Multi-cellular communities are perturbed in the aging human brain and with Alzheimer's disease

Anael Cain[1*], Mariko Taga[2*], Cristin McCabe[4**], Idan Hekselman[3**], Charles C. White[5], Gilad Green[1], Orit Rozenblatt-Rosen[4,11], Feng Zhang[6], Esti Yeger-Lotem[3,8], David A. Bennett[7], Hyun-Sik Yang[5,9], Aviv Regev[4,10,11+], Vilas Menon[2+], Naomi Habib[1,5+], Philip L. De Jager[2,5+]

*** Equal contribution**
**** Equal contribution**
+ Equal contribution
Corresponding Authors: Philip L. De Jager, Naomi Habib, Vilas Menon

[1] Edmond & Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel
[2] Center for Translational & Computational Immunology, Department of Neurology and Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, New York, USA
[3] Department of Clinical Biochemistry and Pharmacology, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[4] Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.
[5] Broad Institute, Cambridge, Massachusetts, USA.
[6] Howard Hughes Medical Institute, McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[7] Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, USA.
[8] National Institute for Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[9] Harvard Medical School, Boston, MA; Center for Alzheimer Research and Treatment, Department of Neurology, Brigham and Women's Hospital, Boston, MA
[10] Howard Hughes Medical Institute, Koch Institute of Integrative Cancer Research, Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.
[11] Current address: Genentech, 1 DNA Way, South San Francisco, CA

Abstract

The role of different cell types and their interactions in Alzheimer's disease (AD) is an open question that we have pursued by mapping the human brain at the single cell level. Here, we present a high resolution cellular map of the aging frontal cortex by single nucleus RNA-sequencing of 24 individuals with different clinicopathologic characteristics; which we used to infer the cellular architecture of 640 individuals from bulk RNA-seq profiles. Powered by this sample of sufficient size to obtain statistically robust results, we uncovered AD associations with neuronal subtypes and oligodendroglial states. Moreover, we uncovered a network of cellular communities, each composed of different neuronal, glial and endothelial cells subpopulations whose frequencies are correlated across individuals. Two of the cellular communities are altered in relation to cognitive decline and tau pathology. Our work provides a roadmap for evaluating cross-cell type differences in the cellular environment of the AD brain.

INTRODUCTION

Over the past decade, our understanding of the molecular landscape of Alzheimer's disease (AD) has advanced rapidly as new experimental and analytic methods synergize to uncover pathways that contribute to the sequence of events that lead to AD dementia. While genetic studies have created a strong foundation of selected loci and genes implicated as causal in susceptibility to AD and its endophenotypes, transcriptomic analyses have sketched a broader outline of molecular changes that capture the dynamic component of the disease and the state of the target organ. However, most prior efforts profiled RNA at the bulk tissue level (*e.g.* [1]), averaging expression measures across a myriad of cell types and states which obscured finer distinctions and contributions from cell subtypes found at low frequency in the target tissue.

Recent efforts to profile single nuclei from post-mortem brain tissue of healthy and AD patients, as well as mouse models, have uncovered specific cell types with different signatures and proportions in AD, especially in certain subsets of microglia, oligodendrocytes and astrocytes[2–8]. While such studies also hint at substantial inter-individual diversity, their limited sample size and moderate number of profiled nuclei per subject yields an incomplete sketch of the architecture of the aging neocortex, and they did not yet tackle the critical question of cellular interactions among the many different cell types and subtypes found in the aging and AD brain. As a result, several key questions remain open: (1) Where in the causal chain of events leading to AD are such changes in the cellular architecture taking place? (2) Are changes in certain cellular subsets coordinated with or independent of one another? Addressing these questions requires relating detailed single cell/nucleus maps to datasets composed of large numbers of deeply phenotyped individuals and using computational methods which characterize both individual cells and cellular communities.

Here, we have deployed a combined approach that integrates single nucleus RNA-seq profiling (snRNA-seq) in a structured subgroup of 24 well-characterized human subjects, together with bulk profiles of a statistically well-powered set of 640 subjects[1]. Each of these subjects is a participant in a prospective longitudinal study of cognitive aging (ROS and MAP[9–11]) with cognitive and neuropathologic characterization, and their dorsolateral prefrontal cortex (DLPFC) was profiled by RNA-seq: 24 participants (profiled by snRNA-seq) span four archetypal categories of older women and men, to ensure the capture of the cellular diversity of this brain region across a range of clinicopathologic states; 640 participants (with previous bulk RNA-seq profiles[1]) are a relatively random sample of the two

cohorts and reflect the distribution of characteristics seen in the older population, enabling robust statistical modeling (**Fig. 1a**). Our deeper sampling of single nuclei profiles and high library quality yielded an enhanced map of cell types, subtypes and cell states of the aging cortex (when compared to previous reports [3–5]). We used our map of single nucleus profiles and a new computational approach, CelMod, to hierarchically estimate broad cell class and subtype proportions across the 640 bulk profiles. This large set of inferred proportions provided power for statistical assessments of cell subpopulations associated with pathophysiology, uncovering an association between cognitive decline and the proportion of specific subsets of glial cells and subtypes of inhibitory neurons, in particular, a relative decrease in somatostatin (SST) neurons. Further, we used our single nucleus data to infer a map of multi-cellular communities whose proportions are correlated across individuals, suggesting the existence of possible micro-environments in the aging brain. Associating each of these cellular communities to AD related traits identified two anti-correlated communities, which each display opposite associations with cognitive decline in AD and tau pathology. Finally, causal modeling suggests that the differences in cell subtypes and signatures occur downstream of the accumulation of tau pathology and might mediate downstream effects that accelerate cognitive decline. Our model can inform further investigation and therapeutic development, by identifying those cellular factors that may most proximally and directly contribute to loss of cognitive function with advancing age and AD.

RESULTS

Inter-individual diversity drives cellular landscapes in the aging DLPFC

To build a map of the aging DLPFC (BA9), we generated snRNA-seq profiles from frozen tissue samples obtained from 24 ROSMAP[10,11] participants (average age of 87.9 years at the time of death). ROS (Religious Order Study) and MAP (Memory and Aging project) are two harmonized longitudinal studies of cognitive aging with prospective autopsy and deep cognitive and neuropathologic characterization that are run by a single group of investigators and designed to be analyzed together[10,11]. To sample a wide variety of cell subtypes and states, we selected participants that represent four major archetypes of the aging population (**Fig. 1a, Supplementary Table 1**): (1) a reference group of cognitively non-impaired individuals with minimal AD pathology, (2) a resilient group of cognitively non-impaired individuals with a pathologic diagnosis of AD, (3) an AD group diagnosed who fulfill diagnostic criteria for both clinical AD dementia and pathologic AD, and (4) a clinical-AD group diagnosed with AD dementia but only minimal AD pathology post-mortem. Each

group consists of 6 individuals (50% men/women). This sampling strategy ensures that we capture a wide range of cellular signatures across the different pathological and clinical manifestations of the disease within the 24 individuals profiled by snRNA-seq. Following rigorous pre-processing, we retained 162,562 DLPFC nuclei for analysis, with an average of 7,094 nuclei per participant after doublet filtering and quality control, **Fig. 1b, Supplementary Fig. 1, Methods**).

Using an unsupervised clustering approach, we identified distinct groups of nuclei spanning all 8 major expected cell types (**Fig. 1c, Supplementary Fig. 1b, Supplementary Table 2**), with all subsets present in each individual, but with considerable inter-individual variation in their relative proportions (**Fig. 1c)**. Cell proportions varied in coordinated ways across the 24 individuals, forming two cell groupings (**Fig. 1d**), with intra-group positive correlation (mean Spearman ρ=0.37 and 0.29) and inter-group negative correlation (mean Spearman ρ=-0.1). Group 1 includes microglia, oligodendrocytes, pericytes and endothelial cells, and Group 2 includes astrocytes, oligodendrocyte progenitor cells (OPCs), and inhibitory neurons. Excitatory glutamatergic neurons do not correlate with either group (**Fig. 1d**). The association trends agree with previously-reported changes in broad glial and neuronal categories in aging human brains [3–5].

Moreover, there were differences between the four patient groups within the major cell classes, such that the spectrum of cells within a category shifted between patient groups (**Fig. 1e**). For example, while all participants had astrocyte cells across a spectrum of subtypes, the distribution of astrocytes along this spectrum varied between the four patient groups. This distribution was more similar between the two groups with high AD pathology and more distinct in the reference group (**Fig. 1e**).

A high-resolution atlas of cell subtypes in the aging neocortex

To explore finer distinctions in cellular diversity, we assembled our 162,562 nuclei to characterize the diversity of cell subsets (subtypes, states and expression programs) in the neocortex. We analyzed profiles from each of the 8 major cell types separately and used one of two approaches: we used either a standard clustering of cells into subsets or recovery of co-varying gene programs (topic models), depending on the specific properties of each cell type.

We used clustering to identify discrete subsets for 4 of the cell types (**Methods**, **Fig. 2a, Supplementary Fig. 2)**; this yielded 5 astrocytic, 6 microglial, 4 endothelial, and 16 neuronal subsets (**Fig. 2b, Supplementary Table 2**). We did not identify robust subdivisions within the OPCs and pericytes, possibly due to their relatively lower numbers. Among neurons, 9 excitatory and 7 inhibitory subsets (**Fig. 2b**) capture the diversity of the neocortex which are consistent with prior characterizations [12] (**Fig. 2c)**. They include pyramidal neurons across the different layers of the cortex and known GABAergic inhibitory subtypes, such as those expressing vasoactive intestinal peptide (VIP), somatostatin (SST), or parvalbumin (PV) (**Fig. 2c**).

On the other hand, oligodendrocytes appear to be distributed along several different gradients of expression, without clear boundaries. As a result, we applied topic modeling using Latent Dirichlet Allocation (LDA)[13–16] (**Fig. 2d, Supplementary Table 2, Methods**), to recover gene programs (called "topics") based on co-variation patterns of gene expression across cells. These topics are weighted differently across each cell/nucleus, and thus this approach simultaneously models every cell as a mixture of these gene programs. Each gene has a score for each topic, indicating its contribution to the topic, and each nucleus profile has a weight for each topic, indicating the extent to which that topic (program) is present in a particular nucleus. In oligodendrocytes, we found four major topics (**Fig. 2c**), and some nuclei are weighted for more than one topic. We annotated the topics by their highly scoring genes per topic, using a score based on the KL divergence to compare the distribution of topic weights and the expression level of genes across cells[15] (**Methods**, **Supplementary Table 3**). For example, genes highly weighted in topic Olig.3 include insulin growth factor receptor 1 (IGFR1, KL divergence score=0.089, **Methods**) previously linked to Aβ degradation[17], and the calcium homeostasis gene ITPR2 (KL divergence score=0.047). Another notable gene is Potassium Voltage-Gated Channel KCNAB1, which had a high weight in Topic Olig.1 and plays a role in extracellular potassium homeostasis (KL divergence score=0.11). Examining the relative distribution of cells from the 24 subjects across the four archetype groups suggests that oligodendrocytes distribute differently between groups, in a manner that is partially mirrored by the distribution of topics (**Fig. 2e**). Thus, such shifts in cellular states may be related to differences in gene programs associated with pathology and cognitive decline; however, the small sample size of this dataset is underpowered for meaningful statistical analysis (**Figure 4D**).

We thus subsequently deployed this refined atlas of cell subsets and topics to assess inter-individual diversity and inter-group variation in the context of demographic and clinico-pathological measures.

A diversity of cell states in microglia and astrocytes

The 3,916 microglial nuclei profiled were partitioned into 6 major subsets (**Fig. 3a, Supplementary Fig. 3a-d**), characterized by expression of known marker genes (**Fig. 3a,b**) as homeostatic microglia (Micr.1), two subsets of active microglia (Micr.2, Micr.3), active interferon-associated (Micr.4), and proliferating (Micr.5) microglia subsets, and a small cluster of myeloid-like cells (Micr.6, *k*=67 nuclei). We then used our reference map[8] of 9 microglial subsets generated from 16,096 living human microglia isolated from the same brain region (DLPFC) to further annotate and validate each of the 6 nucleus-derived clusters (**Fig. 3c,d**). Most nuclei subsets mapped clearly to a defined set of clusters of live microglia cells: Micr.1 mapped to homeostatic cells-clusters; Micr.2 to putative stress response, the anti-inflammatory and the antigen-presenting cells-clusters; Micr.3 to the putative enhanced redox and the interferon response cells-clusters; Micr.4 to the putative interferon-response cells-cluster; Micr.5 to the proliferation cells-cluster. Not all mapping between microglia and nuclei subsets were one to one: Both Micr.3 and Micr.4 contain nuclei that mapped to the living microglial interferon-response cluster. Micr.2 contains nuclei mapped to more than one cluster of living microglia annotated as activated - one defined as stress response and another as anti-inflammatory response - therefore Micr.2 captured at least two different reactive states of microglia cells that could be further partitioned when additional cells are available (as observed by their projection to two distinct domains on the UMAP, **Fig. 3c**). While there are differences in individual gene levels in RNA profiles from nuclei *vs*. living microglia (consistent with earlier reports[18,19], we detect expression of known markers of microglial subtypes in nuclear-derived profiles (**Fig. 3b**). Our nuclear-derived model is also consistent with earlier, lower-resolution models of nuclear-derived microglia subtypes, showing the consistency of data generated independently from frozen cortex[3–5,7]. Thus, the overall population structure of microglial cells[8] is well captured in our nuclear data and allows us to evaluate the inter-relation of microglia with other cell types.

The astrocytes (30,078 nuclei) included five major subsets, each with a distinct RNA profile (**Fig. 3e, Supplementary Fig. 3e-h**): homeostatic protoplasmic-like astrocytes (Astr.1, expressing SLC1A2, SLC1A3, MFGE8[2]), two non-homeostatic subsets (Astr.2 and Astr.3,

expressing reactive markers genes GFAP and OSMR [2]), fibrous-like astrocytes (Astr.4, expressing reactive marker GFAP and fibrous markers CD44[20] and endfeet marker AQP4[21]) and a small subset expressing interferon-response genes (Astr.5, *e.g.* IFI44L, IFI6) (**Fig. 3e,f, Supplementary Fig. 3f**), detected in samples from only two individuals. Interestingly, the AD susceptibility gene *APOE* [22] is upregulated in the homeostatic subset of astrocytes (Astr.1, **Fig. 3f**), while it is upregulated in non-homeostatic microglia cells (**Fig. 3b**). Genes induced in Astr.4 nuclei were enriched (FDR<0.01, hypergeometric test) in cholesterol metabolism and stress pathways, including heat shock response, oxidative stress and senescence (*e.g.*, ID3, **Fig. 3g**). Astr.4 and Astr.3 expressed genes recently reported in astrocytes enriched in a murine amyloid proteinopathy model[2,7] (ID3, VIM, FOS in Astr.4 and OSMR in Astr.3, **Fig. 3f**). However, Astr.3 profiles had higher expression of the FGF signaling pathway (FGF2 and its receptor NRP1, FDR<0.01), associated with a protective in an *in vitro* amyloid toxicity model [24] (**Fig. 3f**), and not expressed by Astr.4 nuclei (**Fig. 3f**).

Distinct populations of glial cells and a GABAergic neurons are associated with AD traits

We next turned to explore the association of cell subtype proportions with AD-related traits available from the deep ante-mortem and post-mortem characterization of ROSMAP participants[9–11]. We prioritized three outcomes that capture distinct critical aspects of AD: quantitative measures of (1) β-amyloid and (2) tau, the two defining pathologic characteristics of AD, as well as (3) the slope of aging-related cognitive decline over up to 20 years before death. This last measure captures the progressive cognitive impairment that leads to dementia. β-amyloid generally accumulates earlier than tau pathology but, while both pathologic features are found in the DLPFC of most ROSMAP participants, tau burden is more closely associated with cognitive impairment and dementia [25,26].

