

A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation

Alexandra Grubman^{1,2,3,9}, Gabriel Chew^{4,9}, John F. Ouyang^{1,4,9}, Guizhi Sun^{1,2,3}, Xin Yi Choo^{1,2,3,5}, Catriona McLean^{1,6}, Rebecca K. Simmons^{7,8}, Sam Buckberry^{7,8}, Dulce B. Vargas-Landin^{1,7,8}, Daniel Poppe^{7,8}, Jahnvi Pflueger^{7,8}, Ryan Lister^{1,7,8}, Owen J. L. Rackham^{1,4*}, Enrico Petretto^{1,4*} and Jose M. Polo^{1,2,3*}

There is currently little information available about how individual cell types contribute to Alzheimer's disease. Here we applied single-nucleus RNA sequencing to entorhinal cortex samples from control and Alzheimer's disease brains ($n=6$ per group), yielding a total of 13,214 high-quality nuclei. We detail cell-type-specific gene expression patterns, unveiling how transcriptional changes in specific cell subpopulations are associated with Alzheimer's disease. We report that the Alzheimer's disease risk gene *APOE* is specifically repressed in Alzheimer's disease oligodendrocyte progenitor cells and astrocyte subpopulations and upregulated in an Alzheimer's disease-specific microglial subpopulation. Integrating transcription factor regulatory modules with Alzheimer's disease risk loci revealed drivers of cell-type-specific state transitions towards Alzheimer's disease. For example, transcription factor EB, a master regulator of lysosomal function, regulates multiple disease genes in a specific Alzheimer's disease astrocyte subpopulation. These results provide insights into the coordinated control of Alzheimer's disease risk genes and their cell-type-specific contribution to disease susceptibility. These results are available at <http://adsn.ddnetbio.com>.

Alzheimer's disease is the most common form of dementia in the elderly and, as there are currently no effective treatments, is one of the leading contributors to global disease burden. Genetic mapping approaches such as genome wide association studies (GWAS) have uncovered Alzheimer's disease susceptibility loci, which are enriched in genes involved in endocytic and microglial pathways^{1,2}. Analyses of late-onset Alzheimer's disease (LOAD) bulk brain transcriptomes³ indicated changes in the underlying network structure controlled by gain of microglial gene connectivity and loss of neuronal connectivity in Alzheimer's disease. Although they have advanced understanding of the transcriptional network dynamics of Alzheimer's disease, bulk analyses do not enable the resolution of the precise cellular changes that contribute to Alzheimer's disease. For example, the effects of microglial niche expansion⁴ or neuronal loss⁵ on the observed gene network expression changes in Alzheimer's disease remain to be elucidated at the single-cell level. As such, there is a clear need to examine the single-cell landscape of Alzheimer's disease brains to understand how gene regulatory networks drive specific transcriptional changes in different cell types underlying Alzheimer's disease. Studies have examined the single-cell transcriptional landscape during human brain development⁶ and, more recently, Mathys and colleagues profiled single prefrontal cortex nuclei in a subset of Alzheimer's disease patients from the ROSMAP cohort⁷. In light of the complex genetic

contributions to LOAD, it is important to understand LOAD GWAS gene involvement in cell-type-specific transcription factor networks that drive the transitions of cells from healthy to Alzheimer's disease states (Alzheimer's disease cell transitions).

Here we applied an unbiased approach using single-nucleus RNA sequencing (DroNc-Seq) to characterize the transcriptional changes and cellular heterogeneity in the entorhinal cortex of Alzheimer's disease patient brains. We identified novel subpopulations of cells that were present only in Alzheimer's disease brains with common and distinct networks of coregulated genes and functions across different cell types. Importantly, as a proof-of-principle for the utility of this data resource, we mapped GWAS genetic data onto these regulatory modules and identified how GWAS genes may functionally influence Alzheimer's disease susceptibility in specific cell subclusters to aid the discovery of possible pathogenic subclusters and cell state transitions underlying Alzheimer's disease. Lastly, we created a web interface (<http://adsn.ddnetbio.com>) for this Alzheimer's disease brain cell atlas, providing a useful resource to dissect mechanisms of cell heterogeneity and dysfunction in Alzheimer's disease.

