig. 40). Broadly shared genes across cell types showed region-specific enrichment, including for DNA damage (EC), amyloid-β binding and iron transport (HC) and glycolysis (thalamus) in upregulated genes as well as for phospholipid biosynthesis and autophagy in downregulated genes (Extended Data Fig. 8e). Gene sets based on DEGs for global AD pathology burden in the PFC across 427 individuals changed consistently in each region and glial cell type across global pathology, indicating that a significant component of the glial AD response is consistent across regions[27] (Extended Data Figs. 8b–e and 9a,b). The remaining regional DEGs (on average, 48% of DEGs) highlighted region- and cell-type-specific changes. In microglia, these included *PPARG* and *MSR1*, upregulated in the HC, each associated with microglia polarization, as well as upregulation of lipoprotein modifier *APOC1* and downregulation of transcription factor *FOXP2* in the EC (Extended Data Fig. 8f and Supplementary Fig. 40).

We next examined which cell types and regions were most enriched for genes identified in genome-wide association studies (GWASs) of AD by computing GWAS scores for each cell using single-cell disease-relevance score (scDRS)[54,55]. Microglia and immune cells showed consistently high scores across regions, with the top scores for the microglia *TPT1*$^+$ subtype and macrophages in the HC, thalamus and AG (Extended Data Fig. 9c). We examined whether GWAS genes showed region-specific differences in expression that might be linked to the region specificity of AD progression. We identified eight GWAS genes with region-specific expression in microglia, including *PLCG2* (EC), *APOE* and *SORL1* (thalamus), and *MS4A4A* (midtemporal cortex) (Extended Data Fig. 9c–f).

To determine whether GWAS-identified genes have regional associations with Alzheimer's pathology, we intersected DEGs for regional pathology measurements with 149 identified familial AD and GWAS locus genes[56–58] (Extended Data Figs. 10 and 11a). We found that 74 genes (49%) were differentially expressed for at least one cell type, and multiple genes showed region-specific expression, including the lipid transporter *ABCA7* (enriched in thalamus), the zinc-finger protein *ZNF655* (EC) and the complement receptor CR1 (neocortex)[56] (Extended Data Fig. 10). GWAS and familial AD genes were maximally expressed (75 genes) and differentially expressed (30 genes) in microglia, and 25 genes were differentially expressed in at least three cell types, including upregulated *CLU*, *PLCG2* and *SORT1*, and downregulated *DENND6A* (Extended Data Fig. 10). Among all of the cell types, astrocytes and microglia showed the largest differential changes for these genes in regions with high neuritic plaque density, for example, for *APOE*, *HLA-DRA*, *PILRA* and *SORT1*, and showed the most response to diffuse plaque. Neurons and oligodendrocyte-lineage cells showed stronger differences for these genes, including for *PLCG2*, *CLU* and *MAF*, in regions with high NFT density (Extended Data Fig. 10).

Pathology-specific expression changes

To determine whether different pathologies induce distinct transcriptional responses, we computed DEGs for region-specific measurements of NFT and neuritic amyloid-β plaque burden (measured in each region except thalamus) (Fig. 4a, Extended Data Fig. 11a,b and Supplementary Fig. 41). DEGs for AD pathology showed a high overlap with DEGs for pathologic diagnosis (NFT: 45% and plaque: 53% on average) (Fig. 4a). Agreement between NFT and plaque DEGs was highest in the EC and HC for all cell types (average adjusted R^2 of 67% in both) and lowest in the PFC (43%) and AG (21%), consistent with late-AD NFT appearance in the neocortex (Fig. 4b).

We next identified genes with higher differential effects in either NFTs or neuritic plaques (Fig. 4c, Extended Data Fig. 11c–f and Supplementary Table 10). Consistently, NFT-associated genes (374 genes, differentially expressed in 2+ cell types) included *PLCG2*, *CLU* and *CTNNA2* (in oligodendrocytes and OPCs) and mitochondrial subunits, and were enriched for ER protein processing, electron transport and cadherin binding (Fig. 4d,e). Neuritic-plaque-associated genes (190 genes) included the energy-homeostasis-regulating genes *IRS2*, *PDK4* and *HIF3A*, and genes enriched for immune response, chromatin regulation and lipid droplets. Notably, in excitatory neurons, plaque-associated and upregulated DEGs were strongly enriched for aerobic transport chain components (including *NDUFA4* and *COX6B1*) (Fig. 4f,g). On the other hand, NFT-associated and downregulated DEGs were enriched for TCA cycle genes, whereas upregulated DEGs were enriched for unfolded protein response and lysosome-linked genes. Finally, astrocytes contained more plaque-associated DEGs compared with other cell types, and their pathology-associated DEGs were enriched in our expression modules, including in metallostasis (M28) for plaque-associated DEGs and oxidative phosphorylation (M27) and chaperones (M8) for NFT-associated DEGs (Fig. 4c,h–j). Interestingly, a reactive astrocyte module (M3) was enriched in upregulated genes for plaques but in downregulated genes for NFTs (Fig. 4j).

Given the enrichment of NFT-associated or plaque-associated DEGs in expression modules, we next examined whether gene modules were enriched for AD DEGs (for AD pathology or for AD diagnosis) (Fig. 4k, Extended Data Fig. 11g and Supplementary Fig. 42). The same modules enriched for pathology-associated astrocyte DEGs were also enriched for the full sets of DEGs, including metallostasis (M28) with neuritic plaque DEGs and oxidative phosphorylation (M27) with NFT DEGs (Fig. 4k). Modules including ECM, adhesion and neurogenesis-related genes were much lower in AD (M1 and M11), while the modules for specific astrocyte subtypes (M7, *DCLK1*$^+$; and M24, *DPP10*$^+$) were enriched for upregulated DEGs (Fig. 4k).

We independently identified modules for heat-shock chaperones, glycolysis and oxidative phosphorylation in multiple cell types, which were correlated across cell types and were enriched for upregulated DEGs (Fig. 4k, Extended Data Figs. 6a,b and 11g and Supplementary Fig. 42a,b). The glycolysis modules were enriched among diffuse plaque DEGs in microglia and astrocytes and shared a set of genes that included canonical glycolysis genes (*PDK1/3/4*, *PFKL/P*, *PKM* and *PGK1*), anaerobic glycolysis enzymes (*TPI1* and *LDHA*) and stress-induced genes (*EGLN1*, *DDIT4*, *VEGFA* and *BNIP3L*) (Fig. 4k–m and Extended Data Fig. 6a,b). All glial types upregulated core glycolysis driver *GAPDH* and mitophagy-regulating *BNIP3L* in response to NFT burden and in individuals with cognitive impairment[59] (Fig. 4m). In regions with high diffuse plaque, astrocytes upregulated glycolysis enzymes converting glucose-6-phosphate to pyruvate, while downregulating *MPC1*, the mitochondrial pyruvate transporter[60] (Fig. 4n). In parallel, astrocytes uniquely upregulated *DDIT4*, *PFKP* and *ADCY8*, along with genes that suppress fatty acid metabolism (*ANGPTL4*) and promote lipid droplet storage of fatty acids (*HILPDA*), while microglia upregulated multiple glycogen-related genes (*GBE1*, *UGP2* and *PYGL*)[48] (Fig. 4m).

To validate differential expression of *ADCY8* and *PFKP*, we performed in situ hybridization (RNAscope) in AG tissue samples from patients with AD and control individuals without AD and found a significant increase in transcripts of both genes in AQP4$^+$ astrocytes (Fig. 4o,p). Finally, we noticed that the glycolysis pathway genes were maximally expressed at different points in global AD progression for each region (pathology diagnosis by ABC score)[29,30]. The pathway peaked very early

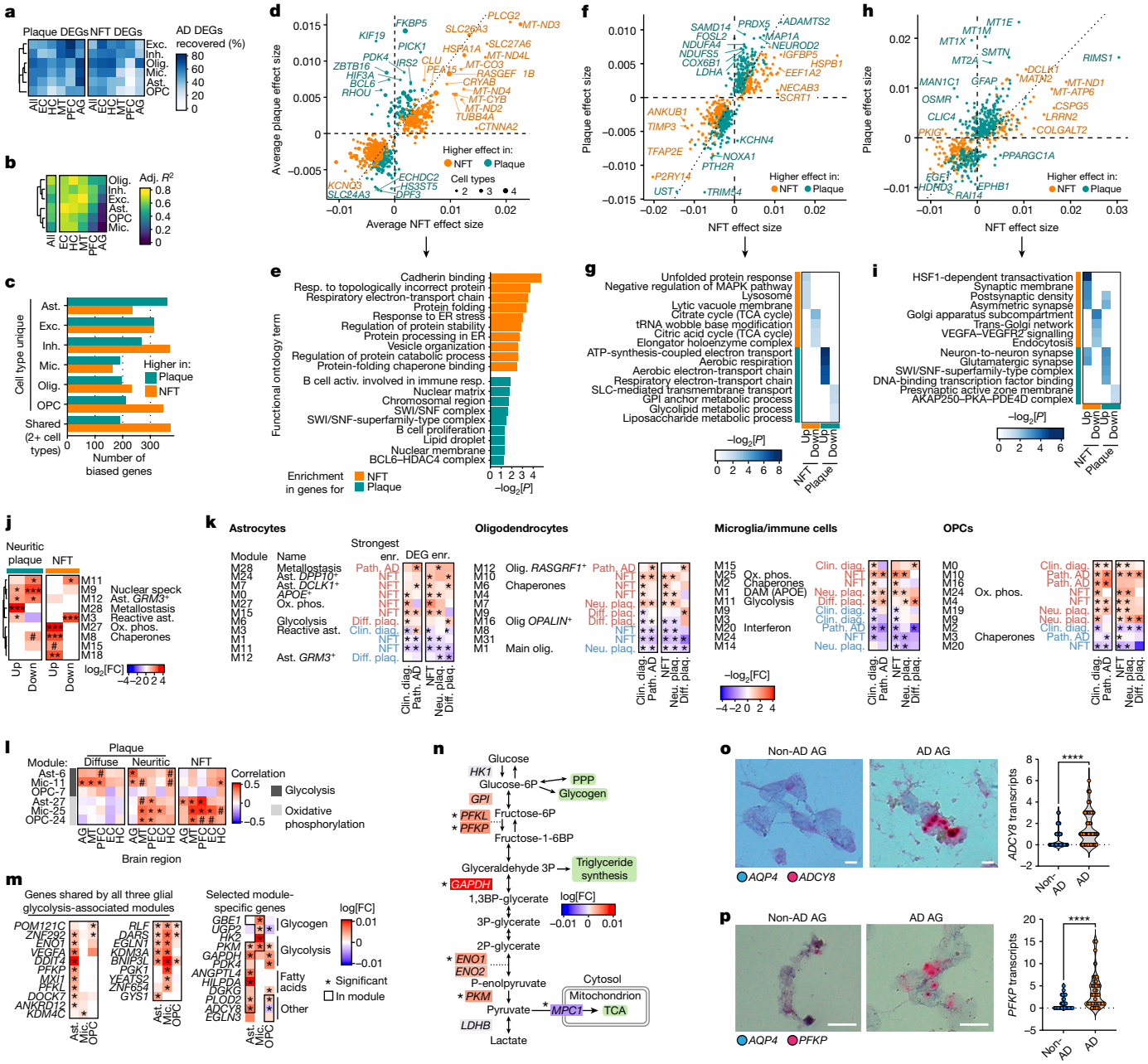


Fig. 4 | Gene expression modules annotate and separate AD changes across pathology. a, The percentage of AD DEGs (pathologic diagnosis) overlapping with DEGs for neuritic plaques (neu. plaq.) and NFTs in each major cell type and region. **b**, Concordance of effect-sizes between neuritic plaque and NFT DEGs. Adjusted R^2 of log-transformed fold changes between neuritic plaque and NFT DEGs in each major cell type and region. **c**, The number of neuritic-plaque- or NFT-biased DEGs (≥3 DEGs for one of plaques or NFTs, and ≤2 for the other) for each major cell type or shared across 2+ cell types. **d**–**i**, The average effect sizes for NFTs and neuritic plaques for DEGs with biased differential effect sizes (**d**,**f**,**h**) and their respective functional enrichments (**e**,**g**,**i**), for DEGs shared across multiple cell types (**d**,**e**), in excitatory neurons (**f**,**g**) or in astrocytes (**h**,**i**). **j**, Enrichments (hypergeometric test) of pathology-biased DEGs in astrocyte modules. **k**, Enrichments (enr.) of AD DEGs in glial gene expression modules ($*P_{adj} < 0.05$, signed $\log_2$[fold change], only significant modules are shown). **l**, Pearson correlation of module scores in each region with region-level pathology measures for glycolysis and oxidative phosphorylation modules in astrocytes, microglia and OPCs ($^{\#}P < 0.1$). **m**, Core and selected diffuse plaque (diff. plaq.) DEGs in glial glycolysis-associated modules. **n**, Schematic of the glycolysis pathway, annotated by astrocyte diffuse plaque DEGs. Significant DEGs for diffuse plaques across all regions are indicated by asterisks. **o**,**p**, RNAscope validation of astrocyte energy metabolism DEGs in the AG of individuals with AD relative to control individuals without AD (pathologic diagnosis of AD). Representative images (left) show *AQP4* transcripts (blue puncta) and *ADCY8* (**o**) or *PFKP* (**p**) transcripts (red puncta). Scale bars, 20 µm (**o** and **p**). Quantification (right) was performed using unpaired two-tailed Student's *t*-tests (*ADCY8*: $n = 117$ (non-AD) and $n = 76$ (AD) cells; *PFKP*: $n = 43$ (non-AD) and $n = 40$ (AD) cells). The dots represent individual cells, pooled from eight samples (four individuals; each had one PFC and one thalamus sample). Activ., activation; DAM, disease associated microglia; ox. phos., oxidative phosphorylation; resp., response.

in the EC (ABC score of 1, low levels of AD pathology), later in the HC and midtemporal cortex (intermediate levels), and very late in the PFC (high levels) (Supplementary Fig. 42c), suggesting that the glial metabolic response to AD may not be coordinated globally.

Astrocytes and cognitive resilience

In addition to understanding cellular alterations associated with specific pathological measures in AD, we investigated what transcriptional

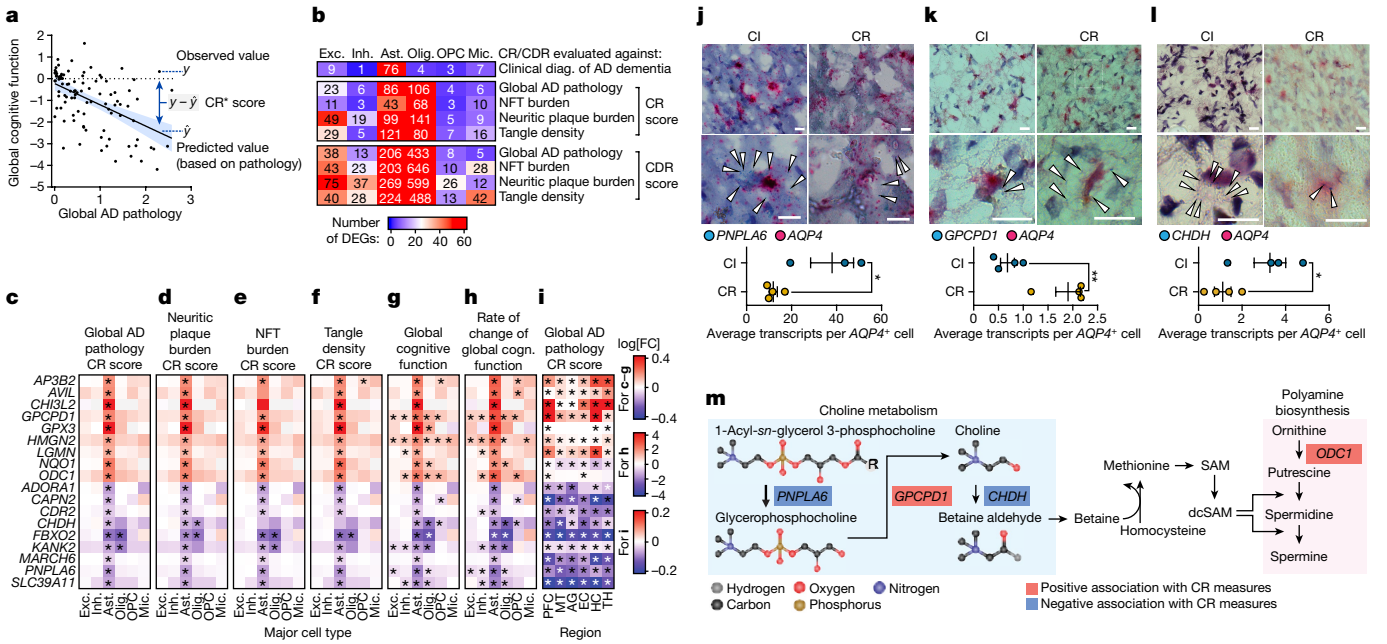


Fig. 5 | Molecular correlates of CR to AD pathology. a, The concept of CR and CDR scores. Pathology measurements are used to predict global cognitive function, for CR scores, or rate of cognitive decline, for CDR scores. **b**, The number of significant DEGs in major cell types across nine measures of CR. **c**–**h**, Association of astrocyte CR genes with measures of CR (global AD pathology CR score (**c**), neuritic plaque burden CR score (**d**), NFT burden CR score (**e**), tangle density CR score (**f**), global cognitive (cogn.) function (**g**) and rate of change of global cognitive function (**h**)) across six major cell types in the PFC (427 individuals, DEGs were computed using muscat). **i**, The association between the expression of CR genes in astrocytes across six brain regions and CR to global AD pathology (48 individuals; DEGs were computed using MAST). **j**–**l**, RNAscope validation of the differentially expressed astrocyte CR genes *PNPLA6* (**j**), *GPCPD1* (**k**) and *CHDH* (**l**) in the PFC of individuals with cognitive impairment (CI) relative to cognitively resilient (CR) individuals. Representative images (top) show *AQP4* transcripts (red puncta) and CR gene transcripts (blue puncta). Scale bars, 20 µm (**j**–**l**). Quantification (bottom) was performed using unpaired two-tailed Student's *t*-tests; $P = 0.0249$ (**j**), $P = 0.0052$ (**k**), $P = 0.0375$ (**l**). Data are mean ± s.e.m. *PNPLA6*: $n = 3$ (CI) and $n = 4$ (CR) individuals; *GPCPD1* and *CHDH*: $n = 4$ individuals per group. **m**, Schematic of choline metabolism and polyamine biosynthesis; significant astrocyte CR genes are highlighted.

changes are associated with cognitive resilience (CR) in AD, cases in which individuals with AD brain pathology display much less cognitive impairment than expected[3,25–27]. To identify potential molecular mediators that confer CR to AD pathology, we defined CR either categorically as the absence of cognitive impairment despite a pathologic diagnosis of AD (clinical diagnosis condition), or continuously, as the difference between observed cognition and the cognition expected on the basis of pathology level (Fig. 5a). We computed both scores for CR based on global cognitive function and for cognitive decline resilience (CDR) based on the rate of change of global cognitive function over time, and used four different measures of AD pathology: global AD pathology, neuritic plaque burden, NFT burden and tangle density (Fig. 5a).