Although our sample size has limited statistical power, we found a clear trend of positive or negative correlations between the proportion of certain cell subsets (*within* each broad cell class) and cognitive decline (**Fig. 4a**) or measures of amyloid and tau accumulation that are defining characteristics of AD (**Supplementary Fig. 4a**). For example, the proportion of the Olig.1 subset among oligodendrocytes and of the Inh.5 subtype among inhibitory neurons are both negatively correlated with cognitive decline (r=-0.56 and -0.3 respectively), while Olig.3 and Inh.3 proportions are positively correlated (r=0.24 and 0.15 respectively) (**Fig. 4a,b**), yet more samples are required for significant statistical validations.

Deconvolution of cell subtype composition in bulk RNAseq profiles of 640 individuals validates association of cell populations with AD traits

Given that our 24 profiled participants were selected to capture different archetypes, they are not representative of the distribution of traits in aging brains captured by the parent ROSMAP studies; there is a lot of heterogeneity in the clinicopathologic characteristics of the ROSMAP participants. We thus used the single nucleus profiles to infer cell proportions in bulk DLPFC RNA profiles from 640 ROSMAP participants (**Supplementary Table 1**). To this end, we developed **CelMod** (Cellular Landscapes Modeling by Deconvolution), a method that builds and validates a model of cell type and cell states proportions from matched sc/snRNA-seq and bulk RNA-seq from the same individuals (**Fig. 1a**). CelMod relies on a consensus of gene-wise regression models, with cross validation to estimate accuracy (**Methods, Fig. 1**). Unlike most methods, CelMod is sensitive enough to detect the proportions of the different cell subsets (subtypes and states) *within* each broad cell class, and it can also deconvolve the relative contribution of expression programs (topics) in bulk profiles using program-specific weights (as opposed to proportions of discrete cell populations). We applied CelMod first at the broad cell class level and then subsequently within each class. CelMod identifies a large set of informative genes for each cell subset, ensuring that a small set of overlapping gene markers from different cell groups are not skewing the proportion estimates for cell types as well as for subsets within each cell class. Ultimately, this resulted in inferred relative proportions of each of the 35 cell subsets and 4 oligodendrocyte topics in the bulk DLPFC RNA-seq data from each of the 640 ROSMAP participants (**Supplementary Table 4**).

CelMod accurately inferred proportions of cell types, subsets and programs (topics) as validated by comparing the inferred to the matched empirically defined proportions from snRNA-seq data in a four-fold cross-validation (mean r=0.79 , stdev=0.13 across cell types and mean r = 0.78, stdev = 0.15 across cell subsets *within* each cell type, **Fig. 4c,d, Supplementary Fig. 4b,c**, **Methods**). We also compared the inferred proportions to matched proportions derived from immunofluorescence data on DLPFC tissue sections from 48 ROSMAP participants (**Fig. 4e,f**) that have both data: we compared proportions of neurons (NeuN+), astrocytes (GFAP+) and microglia (IBA1+) (**Methods**). The estimated proportions correlated well with histology-derived proportions for neurons (r=0.45) and microglia (r=0.34, **Fig. 4e**) but less strongly for astrocytes (r=0.18). Re-examination of the staining of the astrocyte marker GFAP, showed that the correlation is higher (r=0.5) with the inferred

proportions of the of GFAP+ astrocyte subsets (Astr.2, Astr.3 and Astr.4) out of all other cells (**Fig. 4e**).

Testing for association of the three AD-related traits with the inferred cellular proportions (*within* each broad cell class) was performed using linear regression (with correction for age, RIN score and sex, **Methods**, **Supplementary Table 5**) and revealed significant associations (FDR<0.05) between the proportions of neuron and glial cell subsets (*within* each cell class) and pathologic and/or cognitive traits (**Fig. 4g-i**). Interestingly, variation in proportions of many cell subsets were significantly associated with β-amyloid pathology (FDR<0.05), including negative association with most (5 out of 9) subtypes of glutamatergic neurons, inhibitory neuron subtypes Inh.5/13, and oligodendrocyte topics Olig.2/4, as well as positive association with non-homeostatic microglia states (Micr.3/4/5) and oligodendrocyte topic Olig.3 (**Fig. 4h**). However, only a minority (19%) of these β-amyloid-associated populations were also significantly associated with cognitive decline (*e.g.* Olig.3, Micr.3 and Inh.5, **Fig. 4g**). In contrast, cell subsets and topics associated with tau pathology also showed significant associations to cognitive decline (FDR<0.05, **Fig. 4h**); these include a positive association with cognitive decline for Olig.3, Inh.3 and Micr.3 and a negative association for Olig.1 and Inh.5. This overall finding is consistent with the stronger association between tau (vs. β-amyloid) pathology and cognitive decline. β-amyloid pathology, while correlated with Tau pathology (r=0.48), has a broader range of cellular associations, only some of which are also correlated with tau pathology. So, the two pathologies have a largely distinct set of associations with cell subtypes/states, consistent with our earlier report of distinct microglial transcriptional programs being associated with amyloid and tau pathology[27].

The inferred associations with cognitive decline and tau-pathology showed that oligodendrocyte topic Olig.1 appears to be more prominent in non-impaired individuals while topic Olig.3 is enriched in individuals with cognitive decline. In parallel, the relative proportion of inhibitory neuronal subtype Inh.5 (SST+) is higher in healthy individuals, in contrast to inhibitory subtype Inh.3 (PV+), which is relatively higher in individuals with cognitive decline (**Fig. 4f-h**). This suggests a potential difference in the vulnerability of SST+ vs. PV^+ GABAergic neurons in AD. Additional subsets positively associated with cognitive decline include Micr.3 which is enriched for genes related to reactive oxygen pathways and interferon response (**Fig. 3**) and the endothelial subset Endo.3 (**Fig. 4f-h**).

Inter-individual variation in composition of cell states and subtypes is confirmed in 640 individuals

Assessing the extent of inter-individual variation in cell subsets (subtypes, states and programs) composition across the 24 subjects revealed differences *within* each of the different broad cell classes separately (**Fig. 5a**), which was beyond the association of specific cell subsets to the AD subgroup of individuals. For example, among the astrocytes, most (70.83%) of the individuals had predominantly homeostatic astrocytes (Astr.1 and Astr.5), and the remainder predominantly had astrocyte from one of the non-homeostatic subsets (Astr.2 & Astr.3) but not both (**Fig. 5a**). Similarly, among microglial cells, individuals were distinguished by having either a predominant homeostatic signature (Micr.1, 58.3%) or an active stress/anti-inflammatory associated signature (Micr.2) (**Fig. 5a**). A similar but less pronounced trend of a predominant signature per individual was present in oligodendrocytes and endothelial cells (**Fig. 5a**). Finally, while all neuronal subtypes are detected in all individuals, we found differences in their relative proportions across groups of individuals (**Fig. 5b**), including a systematic anti-correlation within the proportions of the SST+ (Inh.5) and PV+ (Inh.3) GABAergic subtypes (as a fraction of all GABAergic neurons, **Fig. 5b**).

To assess the robustness of these correlated differences in composition of cell subsets across individuals, we used the CelMod-inferred proportions of each of the 31 cell subsets and 4 oligodendrocyte topics from the bulk DLPFC RNA-seq across 640 individuals (**Supplementary Table 4**). We observed the same co-variation in broad cell class proportions across 640 individuals (**Fig. 1e, Supplementary Fig. 4d**), which we then confirmed by immunohistochemistry across 48 individuals (**Supplementary Fig. 4e**). We also confirmed the variation in cell subsets across individuals *within* each broad cell class separately (**Fig. 5c,d**, **Supplementary Fig. 5a**). For example, we again observed the mutually exclusive predominance of homeostatic (Astr.1) *vs*. non-homeostatic (Atsr2 and Astr.3) astrocyte signatures (R=-0.95), and of homeostatic (Micr.1) *vs*. non-homeostatic (Micr.2) microglial subtypes (R=-0.88) in the 640 individuals (**Fig. 5c,d**). Similarly, we continued to observe a strong anti-correlation between Olig.1 and Olig.3 oligodendrocytes topics (R=-0.81) as well as between the relative proportions of PV+ and SST+ inhibitory neuronal subtypes as a fraction of total GABAergic neurons (R=-0.58) (**Fig. 5d**).

Correlated proportions of cell subsets across individuals suggest the existence of distinct cellular communities in aging brains

We next asked how the variation across all the individuals in the proportion of each cell subsets was correlated with variation in other cell subsets, particularly those from a different broad cell class. To this end, we calculated the Spearman correlation coefficient for each pairwise combination of cell subset proportions, and clustered the subsets into communities by these correlations (**Fig. 6a, Methods**). The cell subsets exhibited significantly similar co-variation structures across both the 24 participants with empirically determined proportions (**Fig. 6b**) and across the 640 participants with inferred proportions (**Fig. 6c**), as assessed by a permutation test (**Fig. 6d**, p-value<0.001, **Methods**). A network of cell subsets, connecting each pair of subsets (nodes) with significant correlations and anti-correlations (edges) was also found to be similar between the 24 participants network (**Fig. 6e**) and 640 participant network (**Fig. 6f, Supplementary Fig. 5b,c**).

These networks identified five cellular communities (groups of positively correlated cell subsets), defined by community detection using eigenvector matrices[28] on all (significant) positive edges (**Fig. 6g**, **Methods**) across the 640 participants (and matched the 24 participant network, **Supplementary Fig. 5b,c**). These included two pairs of coupled, anti-correlated communities. The first pair includes a putative **homeostatic community** (*Community 1*, mean intra-community R=0.73, stdev = 0.13), composed of a microglia (Micr.1), astrocyte (Astr.1), and endothelial (Endo.1) subsets; and a putative **non-homeostatic community** (*Community 2,* mean intra-community R=0.67, stdev = 0.15), composed of a microglia (Micr.2) and two astrocytes (Astr.2, Astr.3) subsets. These two communities are strongly anti-correlated (mean inter-community R=-0.79, stdev = 0.06) with one another across individuals. The second anti-correlated pair of communities consists of two mixed **neuronal-glial communities**, *Community 3* (including Olig.1, Inh.5/4/9/10, Exc.0/1, Astr.4/5, and Endo.2 subsets) and *Community 4* (including Olig.3, Micr.3, Inh.3, and End.3/4 subsets). In addition, there exists a community of **predominantly glutamatergic neurons** (*Community 5*, mean intra-community R=0.57, stdev = 0.23). Notably, the representation of neuronal-glial community 3 was correlated with homeostatic community 1 (mean inter-community R=0.17, stdev = 0.14) across individuals, whereas neuronal-glial *Community 4* was correlated with both non-homeostatic-community 2 (mean inter-community R=0.21, stdev = 0.07) and glutamatergic community 5 (inter-community mean R=0.28, stdev = 0.27). Although there were minor differences between the community networks from the 24 and 640 participants (**Supplementary Fig. 5b,c**), we focused on their shared elements in all subsequent analyses.

Distinct ligand-receptor pairs identified in the homeostatic and non-homeostatic cellular communities

We hypothesized that the anti-correlation between homeostatic *community 1* and non-homeostatic *community 2* (**Fig. 6h**), may reflect distinct signaling patterns between cell subtypes *within* each community[29,30]. We thus searched for ligand-receptor pairs (LRPs) [31,32] that putatively connect the microglia and astrocyte signatures *within* each of the two communities (Micr.1 with Astr.1, or Micr.2 with combined Astr.2/Astr.3) (**Methods**). We identified 452 *expressed* LRPs (2.05% of all possible LRPs, **Methods**) in a cellular community, with at least one cell subset expressing the ligand and one expressing the receptor within the community (**Methods**). The number of expressed LRPs distributed equally across all possible cell subsets ($p = 0.13$, Chi Square test). Next, we searched for *community-specific* LRPs, defined as expressed LRPs in which either the ligand, the receptor, or both, were differentially expressed within a cell type between the two communities (FDR<0.01, comparing expression in Micr.1 to Micr.2, and Astr.1 to combined Astr.2/Astr.3; **Methods**). We identified 25 *community-specific* LRPs (**Fig. 6h-j**), considering, within each community, putative signaling from microglia to astrocytes, astrocytes to microglia, and auto-signaling. *Community-specific* LRPs were found between all pairs of cell populations, with no significantly favored pair of cell subsets ($p > 0.5$, Chi Square test). Yet, the overall number of *community-specific* LRPs within the non-homeostatic *community 2* was higher (21 LRPs) than in the homeostatic *community 1* (4 LRPs; **Fig. 6j**).

For example, within the non-homeostatic *community 2*, we found upregulation of Transforming Growth Factor Beta (TGF-β) signaling. The TGFB2 ligand is upregulated in Astr.2/Astr.3 (compared to Astr.1), and its receptors TGFBR1 and TGFBR3 are upregulated in Micr.2 (compared to Micr.1) and Astr.2/Astr.3 (compared to Astr.1, **Fig. 6g-i**), respectively. TGF-β signaling was shown to be upregulated during infection in the CNS in astrocytes [33], and it has a known anti-inflammatory effect [33]. There are also indications for its dysfunction in AD brains[34]. Within the homeostatic *community 1*, we find an up-regulation of APOE-associated signaling, with higher expression in Astr.1 of APOE, a major AD susceptibility gene [35], and its putative receptor SORL1 (based on previous reports [36–38]), a susceptibility locus for AD associated with cognitive decline levels [39], upregulated in homeostatic Micr.1 (**Fig. 6g-i**). An increase in SORL1 levels has been associated with reduced β-amyloid accumulation [34,40].

A mixed neuro-glial cellular community is associated with cognitive decline

We next asked whether the changes in the composition of cell subsets found to be associated with AD (**Fig. 4g-h**) reflect multiple independent effects that contribute to cognitive decline, or a coordinated set of changes. We found that certain cellular communities were significantly enriched for cell subsets associated with AD related traits (**Fig. 7a,b,c**), suggesting that these cellular populations may be engaged in a common disease-related process. In particular, the cognitive decline-associated cell subsets segregated into two anti-correlated communities, each composed of highly interconnected neuro-glial-endothelial cell subsets: *Community 4* includes most (4 out of 6) of the cell subsets that have a significant positive association with more rapid cognitive decline (Hypergeometric *p-value* < 0.001) and tau pathology load (4 out of 4, *p-value* < $9.4e^{-5}$) including the Olig.3 signature, Micro.3, Endo.3 and GABAergic PV+ neuronal subtype Inh.3 (**Fig. 4g,h, Fig. 7a,b**). Conversely, *Community 3* includes all four cell subsets whose relative proportions were significantly higher in individuals with little or no cognitive decline (*p-value* < 0.004), including the Olig.1 signature and GABAergic SST+ neuronal subtype Inh.5 (**Fig. 4g,h**, **Fig. 7a-b**). Finally, *Community 5,* was enriched for most (9 out of 10) cell subsets and programs negatively associated with β-amyloid load (*p-value* < $7.2e^{-6}$), including most glutamatergic neuronal subtypes (**Fig. 4i, Fig. 7c**). We therefore identify several sets of coordinated cellular responses that are associated with distinct aspects of AD, with Community 4 appearing to be associated with accumulation of both proteinopathies as well as cognitive decline.