We calculated DEGs for both CR and CDR in each major cell type in the PFC (snRNA-seq from 427 ROSMAP study participants)[27]. Astrocytes were the only cell type with a consistently high number of genes associated with CR across all of the measures tested (Fig. 5b). To identify specific molecular pathways within astrocytes that may contribute to CR, we focused on genes that are consistently associated with multiple measures of CR in astrocytes (termed CR-associated genes). Several CR-associated genes, including *GPX3*, *HMGN2*, *NQO1* and *ODC1* (encoding a rate-limiting enzyme of polyamine biosynthesis), possess or promote antioxidant activities[61–66] (Fig. 5c–f and Supplementary Fig. 43a–d). The expression of *HMGN2*, *NQO1*, *ODC1* and *GPX3* in astrocytes was also positively associated with cognitive function (Fig. 5g and Supplementary Fig. 43e), and these genes exhibited the highest expression in astrocytes isolated from those individuals with the least cognitive decline over time (Fig. 5h and Supplementary Fig. 43f). Analysis of bulk RNA-seq data from the ROSMAP cohorts ($n = 638$) confirmed a significant positive association between the expression level of *HMGN2*, *ODC1* and *GPX3* and multiple measures of cognitive function and CR to AD pathology (Supplementary Fig. 44a–d).

Furthermore, we noticed that several CR-associated genes within astrocytes encode enzymes that catalyse metabolic reactions that are involved in choline formation and breakdown. The expression of *GPCPD1*, which encodes glycerophosphocholine phosphodiesterase 1, an enzyme that is critical for cleaving glycerophosphocholine (GPC) to produce choline, was positively associated with measures of CR in astrocytes (Fig. 5c–f and Supplementary Fig. 43a–d). Conversely, *PNPLA6*, which encodes a phospholipase that catalyses the hydrolysis of intracellular phosphatidylcholine, a major membrane lipid, generating GPC, and *CHDH*, which encodes choline dehydrogenase, an enzyme that catalyses the conversion of choline to betaine aldehyde, were both negatively associated with multiple measures of CR in astrocytes (Fig. 5c–f and Supplementary Fig. 43a–d). Many of the CR-associated genes identified in PFC astrocytes were also associated with CR in astrocytes from other regions of the human brain (Fig. 5i and Supplementary Fig. 45), corroborating a link between astrocytes and CR beyond the PFC.

To validate the choline pathway genes *PNPLA6*, *GPCPD1* and *CHDH*, we selected PFC samples from individuals with high amyloid and tau pathology and compared transcript levels between individuals with intact cognition (that is, cognitively resilient) to those with cognitive impairment, and performed in situ hybridization (RNAscope) analysis of these genes with *AQP4* as a marker for astrocytes. We found a significant decrease in *PNPLA6* and *CHDH* transcripts and a significant increase in *GPCPD1* transcripts in cognitively resilient individuals, in agreement with the differential expression results (Fig. 5j–l). Notably, choline oxidation to betaine generates a labile methyl group that can be used for homocysteine remethylation, resulting in methionine

formation, which is subsequently transformed into the universal methyl donor *S*-adenosylmethionine[67]. *S*-adenosylmethionine is involved in the biosynthesis of spermidine, linking choline metabolism and polyamine biosynthesis in astrocytes in CR to AD pathology (Fig. 5m).

Discussion

Here we present a transcriptomic atlas of the aged human brain—spanning six brain regions from 48 individuals with and without a diagnosis of AD—that we used to annotate regional cellular diversity, identify gene expression programs and differences in AD across cell types, and pinpoint region-specific cell populations that are vulnerable to AD. We provide an interactive website for exploring the atlas and these annotations, markers, functional modules and differences in AD at both the single-cell and pseudo-bulk levels (http://compbio.mit.edu/ad_multiregion).

By annotating neuronal and glial subtypes by brain region, we found significant compositional differences between regions, including a subtype of thalamic GABAergic neurons (*MEIS2*$^{+}$*FOXP2*$^{+}$) that is molecularly distinct from the canonical subclasses of inhibitory neurons in the neocortex. We used region-specific measurements of AD pathology to identify changes in gene expression associated with neurofibrillary tau tangle or amyloid-β plaque burden, including plaque-associated upregulation of metallostasis in astrocytes and of the electron-transport chain in excitatory neurons. We found that AD-risk genes were highly perturbed in AD—in particular for microglia, consistent with their enrichment for GWAS signal[68]—but few risk genes showed region-specific expression. To further examine the cellular and regional heterogeneity of the human brain, we developed a scalable method, scdemon, which uses sample decorrelation to annotate both ubiquitous and rare gene expression programs in each major cell type, and used annotated modules to identify functional programs associated with specific pathological variables, including a glycolysis- and energy-metabolism-linked program in glia[48,60] associated with diffuse plaque burden.

We identified five excitatory neuron subtypes that were reduced in patients with AD (vulnerable subtypes) in the early-affected EC and HC[1,17,18], including EC layer II (L2), *RORB*-positive L5 (*AGBL1*$^{+}$*GPC5*$^{+}$)[19] and hippocampal CA1 subfield neurons[20–23]. Notably, vulnerable excitatory neurons shared expression of genes involved in Reelin signalling and heparan sulfate proteoglycan biosynthesis, both of which were also predictive of inhibitory neuron vulnerability to AD. Recent case studies have identified variants in *RELN* and *APOE* as potentially mediating CR to autosomal-dominant AD. Notably, the *RELN* variant enhanced its binding to glycosaminoglycans (GAGs) and *NRP1*, and the *APOE* variant decreased binding to GAGs, potentially affecting their ability to compete for receptor binding[69,70]. Thus, our findings suggest a convergence of factors associated with cellular vulnerability in sporadic AD, and resilience to autosomal-dominant AD.

Finally, we analysed the transcriptomic correlates of cognition and pathology in AD, and identified a set of astrocyte genes linked to CR to AD pathology. Notably, these genes converged on the pathways of choline metabolism and polyamine biosynthesis. This finding aligns with studies showing benefits of dietary choline intake and supplementation on cognitive performance in human individuals and in animal models[71–78]. Similarly, dietary supplementation with the polyamine spermidine prolongs life span and health span in several animal models[66], and spermidine has also been shown to enhance memory performance and counteract age-related cognitive decline[79–81]. Our findings support choline metabolism and polyamine synthesis as attractive targets for promoting CR in AD.

Our study has several limitations: isotropic fractionation and read depth cut-offs may bias cell recovery based on their nuclear content; nuclear RNA may not fully capture microglial states[82] or localized transcriptomic changes; and pathology burden is based on per-sample averages instead of on the spatial context of each cell. Additional individuals and data modalities will strengthen future analyses of region-specific alterations in AD, and spatial data may help to further separate pathology-associated changes.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-07606-7.

1. Braak, H. & Braak, E. Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiol. Aging* **16**, 271–278 (1995).
2. Wilson, R. S. et al. Temporal course and pathologic basis of unawareness of memory loss in dementia. *Neurology* **85**, 984–991 (2015).
3. Stern, Y. et al. Whitepaper: defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimers Dement.* **133**, 484–491 (2020).
4. Sonnen, J. A. et al. Pathological correlates of dementia in a longitudinal, population-based sample of aging. *Ann. Neurol.* **62**, 406–413 (2007).
5. Bennett, D. A. et al. Natural history of mild cognitive impairment in older persons. *Neurology* **59**, 198–205 (2002).
6. 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).
7. Braak, H. & Braak, E. Alzheimer's disease affects limbic nuclei of the thalamus. *Acta Neuropathol.* **81**, 261–268 (1991).
8. Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
9. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
10. 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).
11. Grubman, A. et al. A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat. Neurosci.* **22**, 2087–2097 (2019).
12. Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
13. Franjic, D. et al. Transcriptomic taxonomy and neurogenic trajectories of adult human, macaque, and pig hippocampal and entorhinal cells. *Neuron* **110**, 452–469 (2022).
14. Siletti, K. et al. Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046 (2023).
15. Gabitto, M. I. et al. Integrated multimodal cell atlas of Alzheimer's disease. Preprint at *bioRxiv* https://doi.org/10.1101/2023.05.08.539485 (2023).
16. Sziraki, A. et al. A global view of aging and Alzheimer's pathogenesis-associated cell population dynamics and molecular signatures in human and mouse brains. *Nat. Genet.* **55**, 2104–2116 (2023).
17. Chin, J. et al. Reelin depletion in the entorhinal cortex of human amyloid precursor protein transgenic mice and humans with Alzheimer's disease. *J. Neurosci.* **27**, 2727–2733 (2007).
18. Kobro-Flatmoen, A. et al. Re-emphasizing early Alzheimer's disease pathology starting in select entorhinal neurons, with a special focus on mitophagy. *Ageing Res. Rev.* **67**, 101307 (2021).
19. Leng, K. et al. Molecular characterization of selectively vulnerable neurons in Alzheimer's disease. *Nat. Neurosci.* **24**, 276–287 (2021).
20. Kordower, J. H. et al. Loss and atrophy of layer II entorhinal cortex neurons in elderly people with mild cognitive impairment. *Ann. Neurol.* **49**, 202–213 (2001).
21. Van Hoesen, G. W., Hyman, B. T. & Damasio, A. R. Entorhinal cortex pathology in Alzheimer's disease. *Hippocampus* **1**, 1–8 (1991).
22. Gómez-Isla, T. et al. Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. *J. Neurosci.* **16**, 4491–4500 (1996).
23. Price, J. L. et al. Neuron number in the entorhinal cortex and CA1 in preclinical Alzheimer disease. *Arch. Neurol.* **58**, 1395–1402 (2001).
24. Kampmann, M. Molecular and cellular mechanisms of selective vulnerability in neurodegenerative diseases. *Nat. Rev. Neurosci.* **25**, 351–371 (2024).
25. Arenaza-Urquijo, E. M. & Vemuri, P. Resistance vs resilience to Alzheimer disease: clarifying terminology for preclinical studies. *Neurology* **90**, 695–703 (2018).
26. Wagner, M. et al. Quantifying longitudinal cognitive resilience to Alzheimer's disease and other neuropathologies. *Alzheimers Dement.* **18**, 2252–2261 (2022).
27. Mathys, H. et al. Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* **186**, 4365–4385 (2023).
28. Bennett, D. A. et al. Religious orders study and rush memory and aging project. *J. Alzheimers Dis.* **64**, S161–S189 (2018).
29. Agrawal, S. et al. Brain autopsies of critically ill COVID-19 patients demonstrate heterogeneous profile of acute vascular injury, inflammation and age-linked chronic brain diseases. *Acta Neuropathol. Commun.* **10**, 186 (2022).
30. Hyman, B. T. et al. National Institute on Aging—Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimers Dement.* **8**, 1–13 (2012).
31. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
32. Sorrells, S. F. et al. Human hippocampal neurogenesis drops sharply in children to undetectable levels in adults. *Nature* **555**, 377–381 (2018).

33. Hodge, R. D. et al. Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
34. Kotliar, D. et al. Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *eLife* **8**, e43803 (2019).
35. Van de Sande, B. et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* **15**, 2247–2276 (2020).
36. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
37. Pakkenberg, B. & Gundersen, H. J. Total number of neurons and glial cells in human brain nuclei estimated by the disector and the fractionator. *J. Microsc.* **150**, 1–20 (1988).
38. Azevedo, F. A. C. et al. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J. Comp. Neurol.* **513**, 532–541 (2009).
39. Saunders, A. et al. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* **174**, 1015–1030 (2018).
40. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014 (2018).
41. Franjic, D. et al. Transcriptomic taxonomy and neurogenic trajectories of adult human, macaque, and pig hippocampal and entorhinal cells. *Neuron* **110**, 452–469.e14 (2022).
42. Ortiz, C. et al. Molecular atlas of the adult mouse brain. *Sci. Adv.* **6**, eabb3446 (2020).
43. Yao, Z. et al. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222–3241 (2021).
44. Witter, M. P. & Moser, E. I. Spatial representation and the architecture of the entorhinal cortex. *Trends Neurosci.* **29**, 671–678 (2006).
45. Amunts, K. et al. Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. *Anat. Embryol.* **210**, 343–352 (2005).
46. Fujita, M. et al. Cell subtype-specific effects of genetic variation in the Alzheimer's disease brain. *Nat. Genet.* **56**, 605–614 (2024).
47. Lau, S.-F., Cao, H., Fu, A. K. Y. & Ip, N. Y. Single-nucleus transcriptome analysis reveals dysregulation of angiogenic endothelial cells and neuroprotective glia in Alzheimer's disease. *Proc. Natl Acad. Sci. USA* **117**, 25800–25809 (2020).
48. Zhou, Y. et al. Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat. Med.* **26**, 131–142 (2020).
49. Su, Y. et al. A single-cell transcriptome atlas of glial diversity in the human hippocampus across the postnatal lifespan. *Cell Stem Cell* **29**, 1594–1610 (2022).
50. Morabito, S. et al. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat. Genet.* **53**, 1143–1155 (2021).
51. Johnson, E. C. B. et al. Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nat. Med.* **26**, 769–780 (2020).
52. Zeng, H. et al. Integrative in situ mapping of single-cell transcriptional states and tissue histopathology in a mouse model of Alzheimer's disease. *Nat. Neurosci.* **26**, 430–446 (2023).
53. Otero-Garcia, M. et al. Molecular signatures underlying neurofibrillary tangle susceptibility in Alzheimer's disease. *Neuron* **110**, 2929–2948 (2022).
54. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
55. Zhang, M. J. et al. Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. *Nat. Genet.* **54**, 1572–1580 (2022).
56. Lambert, J.-C. et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat. Genet.* **41**, 1094–1099 (2009).
57. Wightman, D. P. et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat. Genet.* **53**, 1276–1282 (2021).
58. Prokopenko, D. et al. Whole-genome sequencing reveals new Alzheimer's disease-associated rare variants in loci related to synaptic function and neuronal development. *Alzheimers Dement.* **17**, 1509–1527 (2021).
59. Zhang, J. & Ney, P. A. Role of BNIP3 and NIX in cell death, autophagy, and mitophagy. *Cell Death Differ.* **16**, 939–946 (2009).
60. Magistretti, P. J. & Allaman, I. A cellular perspective on brain energy metabolism and functional imaging. *Neuron* **86**, 883–901 (2015).
61. Liu, K. et al. Nuclear protein HMGN2 attenuates pyocyanin-induced oxidative stress via Nrf2 signaling and inhibits *Pseudomonas aeruginosa* internalization in A549 cells. *Free Radic. Biol. Med.* **108**, 404–417 (2017).
62. Stringer, J. L. et al. Presence and induction of the enzyme NAD(P)H: quinone oxidoreductase 1 in the central nervous system. *J. Comp. Neurol.* https://doi.org/10.1002/cne.20048 (2004).
63. Jaber, S. M. et al. Idebenone has distinct effects on mitochondrial respiration in cortical astrocytes as compared to cortical neurons due to differential NQO1 activity. *J. Neurosci.* https://doi.org/10.1523/JNEUROSCI.1632-17.2020 (2020).
64. Ross, D. & Siegel, D. The diverse functionality of NQO1 and its roles in redox control. *Redox Biol.* **41**, 101950 (2021).
65. Huang, J. et al. Spermidine exhibits protective effects against traumatic brain injury. *Cell. Mol. Neurobiol.* https://doi.org/10.1007/s10571-019-00783-4 (2020).
66. Madeo, F., Eisenberg, T., Pietrocola, F. & Kroemer, G. Spermidine in health and disease. *Science* **29**, 1243–1252 (2018).
67. Korsmo, H. W. et al. Maternal choline supplementation and high-fat feeding interact to influence DNA methylation in offspring in a time-specific manner. *Front. Nutr.* **9**, 841787 (2022).
68. Gjoneska, E. et al. Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature* **518**, 365–369 (2015).
69. Arboleda-Velasquez, J. F. et al. Resistance to autosomal dominant Alzheimer's disease in an APOE3 Christchurch homozygote: a case report. *Nat. Med.* **25**, 1680–1683 (2019).
70. Lopera, F. et al. Resilience to autosomal dominant Alzheimer's disease in a Reelin-COLBOS heterozygous man. *Nat. Med.* **29**, 1243–1252 (2023).
71. Teather, L. A. & Wurtman, R. J. Dietary cytidine (5′)-diphosphocholine supplementation protects against development of memory deficits in aging rats. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **27**, 711–717 (2003).
72. Poly, C. et al. The relation of dietary choline to cognitive performance and white-matter hyperintensity in the Framingham Offspring Cohort. *Am. J. Clin. Nutr.* **94**, 1584–1591 (2011).
73. Teather, L. A. & Wurtman, R. J. Dietary CDP-choline supplementation prevents memory impairment caused by impoverished environmental conditions in rats. *Learn. Mem.* **12**, 39 (2005).
74. Nurk, E. et al. Plasma free choline, betaine and cognitive performance: the Hordaland Health Study. *Br. J. Nutr.* **109**, 511–519 (2013).
75. Blusztajn, J. K., Slack, B. E. & Mellott, T. J. Neuroprotective actions of dietary choline. *Nutrients* **9**, 815 (2017).
76. Liu, L. et al. Choline intake correlates with cognitive performance among elder adults in the United States. *Behav. Neurol.* **2021**, 2962245 (2021).
77. Yuan, J. et al. Is dietary choline intake related to dementia and Alzheimer's disease risks? Results from the Framingham Heart Study. *Am. J. Clin. Nutr.* **116**, 1201–1207 (2022).
78. Sienski, G. et al. APOE4 disrupts intracellular lipid homeostasis in human iPSC-derived glia. *Sci. Transl. Med.* **13**, eaaz4564 (2021).
79. Sigrist, S. J. et al. Spermidine-triggered autophagy ameliorates memory during aging. *Autophagy* **10**, 178–179 (2014).
80. Schroeder, S. et al. Dietary spermidine improves cognitive function. *Cell Rep.* **35**, 108985 (2021).
81. Pekar, T. et al. The positive effect of spermidine in older adults suffering from dementia: first results of a 3-month trial. *Wien. Klin. Wochenschr.* **133**, 484–491 (2021).
82. Thrupp, N. et al. Single-nucleus RNA-seq is not suitable for detection of microglial activation genes in humans. *Cell Rep.* **32**, 108189 (2020).

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2024

Methods

Data reporting

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

snRNA-seq

Sample selection from ROSMAP. We selected 48 individuals from ROSMAP, both ongoing longitudinal clinical–pathological cohort studies of ageing and dementia, in which all of the participants are brain donors. The studies include clinical data collected annually, detailed post-mortem pathological evaluations, and extensive genetic, epigenomic, transcriptomic, proteomic and metabolomic bulk-tissue profiling[28]. For the purpose of this study, individuals were selected based on the modified NIA-Reagan diagnosis of AD and the Braak stage score (Braak stages 0, 1 and 2, $n = 20$; Braak stages 3 and 4, $n = 14$; Braak stages 5 and 6, $n = 14$), with 26 individuals having a positive pathologic diagnosis of AD and 22 individuals having a negative pathologic diagnosis of AD[83]. Details of the clinical and pathological data collection methods have been previously reported[2,5,6,28,84]. Individuals were balanced between sexes (male:female ratios 13:13 in AD, 11:11 in NoAD), matched for age (median, 86.6 years (AD) and 86.0 years (no AD)) and post-mortem interval (median, 5.9 h (AD) and 6.3 h (no AD)). Informed consent was obtained from each participant, and the Religious Orders Study and Rush Memory and Aging Project were each approved by an Institutional Review Board (IRB) of Rush University Medical Center. The participants also signed an Anatomic Gift Act, and a repository consent to allow their data to be repurposed.

Dissection criteria. All dissections were done on a bed of dry ice using either a fine-toothed razor saw (for cortical regions) or a jewellers saw with diamond wire (for subcortical regions). Region-specific descriptions are as follows. (1) AG: full thickness cortex from the AG (Brodmann area: BA 39/40); take from the first or second slab posterior to the end of the HC. Minimize white matter. (2) MT: full thickness cortex from the middle temporal gyrus (BA 22); take as close to the level of the anterior commissure as possible. Minimize white matter. (3) PFC: full thickness cortex from the frontal pole (BA 10); take from the lateral side of the first or second slab. Minimize white matter. (4) EC: full thickness cortex from the EC (BA 28); take at the level of the amygdala. Avoid amygdala. Minimize white matter. (5) Posterior HC: take from the last slab containing HC. If the last slab has less than 5 mm of HC, take from the next slab anterior. Collect a full cross section. (6) TH: take from the first slab with thalamus. Include the most medial aspect.

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[12]. All of the procedures were performed on ice or at 4 °C. In brief, post-mortem brain tissue was homogenized in 700 µl homogenization buffer (320 mM sucrose, 5 mM $CaCl_2$, 3 mM $Mg(CH_3COO)_2$, 10 mM Tris HCl pH 7.8, 0.1 mM EDTA pH 8.0, 0.1% IGEPAL CA-630, 1 mM β-mercaptoethanol and 0.4 U µl^{-1} recombinant RNase inhibitor (Clontech)) using a Wheaton Dounce tissue grinder (15 strokes with the loose pestle). The homogenized tissue was then filtered through a 40 µm cell strainer, mixed with an equal volume of working solution (50% OptiPrep density gradient medium (Sigma-Aldrich), 5 mM $CaCl_2$, 3 mM $Mg(CH_3COO)_2$, 10 mM Tris HCl pH 7.8, 0.1 mM EDTA pH 8.0 and 1 mM β-mercaptoethanol) and loaded on top of an OptiPrep density gradient (750 µl 30% OptiPrep solution (30% OptiPrep density gradient medium, 134 mM sucrose, 5 mM $CaCl_2$, 3 mM $Mg(CH_3COO)_2$, 10 mM Tris HCl pH 7.8, 0.1 mM EDTA pH 8.0, 1 mM β-mercaptoethanol, 0.04% IGEPAL CA-630 and 0.17 U µl^{-1} recombinant RNase inhibitor)) on top of 300 µl 40% OptiPrep solution (40% OptiPrep density gradient medium, 96 mM sucrose, 5 mM $CaCl_2$, 3 mM $Mg(CH_3COO)_2$, 10 mM Tris HCl pH 7.8, 0.1 mM EDTA pH 8.0, 1 mM β-mercaptoethanol, 0.03% IGEPAL CA-630 and 0.12 U µl^{-1} recombinant RNase inhibitor). The nuclei were separated by centrifugation (5 min, 10,000*g*, 4 °C). A total of 100 µl of nuclei was collected from the 30%/40% interphase and washed with 1 ml of PBS containing 0.04% BSA. The nuclei were centrifuged at 300*g* for 3 min (4 °C) and washed with 1 ml of PBS containing 0.04% BSA. The nuclei were then centrifuged at 300*g* for 3 min (4 °C) and resuspended in 100 µl PBS containing 0.04% BSA. The nuclei were counted and diluted to a concentration of 1,000 nuclei per µl 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 v3 according to the manufacturer's protocol (10x Genomics). The generated snRNA-seq libraries were sequenced using NextSeq 500/550 High Output v2 kits (150 cycles) or NovaSeq 6000 S2 reagent kits.

snRNA-seq processing, QC, and annotation

snRNA-seq data preprocessing. Gene counts were obtained by aligning reads to the GRCh38 genome using Cell Ranger software (v.3.0.2) (10x Genomics)[85]. To account for unspliced nuclear transcripts, reads mapping to pre-mRNA were counted. After quantification of pre-mRNA using the Cell Ranger count pipeline, the Cell Ranger aggr pipeline was used to aggregate all libraries (without equalizing the read depth between groups) to generate a gene–count matrix. The Cell Ranger v.3.0 default parameters were used to call cell barcodes. We used SCANPY[86] to process and cluster the expression profiles and infer cell identities. We retained only protein-coding genes and filtered out cells with over 20% mitochondrial or 5% ribosomal RNA, leaving 1.47 million cells over 48 individuals and 283 samples across all regions. We further filtered the dataset to the top 5,000 most variable genes and used them to calculate the low dimensional embedding of the cells (UMAP) (default parameters, using 50 principal components and 15 nearest neighbours) and clusters using the Leiden clustering algorithm at a high resolution (15), giving 337 preliminary clusters[87]. We separately called doublets using DoubletFinder and flagged and removed clusters with strong doublet profiles and clusters showing strong individual-specific batch effects, leaving a final dataset of 1.35 million cells[88].

Cell type annotations. For the UMAP visualization of individual major cell type classes (excitatory neurons, inhibitory neurons, astrocytes, oligodendrocytes, OPCs, immune cells), the SCTransform-based integration workflow of Seurat was used to align data from individual samples, using the default settings[89,90]. We selected the set of relevant principal components on the basis of Elbow plots. We annotated cell types using previously published marker genes and single-cell RNA-seq data[9,12,33,91–93]. To annotate cell types on the basis of previously published single-cell RNA-sequencing data (Allen Institute's cell types database; https://portal.brain-map.org/atlases-and-data/rnaseq/human-multiple-cortical-areas-smart-seq), we used three separate approaches. First, Spearman rank correlation coefficients between the average expression profiles of neuronal subpopulations previously defined by the Allen Brain Institute[33] and the neuronal subtypes identified in this study were computed using the cor function in R. Second, to project annotations of neuronal subpopulations previously defined by the Allen Brain Institute onto the neuronal cells analysed in this study, we followed the integration and label transfer workflow of Seurat[90]. Third, we determined cell type marker genes based on data published by the Allen Brain Institute[33] using the FindAllMarkers function from Seurat (Wilcoxon rank-sum test with Bonferroni correction for multiple testing; $P_{adj} < 0.05$) and computed module scores for each cell type marker gene set across all neuronal cells analysed in this study using the AddModuleScore function of Seurat. To further annotate cell types, we determined marker genes using the FindAllMarkers function from Seurat (Wilcoxon rank-sum test with Bonferroni

correction for multiple testing; $P_{adj} < 0.05$). We tested only genes that were detected in a minimum of 25% of the cells within the cell type (min.pct = 0.25) and that showed, on average, at least a 0.25-fold difference (log-scale) between the cells of the cell type and all remaining cells (logfc.threshold = 0.25). Marker genes of the high-resolution cell types or states were determined separately for each major cell type class. We additionally compared the EC excitatory neuron subtypes to cell type annotations previously reported previously[94], which were computed using ACTIONet[95], and compared microglial markers to previously reported subtypes[96,97].

Cell cycle scores and global properties of gene expression. G2/M and S phase cell cycle scores were determined using the function CellCycleScoring in Seurat. Histograms showing the distribution of the G2/M- and the S phase module scores in each major cell class were generated using Prism 9 software. The statistical analyses comparing the number of genes detected per cell and the number of unique transcripts (UMIs) detected per cell between cell types was performed using Prism 9 software.

Integration of external data sources. Single-cell transcriptomic data from the human dLGN[98] were obtained from the Allen Brain Institute (https://portal.brain-map.org/atlases-and-data/rnaseq/comparative-lgn). Single-cell transcriptomic data from multiple cortical areas and the hippocampal formation of the mouse brain[43] were obtained from the Allen Brain Institute (https://portal.brain-map.org/atlases-and-data/rnaseq/mouse-whole-cortex-and-hippocampus-10x). Single-cell transcriptomic data across nine regions in the adult mouse brain[39] were obtained from the McCarroll and Macosko Labs (http://dropviz.org/). Single-cell transcriptomic data from the mouse nervous system[40] were obtained from the Linnarsson laboratory (http://mousebrain.org/adolescent/downloads.html). The external datasets and the human multiregion data presented in this study were integrated using the reciprocal PCA (RPCA) pipeline with the default parameters in Seurat (https://satijalab.org/seurat/articles/integration_rpca.html). The integration of single-cell data was performed separately for astrocytes, excitatory neurons and inhibitory neurons. For the integration of GABAergic neurons, the single-cell transcriptomic data from multiple cortical areas and the hippocampal formation of the mouse brain[43] were downsampled to 50,000 GABAergic neurons. For the integration of excitatory neurons, the human multiregion dataset was downsampled to 5,000 neurons per high-resolution cell type. The mouse cortical data[43] were downsampled to 50,000 excitatory neurons. The frontal cortex, posterior cortex, HC and thalamus data of the DropViz dataset were combined and downsampled to 50,000 neurons. Downsampling of the data was performed using the Seurat function subset with the default parameters. The comparison of microarray data from different human brain regions was performed using the Differential Search tool of the Allen Brain Atlas data portal (https://human.brain-map.org/microarray/search). The thalamus was selected as the target structure and compared to the cerebral cortex as the contrast structure. The differential search results including the fold change values and *P* values of the top 2,000 genes were downloaded from the data portal.

Gene expression and regulon modules

Gene expression modules using ZCA (scdemon framework). We would like to find gene expression modules by calculating gene–gene correlations in single-cell data and using these to detect communities of similarly expressed genes. However, in single-cell data, which often contain an unbalanced composition of cell types, modules computed using this approach will be dominated by the most common cell type markers and pathways. Moreover, correlation values will often be inflated for pairs of sparsely expressed genes. We developed a method which accounts for these pitfalls to call multiresolution gene expression programs from single-cell data using an SVD-based approximation of zero-phase component analysis (ZCA) and gene sparsity-dependent thresholding[99,100].

scdemon (single-cell decorrelated module networks) method. The preprocessing transformations alternately called decorrelation, whitening, or sphering, transform a matrix X with a matrix W such that the covariance of XW is the identity matrix[101]. In particular, ZCA is the transformation which maximizes the similarity of the transformed data to the original, which is achieved by setting $W = C^{-1/2}$, where C is the covariance of X.

In single-cell data, given a count matrix X with n cells (rows) and g genes (columns), we would like to perform ZCA decorrelation on the samples as a preprocessing step for calling modules. Computing and storing the n by n sample-wise covariance $C_n = XX^T/g$ is prohibitively expensive for modern datasets (with $n > 1 \times 10^6$), even without centring X. Instead, we can analytically approximate the covariance with the SVD of $X^T = U_n S_n V_n^T$ as $C_n \approx (U_n S_n V_n^T)^T (U_n S_n V_n^T)/g = V_n S_n^2 V_n^T/g$ and therefore $C_n^{-1/2} = g^{1/2} V_n S_n^{-1} V_n^T$. The ZCA transformation can then be computed as $X_{ZCA} = C_n^{-1/2} X = g^{1/2} V_n S_n^{-1} V_n^T X$ before calculating the covariance of X_{ZCA} for downstream analysis. While this approximation is already tractable for small single-cell datasets, we may not be able to compute the matrix multiplications or centred SVD for larger datasets. Here, we can use the SVD of $X = USV^T$, which is commonly calculated in single-cell analyses, to approximate $C_n^{-1/2}$ as $g^{1/2} U S^{-1} U^T$ and $X_{ZCA} = g^{1/2} U S^{-1} U^T X$. From this, the non-centred covariance of X_{ZCA} is $C_{ZCA} = X_{ZCA}^T X_{ZCA}/n = g \times (US^{-1}U^T X)^T (US^{-1}U^T X)/n$. By substituting the SVD in for X, this reduces to $C_{ZCA} = g \times VV^T/n$, which is a simple and efficient approximation for very large single-cell data. As this approximation commonly drops out the largest identity program in the data due to the decorrelation approach, we allow computing $C_{ZCA} = g \times VS^p V^T/n$, for any p, to tune the relative involvement of the larger eigenvalue components of the SVD.

To control for inflated correlation estimates in highly sparse genes, we bin the estimated correlations (C_{ZCA}) for each pair of genes according to their sparsities (fraction of cells expressing the gene, binned on a $\log_{10}$ scale). We calculate the mean and s.d. for each 2D bin and smooth the estimates by fitting bivariate splines to the binned statistics, weighted by the log number of examples in the bin. We use the smoothed estimates to z-score the correlation matrix (C_{ZCA}), which we then threshold with a single z-score cut-off to create an adjacency matrix for a gene–gene graph. Graphs are laid out using the Fruchterman–Reingold algorithm and we remove connected components with fewer than four genes[102]. We then use the leidenalg package and the Leiden algorithm with an RBConfiguration vertex partition to cluster the graph into gene modules[87]. To robustly estimate modules for each set of cells in our analyses, we first performed a grid search for the optimal number of SVD components for cells of that type. We then computed the z-scored matrices for each of 10 bootstraps, selecting 90% of batches for each bootstrap and using only genes expressed in over 5% of cells in the full dataset for the cell type. We thresholded the average of the bootstrapped z-score estimates with $z = 4.5$ to build a graph. To balance the contributions of modules across the compositional spectrum, we calculated and thresholded separate graphs for eigenvalue powers $p = 0, 0.25, 0.5, 0.75$ and 1 and combined them using multigraph Leiden clustering to call modules with leiden resolution = 3. Although we identify smaller modules, here we only report modules with at least 10 genes. We also ran the modules method on three published datasets, for which we ran the method with the same parameters on each dataset ($k = 100$, $z = 4.5$, resolution = 2.5), used genes with >5% sparsity for the COVID[103] and brain[16] datasets and genes with >10% sparsity for the Tabula Sapiens dataset[104], and report modules with 10 genes or more.

Module enrichments, network and contour plots. Module enrichments for cell subtypes and brain regions were performed using the hypergeometric test by calculating whether cells with a module score above 1 s.d. from the mean score were significantly enriched in a specific

subtype or region. For plotting scores against other modules, averaged modules scores to either the subtype by sample level (within the same major cell type) or at the sample level alone (across cell types) and calculated Pearson correlations and *P* values using the cor.test function of R. To create the module–module network across microglial and immune modules, we calculated module–module score Pearson correlations using the logged module scores at the aggregated subtype by sample level, using a one-sided test with $P_{adj} < 0.01$ as a cut-off[105]. To generate contour plots of module expression on a UMAP, we first smoothed cell-level expression on a 500 × 500 grid with a 2D Gaussian kernel (size = 25 × 25 and $\sigma = 1$) and then plot contours for smoothed values (0.1 to 0.8).

Gene expression programs using cNMF. Gene expression programs underlying both cell type identity and cellular activities were determined according to the consensus NMF (cNMF) analysis pipeline established previously using the default parameters[34]. The number of components (*K*) to use for cNMF was determined on the basis of a diagnostic plot showing the stability of the solution and the Frobenius reconstruction error as a function of *K*. To reduce runtime and working memory requirements, the data were downsampled using the Seurat function subset with the default parameters. The data were downsampled to 200 cells per major cell type. For the cNMF analysis at the level of high-resolution cell types, the analysis was performed separately on excitatory neurons, inhibitory neurons and astrocytes. For these analyses, the data were downsampled to 2,000 cells per astrocyte subtype and 1,000 cells per excitatory and inhibitory neuron subtype. Statistical significance of the overlap between the top 200 genes of a gene expression program and cell type marker genes was computed using Fisher's exact tests.

SCENIC analysis and computation of regulon module scores. The gene regulatory network analysis was performed using pySCENIC with the default parameters[35]. To reduce runtime and working memory requirements, the data were downsampled to 2,000 cells per major cell type. For the SCENIC analysis at the level of high-resolution cell types, the analysis was performed separately on excitatory neurons, inhibitory neurons and astrocytes and the data were downsampled to 1,000 cells per high-resolution cell type. To identify the top cell-type-specific regulons, we calculated regulon specificity scores as described by previously and ranked the regulons based on their regulon specificity score[106]. Finally, we calculated the activity of each regulon in each cell using the AddModuleScore function of Seurat. The calculation of regulon module scores for major cell types was performed on a random sample of 50% of the cells (676,537 cells). For the analysis at the level of high-resolution cell types, the regulon module scores were determined based on all the cells of a major cell type class. For the statistical analysis of differences in the activity of regulons between cell types, the average regulon module score per individual and major cell type or high-resolution cell type was computed, respectively. The statistical analyses comparing the regulon module score activity was performed using Prism 9 software.

Analysis of GABAergic and glutamatergic module scores. GABAergic and glutamatergic module scores across all neuronal cell types were determined on the basis of a set of GABAergic and glutamatergic neuron marker genes, respectively, using the AddModuleScore function of Seurat. The sets of GABAergic and glutamatergic neuron marker genes were determined based on the human multiple cortical areas SMART-seq dataset from the Allen Brain Institute (https://portal.brain-map.org/atlases-and-data/rnaseq/human-multiple-cortical-areas-smart-seq). We identified marker genes using the FindAllMarkers function from Seurat (Wilcoxon rank-sum test with Bonferroni correction for multiple testing; $P_{adj} < 0.05$). We tested only genes that were detected in a minimum of 25% of the cells within the cell type (min.pct = 0.25) and that showed, on average, at least a 0.25-fold difference (log-scale) between the cells of the cell class of interest and all remaining cells (logfc.threshold = 0.25). To quantify the intermediate character of thalamic excitatory and inhibitory neurons, we first computed the average GABAergic and glutamatergic module score values for each neuronal cell type and for each individual. We then used the resulting data to determine the first principal component (PC1) scores (the coordinates of the individual observations on the first principal component axis) using the princomp function in R. The ridgeline plot showing the distribution of PC1 score for each neuronal cell type was generated using the ggplot2 package in R. To determine the association between the average glutamatergic and the GABAergic module score across neuronal cell types, we performed a simple linear regression analysis using Prism 9 software.