The proportion of inhibitory neurons and oligodendrocytes subsets partially mediate the association between tau pathology and cognitive decline

While our autopsy-based cross-sectional data cannot formally determine the causal chain of events leading to AD, rigorous modeling can propose a most likely scenario that will inform the design of validation studies. The strongest common genetic risk factor of late onset AD, the *APOEε4* haplotype, explains 5.6% of the variance in cognitive decline in the ROSMAP participants (adjusted for age, sex and RIN values, **Methods**), and this effect size is comparable to the 5.3% variance explained (VE) by the frequency of Inh.5 (SST+ neurons) and the 4.7% VE explained by the Olig.1 topic. Micr.3, Inh.3 and Olig.3 have slightly smaller VE: 3.3%, 2.8% and 2.7%, respectively. Thus, the cell subtypes that we have identified explain a substantial fraction of the variance of cognitive decline, but this analysis does not in itself provide any inference as to the sequence of events.

To model the sequence of events from risk factors to dementia, we conducted mediation analyses on the variability of the proportions of the two most significant subsets associated with cognitive decline: Olig.3 (positive association) and Inh.5 (SST+, negative association). We first leveraged a Mendelian randomization approach [41] using *APOEε4* as an instrumental variable to infer the causal relationship between tau pathology and cell subset proportions. *APOE*ε4 is a genetic risk factor that is not affected by reverse causation, and has a known strong association with tau pathology. *APOE*ε4 was also associated with both Inh.5 (β=-7.9×10-3, 95% CI -0.015 to -6.4×10-4, p=0.033) and Olig.3 (β=0.020, 95% CI 0.002 to 0.038, p=0.028), but these associations were no longer significant once we adjusted for tau pathology (Inh.5 p=0.31, Olig.3 p=0.24). By contrast, the association between *APOEε4* and tau pathology remained similarly strong even after adjusting for the relative proportions of Inh.5 and Olig.3 (before adjustment: β=0.63, 95% CI 0.42 to 0.84, $p=4.6\times10^{-9}$; after adjustment: β=0.58, 95% CI 0.38 to 0.79, p=3.8×10-8), ruling out the scenario that the *APOEε4* – tau link is mediated by changes in frequency of these two cell subsets. Thus, *APOEε4* is a valid instrumental variable to assess the causal relationship between tau pathology and cell type proportions. Based on these results, we performed two Mendelian randomization analyses for each cell subset (Inh.5 and Olig.3) separately, with APOE-ε4 as an instrumental variable, tau as an exposure of interest, and Inh5 or Olig3 as the outcome of interest (**Methods**, **Fig. 7d**). The results are consistent with a model in which there is a significant causal effect of tau pathology on cell subset proportions. We note that Inh.5 and Olig.3 were highly collinear (r=-0.76), and thus we could not statistically assess whether the proportion variation of one cell subset is upstream of the other.

Next, we examined whether the downstream variation in proportions of cell subsets, mediate the relationship between tau pathology and cognitive decline. We performed mediation analysis for each cell subset (Inh.5 and Olig.3). In the resulting causal models, cell subset proportions partially mediated the association between tau pathology and cognitive decline: both Inh.5 and Olig.3 proportions differences mediate a small but significant part of the association between tau pathology and cognitive decline (**Fig. 7e**). Thus, our analysis suggests that differences in Inh.5 and Olig.3 proportions may reflect some of the consequences of tau pathology accumulation that lead to impaired cognition. This model is consistent with the current biomarker-based model of human AD progression, in which

neuropathology precedes alterations of the brain parenchyma (such as neurodegeneration), which is then followed by cognitive decline and functional deficits [42].

DISCUSSION

In this study, we constructed a high-resolution cellular map of the DLPFC from 24 individuals, which enabled the characterization of the diversity of neuronal, glial and endothelial cell subsets and topic models at unprecedented resolution in order to extract new insights about intra- and inter-individual diversity in the aging brain. Our optimized snRNA-seq protocol and analytic pipelines uncovered a greater diversity of cellular populations in the aging neocortex (**Fig. 2 and 3**) than has been previously reported [3,7]. Further, using a new method (CelMod), we integrated this detailed cellular map with bulk RNA-seq data from the same brain region in the same 24 individuals, to estimate the proportions of the major cell classes and cell subsets in a cohort of 640 aging individuals. The general approach can be readily applied to other matched tissue-level datasets. The inferred cellular composition data was statistically powered for disease association analyses, which uncovered key players with potential involvement in the cellular cascade leading to cognitive decline, including an anti-correlated pair of GABAergic neuronal subtypes as some of the most prominent findings (**Fig. 4 and 7**). Specifically, our analysis suggests a striking inverse association between PV+ (Inh.5) and SST+ (Inh.3) GABAergic neuron proportions in healthy and diseased participants, suggesting that SST neurons may be more vulnerable or PV neurons relatively more resilient to tau pathology (**Fig. 4 and 7**). Such changes were not reported in previous studies [3,7], which have profiled a lower total number of neurons. Next, we applied a new computational framework to expand our analysis across all cell types, which led to the discovery of distinct *cellular communities* (**Fig. 6**) that we linked to AD associated traits (**Fig. 7**).

In microglia, we mapped our snRNA-seq-based clusters to those found in single-cell analysis of live human microglia profiled from fresh autopsy and surgically resected tissue [8], finding a good match with only few differences in subtype composition between the two experimental approaches. Although the numbers of microglial nuclei in our single nucleus data set remains small (<4,000 nuclei), we nonetheless identify clearly distinct signatures of activated microglia, which was not necessarily guaranteed with a snRNA-seq approaches given the low and unique content of nuclear RNA that have been previously noted [18]. Oligodendrocytes emerge as an interesting cell class for further evaluation. We highlight the need to deploy a

distinct analytic approach when studying this cell type: We represented the heterogeneity within these cells utilizing a continuous approach by topic models [13–16] to capture expression programs instead of discrete cell clusters. One such topic (Olig.3) is very strongly associated with tau pathology and cognitive decline, while another (Olig.1) is prevalent in cognitively non-impaired brains, suggesting a clear oligodendrocytic signature associated with AD.

β-amyloid pathology was associated with a much broader set of cell subsets than tau pathology, yet most of the variability in β-amyloid does not appear to relate directly to cognitive decline. This can help de-prioritize those cell types, subtypes and expression programs that may not contribute to clinical symptoms, namely cognitive decline. Since most studies focus on a pathology-based diagnosis of AD and/or AD dementia, or a low-resolution measure of pathology such as Braak score, it is challenging to resolve the effects of the distinct pathologies involved in the course of AD. Our extensive, high resolution profiling of deeply characterized individuals provides a more informed prioritization of candidate gene signatures and cellular subsets for further investigation.

Finally, a major innovation of our study is the definition of cellular *communities* defined by correlated changes in the frequency of different cell subsets across individuals (**Fig. 6 and 7f**). The correlation structure initially observed in our snRNA-seq data from 24 individuals (**Fig. 6b,e**) was reproduced in the 640 ROSMAP participants who have inferred cell type proportions (**Fig. 6c,f**). Our modeling further shows that the composition of cell subsets and topics in these communities may reflect an underlying shared process, which is supported by finding *community-specific* ligand-receptor pairs(**Fig. 6h-j**), such as (TGF-β) signaling that is enhanced in *Community 2* (**Fig. 6i**). Moreover, this community structure is clinically relevant as *Community 4* is enriched for associations with both proteinopathies and cognitive decline (**Fig. 7a-c**). These results fit within our conceptual understanding that AD is a distributed pathophysiologic process involving multiple interacting cell types. Spatial transcriptomics methods will help to resolve whether these communities represent groups of co-located cells or distributed communities responding to an underlying, shared signal.

Finally, because profiling post-mortem brain tissue is by definition a cross-sectional study, we have to infer the temporal links between cellular communities, appearance of pathology, and cognitive symptoms. Mediation analysis suggests that the strong known link between tau pathology and cognitive decline is mediated in part through the cell type proportion changes that we observed, specifically through an oligodendrocyte state changes and and differences

in SST+ and PV+ interneuron proportions (**Fig. 7d,e**). Our modeling thus presents a clear hypothesis that the observed alterations in cell frequency occur downstream of the accumulation of tau pathology but upstream of cognitive deficits (**Fig. 7d,e**). This can help the design of future studies to validate these observations and, later, potential avenues to mitigate cognitive decline by targeting the molecular processes leading to these cellular changes.

Overall, our work highlights the importance of a unified, cross-cell type view of the cellular ecosystems of the brain, beyond a cell-type-specific focus, in the study of Alzheimer's Disease and other complex neurodegenerative disorders. Embracing the complexity of this heterogeneous parenchymal tissue, network approaches can uncover new insights, such as key members of each cellular community that are involved in different aspects of the aging brain.

Figures and Legends

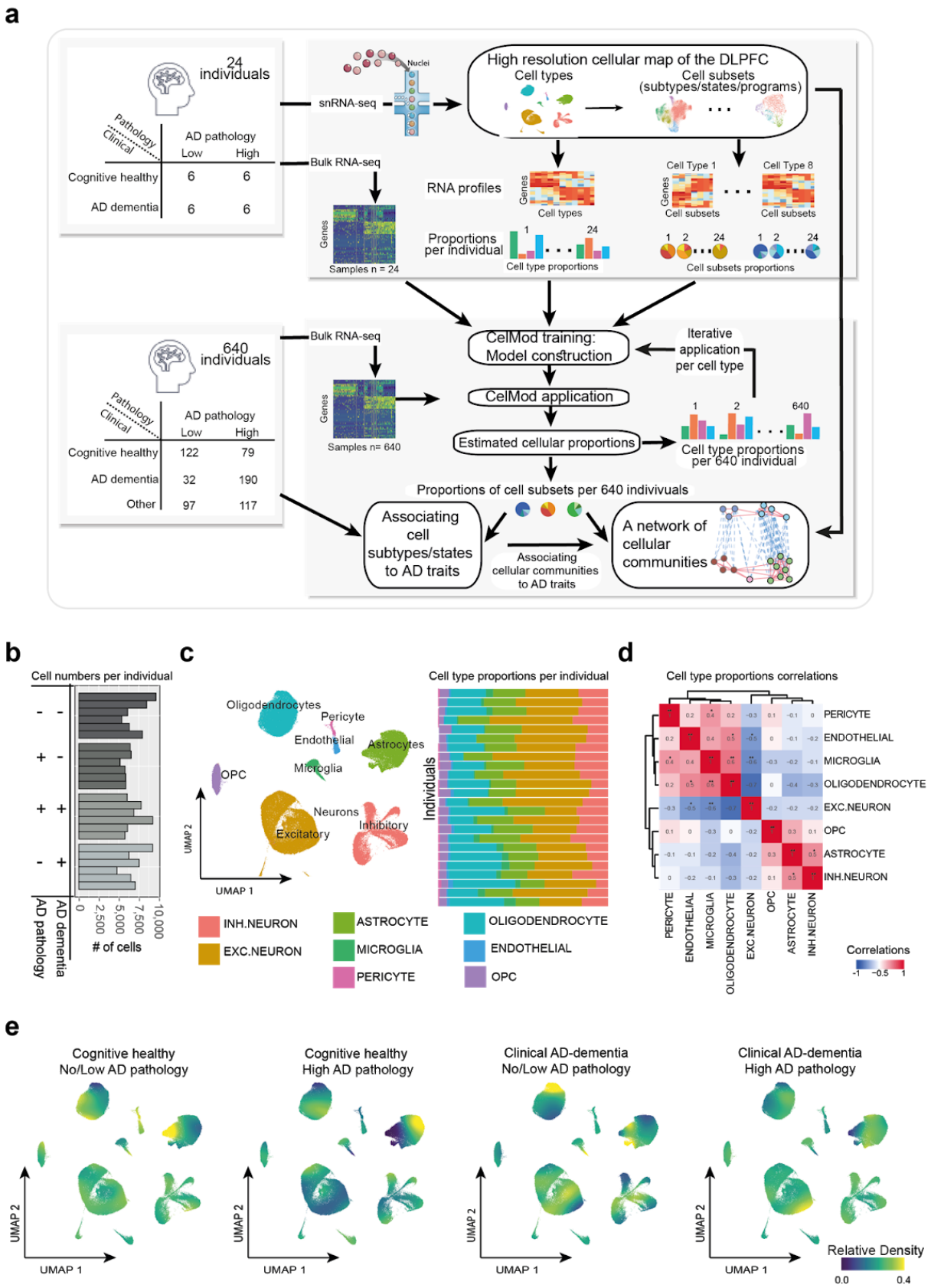

Figure 1. A cellular-molecular map of the human aging DLPFC shows coordinated changes in cell type proportions across individuals. (**a**) Overview of the experimental scheme and analysis. 24 individuals with clinicopathologic characteristics were profiled by single nucleus RNA-seq (snRNA-seq) to generate a high resolution *cellular map* of the aging DLPFC brain region, and used as

input to our CelMod deconvolution algorithm to iteratively estimate *cellular compositions* in 640 individuals. Network analysis uncovered *cellular communities*, cell subsets coordinately varying across individuals, and statistical modeling *associated AD traits* to cell subsets and to cellular communities. (**b**) High quality 162,562 libraries generated across 24 post-mortem samples of the DLPFC brain region of aging individuals. The number of cell profiles for each individual, divided by the four major archetypes of the aging population: No clinical and minor pathological AD, pathological AD only, clinical & pathological AD, and clinical AD only. (**c**) A detailed cellular-molecular map of the human aging DLPFC. Left: Umap embedding of 162,562 single-nuclei RNA profiles from the DLPFC brain region of 24 individuals; colored by cell type (or by cluster in **Supplementary Fig. 1b**). Right: Diversity of cell type proportions across individuals. The proportion of cell types (color coded as in the Umap) for each individual (columns). (**d**) Inter-individual variations in cell type proportions across individuals is correlated across cell types. A heatmap of the pairwise Spearman's correlation coefficient of cell type proportions across 24 individuals. (**e**) Cellular landscapes diverge between groups of individuals from four archetypes of the aging population. Umap embedding (as in **c)**, split to the four major archetypes of the aging population (as in **b**), colored by the local relative density of cells from each archetype (**Methods**).