Analysis of extratelencephalic projection neuron module scores. Marker genes significantly upregulated in extratelencephalic projection neurons (exc. L5 ET) compared with near-projecting excitatory neurons in layers 5 and 6 (exc. L5/6 NP) were determined using the FindAllMarkers function from Seurat (Wilcoxon rank-sum test with Bonferroni correction for multiple testing; $P_{adj} < 0.05$). We tested only genes that were detected in a minimum of 25% of the cells within the cell type (min.pct = 0.25) and that showed, on average, at least a 0.25-fold difference (log-scale) between exc. L5 ET cells and exc. L5/6 NP cells (logfc.threshold = 0.25). The exc. L5 ET module score was computed based on the identified marker genes using the AddModuleScore function of Seurat. To determine the exc. L5 ET module scores across excitatory and inhibitory neurons, cells were downsampled to 2,000 cells per high-resolution cell type.

Cell–cell communication analysis

Cell–cell communication events were predicted using the LIgand-receptor ANalysis frAmework (LIANA)[107] in R. Specifically, the ligand–receptor analysis was performed using liana_wrap(). The methods included were CellPhoneDB[108], NATMI[109] and SingleCellSignalR[110]. liana_aggregate() with the argument 'aggregate_how' set to 'magnitude' was run to find consensus ranks of different methods. Only interactions (ligand–receptor pairs) with a robust rank aggregation (RRA) score smaller than 0.05 (aggregate_rank < 0.05) were considered in downstream analyses. The interaction score of ligand–receptor pairs was calculated by applying $-\log_{10}$ transformation to the RRA score (aggregate_rank). To determine the number of interactions and the overlap of interactions between regions, liana_wrap() was run on the pool of cells isolated from all individuals, with separate analyses conducted for each brain region. To determine cell–cell communication events that are brain-region specific, liana_wrap() was run separately for each individual. We then used a linear mixed-effects model to evaluate the association between the interaction scores of individual ligand–receptor pairs obtained from LIANA and the respective brain region serving as the predictor variable. To account for potential confounding factors and individual variability, we included age, sex and post-mortem interval as covariates in the linear mixed-effects model. These variables were added as fixed effects to the model. Moreover, we included a random effect for the individual to capture the participant-specific variability in the data. Linear mixed-effects models were implemented using the R software packages lme4[111] and lmerTest[112]. The lmer() function from the lme4 package was employed to fit the models. To obtain *P* values for the fixed effects in these models, we used the lmerTest package, which incorporated Satterthwaite's degrees of freedom approximation. To account for multiple hypothesis testing, the obtained *P* values were further adjusted using the Bonferroni method.

Cell type composition

Analysis of cell type composition differences between brain regions. For comparing the relative abundance of major cell types across brain

regions, the fraction of cells of a major cell type class was computed relative to all the cells isolated from a region. For the statistical analysis of cell type composition differences between brain regions, we also computed the relative abundance of major cell type classes separately for each study participant. To this end, the fraction of cells of a major cell type class was computed relative to all the cells isolated from a brain region of an individual. At the level of high-resolution cell types or cell states, two distinct measures of relative abundance were computed. First, the relative abundance of each subtype of a major cell class was computed as the proportion of a subtype relative to all cells of the corresponding major cell class isolated from a brain region. Second, for the statistical analysis of differences between brain regions, the fraction of cells of a subtype was computed relative to all the cells isolated from a brain region of an individual. The statistical analyses comparing the relative abundance of major cell types and subtypes between brain regions was performed using Prism 9 software.

Analysis of cell type composition alterations in AD. We calculated compositional differences in individuals with AD versus individuals without AD (or AD dementia versus no dementia) by modelling the number of cells of a certain cell type or subtype in a specific sample (individual by region) relative to the total number of cells using a quasi-binomial regression model. We modelled AD status by binary ascertainment variables (cogdx 4–5, NIA-Reagan score 1–2, Braak stage 5–6 versus others, as well as any detected presence of NFTs, neuritic plaque or diffuse plaques in the region) while adjusting for brain region and sex. We used the emmeans package in R to assess the significance of the regression contrasts and used p.adjust with the fdr method to adjust *P* values. We modelled the effects of fraction of cells on cognitive performance in multiple domains with gaussian linear regression of cognitive performance on last visit versus the $\log_{10}$-transformed fraction of cells in the subtype or major cell type jointly with covariates for age, sex, APOE-ε4 and post-mortem interval, with false-discovery rate *P*-value correction (p.adjust in R). We compared the fractional abundances of pairs of neuronal subtypes between two regions using Kendall's τ only in individuals with AD (NIA-Reagan score). Significance was assessed using beta regression (R library betareg) controlling for sex, *APOE* genotype, post-mortem interval and age of death, and we adjusted *P* values using p.adjust in R with the fdr method.

Differential gene expression

DEGs in AD. We performed differential expression analyses with three separate methods: MAST, Nebula and Wilcoxon testing[113,114]. For all methods, we subset the tested genes to only genes present in over 20% of cells. For MAST and Nebula, we calculated and included in the regression the top 10 components of unwanted variation calculated using RUV on the pseudo-bulk-level data (individuals by regions). For these methods, we also included as covariates the individual's sex, age of death and post-mortem interval, each cell's counts per gene and number of captured genes and, where applicable, the high-resolution cell subtypes and the brain region. For Nebula we used a Poisson mixed-model on the counts data with an offset of the $\log_{10}$-transformed total counts per cell. For MAST and Wilcoxon, we normalized each cell to a total library size of 10,000 counts. We ran Wilcoxon tests on both the cell and individual levels. We adjusted *P* values for multiple testing in all cases by using the p.adjust function in R with the fdr method. For our final set of differential genes in each analysis, we took all genes that were significant ($P_{adj} > 0.05$) and concordant in both the MAST and Nebula results. We also provide the results for Wilcoxon tests, but did not use these to determine concordant results as they do not correct for many covariates. We computed differential expression against five AD ascertainment variables: continuous measurements of NFT, plaq_n, and plaq_d measured in each region except the thalamus (excluded from these analyses) and binary cognitive impairment (cogdx no dementia = 1 and 2 versus AD dementia = 4 and 5) and NIA-Reagan score classifications (non-AD = 3 and 4 versus AD = 1 and 2). We provide differential expression results for each of the 14 major cell types (with T cells, CNS macrophages, and each vascular subtype separately) for all regions jointly and for each region separately. We also provide results for each of the excitatory subtypes either in its most prevalent region for EC, HC or TH subclasses, or across the neocortex for neocortical subtypes (Supplementary Table 9). We also computed DEGs for the interaction between pathologic diagnosis of AD and sex in each major cell type, both across all regions and in each region separately. For the glial energy metabolism analyses we recomputed all DEGs in glial glycolysis-associated modules separately (keeping all genes, with no cell percentage cut-off). Glycolysis pathway diagram is from the glycolysis and gluconeogenesis pathway from WikiPathways[115].

Pathology-biased DEGs. Pathology-biased DEGs were based on neuritic plaque or NFT pathology measurements in each region and were computed in each major cell type across all regions and in each region (except for the TH). Genes were ordered by the residual between NFT effect size and predicted NFT effect size from a regression using plaque effect size and region. Genes were retained if they were consistently up (or down) in 3+ regions for either NFT or plaque but in fewer than 2 regions for the other pathology measurement (shared genes are genes found in 2+ cell types).

Comparison with published DEGs. We compared our differential expression results to results from seven different previously published studies[11,12,19,47–50]. We compared the published DEGs both to: (1) cross-region DEGs calculated in each major cell type for individual-level AD status (NIA-Reagan score or clinical diagnosis of AD) and for quantitative measurements of AD pathology (neuritic plaques, diffuse plaques and NFTs); and (2) region-specific DEGs calculated in each major cell type and in endothelial cells, computed relative to pathologic diagnosis of AD (NIA-Reagan score, AD, 1–2; non-AD, 3–4). As some studies reported only the significant genes, we compared the log-transformed fold change estimates for our DEGs and reported DEGs by a Pearson correlation test.

DEG module enrichments. To assess the enrichment of upregulated, downregulated non-differentially expressed genes in each module, we first assigned each tested DEGs to its closest module by correlation to the module's average expression profile. We then performed a hypergeometric enrichment test for the number of genes in a category (upregulated, downregulated, not differentially expressed) assigned to the module, against the total number of tested genes assigned to the module, the total number in the category and the total number tested and corrected *P* values using p.adjust (Benjamini–Hochberg). Enrichments of pathology-biased DEGs in modules were performed in the same manner.

Neuronal DEG partitions. To partition neuronal DEGs into non-vulnerable and vulnerable-associated subclasses, we calculated each genes' average expressions and differential expression effect sizes at the subtype level and compared these to the relative depletion of the subtypes. For each gene that was differentially expressed in late-AD (stratified by Braak stage, late AD, 5–6 versus non-AD or early-AD, 1–4) in at least 25% of all neuronal subtypes, we calculated the correlation of its average subtype expression in late-AD with each subtype's compositional stability ($\log_2$[OR] in late AD) across excitatory subtypes, separating non-vulnerable-associated genes (correlation > 0.2) from vulnerable-associated genes (correlation < −0.2). We calculated functional enrichments on neuronal DEG partitions using the top 250 genes ordered by effect size in each category. We further separated DEGs with higher effect sizes in vulnerable subtypes from those with similar effect sizes across all neuronal subtypes by calculating the correlation of their differential effect sizes in each subtype with that subtype's depletion

($\log_2$[OR] in late AD). To perform enrichments along the continuum of genes associated with vulnerability to non-vulnerability, we kept only genes with biased effect sizes (effect size correlation < −0.2) and binned them along the axis of expression correlation (window size 0.2 for at 0.05 intervals) and performed functional enrichments for all bins jointly.

DEG and module pathway enrichments. We performed DEG enrichments for each differential expression run using the gprofiler2 package in R, with multi-query for upregulated and downregulated genes, as unordered queries, a *P*-value cutoff of 0.05, and using GO, REAC, WP, KEGG and CORUM as annotation sources, and retained enriched terms with fewer than 500 genes. Module and module cluster enrichments were performed in the same manner, using the core genes identified for each module and for genes found in more than two modules within a module cluster.

Markers of neuronal vulnerability. We identified markers associated with excitatory neuron subtype vulnerability by performing linear regression to predict the $\log_2$[OR] of each subtype's depletion in late AD based on its expression at the subtype-aggregate level ($\log[X+1]$, averaged normalized expression in each subtype by region by individual batch), controlling for age, sex and post-mortem interval and adjusting *P* values with p.adjust (fdr).

Identification of genes associated with cellular vulnerability in inhibitory neurons. Processed snRNA-seq data (DLPFC, experiment 2) were obtained from Synapse (syn51123521) and integrated with our own PFC snRNA-seq dataset comprising 427 individuals. To identify vulnerable inhibitory neuron subtypes, we examined the association between the relative abundance of cell types and the measure of NFT density (variable tangles). We used a quasi-binomial regression model to model the number of cells belonging to a specific cell type in a given sample (individual study participant), relative to the total number of cells in that sample. We fitted the regression model using the glm function in R, including age, sex and post-mortem interval as covariates. *P* values were corrected for multiple testing using the Benjamini–Hochberg procedure as implemented in the R function p.adjust. The results are presented in the form of association scores (signed $-\log_{10}$-transformed Benjamini–Hochberg-adjusted *P* value, where the sign was determined by the direction (positive or negative) of the association). Inhibitory neuron subtypes demonstrating a significant negative association with tangle density (Benjamini–Hochberg-adjusted *P* value < 0.05) were classified as vulnerable subtypes, while all other subtypes were categorized as non-vulnerable. Genes exhibiting differential expression between vulnerable and non-vulnerable inhibitory neuron subtypes were identified on the basis of our PFC snRNA-seq dataset. This analysis was restricted to individuals without a pathologic diagnosis of AD. The differential expression analysis comparing vulnerable to non-vulnerable inhibitory neuron subtypes was performed using the R package dreamlet (https://diseaseneurogenomics.github.io/dreamlet/). We used the dreamletCompareClusters function with the argument 'method' set to 'fixed' for this analysis. Adjusted *P* values for multiple testing were obtained using the topTable function of dreamlet, with the 'adjust.method' set to 'BH'.

GWAS analyses. Intersection of regional expression and pathology-specific DEGs (across all regions) was performed for 149 annotated AD GWAS familial and AD risk loci from recent GWAS[54,56–58]. We calculated the disease-relevance score of each cell in the dataset against a recent Alzheimer's GWAS, using scDRS (based on MAGMA)[54,55,116]. For the scDRS results, we counted the fraction of cells with significant scDRS scores (FDR < 0.05) in each cell type, subtype and region. To test for overlap with microglia/immune modules, we compared the set of immune cells with significant expression of each module (*z* score > 2.5) and with the set of cells with significant scDRS scores (FDR < 0.05) and performed a hypergeometric test for significance of the overlap ($P_{adj} < 0.01$, Benjamini–Hochberg correction). To identify region-specific GWAS genes, we performed an analysis of variance for the effect of region on average gene expression at the pseudobulk level.

Identification of genes associated with CR. To quantify CR, we computed a CR score as the difference between the observed cognition and the cognition predicted by a linear regression model, given the level of pathology (Fig. 5a). Using this approach, we computed cognitive resilience (CR) scores based on the measure of global cognitive function and CDR scores based on the measure estimating the rate of change of global cognitive function over time (Fig. 5a). Four distinct CR and CDR scores were derived using four distinct measures of AD pathology, namely global AD pathology and, separately, neuritic plaque burden, NFT burden and tangle density.

We performed differential expression analyses using the R package muscat to identify genes associated with CR in the PFC[117]. Low-expressed genes were excluded and only genes with more than one count in at least ten cells were considered. To take advantage of robust bulk RNA-seq differential expression frameworks, such as edgeR[118], in a first step, muscat aggregates measurements for each sample (in each cluster) to obtain pseudobulk data. Using this approach, single-cell measurements were aggregated per study participant and cell type using the sum of raw counts option. Differential expression analysis was run using the edgeR method as implemented in muscat. We included as covariates the individual's age at death and post-mortem interval. We report adjusted *P* values for multiple testing in all cases by using the p.adjust function with the Benjamini–Hochberg method as implemented in muscat. The multiple testing correction was performed locally, that is, on each of the cell types separately with the number of tests equal to the number of genes considered. These differential expression analyses were performed on the entire set of 427 individuals except for the group-based differential expression analysis based on our categorical definition of CR. In this case, we focused on comparing two distinct groups determined by their pathologic and clinical diagnoses of AD. First, we identified individuals with a pathologic diagnosis of AD, using the NIA-Reagan pathology criteria. Subsequently, these individuals were further categorized on the basis of their clinical consensus diagnosis of cognitive status at the time of death. Specifically, we compared individuals with no cognitive impairment (NCI, final consensus cognitive diagnosis (cogdx) value of 1) against individuals with a cognitive diagnosis of AD dementia and no other cause of cognitive impairment (cogdx value of 4) among individuals with a pathologic diagnosis of AD.

To confirm the differential gene expression results based on the CR and CDR scores, we also evaluated the association between gene expression and global cognitive function or the rate of change of global cognitive function adjusting for AD pathology as a covariate. The AD pathology measures considered as a covariate were global AD pathology (gpath), neuritic plaque burden (plaq_n), NFT burden (nft), or tangle density (tangles). Thus, together with the DGE analysis based on CR and CDR scores, we performed a total of 16 different tests assessing the association between gene expression and CR.

We used the model-based analysis of single-cell transcriptomics (MAST) tool to investigate whether the CR genes identified in PFC astrocytes were also associated with CR in astrocytes from other regions of the human brain. To ensure robust analysis, we initially filtered the genes under investigation, selecting only those with more than one count in at least 10 cells. The analytical model incorporated the condition variable of interest, as well as several covariates known to influence gene expression. These covariates included the cellular detection rate (cngeneson), age at death (age_death), post-mortem interval (pmi), and sex (msex). We also accounted for potential participant-specific variation in the data by incorporating a random effect term for the individual (1|individual). To account for multiple comparisons, the

P values were adjusted using the FDR method as implemented in the p.adjust function.

Bulk RNA-seq differential expression analysis. Differential expression analysis of bulk RNA-seq data from the ROSMAP cohorts was performed using DESeq2[119] (plotted) and edgeR[118]. Age at death and post-mortem interval were converted into *z* scores and included as covariates in the regression equation. Both approaches (DESeq2 and edgeR) provided similar results.

Permutation test for evaluating the significance of the overlap of DEGs between our dataset and the SEA-AD dataset. The average expression level of each gene within each major cell type was determined using the 'AverageExpression' function from the Seurat R package. The genes considered in the differential expression analysis for each major cell type were categorized into ten subsets based on their average expression level within the corresponding cell type. We next intersected the genes in each of the ten subsets with genes identified as significantly associated with either neuritic plaque burden (in our dataset) or the CPS score (in the MTG SEA-AD dataset). This intersection enabled us to determine the number of significant DEGs in each subset. The process was performed separately for genes positively and negatively associated with AD pathology. Subsequently, we randomly sampled the determined number of significant DEGs from each of the 10 subsets, ensuring that the expression level distribution of the DEGs was preserved in the random samples. This random sampling step was repeated for a total of 1,000 iterations. These steps were performed separately for both our dataset and the SEA-AD dataset. For each of the 1,000 random samples, we determined the overlap of genes between datasets and compared it to the observed overlap between the two datasets. To assess the significance of the observed overlap, we computed *z* scores, which represent the difference between the observed value of overlap and the mean value of overlap based on the permutation results, divided by the s.d. of the permutation results.

Comparison with previously published proteomics study of AD. To further validate our differential expression results, we evaluated the correlation between the effect sizes of gene expression changes observed in our study and those identified through quantitative proteomics[51]. We specifically examined the correlation between the effect sizes of genes associated with neuritic plaque burden in our study and the effect sizes of overlapping differentially expressed proteins in the quantitative proteomics analysis of bulk tissue. The correlation was computed using the cor.test function in R with the argument 'alternative' set to 'two.sided' and the argument 'method' set to 'pearson'. *P* values were adjusted for multiple testing using the Benjamini–Hochberg method as implemented in the R function p.adjust.

Inter-regional comparison of AD pathology-associated gene sets in glial cells along the spectrum of global AD pathology burden. We determined genes significantly associated with global AD pathology for each glial cell type, using single-nucleus RNA sequencing data from the PFC of 427 participants in the ROSMAP study. We then calculated module scores for these gene sets in astrocytes, microglia, oligodendrocytes, and OPCs using the Seurat 'AddModuleScore' function. The module scores were determined separately for genes positively and negatively associated with global AD pathology. To assess the progression of these scores across the spectrum of global AD pathology burden, we averaged the module scores of all cells of a specified cell type isolated from the brain region of interest of an individual. For visualizing the relationship between the global AD pathology burden and mean module scores, we employed Locally Estimated Scatterplot Smoothing (LOESS) using the ggplot2 package in R, with the 'geom_smooth' function and the 'method' parameter set to 'loess'. The correlation of mean module scores between regions was determined using the cor.mtest function of the R package corrplot. *P* values were adjusted for multiple hypotheses testing using the Benjamini–Hochberg method as implemented in the R function p.adjust.