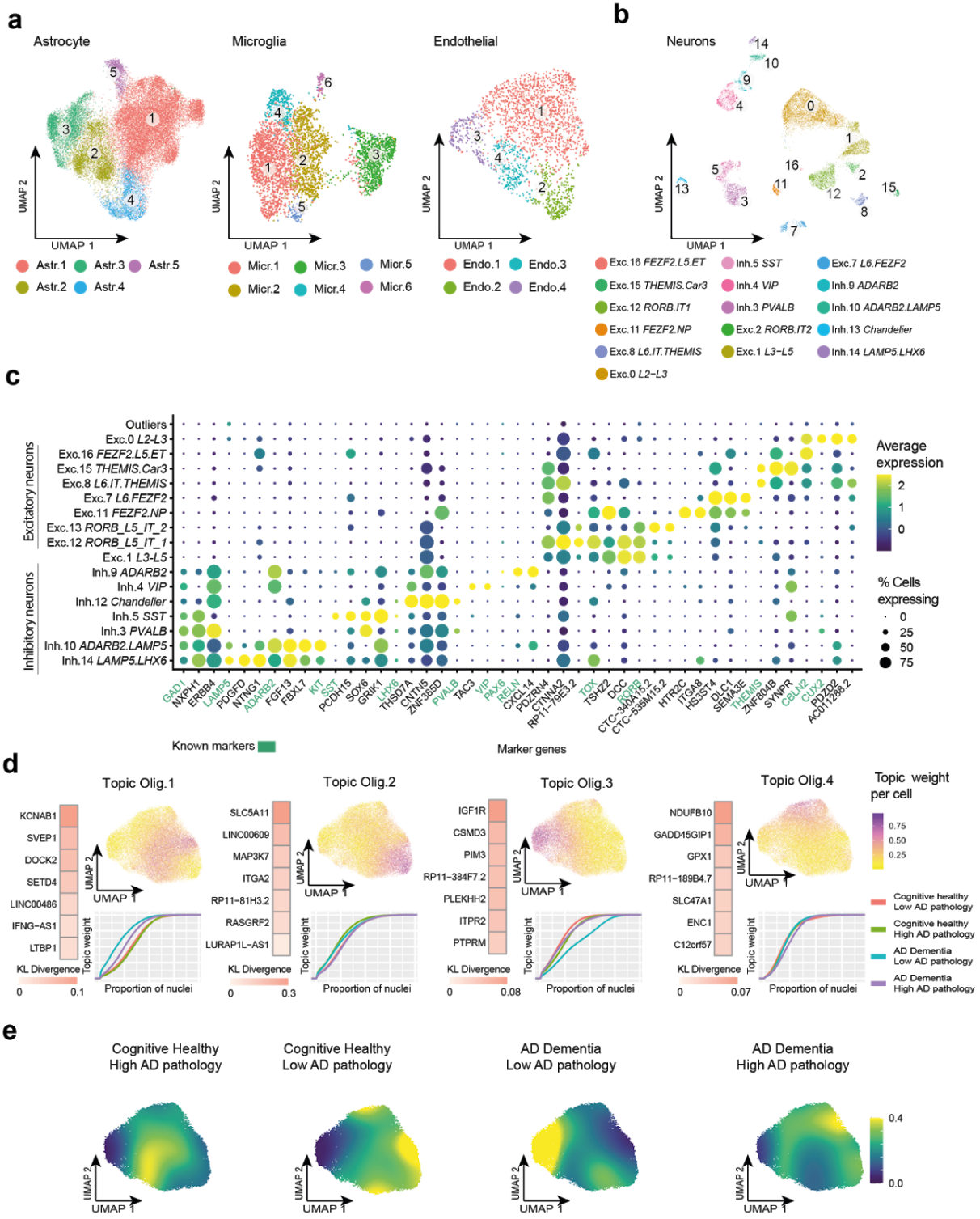

Figure 2. A high resolution map of cell subtypes and states in the aging DLPFC. (**a**) Diversity of non-neuronal cell states in the aging DLPFC. Umap embedding of individual cell types, colored by clusters capturing distinct cell states in (from left to right): astrocytes (30,078 cells, 5 clusters), microglia (3,916 cells, 6 clusters), and endothelial cells (1,988 cells, 4 clusters). (**b**) Neuronal diversity in the DLPFC. Umap embedding of excitatory (58,359 cells, 9 clusters) and inhibitory neuronal subtypes (24,805 cells, 7 clusters), colored by clusters capturing distinct subtypes. Annotated by Allen Brain Atlas [12] (**Methods**). (**c**) Distinct marker genes per neuronal cluster. Mean expression level in expressing cells (color) and percent of expressing cells (circle size) of selected markers in

each cluster of neuronal subtypes (rows) of novel marker genes and known marker genes (marked in green, from Allen Brain Atlas [12]), including pyramidal neurons across cortical layers and GABAergic subtypes. (**d**) Diverse continuum of expression programs in oligodendrocyte cells inferred by topic modelling [13–16]. For each topic model: Umap embedding of oligodendrocytes cells, colored by the weight of each topic per cell (right); The top scoring genes (colored by the score), computed as the Kullberg-Leibler divergence between the expression level and the topic's weight across cells (left); and The cumulative distribution function of topic weights for cells split by the sample of origin to four major archetypes of the aging population (as in **Fig. 1e**, bottom). (**e**) Oligodendrocytes composition diverge between four archetypes of the aging population. Umap embedding of oligodendrocyte cells (as in **d**), split to four major archetypes of the aging population (as in **Fig. 1e**), and colored by the local relative density of cells from each archetype (**Methods**).

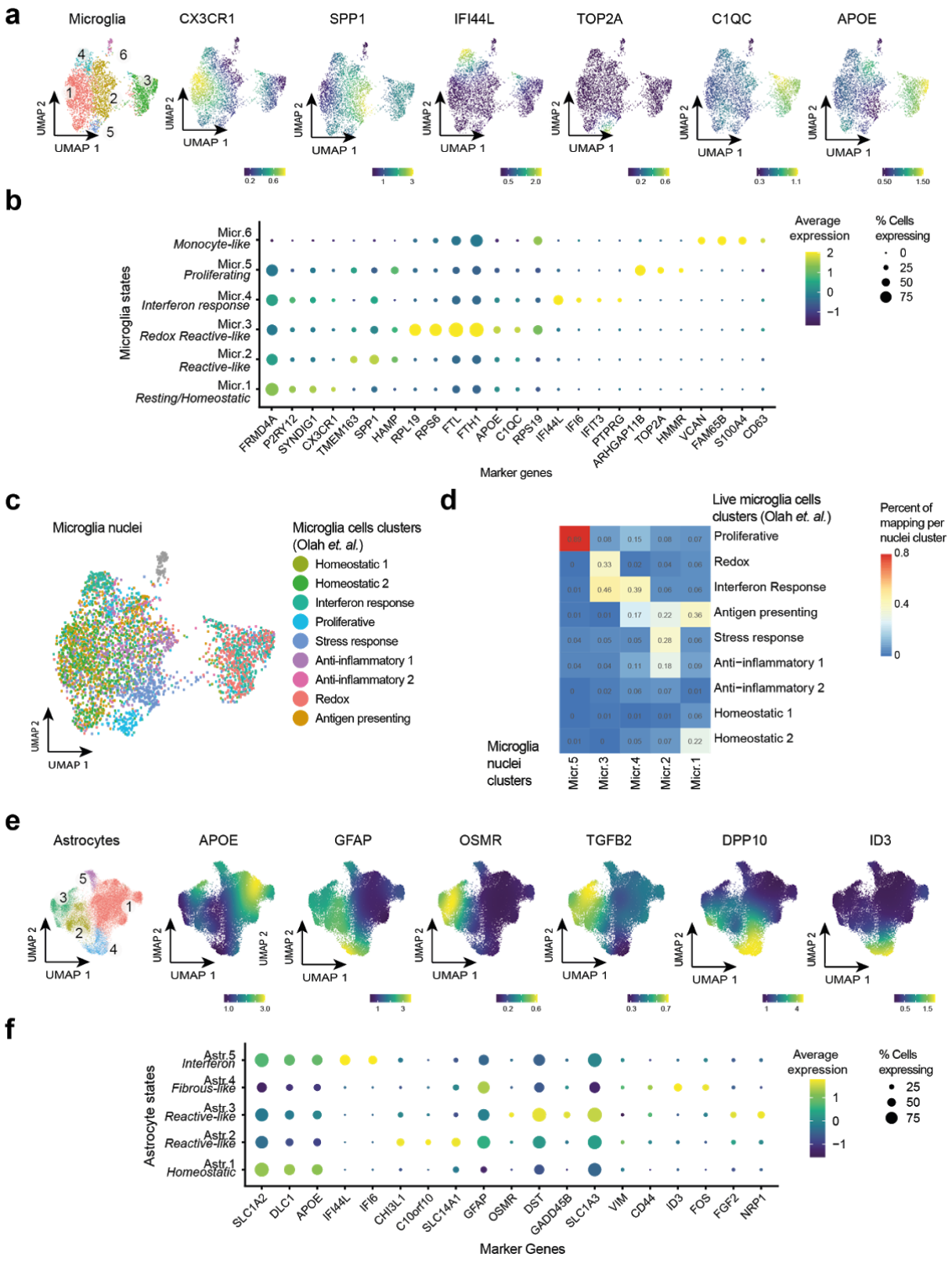

Figure 3. Diverse cell states in astrocytes and microglia cells. (**a**) Microglia subclusters are associated with distinct marker genes. Umap embedding of microglia cells colored by cell cluster (as in **Fig. 2b**, top left), or by the local density of expression levels (**Methods**) of marker genes for: microglia activation (SPP1), interferon signaling (IFI44L), AD-risk (APOE) and proliferation (TOP2A). (**b**) Dot plot of the mean expression level in expressing cells (color) and percent of expressing cells (circle size) of selected marker genes across microglia subsets (rows). (**c-d**) Clusters

of microglia nuclei from snRNA-seq match published live microglia cell clusters from scRNA-seq [8]. Umap of snRNA-seq microglia colored by the best predicted scRNA-seq cluster ID (**Methods**, **c**). The proportions (color scale, scaled per column) of nuclei per cluster (columns) mapped to each scRNA-seq cell cluster (rows, **d**). (**e**) Diverse astrocyte subsets are associated with distinct marker genes. Umap embedding of astrocyte cells, colored by subcluster (as in **Fig. 2b**, top left), or by the density of expression levels of marker genes (**Methods**): astrocyte activation marker (GFAP), senescence regulator (ID3), AD-risk gene (APOE). (**f**) Dot plot showing the mean expression level in expressing cells (color) and percent of expressing cells (circle size) of selected genes in each astrocyte cluster.

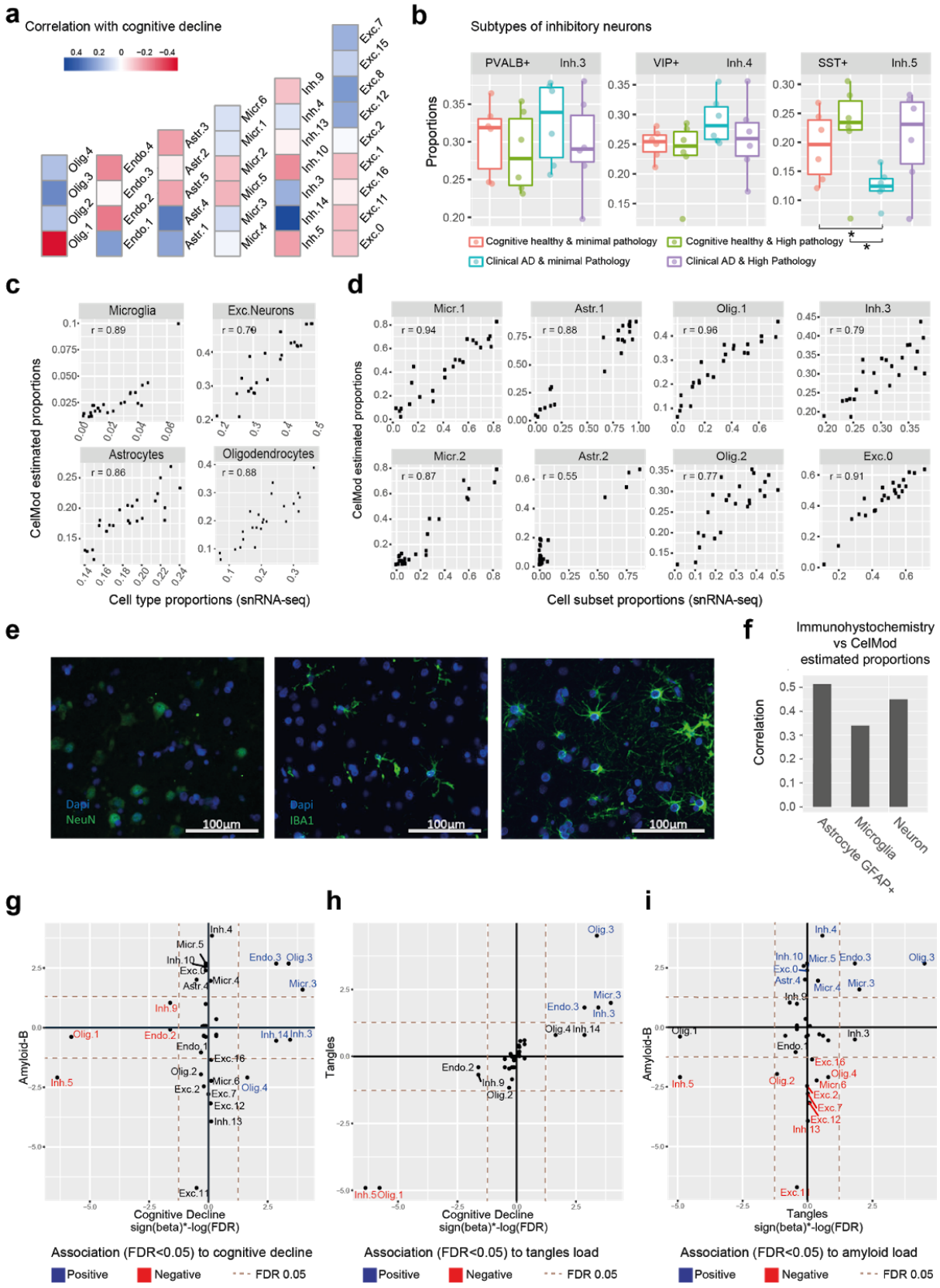

Figure 4. The proportion of neuronal subtypes and glial subsets are associated with AD-traits in a cohort of 640 individuals. (**a**) Proportions of cellular subsets within cell classes correlate with cognitive decline. Correlation (color scale) of proportions of cell subsets *within* each cell type (individual columns) to cognitive decline. Associations to other AD traits in **Supplementary Fig. 4a**.

(**b**) GABAergic subtypes proportions vary in association with cognitive decline. Box plots of the proportion of three GABAergic subtypes (out of total GABAergic neurons) across four cohorts: reference group, pathological AD only, clinical AD only, and clinical with pathological AD. Box, 25% and 75% quantiles; line, median; Dots, individual samples. (**c**) Estimated cell type proportions by CelMod algorithm match snRNA-seq data. Scatter plots of the estimated proportions (Y-axis) compared to the snRNA-seq measured proportions (X-axis) across the 24 individuals, for different cell types: microglia, excitatory (Exc) neurons, astrocytes and oligodendrocyte cells (other cell types in **Supplementary Fig. 4b**). (**d**) Estimated cell subtypes and states proportions by CelMod algorithm match snRNA-seq data. Scatter plots as in C, for selected cell subsets (all other subsets in **Supplementary Fig. 4c,d**). (**e-f**) Cell type proportions measured by histology match CelMod estimates. (**e**) Immunohistochemistry in DLPFC sections of 48 individuals (24 healthy, 24 declined), stained for marker for (GFP): neurons (anti-NeuN), microglia (anti-IBA1), astrocytes (anti-GFAP). Representative immunofluorescence images. DAPI, nuclei. Scale bar = 100 μm. Correlation of (**f**) Pearson correlation coefficient of CelMod and immunofluorescence estimations of proportion of cell populations (out of the total number of cells), for: neurons, microglia, and GFAP+ astrocyte populations (Astr.2, Astr.3, Astr.4). (**g-i**) Association of cell population to AD traits. Association scores for the CelMod estimated proportions across 640 individuals of all cell subsets (cell subtypes, states or topic models) to cognitive decline rate (X-axis, in **g** and **i**), β-amyloid load (Y axis, in **g** and **i**), tangle load (Y-axis in **h**, X-axis in **i**). Association score = -log(FDR)*sign(beta), from multivariable linear regression analysis correcting for age, sex and RIN values. Statistically significant subsets with positive association (red) or negative association (blue) with cognitive decline (in **g,h**) or β-amyloid load (in **c**).

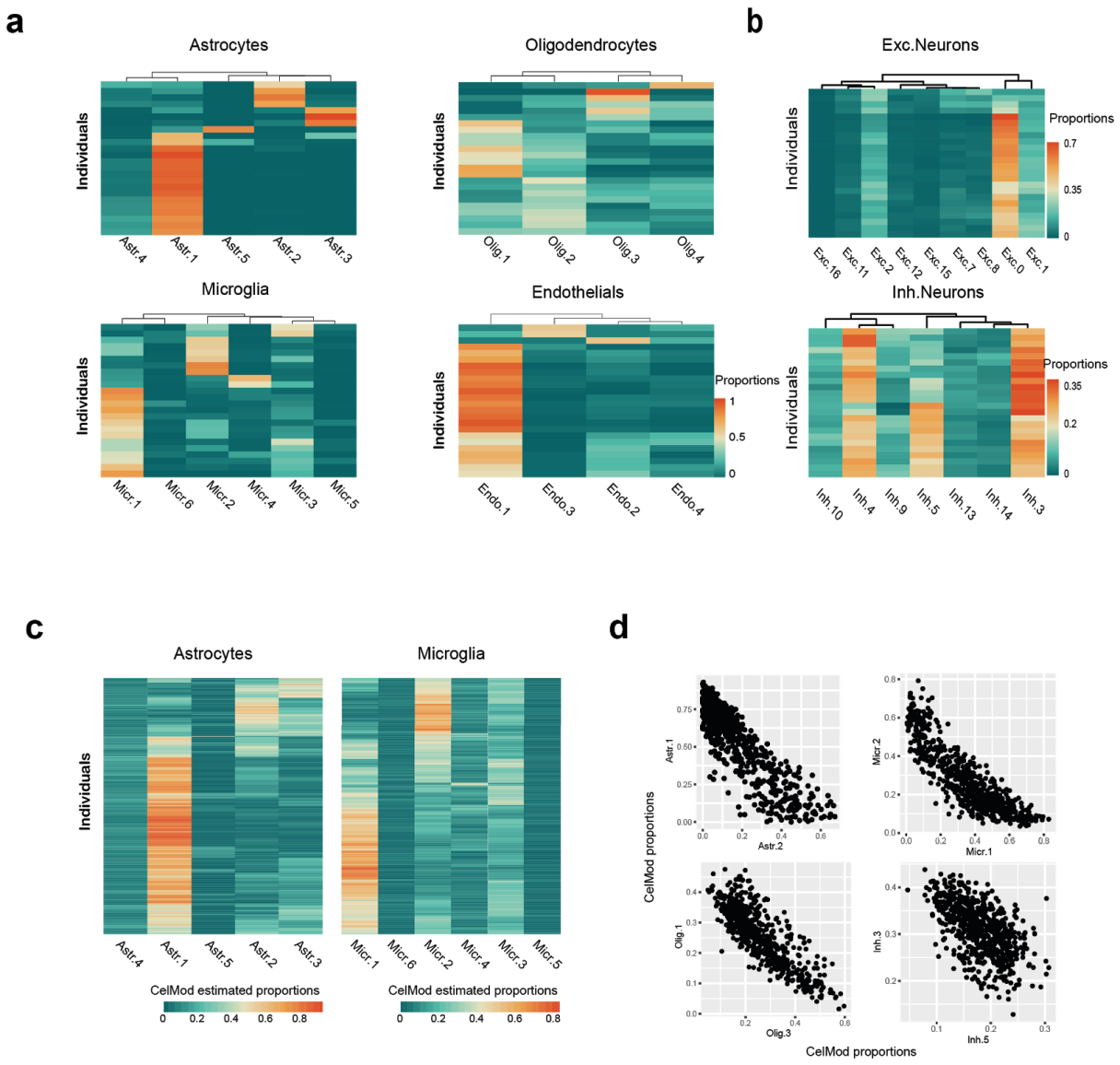

Figure 5. Inter-individuals diversity of cell subsets within each cell type validated in 640 individuals. (**a**) Diverse proportions of cell subsets across individuals in glial and endothelial cells. The frequency (out of total cells in the class, color scale) of each cell subset (columns) in each individual (rows; n = 24) across different cell classes: Astrocytes, microglia, oligodendrocytes and endothelial cells. (**b**) Diverse proportions of neuronal subtypes across individuals. Frequency of neuronal subtypes, (presented as in **a**) in excitatory and inhibitory neurons. (**c-d**) CelMod estimated cellular proportions in 640 individuals validate changes in cell states across individuals observed in snRNA-seq data. (**c**) The frequency of each cell subset in each individual (n = 640) across astrocytes and microglia. The cell subsets (columns) are ordered as in **a**. (**d**) Scatter plots of selected pairs of cell subsets from the same cell class, confirming high anti-correlations between subsets of astrocytes (Astr.1 and Astr.2), microglia (Micr.1 and Micr.2) and oligodendrocytes (Olig.1 and Olig.3), and inhibitory neuronal subtypes (Inh.3 and Inh.5). Full estimated proportions of cell types and cell states in **Supplementary Fig. 5a**.

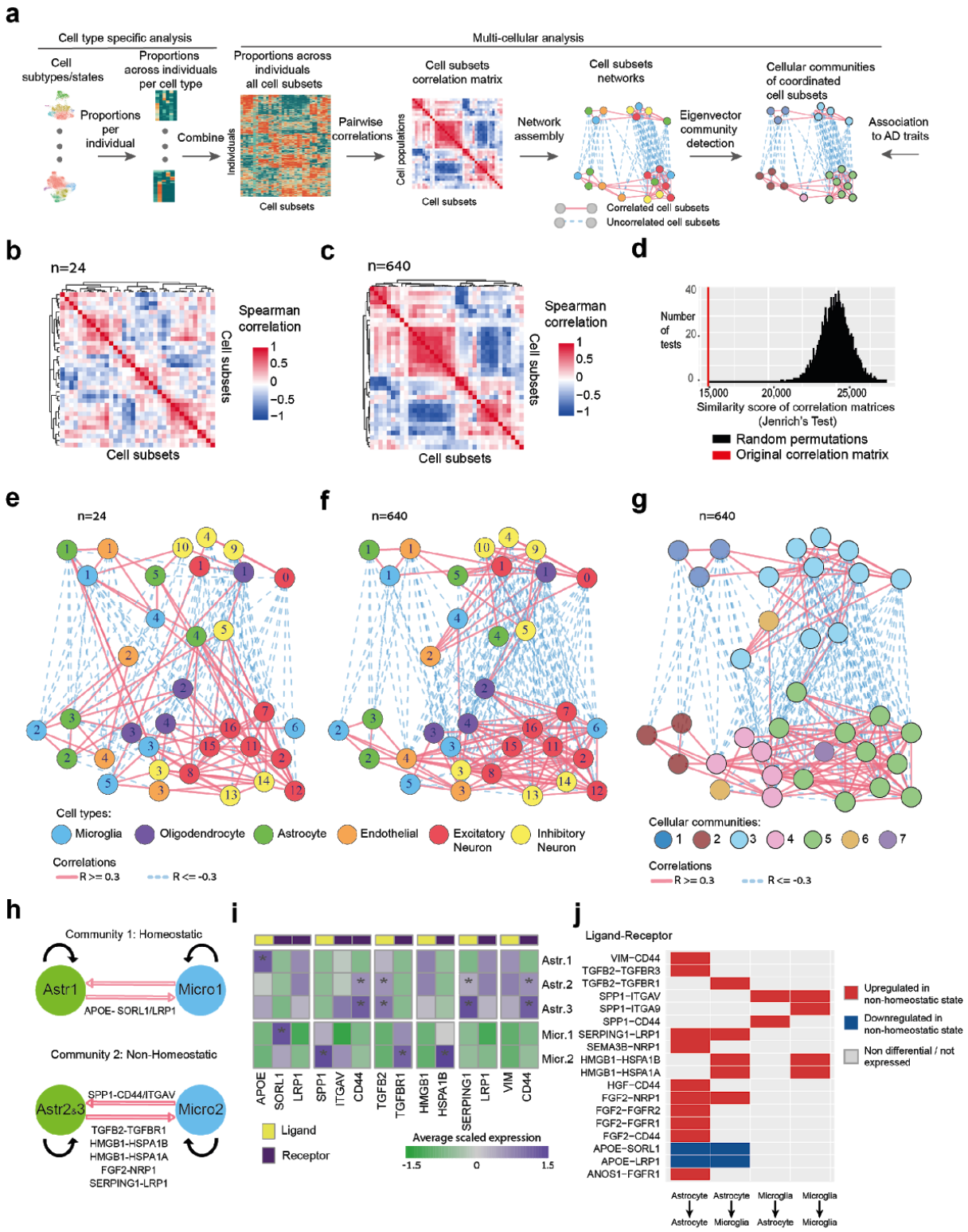

Figure 6. Multi-cellular communities exist in the aging DLPFC brain region. (**a**) Scheme of computational framework for estimating multicellular communities. Proportions of cell subsets across individuals within each cell type are calculated, combined and pairwise correlations between all cellular subsets are computed. A multi-cellular network is derived from the correlation matrix, cellular subsets are assigned to cellular communities using an eigenvectors community detection algorithm,

and associated with AD-traits by statistical analysis (**Methods**). (**b-c**) Coordinated changes in proportions of cell states and subtypes across individuals. A heatmap of the pairwise Spearman's correlation coefficient of the proportions of all cell states and subtypes across 24 individuals (snRNA-seq measurements, **b**) and validated in 640 individuals (CelMod estimated proportions, **c**), exposing a structure of mixed correlated and anti-correlated cellular populations of all cell types. (**d**) CelMod estimated pairwise correlations of cellular populations (n=640 individuals) matches the snRNA-seq measurements (n=24 individuals). Similarity between the two correlation matrices (in **b** and **c**) is statistically significant (p-value<0.001, by permutation test, **Methods**). Histogram of the distribution of similarity scores (Jenrich's test, [43]) of correlation matrices in 10,000 random permutations of the cellular frequencies matrix (in **Supplementary Fig. 5a**) independently per cell type. Red, similarity score of the true matrices in **b** and **c**. (**e-f**) A network of cellular subsets reveals coordinated variation across individuals in multiple cell types. Network of coordinated and anti-coordinated cell subsets (nodes). Edges between pairs of subsets with significantly correlated proportions across individuals (r>0.3, solid red line) or anti-correlated (r<-0.3, dashed blue line), based on snRNA-seq proportions (in **e**, n=24) or CelMod proportions (in **f**, n=640). Nodes are colored by the cell type and numbered as in **Fig. 2a,b**. Comparison of the two networks in **Supplementary Fig. 5b,c.** (**g**) Multi-cellular communities. Cellular community network (as in **f**) colored by the assigned cellular community (n=640 individuals, by eigenvector community detection method [28], **Methods**). (**h-i**) Community-specific ligand and receptor pairs connecting microglia and astrocyte cells in opposing cellular communities. (**h**) Scheme of differentially expressed ligand and receptor pairs putative signaling between astrocytes and microglia states from homeostatic *community 1* (Micr.1 and Astr.1, top) compared to those in non-homeostatic *community 2* (Micr.2, Astr.2/Astr.3, bottom). (**i**) Selected examples of differential ligand and receptor pairs. Scaled mean average expression (color scale) across microglia and astrocyte states. Asterisk, marking cellular populations with a significant up-regulated expression compared between the cell subsets of the same cell type(multiple hypothesis correction p-value <0.001). (**j**) An increase in the community-specific ligand-receptor pairs in the non-homeostatic community 2 compared to the homeostatic community 1. Upregulated (red) or down regulated (blue) or non-differential (grey) ligand-receptor pairs (rows) within pairs of cell subsets (columns).

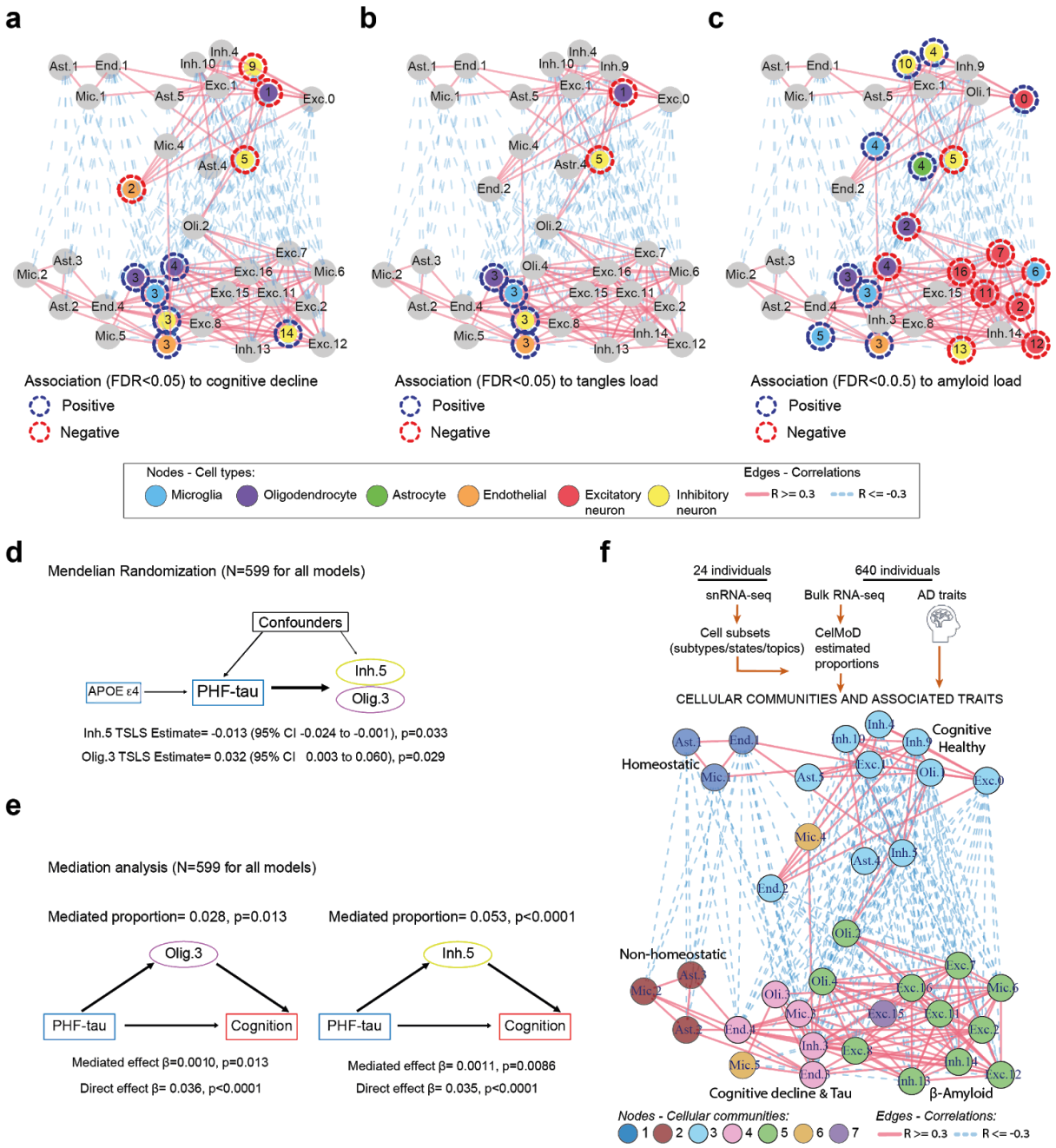

Figure 7. Mapping the cellular environment of the healthy and AD brain. (**a-c**). Cellular communities are linked to AD associated traits. Cellular communities network (as in **Fig. 6f**), of coordinated and anti-coordinated cell subsets (nodes). Edges between pairs of subsets with significantly correlated proportions across individuals (n=640, r>0.3, solid red line) or anti-correlated (r<-0.3, dashed blue line). Nodes are colored by the cell type and numbered as in **Fig. 2a,b** (non-significant subsets in grey). Cellular subsets associated (multivariable linear regression, FDR<0.05) with AD traits (dotted outline: positive, blue; negative, red) for: cognitive decline (**a**), tangles load (**b**), and β-amyloid load (**c**). Other traits in **Supplementary Fig. 6a,b**. (**d**) Tau pathology load is predicted to be upstream of changes in proportion of Inh.5 subset or Olig.3 topic. Mendelian randomization analysis using APOE-ε4 as an intermediate variable, Tau pathology as the exposure, and each cell subset proportion (Inh.5 or Olig.3) as the outcome. TSLS= two-stage least squares. CI=confidence interval. Arrow, the predicted optimal model (**e**) Mediation analysis shows variation in proportions of Inh.5 and Olig.3 partially mediates the association between tau pathology and cognitive

decline. Top: Scheme of mediation analysis results, with tau pathology the continuous independent variable, each cell subset proportion (Inh.5 or Olig.3) the mediator, and cognitive decline the continuous outcome (**Methods**). Bottom: Non-parametric bootstrap was done with 10,000 simulations. (**f**) A scheme of our proposed model of multi-cellular communities of the aging DLPFC brain region and their associations with AD traits. Cellular networks (as in **a**), nodes colored by the community assignments. The significantly enriched associations to AD-traits (hypergeometric p-value) are marked next to the graph.