In situ hybridization (RNAscope)

Frozen human post-mortem brain samples were embedded in Tissue-Tek OCT compound (VWR; 25608-930), sectioned on a Leica cryostat at a thickness of 20 µm, and mounted onto Fisherbrand Superfrost Plus microscope slides (Thermo Fisher Scientific; 12-550-15). Slides were fixed in 4% paraformaldehyde at 4 °C for 30 min, and dehydrated in ethanol. The RNAscope 2.5 HD Chromogenic Duplex Detection Kit and RNAscope Multiplex fluorescent V2 Kit (ACDBio) were then used according to the manufacturer's instructions. Tissue samples were hybridized using the following chromogenic RNAscope probes: GAD2, FOXP2, MEIS2, AQP4, GRM3, ADCY8, PFKP, PNPLA6, GPCPD1 and CHDH (ACDBio). For in situ hybridization of Reelin, tissue samples were hybridized using the following fluorescent RNAscope probes: vGlut and Reelin. Cell nuclei were stained with 50% haematoxylin (for chromogenic experiments) or with Hoechst (for fluorescent experiments). For fluorescence RNAscope analysis, sections were incubated in TrueBlack (Biotium; 23007) for 10 s before Hoechst staining to quench auto-fluorescence. Images were acquired using the Zeiss LSM 900 confocal microscope, with a 63× oil objective. Two images were acquired per tissue sample. For both chromogenic and fluorescence RNAscope experiments, puncta were manually counted by researchers blinded to the experimental group of each image.

Immunohistochemistry

All experiments were performed according to the Guide for the Care and Use of Laboratory Animals and were approved by the National Institute of Health and the Committee on Animal Care at Massachusetts Institute of Technology. Sample sizes were determined on the basis of previous work from our laboratory, without power analysis calculation or randomization. In the experiment comparing *App*-KI (C57BL/6-App<tm3(NL-G-F)Tcs>, RBRC06344) and WT mice, the *App*-KI group consisted of 7 mice (5 male and 2 female) and the WT group included 6 female mice. In the experiment comparing Tau P301S (The Jackson Laboratory, 008169) to WT mice, both groups consisted entirely of male mice, with 6 mice in the Tau(P301S) group and 5 mice in the WT group. Mice were transcardially perfused with ice-cold phosphate-buffered saline, followed by 4% paraformaldehyde for fixation. Brains were dissected out and post-fixed in 4% paraformaldehyde overnight at 4 °C. Brains were sectioned horizontally on the Leica vibratome at a thickness of 40 µm. Slices containing the EC were selected under a dissecting microscope to ensure consistent anatomical structure across all cohorts. Brain sections were incubated in antigen retrieval solution (pH 6, 100 mM sodium citrate buffer, prewarmed to 80 °C) for 20 min, and then cooled to room temperature. The sections were then washed twice with phosphate-buffered saline, and blocked in buffer (0.3% Triton X-100, 2% bovine serum albumin, 10% normal donkey serum in phosphate-buffered saline) for 1 h at room temperature.

The sections were incubated in primary antibodies (anti-Reelin, 1:200; anti-NeuN, 1:200, anti-phospho tau, 1:200; and anti-amyloid-β, 1:500) overnight at 4 °C. After primary antibody incubation, the sections were washed three times with PBS, twice with blocking buffer and incubated in secondary antibody (1:1,000) for 2 h at room temperature. The sections were then washed three times with PBS, incubated with Hoechst (1:1,000) for 10 min and washed once more with PBS.

Confocal tile scans of the EC were acquired on the Zeiss LSM 900 using a 20× objective, with consistent laser setting across all cohorts. Layer II–III EC was identified based on previous criteria[120]. Orthogonal projections of the confocal tile scans were exported to Fiji for signal quantification. In Fiji, layer II–III of the EC was set as a region of interest, and a macro was used to count Reelin-positive cells in the region

of interest and quantify the mean fluorescence intensity for each cell. The signal intensity of the Reelin channel was subsequently normalized to the NeuN signal. Researchers were blinded to animal genotype.

External data sources

Processed snRNA-seq data generated by the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) consortium (SEAAD_MTG_RNAseq_final-nuclei.2023-05-05.h5ad) were obtained from the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) (https://sea-ad-single-cell-profiling.s3.amazonaws.com/index.html#MTG/RNAseq/). The SEA-AD DLPFC data (SEAAD_DLPFC_RNAseq_final-nuclei.2023-07-19.h5ad) were downloaded from https://sea-ad-single-cell-profiling.s3.amazonaws.com/index.html#DLPFC/RNAseq/. Additional processed snRNA-seq datasets (specifically the h5ad files Neurons.h5ad and Nonneurons.h5ad) were obtained from the Linnarsson laboratory (https://console.cloud.google.com/storage/browser/linnarsson-lab-human;tab=objects?authuser=0&prefix=&forceOnObjectsSortingFiltering=false).

Reporting summary

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

Data availability

snRNA-seq profiling data are available from Synapse in coordination with the ROSMAP project. Data are accessible under accession codes syn52293442 (as part of the MIT ROSMAP Single-Nucleus Multiomics Study; Synapse: syn52293417). 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 (https://www.radc.rush.edu/; https://adknowledgeportal.synapse.org/). Additional processed data as well as integrative visualization and exploration of the atlas are available online (http://compbio.mit.edu/ad_multiregion/ and https://ad-multi-region.cells.ucsc.edu/)[121]. We also downloaded the following public single-cell gene expression datasets: human multiple cortical areas SMART-seq (https://portal.brain-map.org/atlases-and-data/rnaseq/human-multiple-cortical-areas-smart-seq), human DLPFC (Synapse: syn51123521), SEA-AD MTG (https://sea-ad-single-cell-profiling.s3.amazonaws.com/index.html#MTG/RNAseq/), SEA-AD DLPFC (https://sea-ad-single-cell-profiling.s3.amazonaws.com/index.html#DLPFC/RNAseq/), human dLGN (https://portal.brain-map.org/atlases-and-data/rnaseq/comparative-lgn), multiple human brain regions (https://console.cloud.google.com/storage/browser/linnarsson-lab-human;tab=objects?authuser=0&prefix=&forceOnObjectsSortingFiltering=false), multiple cortical areas and the hippocampal formation of the mouse brain (https://portal.brain-map.org/atlases-and-data/rnaseq/mouse-whole-cortex-and-hippocampus-10x), nine regions in the adult mouse brain (http://dropviz.org/) and Mouse Brain Atlas (http://mousebrain.org/). Source data are provided with this paper.

Code availability

Code for analysis is available at GitHub (https://github.com/cboix/admultiregion_analysis) and Zenodo (https://doi.org/10.5281/zenodo.11051020). The code for the scdemon method for module detection from single-cell RNA-seq is available at GitHub (https://github.com/KellisLab/scdemon).

83. The National Institute on Aging, and Reagan Institute Working Group on Diagnostic Criteria for the Neuropathological Assessment of Alzheimer's Disease. Consensus recommendations for the postmortem diagnosis of Alzheimer's disease. *Neurobiol. Aging* **18**, S1–S2 (1997).
84. Bennett, D. A., Schneider, J. A., Tang, Y., Arnold, S. E. & Wilson, R. S. The effect of social networks on the relation between Alzheimer's disease pathology and level of cognitive function in old people: a longitudinal cohort study. *Lancet Neurol.* **5**, 406–412 (2006).
85. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
86. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
87. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep.* **9**, 5233 (2019).
88. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337 (2019).
89. Hafemeister, C. & Satija, R. Analyzing scRNA-seq data with the sctransform and offset models. Preprint at https://satijalab.org/pdf/sctransform_offset.pdf (2020).
90. Stuart et al. Seurat R. package v.2.3.4 and v3.1.3 (2019).
91. Kierdorf, K., Masuda, T., Jordão, M. J. C. & Prinz, M. Macrophages at CNS interfaces: ontogeny and function in health and disease. *Nat. Rev. Neurosci.* **20**, 547–562 (2019).
92. Velmeshev, D. et al. Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685–689 (2019).
93. Vanlandewijck, M. et al. A molecular atlas of cell types and zonation in the brain vasculature. *Nature* **554**, 475–480 (2018).
94. Davila-Velderrain, J. et al. Single-cell anatomical analysis of human hippocampus and entorhinal cortex uncovers early-stage molecular pathology in Alzheimer's disease. Preprint at *bioRxiv* https://doi.org/10.1101/2021.07.01.450715 (2021).
95. Mohammadi, S., Davila-Velderrain, J. & Kellis, M. A multiresolution framework to characterize single-cell state landscapes. *Nat. Commun.* **11**, 5399 (2020).
96. Sun, N. et al. Human microglial state dynamics in Alzheimer's disease progression. *Cell* **186**, 4386–4403 (2023).
97. Tuddenham, J. F. et al. A cross-disease human microglial framework identifies disease-enriched subsets and tool compounds for microglial polarization. Preprint at *bioRxiv* https://doi.org/10.1101/2022.06.04.494709 (2022).
98. Bakken, T. E. et al. Single-cell and single-nucleus RNA-seq uncovers shared and distinct axes of variation in dorsal LGN neurons in mice, non-human primates, and humans. *eLife* **10**, e64875 (2021).
99. Wainberg, M. et al. A genome-wide atlas of co-essential modules assigns function to uncharacterized genes. *Nat. Genet.* **53**, 638–649 (2021).
100. Bell, A. J. & Sejnowski, T. J. The 'independent components' of natural scenes are edge filters. *Vision Res.* **37**, 3327–3338 (1997).
101. Kessy, A., Lewin, A. & Strimmer, K. Optimal whitening and decorrelation. *Am. Stat.* **72**, 309–314 (2018).
102. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal* **1695**, 1–9 (2006).
103. Ren, X. et al. COVID–19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell* **184**, 5838 (2021).
104. Tabula Sapiens Consortium. The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022).
105. Student. Probable error of a correlation coefficient. *Biometrika* **6**, 302–310 (1908).
106. Suo, S. et al. Revealing the critical regulators of cell identity in the mouse cell atlas. *Cell Rep.* **25**, 1436–1445 (2018).
107. Dimitrov, D. et al. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-seq data. *Nat. Commun.* **13**, 3224 (2022).
108. Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* **563**, 347–353 (2018).
109. Hou, R., Denisenko, E., Ong, H. T., Ramilowski, J. A. & Forrest, A. R. R. Predicting cell-to-cell communication networks using NATMI. *Nat. Commun.* **11**, 5011 (2020).
110. Cabello-Aguilar, S. et al. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* **48**, e55 (2020).
111. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. Preprint at arxiv.org/abs/1406.5823 (2014).
112. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: tests in linear mixed effects models (lmer objects of lme4 package). *J. Stat. Softw.* **82**, 1–26 (2017).
113. He, L. et al. NEBULA is a fast negative binomial mixed model for differential or co-expression analysis of large-scale multi-subject single-cell data. *Commun. Biol.* **4**, 629 (2021).
114. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
115. Slenter, D. N. et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.* **46**, D661–D667 (2018).
116. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
117. Crowell, H. L. et al. muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nat. Commun.* **11**, 6077 (2020).
118. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
119. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
120. Witter, M. P., Doan, T. P., Jacobsen, B., Nilssen, E. S. & Ohara, S. Architecture of the entorhinal cortex: a review of entorhinal anatomy in rodents with some comparative notes. *Front. Syst. Neurosci.* **11**, 46 (2017).
121. Speir, M. L. et al. UCSC Cell Browser: visualize your single-cell data. *Bioinformatics* **37**, 4578–4580 (2021).

Acknowledgements We thank the study participants and staff of the Rush Alzheimer's Disease Center; all of the members of the Kellis and Tsai laboratories at MIT for discussions and feedback, including A. Grayson and P. Purcell for editing and feedback on the manuscript; the following members of the Mathys laboratory at the University of Pittsburgh for their assistance

with the computational analysis: K. Ghafari, A. K. Kunisky, H. Zhao, S. Wang, J. I. Beaudway, E. Clark, Y. Ghaffari, W. R. Incorvaia, A. C. Laudenslager, V. N. Lohia and P. R. Patel; and M. Haeussler and B. Wick for their help in hosting the data on the UCSC browser. This work was supported in part by NIH grants RF1 AG062377, RF1 AG054321 and RO1 AG054012 (L.-H.T.), AG058002, AG062377, NS110453, NS115064, AG062335, AG074003, NS127187, MH119509 and HG008155 (M.K.) and the NIH training grant GM087237 (to C.A.B.). This work was partially supported by the Cure Alzheimer's Fund, the JBP Foundation, the Robert A. and Renee E. Belfer Family Foundation, Eduardo Eurnekian and Joseph P. DiSabato. H.M. was supported by an Early Postdoc Mobility fellowship from the Swiss National Science Foundation (P2BSP3_151885). ROSMAP is supported by P30AG10161, P30AG72975, R01AG15819, R01AG17917. U01AG46152 and U01AG61356. ROSMAP resources can be requested at https://www.radc.rush.edu.

Author contributions This study was designed by H.M., C.A.B., L.A.A., D.A.B., M.K. and L.-H.T., and directed and coordinated by M.K. and L.-H.T. The project was initiated by H.M. during his tenure as a postdoctoral associate in the laboratory of L.-H.T.; H.M., A.P.N., X.J., J.M. and K.G. performed sample preparations and single-cell RNA profiling. C.A.B. and H.M. led the computational analysis. L.A.A., Z.X., F.G.-M. and K.L. performed experimental validation experiments. G.A. and S.B. helped with the computational analysis. N.B., B.T.J. and J.D.V. contributed to computational method development and analysis. D.P. and R.E.T. contributed data for GWAS analysis. D.A.B. contributed samples and data. H.M., C.A.B., L.A.A., D.A.B., M.K. and L.-H.T. wrote the manuscript.

Competing interests The authors declare no competing interests.

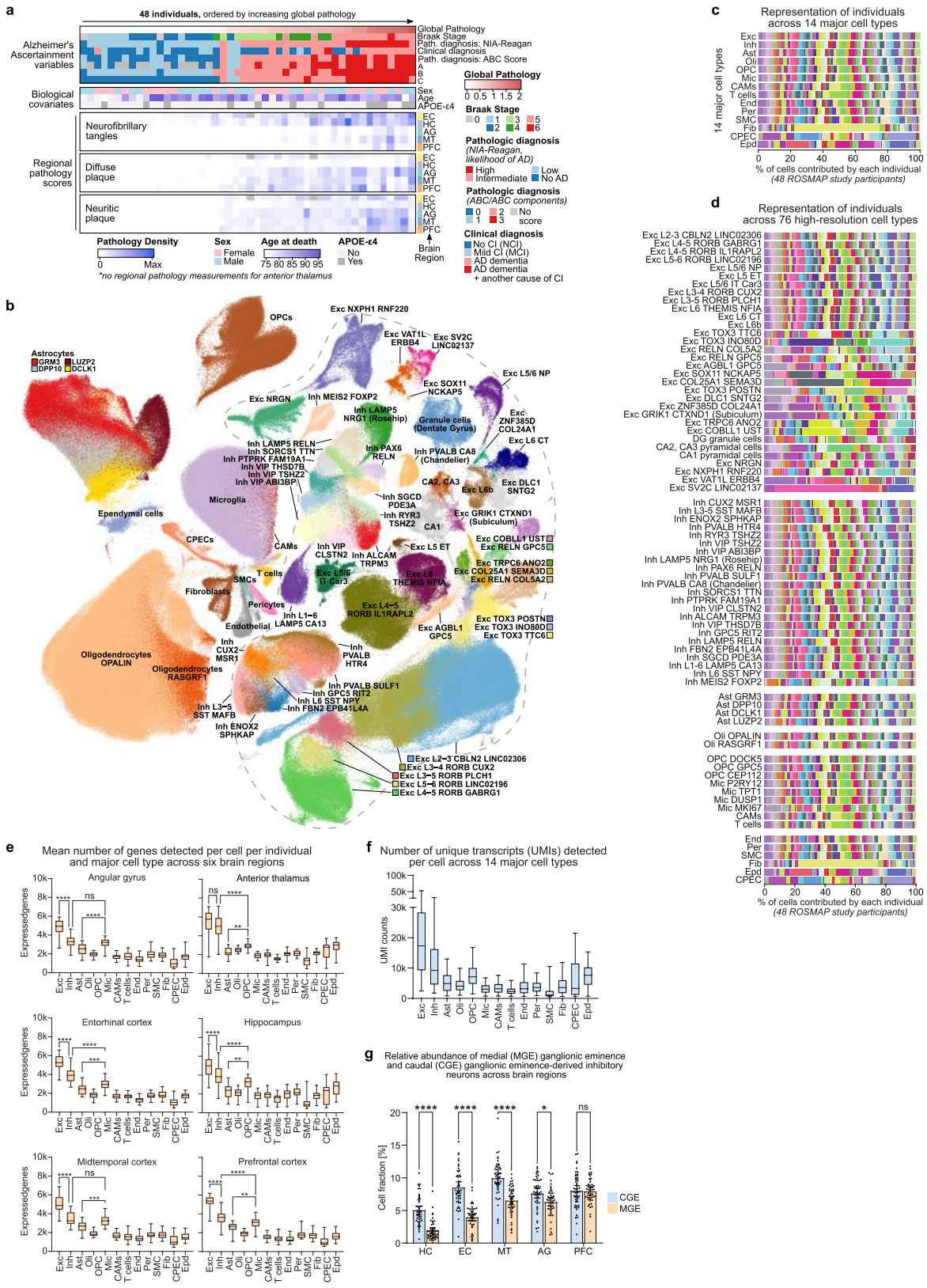
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-07606-7.

Correspondence and requests for materials should be addressed to Li-Huei Tsai or Manolis Kellis.

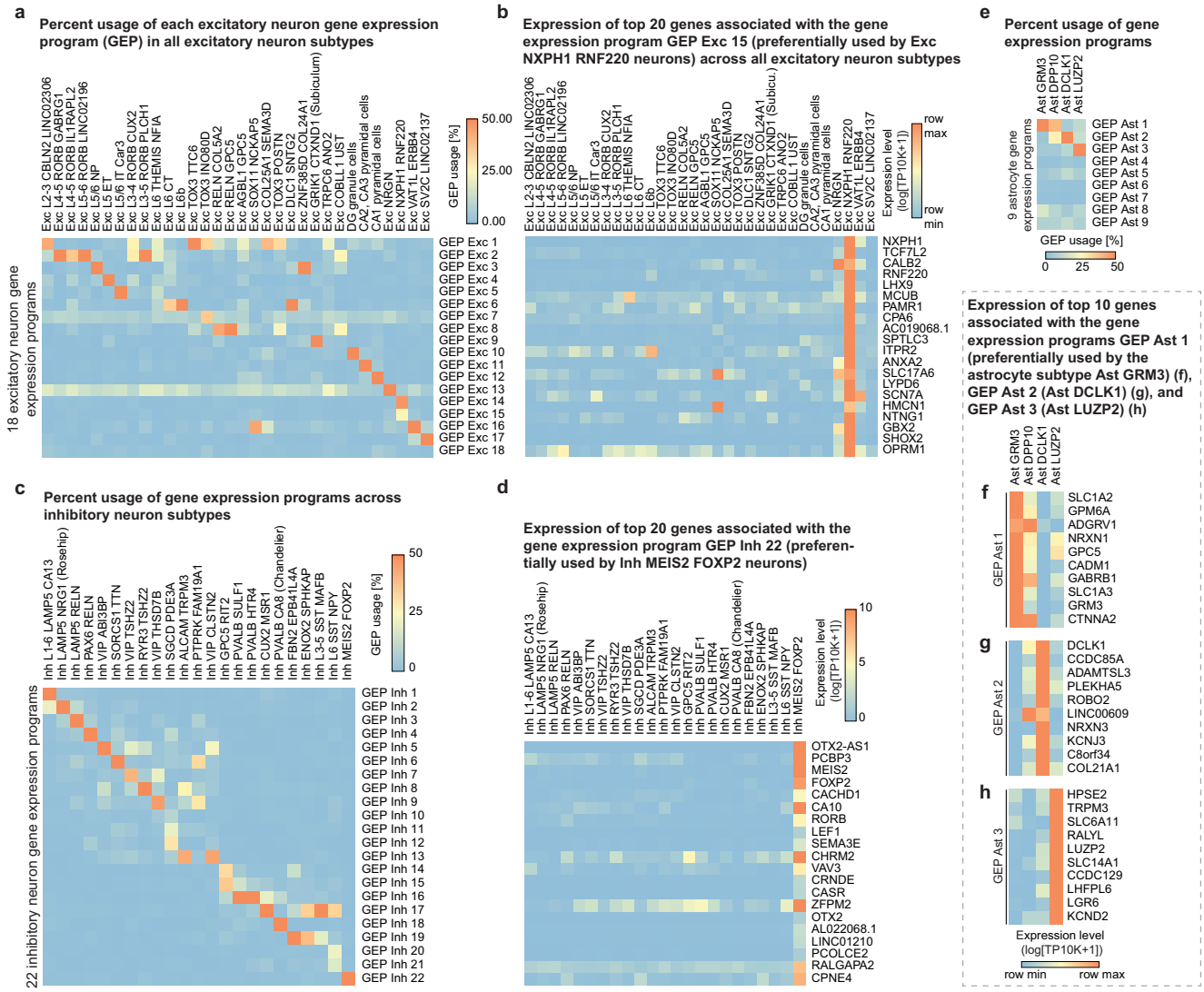
Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work.

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

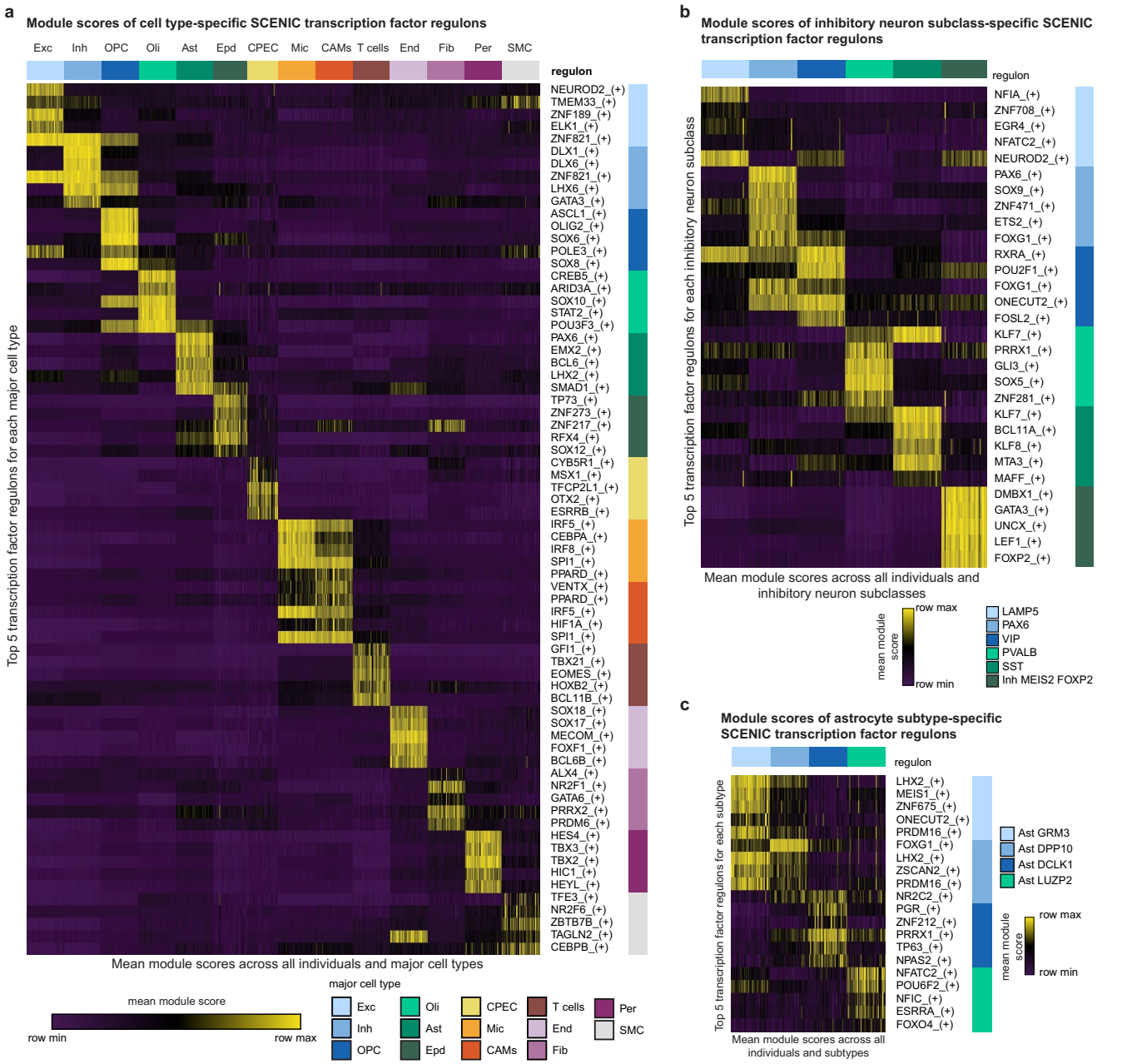


Extended Data Fig. 1 | See next page for caption.

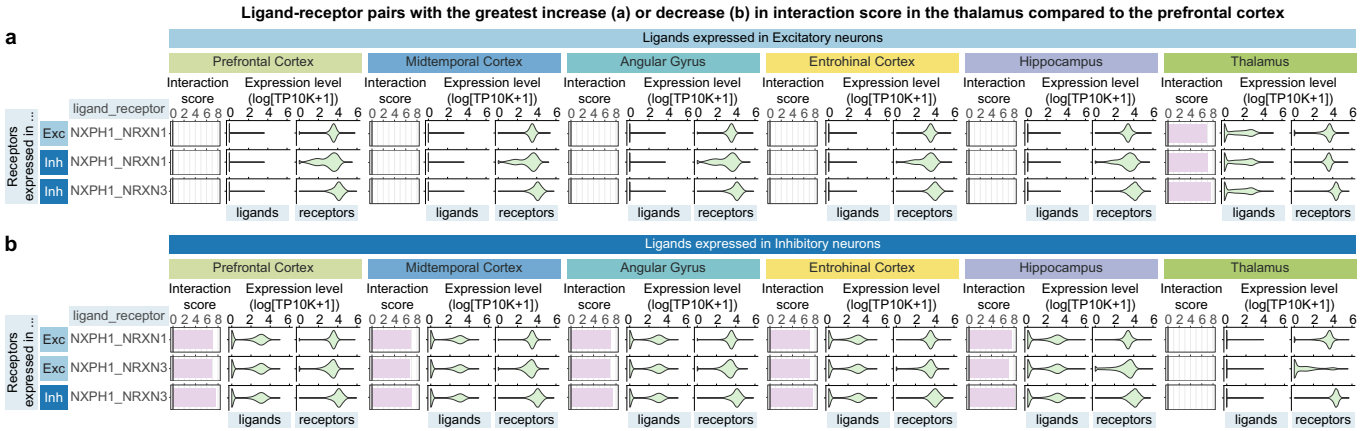
Extended Data Fig. 1 | Overview of the study sample and major cell type annotations. a, Metadata overview: a total of 283 post-mortem brain tissue samples from 24 male and 24 female study participants were analysed across Alzheimer's disease progression (AD). Top two panels show metadata at the individual level and bottom three panels show region-specific pathology measurements of neurofibrillary tangle burden (nft), neuritic plaque burden (plaq_n), and diffuse plaque burden (plaq_d). Individuals (columns) are ordered according to their global AD pathology burden. **b**, Joint UMAP of 1.3 M cells across 14 major cell types coloured and labelled by 76 high-resolution subtypes. **c**,**d**, Representation of individuals across cell types. The stacked bar plots show the proportion of cells contributed by each study participant across 14 major cell types (**c**) and 76 high-resolution cell types (**d**). **e**-**f**, Box plots of the number of genes detected per cell across all major cell types (**e**) and mean number of unique transcripts detected per cell per individual and major cell type across the six brain regions analysed (**f**). Within each box, horizontal lines denote median values; boxes extend from the 25th to the 75th percentile of each group's distribution of values; whiskers extend from the 5th to the 95th percentile. ****P < 0.0001, ***P < 0.001, **P < 0.01; ns, P > 0.05; (ordinary one-way ANOVA corrected for multiple comparisons using Bonferroni's multiple comparisons test). **g**, Relative abundance of inhibitory neurons originating from the medial (MGE) ganglionic eminences (SST and PVALB) and the caudal (CGE) ganglionic eminence (VIP, PAX6, and LAMP5) across brain regions. The bar plots show the mean fraction of cells per individual and brain region (AG, HC, MT, PFC: n = 48; TH: n = 45; EC: n = 46). The fraction of cells was computed relative to all the cells isolated from a brain region of an individual. Data are expressed as mean with 95% confidence intervals and individual data points are shown (two-way ANOVA corrected for multiple comparisons using Bonferroni's multiple comparisons test).



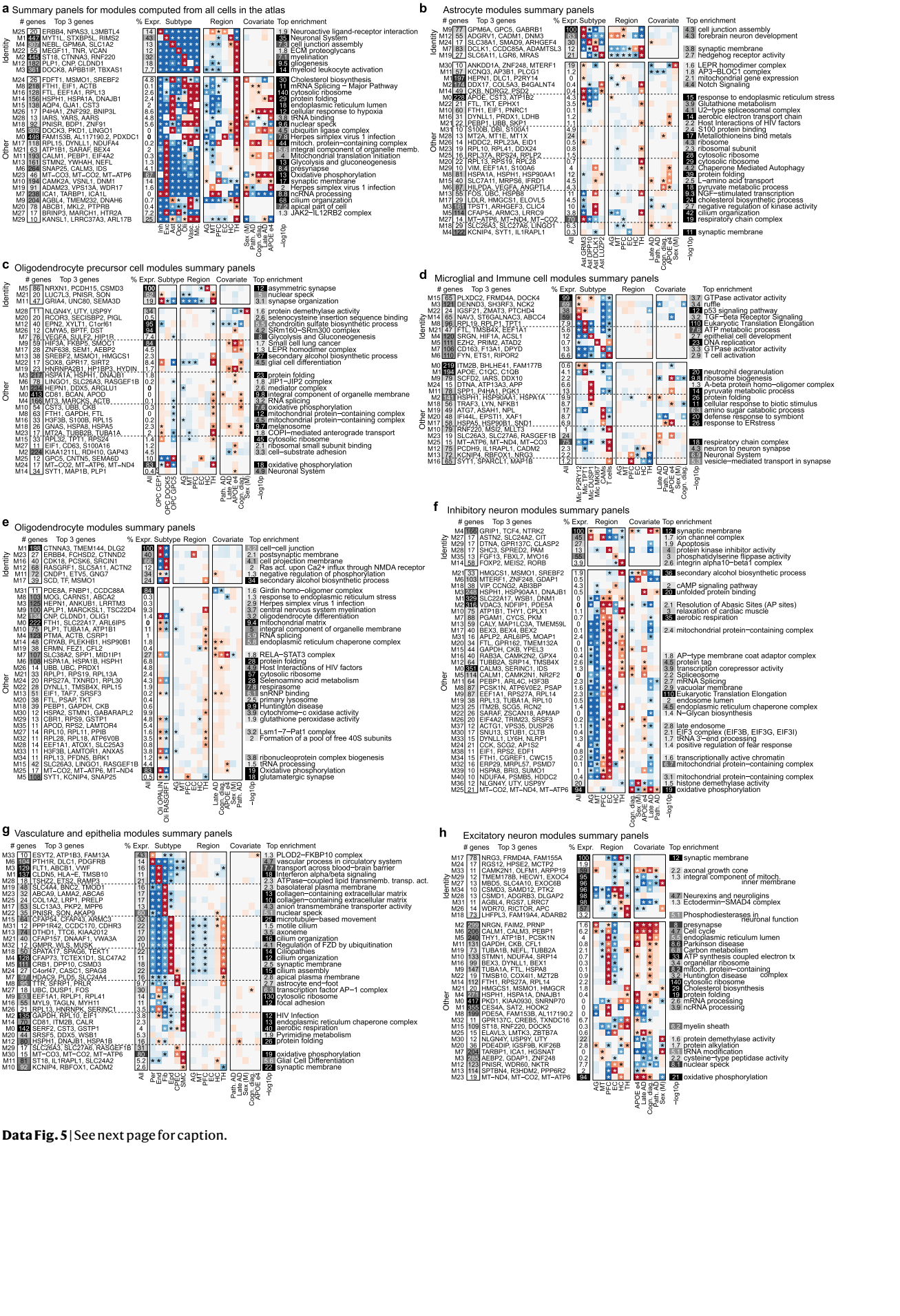
Extended Data Fig. 2 | Gene expression programs. a, Heat map showing percent usage of all excitatory neuron gene expression programs (GEPs) (rows) in all excitatory neuron subtypes (columns). **b**, relative expression level of the top 20 genes associated with the gene expression program GEP Exc 15 (preferentially used by Exc NXPH1 RNF220 neurons) across all excitatory neuron subtypes. **c**, Heat map showing percent usage of all inhibitory neuron gene expression programs (GEPs) (rows) in all inhibitory neuron subtypes (columns). **d**, Expression level of the top 20 genes associated with the gene expression program GEP Inh 22 (preferentially used by Inh MEIS2 FOXP2 neurons) across all inhibitory neuron subtypes. **e**, Heat map showing percent usage of all astrocyte gene expression programs (GEPs) (rows) in all astrocyte subtypes (columns). **f**-**h**, Relative expression level of the top 10 genes associated with the gene expression programs GEP Ast 1 (preferentially used by the astrocyte subtype Ast GRM3) (**f**), GEP Ast 2 (preferentially used by the astrocyte subtype Ast DCLK1) (**g**), and GEP Ast 3 (preferentially used by the astrocyte subtype Ast LUZP2) (**h**) across all astrocyte subtypes.



Extended Data Fig. 3 | Cell and subtype-specific transcription factor regulators. a, Identification of major cell type-specific SCENIC transcription factor regulons. The heat map shows the module score of the top 5 transcription factor regulons (rows) for each major cell type across all individuals and major cell types (columns). **b**, Identification of inhibitory neuron subclass-specific SCENIC transcription factor regulons. The heat map shows the module score of the top 5 transcription factor regulons (rows) for each subclass across all individuals and subclasses of inhibitory neurons (columns). **c**, Identification of astrocyte subtype-specific SCENIC transcription factor regulons. The heat map shows the mean module score of the top 5 transcription factor regulons (rows) across all individuals and astrocyte subtypes (columns).

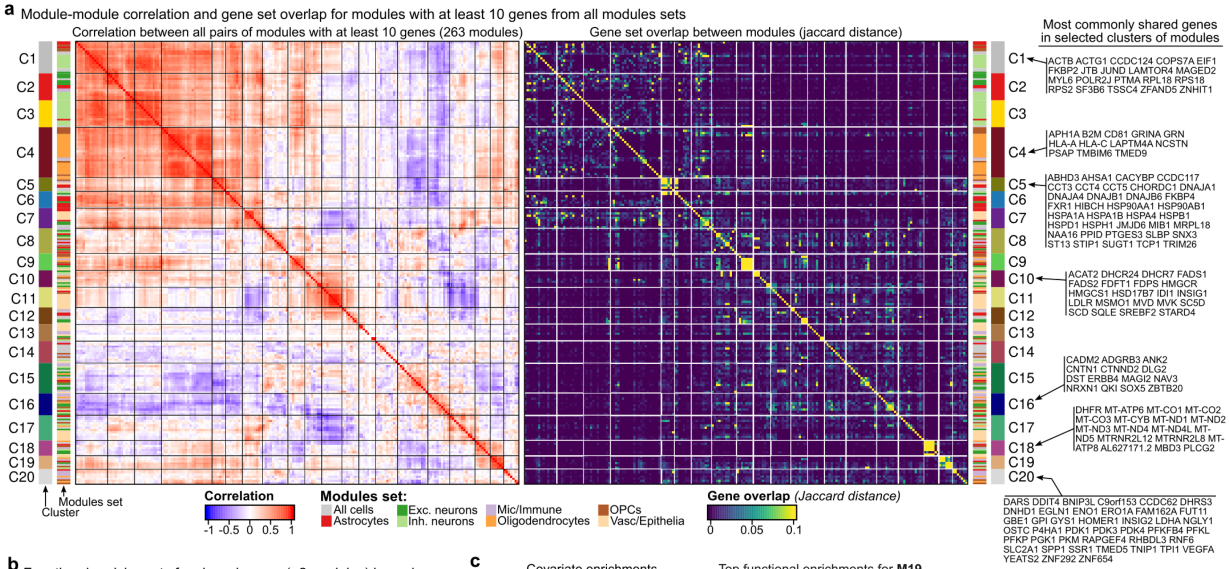


Extended Data Fig. 4 | Region-specific cell-cell communication. a-b, Ligand-receptor pairs with the greatest increase (**a**) or decrease (**b**) in interaction score in the thalamus compared to the prefrontal cortex. Bar plots show the interaction scores for the ligand-receptor pairs indicated. The interaction score was calculated by applying the minus log10 transformation to the robust rank aggregation (RRA) score. A lower RRA score indicates that a ligand-receptor interaction is ranked consistently higher than expected by chance across multiple prediction methods. Violin plots show the expression of the ligand (left) and receptor (right) in the cell types and brain regions indicated.