Methods

Experimental design

Data were derived from subjects enrolled in two clinical-pathologic cohort studies of aging and dementia, the Religious Orders Study (ROS) or the Rush Memory and Aging Project (MAP), collectively referred to as ROSMAP. All participants are not diagnosed with dementia at enrolment, have annual clinical evaluations and agree in advance for brain donation at death. At death, the brains undergo a quantitative neuropathologic assessment, and the rate of cognitive decline is calculated from longitudinal cognitive measures that include up to 20 yearly evaluations [9–11]. For this study, we used data from 24 individuals (12 males and 12 females) chosen to represent the range of pathologic and clinical diagnoses of AD dementia (at the time of death), divided to four groups (**Supplementary Table 1**): (1) a reference group of cognitively non-impaired individuals (cAD =1 in ROSMAP) with minimal AD pathology (pathoAD =0 in ROSMAP), (2) a resilient group of cognitively non-impaired individuals (cAD=1) with a pathologic diagnosis of AD(pathoAD=1), (3) an AD group who fulfill diagnoses for both clinical AD dementia and pathologic AD(cAD=4 and pathoAD=1), and (4) a clinical-AD group of individuals diagnosed with clinical AD dementia but showed only minimal AD pathology upon post-mortem characterization (cAD=4 and pathoAD=0). We excluded samples that had RIN>5 and with post-mortem interval (PMI) < 4, and included only samples that had bulk RNA-sequencing in a previous study [1] as well as whole genome sequencing data [9–11].

Experimental design: AD traits in the ROS/MAP cohorts.

The pathologies were collected as part of the ROS/MAP cohorts (previously described in details [44–47]). We used the following traits in the poise of samples and in the association analysis: ***Cognitive Decline.*** Uniform structured clinical evaluations, including a comprehensive cognitive assessment, are administered annually to the ROS and MAP participants. The ROS and MAP methods of assessing cognition have been extensively summarized in previous publications [11,45,48–50]. Scores from 17 cognitive performance tests common in both studies were used to obtain a summary measure for global cognition as well as measures for five cognitive domains of episodic memory, visuospatial ability, perceptual speed, semantic memory, and working memory. The **summary measure for global**

cognition is calculated by averaging the standardized scores of the 17 tests, and the summary measure for each domain is calculated similarly by averaging the standardized scores of the tests specific to that domain. To obtain a measurement of cognitive decline, the annual global cognitive scores are modeled longitudinally with a mixed effects model, adjusting for age, sex and education, providing person specific random slopes of decline (which we refer to as ***cognitive decline***). The random slope of each subject captures the individual rate of cognitive decline after adjusting for age, sex, and education. Further details of the statistical methodology have been previously described [51]**. Clinical diagnosis of AD at the time of death.** Annual clinical diagnosis of AD dementia follows the recommendation of the joint working group of the National Institute of Neurological and Communicative Disorders and Stroke and the AD and Related Disorders Association [52]. The diagnosis requires a history of cognitive decline and evidence of impairment in memory and at least one other cognitive domain. After a participant had died, a neurologist specializing in dementia reviews all available clinical information and provides a summary opinion with regards to the most likely clinical diagnosis at the time of death. The summary diagnosis was blinded to all neuropathologic data, and case conference are held for consensus as necessary [53]. AD dementia includes persons with probable or possible AD dementia, i.e. AD dementia with or without comorbid conditions that may be affecting cognition. **A pathologic diagnosis of AD** was determined by a board-certified neuropathologist blinded to age and all clinical data and using modified Bielschowsky silver stained 6 micron sections of hippocampus, entorhinal cortex, midfrontal cortex, midtemporal cortex and inferior parietal cortex. The diagnosis follows the recommendation of the National Institute on Aging-Reagan criteria [54]. Briefly, based on the scores of Braak stage for severity of neurofibrillary tangles and CERAD estimate for burden of neuritic plaques, a pathologic AD diagnosis requires an intermediate likelihood AD (i.e., at least Braak stage 3 or 4 and CERAD moderate plaques) or a high likelihood AD (i.e., at least Braak stage 5 or 6 and CERAD frequent plaques). **β-amyloid and tau pathology load.** Quantification and estimation of the burden of parenchymal deposition of β-amyloid and the density of abnormally phosphorylated tau-positive neurofibrillary tangles levels present in the cortex at death (which we refer to as β-amyloid and tau pathology, respectively), tissue was dissected from eight regions of the brain: the hippocampus, entorhinal cortex, anterior cingulate cortex, midfrontal cortex, superior frontal cortex, inferior temporal cortex, angular gyrus, and calcarine cortex. 20μm sections from each region were stained with antibodies to the β-amyloid beta protein and the tau protein, and quantified with image analysis and stereology.

Nucleus isolation and single nucleus RNA library preparation

Dorsolateral Prefrontal Cortex (DLFPC) tissue specimens were received frozen from the Rush Alzheimer's Disease Center. We observed variability in the morphology of these tissue specimens with differing amounts of gray and white matter and presence of attached meninges. Working on ice throughout, we carefully dissected to remove white matter and meninges, when present. We expect this will help to reduce variability between tissue specimens. Working on ice throughout, about 50-100mg of gray matter tissue was transferred into the dounce homogenizer (Sigma Cat No: D8938) with 2mL of NP40 Lysis Buffer [0.1% NP40, 10mM Tris, 146mM NaCl, 1mM $CaCl_2$, 21mM $MgCl_2$, 40U/mL of RNAse inhibitor (Takara: 2313B)]. Tissue was gently dounced while on ice 25 times with Pestle A followed by 25 times with Pestle B, then transferred to a 15mL conical tube. 3mL of PBS + 0.01% BSA (NEB B9000S) and 40U/mL of RNAse inhibitor were added for a final volume of 5mL and then immediately centrifuged with a swing bucket rotor at 500g for 5 mins at 4°C. Samples were processed 2 at a time, the supernatant was removed, and the pellets were set on ice to rest while processing the remaining tissues to complete a batch of 8 samples. The nuclei pellets were then resuspended in 500ml of PBS + 0.01% BSA and 40U/mL of RNAse inhibitor. Nuclei were filtered through 20um pre-separation filters (Miltenyi: 130-101-812) and counted using the Nexcelom Cellometer Vision and a 2.5ug/ul DAPI stain at 1:1 dilution with cellometer cell counting chamber (Nexcelom CHT4-SD100-002). 20,000 nuclei in around 15-30ul volume were run on the 10X Single Cell RNA-Seq Platform using the Chromium Single Cell 3' Reagent Kits v2. Libraries were made following the manufacturer's protocol, briefly, single nuclei were partitioned into nanoliter scale Gel Bead-In-EMulsion (GEMs) in the Chromium controller instrument where cDNA share a common 10X barcode from the bead. Amplified cDNA is measured by Qubit HS DNA assay (Thermo Fisher Scientific: Q32851) and quality assessed by BioAnalyzer (Agilent: 5067-4626). This WTA (whole transcriptome amplified) material was diluted to <8ng/ml and processed through v2 library construction, and resulting libraries were quantified again by Qubit and BioAnalzyer. Libraries from 4 channels were pooled and sequenced on 1 lane of Illumina HiSeqX by The Broad Institute's Genomics Platform, for a target coverage of around 1 million reads per channel.

Pre-processing of snRNA-seq data

De-multiplexing, alignment to the hg38 transcriptome and unique molecular identifier (UMI)-collapsing were performed using the Cellranger toolkit (version 2.1.1, chemistry V2, 10X Genomics, for chemistry Single Cell 3'), and run using cloud computing on the Terra platform (https://Terra.bio). Since nuclear RNA includes roughly equal proportions of intronic and exonic reads, we built and aligned reads to a genome reference with pre-mRNA annotations, which account for both exons and introns. For every nucleus, we quantified the number of genes for which at least one read was mapped, and then excluded all nuclei with fewer than 400 detected genes. Genes that were detected in fewer than 10 nuclei were excluded. Expression values $E_{i,j}$ for gene i in cell j were calculated by dividing UMI counts for gene i by the sum of the UMI counts in nucleus j, to normalize for differences in coverage, and then multiplying by 10,000 to create TPM-like values, and finally computing $\log_2(\text{TP10K} + 1)$ (using the *NormalizeData* function from the *Seurat*[22] package version 3). Next, we selected variable genes (using the FindVariableFeatures function in Seurat, setting the selection method to *vst*) and scaled the data matrix (using the ScaleData function from Seurat 3), yielding the relative expression of each variable gene by scaling and centering. The scaled data matrix was then used for dimensionality reduction and clustering. To rule out the possibility that the resulting clusters were driven by batch or other technical effects, we examined the distribution of samples within each cluster and the distribution of the number of genes detected across clusters (as a measure of nucleus quality). Additional filtration of low-quality cells and clusters was done following this initial clustering analysis, to remove low quality neuronal cells that could not be filtered out initially based on the number of genes detected, due to the cellular heterogeneity of the tissue, that includes both high RNA content cells (such as neurons) and low RNA content cells (such as microglia cells).

Dimensionality reduction and clustering

We used the scaled expression matrix restricted to the variable genes for Principal Component Analysis (PCA), using *RunPCA* method in Seurat (a wrapper for the irlba function), computing the top 50 PCs. After PCA, significant principal components (PCs) were identified using the elbow method, plotting the distribution of standard deviation of each PC (*ElbowPlot* in Seurat), choosing 25 PCs for analysis of all cells, 25 PCs for astrocytes, 16 for microglia, 11 for endothelial cells, and 20 PCS for oligodendrocytes. Scores from only these top PCs were used as the input to downstream clustering or visualization. The scaled data was embedded in a two-dimensional space using Uniform

Manifold Approximation and Projection (UMAP) [55], based on the first significant PCs (as listed above). Clustering was performed using a graph-based approach: a *k*-nearest neighbor (*k*-NN) graph over the cells was calculated (using the FindNearestNeighbors function), followed by the Louvain community detection algorithm [55,56], which decomposes an input graph into communities to find transcriptionally similar clusters of cells (using the FindClusters function). The resolution of the clustering was selected using both cell-type markers and visualization of the UMAP embedding (using resolutions ranging between 0.3 to 0.6). For the full dataset across all cell types, cells were hierarchically clustered and clusters were re-ordered accordingly (using *BuildClusterTree*method the *Seurat* package, given the same set of variable genes as input to the PCA analysis), providing the cluster ordering described throughout this study. In addition, we removed a group of nuclei that exhibits a neuronal signature that we suspected to be technically altered (expressed a low number of genes compared to other neurons, with missing expression of basic neuronal markers, and had a high signature of cytoplasmic specific genes). Cell populations were manually matched to cell types based on the expression of known marker genes as previously done [2,57].

Doublet cells removal

For doublet detection and elimination, we clustered our data at high resolution, to generate multiple small clusters, and removed clusters enriched with suspected doublets. We used a combined approach of manual and automatic detection. First, we ran DoubletFinder [58] to identify clusters of doublets and found small clusters that had over 70% of nuclei classified as doublets, which we then excluded from downstream analysis. We validated that these clusters are “doublet clusters” based on manual inspection of expression patterns of cell type marker genes, showing that they co-express markers of at least two different cell types.

In downstream analysis of specific cell types, a second manual inspection for doublets was performed for expression patterns of cell type marker genes, excluding additional small clusters of doublet cells. Some of these suspected doublet cells might have been cells with high content of ambient RNA, and were removed conservatively.

Sub-clustering analysis of glia, endothelial and neuronal cells

For each cell type that had sufficient number of cells: astrocytes, microglia, endothelial cells, GABAergic neurons and glutamatergic neurons, we performed sub-clustering analysis to reveal additional diversity of cell subtypes and cell states. To this end, the pipeline described for all cells was applied separately to each cell type, including: normalizing the count matrix, finding variable genes, scaling the normalized count matrix, dimensionality reduction by PCA, graph clustering and dimensionality reduction by UMAP for visualization in 2D. To determine the individuals that had predominant homeostatic astrocytes (Astr.1) or homeostatic microglia (Micro.1) cells, we set all individuals that had more than 50% of Astr.1 or 40% of Micr.1. For microglia and astrocytes, initial annotations of subsets were done by known marker genes [2,3,5,7,8]. Microglia subsets were further annotated by comparison to a dataset of live microglial cells [8] (see the section *Comparing snRNA-Seq and scRNA-Seq of microglia*). Differentially expressed signatures were calculated using a negative binomial test and controlled false-discovery rates (FDRs) using the Benjamini–Hochberg procedure, to find genes that are upregulated within each cluster compared with the rest of the nuclei in the dataset or between pairs of clusters, including genes with less than 1% FDR. Genes were required to be expressed in at least 10% of nuclei in the given cluster, and at least 0.3-fold less in all other cells. The differential expression signatures were tested for enriched pathways and gene sets, using a hypergeometric test (function enrichment in the bc3net package in R), and corrected for multiple hypotheses by FDR. Results with FDR < 0.05 were reported as significantly enriched pathways. Gene sets and pathways were taken from the MSigDB/GSEA resource (combining data from Hallmark pathways, KEGG, Reactome, Pathway Interaction Database (PID), Canonical Pathways and Gene Ontology (GO) biological processes) [59].

Annotating neuronal subtypes, was first done by known marker genes [12,57], assigning neuronal cells to excitatory and inhibitory classes, to inhibitory subtypes and excitatory subtypes of pyramidal neurons. Next, an automatic annotation was performed by running a logistic regression classifier against a recent fully annotated cortical single nuclei RNA-seq dataset from the Allen Brain Atlas [12]. Briefly, the classifier is based on a logistic regression model (linear_model.LogisticRegression from Python's sklearn package), which was modified to include a calculation of the classification probability for each cell against a dataset of reference pre-annotated cells. Annotations were done on the cluster level: each cell was associated to a cluster, and the annotation of the cluster was done based on the overall highest scoring cell type across all cells. A regression model was learned based on the dataset from

the Allen Brain Atlas, and then applied to our neuronal data. The automatic annotations were validated by the expression of known marker genes.