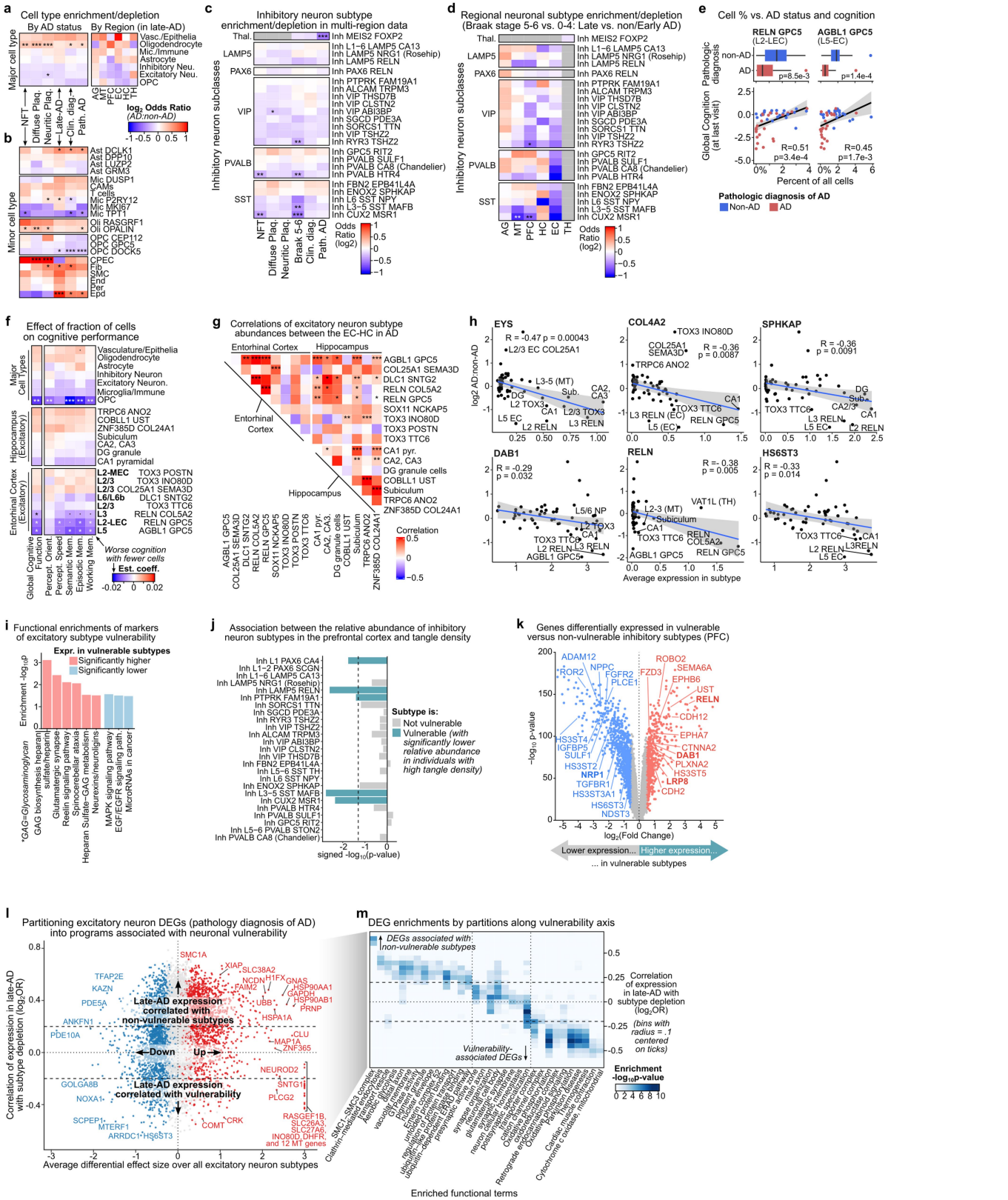
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Module summary panels across modules. **a**-**h**, Overview of gene expression modules with at least 10 genes each across all cells and across major cell types, showing the module name, number of genes, percent expression, top module genes, enrichments by subtype (except for neuron subtypes, see Supplement), covariates, and regions, and the top functional enrichment for each module. Percent expression is the percent of cells whose average expression (log1p, normalized) of the module is above 1. Covariate enrichments are performed by hypergeometric test, comparing the intersection of cells with z-scored module expression of at least 1 vs. with $z < 1$ against a particular level of a covariate of interest (e.g. cells from the entorhinal cortex region or cells of a specific subtype). Panels summarize modules found in all cells (**a**), astrocytes (**b**), OPCs (**c**), microglia and immune cells (**d**), oligodendrocytes (**e**), inhibitory neurons (**f**), vasculature and epithelia (**g**), and excitatory neurons (**h**). All modules except vasculature and epithelia modules are split into identity vs. other, where identity modules are highly enriched in a single subtype and have an average expression greater than 1 (log1p, normalized) for over 50% of the subtype's cells.



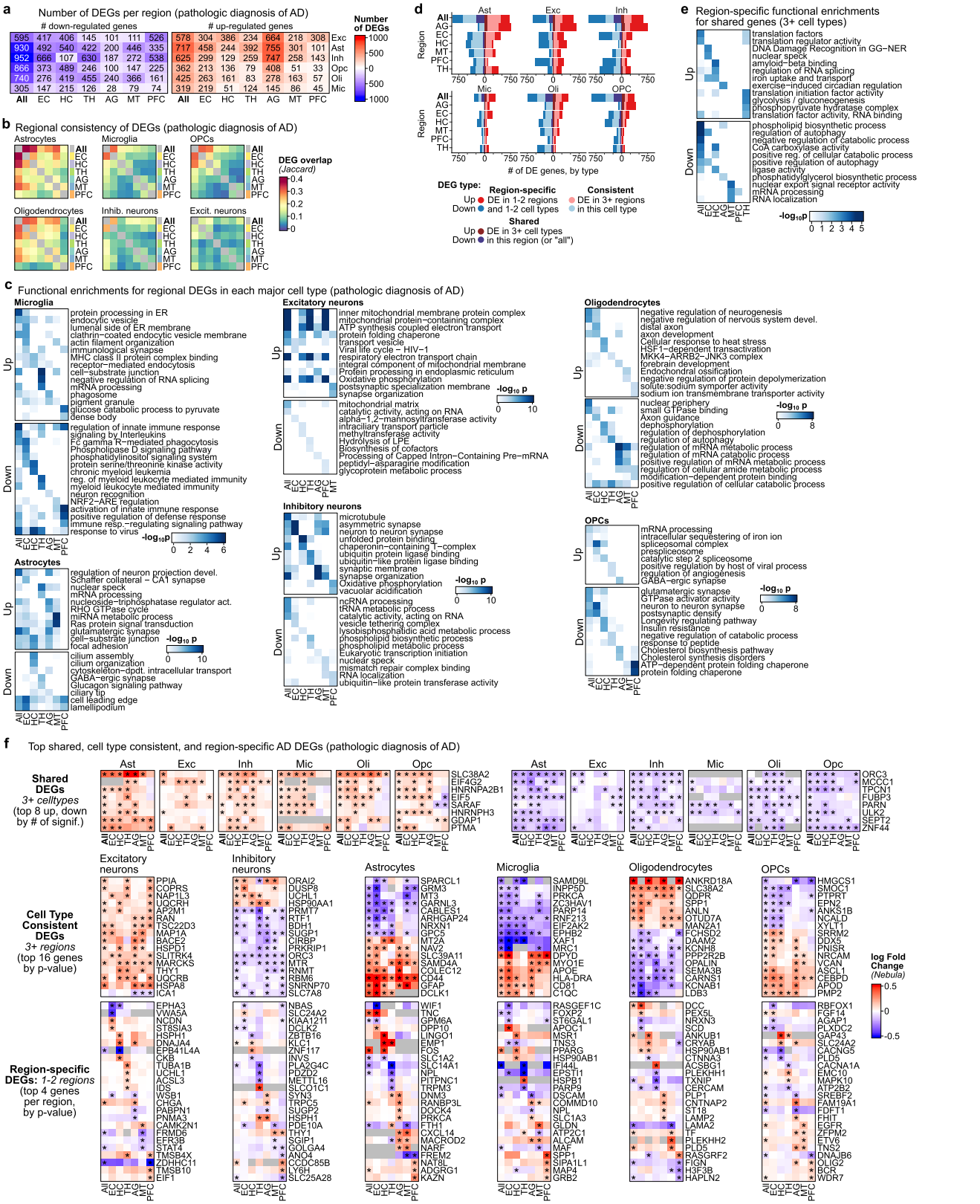
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Cross-module clustering and comparison. a, Module-module correlation (Pearson correlation) and gene set overlap (Jaccard distance) for modules with at least 10 genes from all sets of modules (263 modules in total). Heatmaps are ordered by the hierarchical clustering of the correlation matrix and cuts represent 20 clusters cut from the hierarchical clustering dendrogram. Left and right side bars label rows by their modules set of origin (major cell type colours and grey for all cells). The most commonly shared genes in selected clusters of modules are shown on the right of the gene set overlap heatmap. **b**, Functional enrichments for each cluster of modules for the shared genes (>2 modules) in each cluster (only clusters with significant enrichments shown). Up to 5 enrichments shown, ordered by p-value, labelled by their source and only keeping terms with fewer than 500 genes. **c**, Covariate and functional enrichments for example astrocyte modules M19 (thalamus identity program) and M17 (cholesterol metabolism and biosynthesis program). Region, subtype, and covariate enrichments performed at cell level by stratifying cells with z-score > 1 and testing for regional or subtype enrichment (see Methods). Functional enrichments performed using gprofiler2, keeping terms with fewer than 500 genes. **d**, Scatterplots and correlation of scores for selected pairs of astrocyte modules. Each dot represents the module expression scores for a subtype in a specific sample and is coloured by the astrocyte subtype. Grey area represents the 95% confidence interval around the linear fit. **e**, Functional enrichments for selected astrocyte modules, showing top 10 functional enrichments for each pair of compared correlated modules (and for M6, M13, M27 together). Only terms with fewer than 500 genes shown. **f**, Microglial and immune modules network from correlation of module pairs at the subtype by sample level (edges shown where FDR-adjusted p-value < 0.05). Nodes are coloured by module's relative expression in each of the microglial and immune subtypes and groups highlight sets of subtype-biased modules.

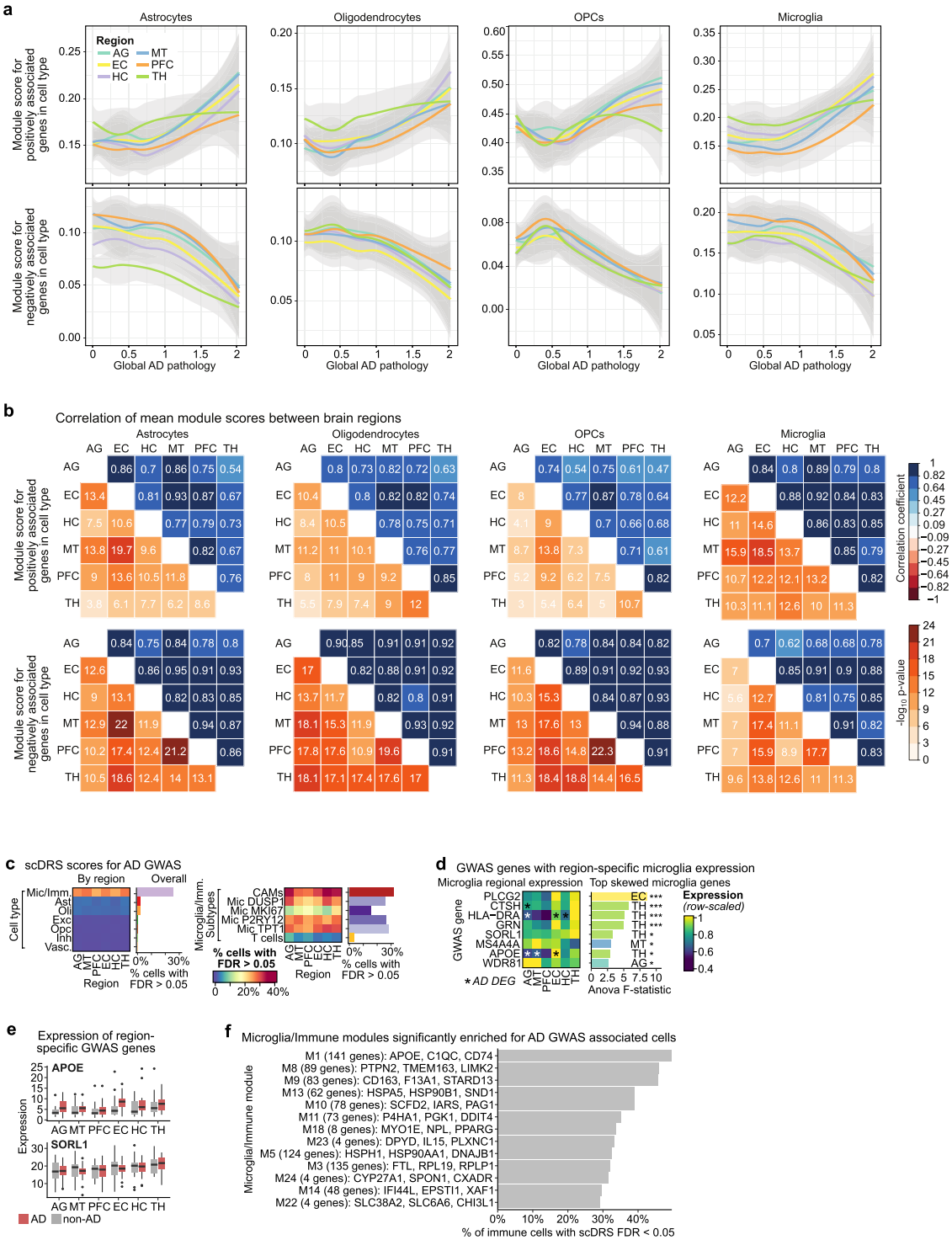


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Neuronal vulnerability, connectivity, and markers of vulnerability. a, Compositional differences for major cell types in AD by quasi-binomial regression with FDR-correction. $\log_2$ OR shown both for each AD variable across regions (left) and for each region in late-AD (right). Analysis performed for individual-level AD status and region-level pathology measurements. Pathologic diagnosis of AD (Path. AD) was stratified by NIA-Reagan score (26 AD and 22 non-AD) and clinical diagnosis was stratified as AD dementia (n = 16) and non-CI (n = 32). **b**-**c**, Compositional differences for glial subtypes and inhibitory neuron subtypes according to individual-level AD status and region-level pathology measurements (as in **a**). Grey boxes indicate interactions that are not computed due to MEIS2 FOXP2 specificity to the thalamus, where we do not have measured regional scores. **d**, Compositional differences in inhibitory neuron subtypes in late AD (Braak Stage 5-6 vs. 1–4) in each region. Grey boxes indicate interactions that are not computed due to subtype regional specificity. **e**, Boxplots (top) of neuronal fraction for two vulnerable EC subtypes, split by AD status (AD: blue, non-AD: red), with p-values from one-sided Wilcoxon test. Scatter plots (bottom) of individuals' global cognition at last visit against cell fraction for two AD-vulnerable entorhinal cortex subtypes, coloured by AD. Linear fit with 95% confidence interval shown in grey. **f**, Estimated effect size of cell fraction ($\log_{10}$) on scores for performance in various cognitive domains at last visit and combined scores from all domains (global). Linear regression FDR-corrected p-values (**adjusted p-value < 0.01, *<0.05, dot is <0.1). **g**, Full correlation matrix between subtype fraction between the hippocampus and entorhinal cortex in the same individuals, as described in the methods (***adjusted p-value < 0.001, **<0.01, *<0.05). **i**, Example genes predictive of subtype vulnerability. Scatterplots show average expression in the subtype across individuals against the effect size of the depletion or enrichment in AD as measured by the $\log_2$ odds-ratio for late-AD, as in the Methods. **i**, Functional enrichments and intersected genes for top 30 markers of subtype vulnerability (terms with <500 genes). **j**, Association (quasi-binomial regression) between the relative abundance of inhibitory neuron subtypes in the prefrontal cortex and the density of neurofibrillary tangles. Association scores (signed negative log10 FDR-adjusted P value, where the sign was determined by the direction (positive or negative) of the association) are shown. The dotted line indicates the significance level threshold of an FDR-corrected P value of 0.05. P values were derived using the glm function in R and adjusted for multiple testing via the Benjamini-Hochberg method. **k**, Volcano plot showing genes differentially expressed in vulnerable versus non-vulnerable inhibitory neuron subtypes (genes significantly higher in vulnerable subtypes in red, lower in blue). FDR-adjusted P values as determined by the R package 'dreamlet' are shown. **l**, Scatter plot of each tested gene's average differential expression effect size in late-AD (y-axis) versus the correlation of its expression in a subtype and that subtype's level of depletion in late-AD (x-axis). Dashed lines separate genes associated with vulnerability and non-vulnerability. **m**, Functional enrichments for each identified class of neuronal DEGs (terms <500 genes) on bins (along x-axis from **l**), from genes associated with vulnerability to those associated with non-vulnerability (only genes with biased effect sizes, see Methods). Dashed lines correspond to the same breaks as in (**l**).