Topic modeling for oligodendrocyte cells

We modeled the cell state diversity in oligodendrocyte cells by topic modelling [13–16], using Latent Dirichlet Allocation (LDA). Topic modeling was performed on the normalized and scaled data matrix. To assign topics and score cells for each topic, we used the CountClust package in R, which calculated the grade of membership for 4 topics (using the GoM function, which was run on the scaled expression matrix with tolerance 0.01). To find the optimal number of topics and tolerance levels, we ran a grid of parameters of topic numbers and tolerance levels, and visualized topic scores on the UMAP. The results were robust to the choice of parameters, yet we favored a small number of topics given the number of cells and samples in our data. A higher number of topics often captured only cells of one individual and would require more individuals for reproducibility. To select the genes highly associated with each topic, we used the Kullback-Leibler divergence (KL), which measures the difference between two probability distributions, applied to the distribution of the topic weights and the distribution of gene expression level across cells (using the ExtractTopFeatures function with default parameters).

Removal of ambient signatures

Removal of ambient RNA signatures was performed prior to topic modeling, as oligodendrocyte cells had a large subset of cells with variable levels of ambient RNA which skewed the analysis. To this end, we identified a set of genes that constitute the *ambient gene signature*, defined as genes that appeared in empty droplets with the highest frequency. In more details, we identified the ambient RNA signature by selecting droplets (cells) that had 10 UMIs or less assigned to them (from the raw CellRanger output) which we considered to be empty droplets. We then selected the top 100 genes that had the overall highest number of UMIs across all the empty droplets, and defined them as the *Ambient RNA signature*. To address potential sample specific technical bias, we performed the analysis for each individual (sample), and united the ambient RNA signatures across individuals. Finally, we

excluded the ambient gene signature from the list of variable genes included in the analysis, and rescaled the data accordingly, and proceeded with the downstream analysis as described.

Densities plots to visualize distribution of cellular populations and gene expression in 2D embeddings

For clear visualization of the distribution of cell populations or distribution of the expression levels of a gene, in dense 2D graphs of multiple cells, we used a Gaussian kernel approach to plot and the density of cells and gene levels. We used this approach to plot groups of individuals divided by AD-associated traits in **Fig. 1 and Fig. 2,** and the expression of marker genes in **Fig. 3**. For discrete traits, we scored each cell by calculating the density of neighbors sharing the same trait, calculating the Gaussian kernel adjacencies of the 2D UMAP embedding for all other cells (using the Gauss kernel from the 'KLRS' package in R, with sigma = 1.5). We filtered out cells with an adjacency measure > 0.0005. For the remaining cells, we calculate the proportion of cells assigned to individuals with the relevant phenotype, weighing them by their Gaussian distance to the cell. Finally, we normalize the obtained measure of the cell by the total distances to the remaining cells. In the continuous case, we use a similar approach, considering all neighboring cells weighted by their individual expression level.

Comparing snRNA-Seq and scRNA-Seq of microglia

We mapped our microglia snRNA-Seq to our recent published dataset of scRNAseq of live microglia cells from fresh autopsy and surgically resected human brain tissue [8]. Mapping was performed using Canonical Correlation Analysis in Seurat with standard parameters (3,000 genes and 20 canonical components) to align the snRNA-Seq and scRNA-seq data. We then used a naive Bayes classifier trained on the scRNA-seq-derived live microglia clusters in canonical correlation component space. This classifier was used to predict the class membership of the CCA-transformed snRNA-Seq nuclei profiles, and each nucleus profile was assigned to the single-cell live microglia cluster with the maximum prediction value.

Estimating major cell class and cell subset proportions from bulk data by CelMod

We developed a regression-based consensus model to extend our snRNA-seq-derived cell class and subset estimates to bulk data, leveraging matched bulk and snRNA-seq data from the same 24 donors.

We train the regression model as follows: (**1**) Perform a linear regression on each gene separately for each cell cluster of interest, using its expression as the dependent variable and the proportion of that cluster in each snRNA-Seq sample in the training set as the independent variable; (**2**) For each gene, use the regression model to calculate the predicted proportion of each cell type, normalizing their sum to 1; (**3**) Rank genes by the 90th percentile of the absolute value of the error between predicted and training proportions, for each cell type; and (**4**) Select the number of top-ranked genes (constant for each cell cluster) to use for deconvolving a new bulk RNA-Seq sample; this number of genes, the only tunable parameter, is selected based on cross-validation, as described below.

To determine the optimal number of genes to use for the prediction, we use 5-fold cross-validation using 80% of the data for training and 20% as the validation set. The validation sets are mutually exclusive, such that after 5 runs, the proportions in every bulk sample have been "predicted" once. This cross-validation is run using 3 to 100 ranked genes (from step 3 above), and the optimal gene number is selected as that which minimizes the mean of the 90th percentile errors for each cell group in all samples. This gene number is then used for deconvolution predictions in the larger bulk RNA-seq data set, for a given set of cell clusters.

We run the algorithm iteratively, starting at the "top level", with the broad cell classes (glutamatergic neurons, GABAergic neurons, astrocytes, oligodendrocytes, OPCs, microglia, endothelial cells, and pericyte), and then again for the subtypes within each of the cell classes. For the broad cell classes, the proportions are based on the total nuclei per sample. For the subtypes, the proportions are normalized to the total nuclei from the broad class of interest - for example, for astrocyte subsets, the proportions for the training (and thus the prediction) are normalized to the total number of astrocyte nuclei per sample. This allows us to directly model both the overall and subtype-level composition of the bulk tissue, especially for cell types that comprise a small fraction of the overall population (such as the endothelial or microglial subtypes). For the oligodendrocyte signatures, which are modeled as topics instead of discrete clusters, we sum the weights for each given signature over all nuclei from a given sample, and then normalize these sums so that they add up to 1 for a given sample. This reflects a "proportion of topic weights" per sample, as opposed to a strict proportion that can be calculated for the discrete cell types.

The performance of this algorithm on the validation set (20% of the data, with 5-fold cross-validation) is shown in **Fig. 4c-d and Supplementary Fig. 4b,c**, for each level, and the validations using immunohistochemistry in **Fig. 4e-f**.

Cellular Communities

To find co-occurring cellular populations across individuals, we generated a network of cellular communities, defined as a set of cell populations (cell types, subtypes, cell states or expression programs) that have coordinated variation of proportions across individuals (for topics we use the proportional weights instead of the subset proportion). We applied this approach to define cellular communities from our snRNA-seq dataset (n=24 individuals) and from the CelMod estimated proportions (n=640 individuals), across 6 broad cell classes (astrocytes, microglia, endothelial cells, oligodendrocytes, inhibitory (GABaergic) and excitatory (glutamatergic) neurons) for which we had a defined sub-clustering (splitting each cell type to subsets capturing distinct cell subtypes or cell states), or defined topic models (splitting each cell type to distinct expression programs, without the need for a discrete assignment of cells to clusters). When applying this approach to the estimated cellular proportions by the CelMod algorithms in 640 individuals, we included the estimations for the oligodendrocyte cells, which allowed for a distribution of weights across the topics for each cell.

To detect cellular communities we followed four steps: **(1) Cellular proportions.** Using the classification of cells to cell subsets (in the current analysis the subsets are cell subtypes for neurons, cell states for glial and endothelial cells, and topic models for oligodendrocytes), we calculated for each cell class and for each individual: the frequency of all subsets *within* each cells class (i.e. the proportion of cells within a subset out of the total number of nuclei in that broad cell class or sum of all topic weights, for oligodendrocytes). Next, we appended the cell subset proportions across *all* cell classes into a *combined frequency matrix*. Of note, for the 640 individuals we used the estimated proportions by CelMod per subset. **(2) Correlation matrix.** We calculated the pairwise Spearman correlation coefficient over the frequency of each cell subset across individuals, created a pairwise correlation matrix and clustered it by hierarchical clustering (using the pheatmap function in R, based on the distance metric of

1-Pearson correlation). **(3) Cellular network.** We transformed the correlation matrix into a graph, where each cell subset is a node, and edges connected every two nodes if their absolute pairwise Spearman correlation value was at least 0.3. Each edge was assigned a positive or negative sign. The graph was laid out using the igraph package (in R), using the function graph_from_data_frame, with positive edges only, and the negative edges were added for visualization (setting the parameter layout = layout_with_fr). We maintained a consistent common layout between the networks derived from 24 and 640 individuals, based on the output of graph_from_data_frame on CelMod data. **(4) Cellular communities.** We assigned cell populations to communities using the eigenvectors of the correlation matrix [28] (cluster_leading_eigen function in igraph, using only the edges with positive weights as input to the function). **(5) Associating cellular communities to AD-traits.** Each cellular subset was first associated to AD-traits by multivariable linear regression (see details in the section *Statistical analysis to assign cell types, subtypes, cell states and topic models to AD related traits*). cellular subset-trait associations with FDR<0.05 were considered significant and assigned as positive (beta>0) or negative (beta<0) associations. For each cellular community we checked the overlap with cellular subsets significantly associated with the various AD traits (cognitive decline, amyloid load and tau pathology load). For cellular communities and traits with overlap of at least 3 subsets, a hypergeometric p-value was calculated, and we termed a cellular community to be *enriched* for a specific trait if the p-value<0.05.

Of note, since we used only 4 topics that are largely disjoint sets of cells for the oligodendrocytes, a hard assignment for an oligodendrocyte to a specific topic (by the maximum weight, thus transforming the data to discrete clusters), did not substantially impact the results compared using the continuous topic weight. However, the framework proposed here applies more generally to any level of topic modeling or other soft assignment of cells to continuous cell states.

To compare the snRNA-Seq network (24 individuals) and the CelMod estimated cellular network (640 individuals, as in **Supplementary Fig. 5b,c**) we did the following: First, we calculated the statistical significance of the similarity between the snRNA-Seq network and the CelMod network by an empirical p-value. Briefly, we performed 10,000 random permutations of the combined frequency matrix over all cell subsets in the 640 individuals (estimated by CelMod), shuffling the values within each cell cluster, such that the frequency of cell subsets will sum to 1 for each cell class. We calculated a correlation matrix for every

permutation, and a similarity score between the snRNA-seq correlation matrix and the permutation matrix using Jenrich's score ([43], ortest.jennrich function in R). An empirical p-value ($p<0.001$) was assigned to the original CelMod estimated proportions, since the non-permutation matrix consistently got the lowest score compared to the permuted-matrix. Other similarity scores, such as the $L2$ and $L1$ norms yielded similar significant p-values, showing the result is robust to the choice of similarity score. Next, we compared the snRNA-seq and CelMod networks directly: for each pair of cell subsets (edges in the graph), we noted whether the edge is preserved, inverted its orientation (negative or positive correlation), specific to the CelMod network (*i.e.* in the CelMod network absolute(r)>0.3 and in snRNA-Seq network absolute(r)<0.3) or specific to the snRNA-seq network (*i.e.* in the CelMod network absolute(r)<0.3 and in snRNA-Seq network absolute(r)>0.3). Of note, in this comparison a Spearman correlation coefficient value higher than absolute(0.3) is considered (anti-)correlated, between -0.15 and 0.15 is considered uncorrelated (or neutral), and between absolute(0.15) and absolute(0.3) is softly correlated. To avoid the effects of harsh thresholds, when in one model the edge is correlated and in the other softly-correlated the edge was considered to be preserved.

Statistical analysis to assign cell types, subtypes, cell states and topic models to AD related traits

We analyzed five major pathological and cognition hallmarks of AD, collected as part of the ROS/MAP cohort, including: primary neuropathological and cognitive phenotypes. The two clinical traits were a clinical diagnosis of AD dementia proximate to death (clinical AD) and a continuous measure of cognitive decline over time quantified as a per-subject slope of the cognitive decline trajectory from a linear mixed-effects model [51]. The three pathology variables include continuous measures of tau tangle pathology density and β-amyloid burden (both averaged over multiple regions) and a binary diagnosis of pathologic AD (as previously described [60]). Details on the traits can be found in the section: *Experimental design: AD traits in the ROS/MAP cohorts.*

To test the statistical associations between neuropathological phenotypes and cell types or cell subsets proportions, we performed multivariable linear regressions, modeling cellular proportions as the outcome, and neuropathology as the independent variable. In each analysis, we adjusted for age, sex, and RIN score as covariates to remove potential confounding, and corrected for multiple testing, by calculating the false discovery rate (FDR)

for each set of pathology analyses. The analysis was done for each of the five traits separately (amyloid, tau, cognitive decline, pathologically determined Alzheimer's disease, and post mortem clinical diagnosis of Alzheimer's disease), across all cell types (astrocytes, oligodendrocytes, endothelial, inhibitory neurons (GABAergic), excitatory neurons (glutamatergic), microglia, pericyte, and oligodendrocyte progenitor cells (OPCs)), or across all cell subsets (5 astrocytes, 4 endocyte, 7 GABAergic neurons, 9 glutamatergic neurons, 6 microglia subsets, and 4 oligodendrocyte topics).

Mediation analysis assigning causal relations to cell states, AD pathology and genetic drivers

To assess the causal relationship between tau burden and inferred cell subtype proportions associated with cognitive decline, we performed Mendelian randomization analysis [41] using *APOE* ε4 allele count (0, 1, or 2) as an instrumental variable. We focused on two cell subtypes/signatures, Olig.3 and Inh.5, whose inferred proportions had the strongest associations with cognitive decline. We first tested whether these cell subtype/signature proportions were also associated with *APOE* ε4 allele count at $p<0.05$, adjusting for age at death, sex, RIN score, among participants with both *APOE* ε4 and PHFtau measurement (n=630). For these cell subtype/signature proportions, we examined whether association between *APOE* ε4 and the respective proportions were no longer significant when PHFtau was controlled in a liner model (adjusted for age at death, sex, and RIN score), to confirm that PHFtau mediates *APOE* ε4 effect on cell subtype/signature proportions, and *APOE* ε4 is a valid instrumental variable. We also assessed whether cell subtype/signature proportions could explain the association between *APOE* ε4 and tau using linear regression models, to rule out an opposite direction of causal link. Finally, we used R package "ivmodel" to run instrumental variable analysis adjusting for age at death, sex, and RIN score, and reported results from two-stage least squares (TSLS) k-class estimator. We note that the limited information maximum likelihood (LIML) estimates were identical to TSLS estimates for the instrumental variable models we ran in this study.

For cell subtype/signature proportions suggested to be downstream of PHFtau per the Mendelian randomization analysis, we examined whether the proportion values mediate the relationship between tau and cognitive dysfunction through mediation analysis using R package "mediation."[41,61]. We set tau as an independent causal variable ("treatment"), each

inferred cell subtype/signature proportion as a continuous mediator, and cognitive decline (the random slope of longitudinal global cognition; already adjusted for age, sex, and education) as a continuous outcome. Mediated (indirect) effect, direct effect, and proportion mediated were estimated using non-parametric bootstrap method with 10,000 simulations, and all models included in the mediation analysis were adjusted for the RIN score.

Immunohistochemistry

Formalin-fixed *post-mortem* brain tissues were obtained from Rush University Medical Center. As part of these studies, all participants consent to brain donation at the time of death. 6µm sections of formalin-fixed paraffin-embedded (FFPE) tissue from the cortex frontal were stained with NEF (Sigma, N2912). Heated-induced epitope retrieval was performed using citrate (pH=6) using microwave (800W, 30% power setting) for 25 min. The sections were blocked with blocking medium (3% BSA) for 30 min at Room Temperature, then incubated with primary antibody anti-NEF prepared in 1% BSA for overnight at 4°C. Sections were washed three times with PBS and incubated with fluochrome conjugated secondary antibodies (Thermo Fisher) for one hour at RT. Anti-fading reagent with Dapi (P36931, Life technology) was used for coverslipping. For each subject, 30 images of cortical grey matter at magnification x20 (Zeiss Axio Observer.Z1 fluorescence microscope) were taken in a zigzag sequence along the cortical ribbon to ensure that all cortical layers are represented in the quantification in an unbiased manner. The acquired images were analyzed using CellProfiler and Cellprofiler Analyst developed by Broad Institute.
We estimated the proportions of a broad cell class or signature (neurons, microglia and GFAP+ astrocytes) from the images as the fraction of nuclei stained with the marker of interest (NeuN, IBA1, or GFAP, respectively for neurons, microglia, and GFAP+ astrocytes) out of all nuclei stained by DAPI.

Ligand-receptor analysis

We searched for logan-receptor interactions linking subsets of microglia and astrocyte cells within two highly anti-correlated cellular communities: *homeostatic* community 1 and *non-homeostatic* community 2. Ligand and receptor interactions were assembled from published resources [31,32]. 1,379 out of the 2,164 known ligand-receptor pairs (LRPs) were measured in our dataset. Receptors or ligands were considered as expressed if they had average normalized counts ≥ 1 in at least one of the relevant astrocytes and microglia subsets

(homeostatic Micr.1 and Astr.1, non-homeostatic Micr.2, Astr.2 and Astr.3). Next, we assigned each of the 1,379 expressed LRPs with all 16 possible pairs of cell subsets (four per ligand and four per receptor), resulting in 22,064 possible LRP tuples, where each tuple is composed of the LRP, the cell subset expressing the ligand and the cell subset expressing the receptor. We denoted LRP tuples where both the ligand and the receptor were expressed in their respective cell subsets as potential LRP tuples, resulting in 452 potential tuples (2.05% of all possible LRP tuples). We scored each potential LRP tuple by the combined expression of the ligand and the receptor, which was calculated by multiplying their average normalized counts in the corresponding populations. Next, we compared the distributions of the combined expression scores of potential LRP tuples and all possible LRP tuples. When scoring all possible LRPs by the combined expression of the ligand and the receptor in their associated community, the 452 expressed LRPs ranked at 93.9 percentile or higher. To estimate if the distribution of the number of potential LRP tuples across the cell pairs was significantly different than expected by chance we used the Chi Square test.

To identify potential LRP tuples that were differentially expressed between homeostatic and non-homeostatic cellular communities, we applied differential expression analysis to all ligands and receptors in the 452 potential LRP tuples, considering within each community putative signaling from microglia to astrocytes, astrocytes to microglia, and auto-signaling. We used the *FindMarkers* function in Seurat V3.2 R package, with a minimum percentage of expressing cells (*min.pct*) and ln(Fold-Change) (*logfc.threshold*) parameters set to 10% and 0.1, respectively. The p-values associated with the differential expression values were adjusted for multiple hypothesis testing by using Bonferroni correction. Only ligands and receptors with adjusted p-value < 0.001 were further considered. LRPs that were both differentially expressed, but in opposite directions (i.e. one upregulated and the other down regulated within the same community) were excluded from the downstream analysis. The differential expression score of each LRP tuple was set to the sum of ln(Fold-Change) of the ligand and the receptor. LRP tuples with an absolute differential expression score ≥ 0.5 were considered as differentially expressed, which we denote as *community-specific LRP*. To test the distribution of the number of LRPs across all the different pairs of cell subsets, within-community and between-community, we used the Chi Square test. LRP expression data for each pair of populations, differential values for each pair of populations are available at **Supplementary Table 6**.

DATA AND SOFTWARE AVAILABILITY

Raw and processed snRNA-seq data is deposited in the SYNAPSE database (https://www.synapse.org). Other ROSMAP data can be requested at the RADC Resource Sharing Hub at https://www.radc.rush.edu. Processed data, CelMod model for deconvolution and code will be available at the time of publication in the Single Cell Portal (https://singlecell.broadinstitute.org/single_cell) and in the GitHub.

SUPPLEMENTAL TABLES

Supplementary Table 1 - Samples details and phenotypes.
Supplementary Table 2 - Cell annotations clustering and projids
Supplementary Table 3 - Oligodendrocytes topic modeling associated genes.
Supplementary Table 4 - CelMod estimated cellular proportions for 640 individuals
Supplementary Table 5 - AD-traits associations to cell types and cell subsets proportions
Supplementary Table 6 - Ligand-Receptor analysis

ACKNOWLEDGEMENTS

We thank the individuals who have generously donated their brain to research through the RUSH University Alzheimer's Disease Center. We thank members of the Habib and DeJager labs, specifically, Adi Ravid and Or Gold; This work was supported by: grants from the National Institute of Aging (AMP-AD U01 AG046152, U01 AG061356, RF1 AG036042)(P.L.D.&D.A.B), the Chan Zuckerberg Initiative's Neurodegeneration Challenge Network (CS-02018-191971)(P.L.D.), the Klarman Cell Observatory (A.R.), Alon Fellowship and Myers Foundation (N.H.), Israel Science Foundation (ISF) research grants no. 1709/19 (N.H.), The European Research Council grant 853409 (N.H.). F.Z. and A.R. are Investigators of the Howard Hughes Medical Institute. N.H is a Goren Khazzam senior lecturer in neuroscience. ROSMAP is supported by NIA grants P30AG10161, R01AG15819, R01AG17917, U01AG46152, and R01AG61356.

AUTHOR CONTRIBUTIONS

P.L.D. and N.H. designed the study; C.M. prepared the single nucleus libraries and performed sequencing; A.C and V.M. performed the computational analysis, with guidance of N.H., V.M., P.L.D. and help from G.G.; I.H. performed the analysis of cellular interactions under the guidance of E.Y.L.; H.Y. and C.W. performed the statistical and modeling analysis under the guidance of P.L.D.; P.L.D., N.H., V.M., A.C. D.A.B., E.Y.L. and A.R. wrote the manuscript, and all co-authors edited it for critical comments. D.A.B. is PI of the parent ROS and MAP studies and obtained funding and performed study supervision. P.L.D., D.A.B and N.H. deposited data in Synapse and the RADC Resource Sharing Hub.

DECLARATION OF CONFLICT OF INTERESTS

A.R. is a founder and equity holder of Celsius Therapeutics, an equity holder in Immunitas Therapeutics and until August 31, 2020 was an SAB member of Syros Pharmaceuticals, Neogene Therapeutics, Asimov and ThermoFisher Scientific. From August 1, 2020, A.R. is an employee of Genentech, a member of the Roche Group.
All other co-authors have no relevant conflicts of interest.

References

1. Mostafavi, S. *et al.* A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat. Neurosci.* **21**, 811–819 (2018).
2. Habib, N. *et al.* Disease-associated astrocytes in Alzheimer's disease and aging. *Nat. Neurosci.* **23**, 701–706 (2020).
3. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
4. Del-Aguila, J. L. *et al.* A single-nuclei RNA sequencing study of Mendelian and sporadic AD in the human brain. *Alzheimers. Res. Ther.* **11**, 71 (2019).
5. 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).
6. Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276–1290.e17 (2017).
7. 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).
8. Olah, M. *et al.* Single cell RNA sequencing of human microglia uncovers a subset associated with Alzheimer's disease. *Nature Communications* vol. 11 (2020).
9. Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *J. Alzheimers. Dis.* **64**, S161–S189 (2018).
10. Bennett, D. A. *et al.* Overview and findings from the rush Memory and Aging Project. *Curr. Alzheimer Res.* **9**, 646–663 (2012).
11. Bennett, D. A., Schneider, J. A., Arvanitakis, Z. & Wilson, R. S. Overview and findings from the religious orders study. *Curr. Alzheimer Res.* **9**, 628–645 (2012).
12. Hodge, R. D. *et al.* Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
13. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
14. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. 3(Jan), 993–1022 (2003)
15. Dey, K. K., Hsiao, C. J. & Stephens, M. Visualizing the structure of RNA-seq expression data using grade of membership models. *PLoS Genet.* **13**, e1006599 (2017).
16. Bielecki, P. *et al.* Skin inflammation driven by differentiation of quiescent tissue-resident ILCs into a spectrum of pathogenic effectors. *Cold Spring Harbor Laboratory* 461228 (2018) doi:10.1101/461228.
17. Werner, H. & LeRoith, D. Insulin and insulin-like growth factor receptors in the brain: physiological and pathological aspects. *Eur. Neuropsychopharmacol.* **24**, 1947–1953 (2014).
18. Thrupp, N. *et al.* Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans. *Cell Rep.* **32**, 108189 (2020).
19. Bakken, T. E. *et al.* Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS One* **13**, e0209648 (2018).
20. Girgrah, N. *et al.* Localization of the CD44 glycoprotein to fibrous astrocytes in normal white matter and to reactive astrocytes in active lesions in multiple sclerosis. *J. Neuropathol. Exp. Neurol.* **50**, 779–792 (1991).
21. Leitão, R. A. *et al.* Aquaporin-4 as a New Target against Methamphetamine-Induced Brain Alterations: Focus on the Neurogliovascular Unit and Motivational Behavior. *Mol. Neurobiol.* **55**, 2056–2069 (2018).
22. Roussotte, F. F. *et al.* Combined effects of Alzheimer risk variants in the CLU and ApoE genes on ventricular expansion patterns in the elderly. *J. Neurosci.* **34**, 6537–6545 (2014).
23. 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).
24. Chen, X., Li, Z., Cheng, Y., Kardami, E. & Loh, Y. P. Low and High Molecular Weight FGF-2

Have Differential Effects on Astrocyte Proliferation, but Are Both Protective Against Aβ-Induced Cytotoxicity. *Front. Mol. Neurosci.* **12**, 328 (2019).
25. Farfel, J. M., Yu, L., De Jager, P. L., Schneider, J. A. & Bennett, D. A. Association of APOE with tau-tangle pathology with and without β-amyloid. *Neurobiol. Aging* **37**, 19–25 (2016).
26. Bennett, D. A. *et al.* Amyloid mediates the association of apolipoprotein E e4 allele to cognitive function in older people. *J. Neurol. Neurosurg. Psychiatry* **76**, 1194–1199 (2005).
27. Patrick, E. *et al.* Deconvolving the contributions of cell-type heterogeneity on cortical gene expression. *PLoS Comput. Biol.* **16**, e1008120 (2020).
28. Newman, M. E. J. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **74**, 036104 (2006).
29. Jha, M. K., Jo, M., Kim, J.-H. & Suk, K. Microglia-Astrocyte Crosstalk: An Intimate Molecular Conversation. *Neuroscientist* **25**, 227–240 (2019).
30. Vainchtein, I. D. & Molofsky, A. V. Astrocytes and Microglia: In Sickness and in Health. *Trends Neurosci.* **43**, 144–154 (2020).
31. Qiao, W. *et al.* Intercellular network structure and regulatory motifs in the human hematopoietic system. *Mol. Syst. Biol.* **10**, 741 (2014).
32. Ramilowski, J. A. *et al.* A draft network of ligand–receptor-mediated multicellular signalling in human. *Nature Communications* vol. 6 (2015).
33. Cekanaviciute, E. *et al.* Astrocytic TGF-β signaling limits inflammation and reduces neuronal damage during central nervous system Toxoplasma infection. *J. Immunol.* **193**, 139–149 (2014).
34. Das, P. & Golde, T. Dysfunction of TGF-beta signaling in Alzheimer's disease. *The Journal of clinical investigation* vol. 116 2855–2857 (2006).
35. Scacchi, R. *et al.* Apolipoprotein E (APOE) allele frequencies in late-onset sporadic Alzheimer's disease (AD), mixed dementia and vascular dementia: lack of association of ϵ4 allele with AD in Italian octogenarian patients. *Neuroscience Letters* vol. 201 231–234 (1995).
36. Jacobsen, L. *et al.* Activation and Functional Characterization of the Mosaic Receptor SorLA/LR11. *J. Biol. Chem.* **276**, 22788–22796 (2001).
37. Yajima, R. *et al.* ApoE-isoform-dependent cellular uptake of amyloid-β is mediated by lipoprotein receptor LR11/SorLA. *Biochem. Biophys. Res. Commun.* **456**, 482–488 (2015).
38. Yano, K. *et al.* Soluble LR11 competes with amyloid β in binding to cerebrospinal fluid-high-density lipoprotein. *Clin. Chim. Acta* **489**, 29–34 (2019).
39. Piscopo, P. *et al.* SORL1 Gene is Associated with the Conversion from Mild Cognitive Impairment to Alzheimer's Disease. *J. Alzheimers. Dis.* **46**, 771–776 (2015).
40. Andersen, O. M., Rudolph, I.-M. & Willnow, T. E. Risk factor SORL1: from genetic association to functional validation in Alzheimer's disease. *Acta Neuropathol.* **132**, 653–665 (2016).
41. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27**, 1133–1163 (2008).
42. Jack, C. R. *et al.* NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia* vol. 14 535–562 (2018).
43. Jennrich, R. I. An Asymptotic | chi 2 Test for the Equality of Two Correlation Matrices. *Journal of the American Statistical Association* vol. 65 904 (1970).
44. Schneider, J. A. *et al.* Cognitive impairment, decline and fluctuations in older community-dwelling subjects with Lewy bodies. *Brain* **135**, 3005–3014 (2012).
45. Bennett, D. A. *et al.* Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology* **66**, (2006).
46. Bennett, D. A., Wilson, R. S., Boyle, P. A., Buchman, A. S. & Schneider, J. A. Relation of neuropathology to cognition in persons without cognitive impairment. *Annals of Neurology* vol. 72 599–609 (2012).
47. Boyle, P. A. *et al.* Much of late life cognitive decline is not due to common neurodegenerative pathologies. *Annals of Neurology* vol. 74 478–489 (2013).
48. Wilson, R. S. *et al.* Individual differences in rates of change in cognitive abilities of older persons. *Psychol. Aging* **17**, 179–193 (2002).
49. Wilson, R. S., Barnes, L. L. & Bennett, D. A. Assessment of Lifetime Participation in Cognitively Stimulating Activities. *Journal of Clinical and Experimental Neuropsychology* vol.

25 634–642 (2003).
50. Wilson, R. S. *et al.* Conscientiousness, dementia related pathology, and trajectories of cognitive aging. *Psychol. Aging* **30**, 74–82 (2015).
51. De Jager, P. L. *et al.* A genome-wide scan for common variants affecting the rate of age-related cognitive decline. *Neurobiol. Aging* **33**, 1017.e1–15 (2012).
52. McKhann, G. *et al.* Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939–944 (1984).
53. Schneider, J. A., Arvanitakis, Z., Bang, W. & Bennett, D. A. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. *Neurology* **69**, 2197–2204 (2007).
54. Hyman, B. T. & Trojanowski, J. Q. Editorial on Consensus Recommendations for the Postmortem Diagnosis of Alzheimer Disease from the National Institute on Aging and the Reagan Institute Working Group on Diagnostic Criteria for the Neuropathological Assessment of Alzheimer Disease. *J. Neuropathol. Exp. Neurol.* **56**, 1095–1097 (1997).
55. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* vol. 3 861 (2018).
56. Lancichinetti, A. & Fortunato, S. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **80**, 016118 (2009).
57. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
58. 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.e4 (2019).
59. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425 (2015).
60. Consensus Recommendations for the Postmortem Diagnosis of Alzheimer's Disease. *Neurobiology of Aging* vol. 18 S1–S2 (1997).
61. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. mediation:RPackage for Causal Mediation Analysis. *Journal of Statistical Software* vol. 59 (2014).