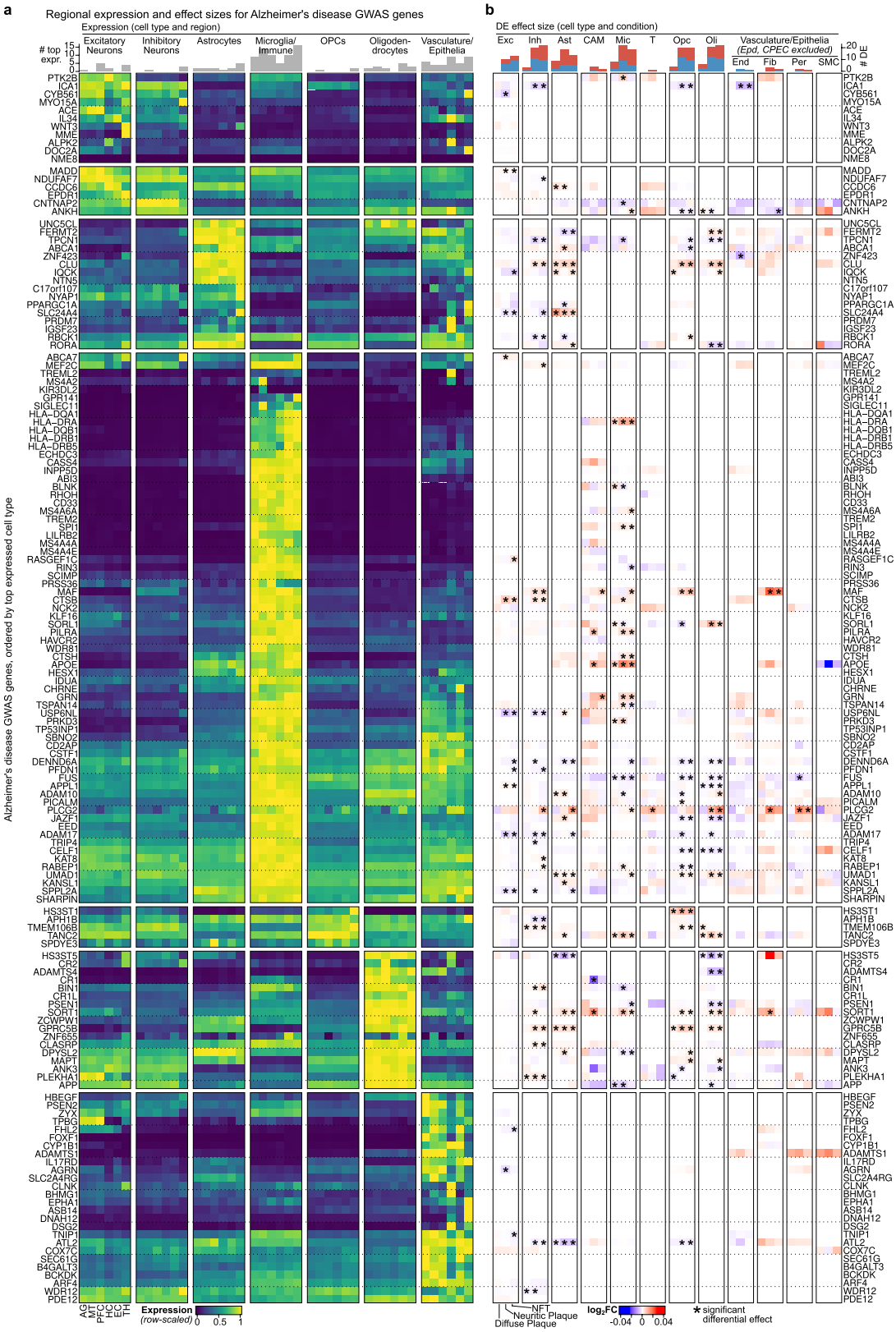


Extended Data Fig. 8 | Regional differential expression and GWAS association. a, Number of up- and down-regulated differentially expressed genes (DEGs) with respect to pathologic diagnosis of AD for each major cell type, calculated in each region separately as well as jointly over all regions. **b**, Heatmaps of Jaccard similarity of DEGs across regions for each major cell type. **c**, Heatmap of -$\log_{10}$ p-values for functional enrichments showing the top pathways for AD DEG shared across 3+ cell types. Enrichments shown for DEGs calculated in each region and in all regions together (up to the top 3 pathways per analysis are shown). **d**, Barplot of number of DEGs per region and cell type, coloured by type of DEG, as determined by its shared differential expression across regions and cell types. **e**, Top functional enrichments for region-specific DEGs and DEGs shared across regions (≥3 regions) with up to the top 2 terms (<500 genes only) shown per region. Panels shown and computed separately for each major cell type. **f**, Heatmap of log fold change for top shared, cell-type consistent, and cell+region-specific DEGs in major cell types. GG-NER: global genome nucleotide excision repair.

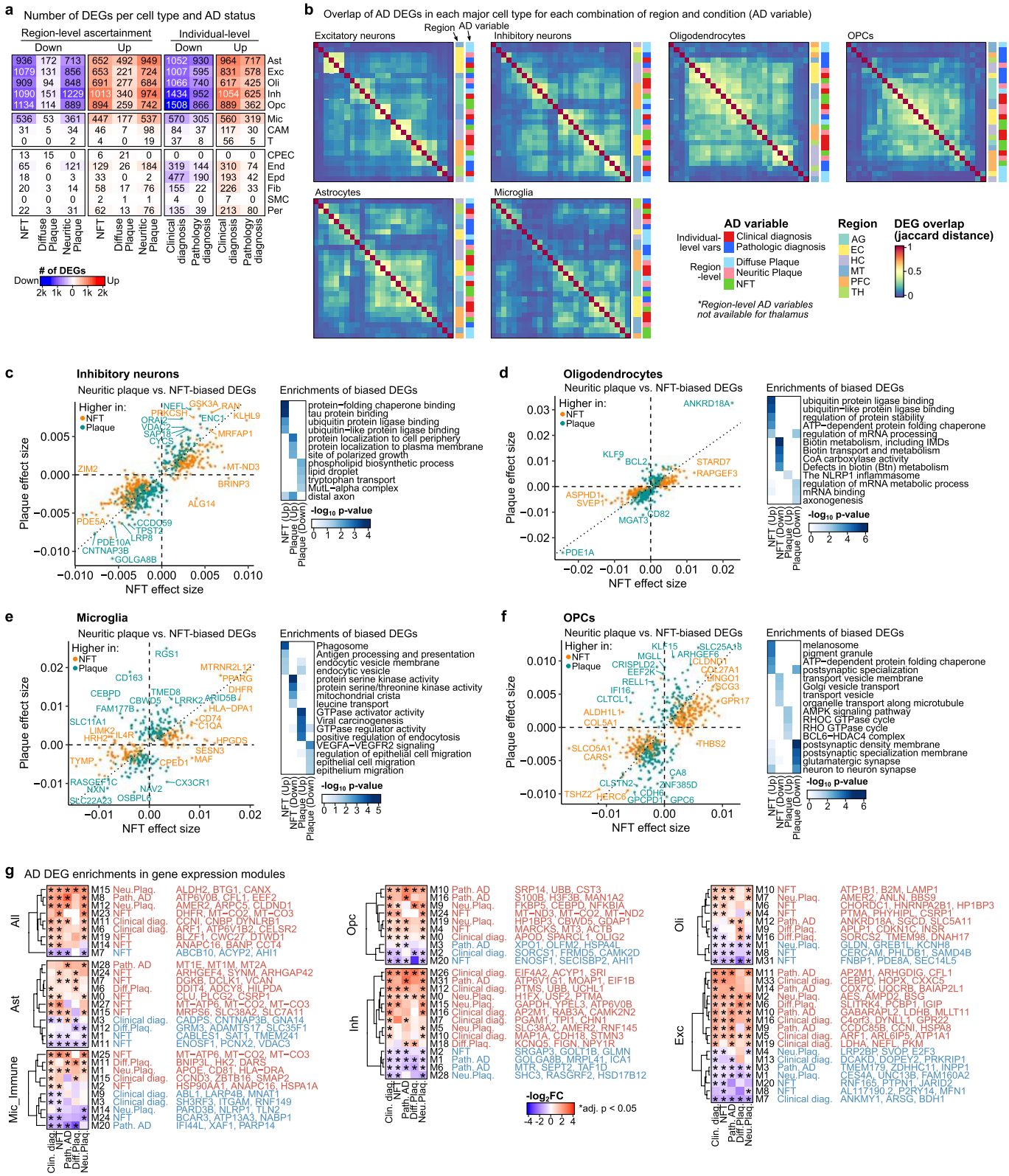


Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Inter-regional comparison of AD pathology-associated gene sets and region-specific GWAS enrichments. a-**b**, Seurat module scores of genes significantly positively (top) or negatively (bottom) associated with the global AD pathology variable in prefrontal cortex for astrocytes, oligodendrocytes, OPCs, and microglia across brain regions and the spectrum of global AD pathology burden. The gene sets used for computing the module scores (genes significantly associated with global AD pathology burden) were determined based on snRNA-seq data derived from prefrontal cortex tissue of 427 ROSMAP study participants. The scatterplots (**a** panels) illustrate the relationship between global AD pathology burden and the mean module score for the specified gene set, with this mean score calculated by averaging the module scores of all cells of the designated cell type isolated from an individual. A LOESS (Locally Estimated Scatterplot Smoothing) regression line with a 95% confidence interval is shown, and the regression lines are coloured by brain region. The central LOESS regression line represents the local measure of central tendency, calculated through locally weighted regression to reflect the smoothed relationship between the module scores indicated and global AD pathology burden. Interregional Pearson's correlation analysis of mean module scores (**b** panels) was performed by first averaging the module scores of all cells of the cell type of interest from an individual study participant. The correlation analysis was then performed between regions based on these averaged scores. P values were calculated using the cor.test function in R and were adjusted for multiple testing using the p.adjust function with the Benjamini-Hochberg method. **c**, Heatmap (by region, left) and barplot (over all regions, right) showing the percentage of cells with significant scDRS (disease relevance scores) for AD GWAS. Rows are split into major cell type groups (top) and microglia and immune subtypes. **d**, Regional expression (heatmap, left) and F-statistic for region in predicting expression (barplot, right) for eight GWAS genes with significantly region-specific expression in microglia. Barplot is coloured by the top expressed region (regression coefficient). Heatmap is labelled with stars if the gene is a DEG for that region. **e**, Boxplots showing expression of two of the region-specific GWAS genes in individuals with and without a pathologic diagnosis of AD. **f**, Microglia/immune modules associated with AD GWAS. Fraction of microglia or immune cells with significant expression of each module (z-score > 2.5) and with significant scDRS scores (FDR < 0.05). Only significant modules are shown (adjusted $p < 0.01$, hypergeometric test with BH correction).



Extended Data Fig. 10 | Alzheimer's disease GWAS-linked genes in the multi-region atlas. a, Expression level by region/subtype and effect sizes of 150 Alzheimer's disease candidate risk genes from Alzheimer's disease GWAS risk loci. **b**, Differential effect sizes and significance for each candidate risk gene in each minor cell type across regional pathology measurements. Ependymal cells and CPEC cells were excluded as the thalamus does not have regional pathology measurements.



Extended Data Fig. 11 | Pathology-biased DEGs for major cell types. **a**, Number of DEGs for each cell type for both region-level pathology measurements and individual-level AD status (DE analysis performed over all regions jointly). **b**, Overlap of AD DEGs in each major cell type for each combination of region and condition (AD variable). DEG overlap computed by Jaccard distance and rows/columns hierarchically ordered by Euclidean distance. **c**-**f**, Scatter plots of average effect sizes for NFT and plaque for DEGs with biased differential effect sizes (left panels) and their respective functional enrichments (right panels), for DEGs specific to inhibitory neurons (**c**), oligodendrocytes (**d**), microglia (**e**), and OPCs (**f**). Genes are coloured by whether they have higher effect size relative to NFT (orange) or plaque levels (teal). **g**, Heatmap of hypergeometric enrichments of up (red) or down (blue) AD DEGs in modules for DEGs in all sets of modules across all regions, by AD condition. Only modules with at least two significant enrichments are shown and rows are clustered hierarchically by Euclidean distance.

nature portfolio

Corresponding author(s): Li-Huei Tsai and Manolis Kellis

Last updated by author(s): Apr 19, 2024

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies 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	Gene counts were obtained by aligning reads to the GRCh38 genome using Cell Ranger software (v.3.0.2) (l0x Genomics). We used SCAN PY (vl.6) to process and cluster the expression profiles and infer cell identities. We called doublets using DoubletFinder (v2.0). The R package Seurat (v.3.2) was used for UMAP visualizations and to determine marker genes. Prism 9 software was used for histogram visualizations and for statistical analyses comparing the number of genes and transcripts detected per cell. The Differential Search tool of the Allen Brain Atlas data portal (v.7) (https://human.brain-map.org/microarray/search) was used to compare microarray data from different human brain regions. The Consensus Non-negative Matrix factorization (cNMF) (vl.3) package implemented in python was used for the NMF analysis. The gene regulatory network analysis was performed using pySCENIC (v0.10.4). MAST (vl.16.0) and Nebula (vl.1.7) were used for the differential expression analyses. The gprofiler2 package in R DEG was used for DEG enrichments. Cell-cell communication events were predicted using the LIgand-receptor ANalysis frAmework (LIANA) (v0.1.12). The differential expression analysis comparing vulnerable to non-vulnerable inhibitory neuron subtypes was performed with the R package dreamlet (v0.99.6). Genes associated with cognitive resilience were identified using the R package muscat (v1.12.1). Bulk RNA-seq differential expression analysis was performed using DESeq2 (v1.38.3). Further statistical analyses and visualizations were implemented in R version 4.0.3 {2020-10-10). Gene expression modules were determined using the scdemon (v0.9.0) framework. The code for the scdemon method for module detection

from single-cell RNA-seq is available at https://github.com/KellisLab/scdemon.
Code for analysis is available at https://github.com/cboix/admultiregion_analysis and Zenodo (11051021, https://doi.org/10.5281/zenodo.11051021).

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Single-nucleus RNA-seq profiling data is available from Synapse (www.synapse.org) in coordination with the ROSMAP project. Data is accessible at https://www.synapse.org/#!Synapse:syn52293442 (as part of The MIT ROSMAP Single-Nucleus Multiomics Study https://www.synapse.org/#!Synapse:syn52293417). The data is 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 (https://www.radc.rush.edu/; https://adknowledgeportal.synapse.org/). Additional processed data as well as integrative visualization and exploration of the atlas are available through http://compbio.mit.edu/ad_multiregion/ and https://ad-multi-region.cells.ucsc.edu/.

We also downloaded the following public single-cell gene expression datasets: Human Multiple Cortical Areas SMART-seq (https://portal.brain-map.org/atlases-and-data/rnaseq/human-multiple-cortical-areas-smart-seq), human DLPFC (https://www.synapse.org/#!Synapse:syn51123521), SEA-AD MTG (https://sea-ad-single-cell-profiling.s3.amazonaws.com/index.html#MTG/RNAseq/), SEA-AD DLPFC (https://sea-ad-single-cell-profiling.s3.amazonaws.com/index.html#DLPFC/RNAseq/), human dLGN (https://portal.brain-map.org/atlases-and-data/rnaseq/comparative-lgn), multiple human brain regions (https://console.cloud.google.com/storage/browser/linnarsson-lab-human;tab=objects?authuser=0&prefix=&forceOnObjectsSortingFiltering=false), multiple cortical areas and the hippocampal formation of the mouse brain (https://portal.brain-map.org/atlases-and-data/rnaseq/mouse-whole-cortex-and-hippocampus-10x), nine regions in the adult mouse brain (http://dropviz.org/), and Mouse Brain Atlas (http://mousebrain.org/).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	We selected 48 individuals from the Religious Orders Study and Rush Memory and Aging Project (ROSMAP), both ongoing longitudinal clinical-pathologic cohort studies of aging and dementia, in which all participants are brain donors. Individuals were balanced between sexes (male:female ratios 13:13 in AD, 11:11 in NoAD) (sex was determined based on self-reporting).
Reporting on race, ethnicity, or other socially relevant groupings	No socially constructed or socially relevant categorization variables were used in this study.
Population characteristics	We selected 48 individuals from the Religious Orders Study and Rush Memory and Aging Project (ROSMAP), both ongoing longitudinal clinical-pathologic cohort studies of aging and dementia, in which all participants are brain donors. For the purpose of this study, individuals were selected based on the modified NIA-Reagan diagnosis of Alzheimer's disease and the Braak stage score (Braak stage 0,1,2, n=20; Braak stage 3,4, n=14; Braak stage 5,6, n=14), with 26 individuals having a positive pathologic diagnosis of AD and 22 individuals having a negative pathologic diagnosis of AD. Individuals were balanced between sexes (male:female ratios 13:13 in AD, 11:11 in NoAD), matched for age (medians 86.6 years (AD) and 86.0 years (NoAD)), and postmortem interval (medians 5.9 hours (AD) and 6.3 hours (NoAD)).
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.

Field-specific reporting

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

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

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

Life sciences study design

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

Sample size	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 11 individuals per group. Therefore we analyzed brain tissue from 26 individuals having a positive pathologic diagnosis of AD and 22 individuals having a negative pathologic diagnosis of AD. The sample size of 48 was chosen based on findings from our previous study, which also included the same number of participants. This previous study demonstrated that a sample size of 48 is sufficient to detect significant differences between individuals diagnosed with and without Alzheimer's Disease.
Data exclusions	Low quality snRNA-seq libraries were excluded and the exclusion criteria are described in the manuscript as follows. We kept only protein coding genes and filtered out cells with over 20% mitochondrial or 5% ribosomal RNA, leaving 1.47M cells over 48 individuals and 283 samples across all regions. We separately called doublets using DoubletFinder and flagged and removed clusters with strong doublet profiles and clusters showing strong individual-specific batch effects, leaving a final dataset of 1.35M cells.
Replication	Verification of the single-nucleus RNA-seq data was performed through validation using RNA in situ hybridization on post-mortem brain tissue. These experiments validated the findings derived from snRNA-seq. The snRNA-seq experiment was performed once. The RNA in situ hybridization (RNAscope) experiments shown in Figure 1i, Figure 2d, Figure 3g, Figure 4o-p, and Figure 5j-l were each performed once. Similarly, the IHC experiments shown in Figure 3h-i were each performed once.
Randomization	The study participants were allocated into groups based on Braak stage.
Blinding	Investigators were not blinded to group allocation. The outcome measures used in our snRNA-seq analysis are objective, relying on standardized computational methods, which reduces the potential for bias that blinding seeks to mitigate.

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	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	• Anti-reelin o Host: monoclonal mouse o Catalogue name: Anti-Reelin Antibody, a.a. 164-496 mreelin, clone G10 o Catalogue number: Millipore Sigma MAB5364 • Anti-NeuN o Host: polyclonal serum from guinea pig o Catalogue name: NeuN antibody o Catalogue number: Synaptic System 266004 • Anti-Amyloid ß o Host: Monoclonal rabbit o Catalogue name: β-Amyloid (D54D2) XP® Rabbit mAb o Catalogue number: 8243S • Anti-Phospho-tau o Host: Polyclonal rabbit o Catalogue name: Phospho-Tau (Ser396) o Catalogue number: 44-752G
Validation	• Anti-reelin: Millipore Sigma application statement: Detect Reelin using this Anti-Reelin Antibody, a.a. 164-496 mreelin, clone G10 validated for use in IH & WB. • Anti-NeuN Silencing CA1 pyramidal cells output reveals the role of feedback inhibition in hippocampal oscillations. Adaikkan C, Joseph J, Foustoukos G, Wang J, Polygalov D, Boehringer R, Middleton SJ, Huang AJY, Tsai LH, McHugh TJ

Nature communications (2024) 151: 2190. 266 004 IHC; tested species: mouse
• Anti-Amyloid ß
Cell Signaling Technology Specificity/Sensitivity statement: β-Amyloid (D54D2) XP® Rabbit mAb recognizes endogenous levels of total β-amyloid peptide (Aβ). The antibody detects several isoforms of Aβ, such as Aβ-37, Aβ-38, Aβ-39, Aβ-40, and Aβ-42. This product detects transgenically expressed human APP in mouse models.
• Anti-Phospho-tau
Invitrogen Advanced Verification statement: This Antibody was verified by Cell treatment to ensure that the antibody binds to the antigen stated.

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.