Results

A molecular survey of the human Alzheimer's disease brain. To investigate cell diversity and disease-related cellular changes in network structure in Alzheimer's disease, we performed DroNc-Seq

¹Department of Anatomy and Developmental Biology, Monash University, Clayton, Victoria, Australia. ²Development and Stem Cells Program, Monash Biomedicine Discovery Institute, Clayton, Victoria, Australia. ³Australian Regenerative Medicine Institute, Monash University, Clayton, Victoria, Australia.

⁴Program in Cardiovascular and Metabolic Disorders, Duke-National University of Singapore Medical School, Singapore, Singapore. ⁵Department of Pathology, The University of Melbourne, Melbourne, Victoria, Australia. ⁶Victorian Brain Bank, Florey Institute of Neuroscience, Parkville, Victoria, Australia. ⁷ARC Center of Excellence in Plant Energy Biology, The University of Western Australia, Perth, Western Australia, Australia. ⁸The Harry Perkins Institute of Medical Research, Perth, Western Australia, Australia. ⁹These authors contributed equally: Alexandra Grubman, Gabriel Chew, John F. Ouyang.

*e-mail: owen.rackham@duke-nus.edu.sg; enrico.petretto@duke-nus.edu.sg; jose.polo@monash.edu

on the 10x platform (Fig. 1a and Extended data Fig. 1). We sampled the entorhinal cortex of six Alzheimer's disease patients and six sex- and age-matched controls (mean age 77.6, range 67.3–91 years; Extended data Fig. 2a and Supplementary Table 1), including a range of *APOE* genotypes (E3/3, E3/4, E4/4, E2/4). We sampled 14,876 cells and, after quality-control filtering (see Methods), we obtained 13,214 cells with a median of 646 detected genes per cell. Visualization of single-nuclei transcriptomes in uniform manifold approximation and projection (UMAP) space was able to separate nuclei into clusters, which we mapped to the six a priori cell types (microglia, astrocytes, neurons, oligodendrocyte progenitor cells (OPCs), oligodendrocytes, and endothelial cells) based on previously established cell-type-specific gene sets (Fig. 1c, see Methods). We defined a cell type score for each gene set and classified cells as 'hybrid' if the two highest cell type scores were within 20% of each other and as 'unidentified' if scores were within the lowest 5% (see Methods for details). Although unlikely, there is a small possibility that, after filtering and quality control (in which doublets were excluded), hybrid cells may still represent doublets. Alternatively, these cells may represent true intermediate cellular states⁸, and hence we retained them for investigation as well as providing the two highest cell scores (Extended data Fig. 2b–e).

In UMAP space, cells from Alzheimer's disease patients segregated from those in controls for each cell type (Fig. 1b). We then identified marker genes for each cell type (Fig. 1d), which included bona fide marker genes such as *HLA-DRA*, *CX3CR1*, *C1QB*, and *CSF1R* for microglia, *AQP4* and *SLC1A2* for astrocytes, *SYT1*, glutamate receptors (for example, *GRIK2*, *GRIA1*, and *GRIN2B*), *RBFOX1* for neurons, *MOBP*, *MBP*, and *PLP1* for oligodendrocytes, *PCDH15* and *MEGF11* for OPCs, and *FLT1* and *CLDN5* for endothelial cells^{9,10}. In addition, our analyses identified several putative novel specific cell-type marker genes (Supplementary Table 2), including *KCNQ3* in microglia, a voltage-gated potassium channel that controls neuronal excitability, previously implicated in benign neonatal epilepsy¹¹. We also identified a novel astrocyte-specific gene, *ADGRV1*, which encodes the nervous system-restricted calcium-binding G-protein-coupled receptor GPR98, mutations in which cause Usher disease¹². Finally, we confirmed that the functional annotation of the cell-type-specific gene sets was largely consistent with the biological function of those cells (Fig. 1e).

We then set out to examine the cell-type-specific gene expression changes between control and Alzheimer's disease nuclei in the context of the existing literature. We first performed a detailed comparison with the recently published single-cell data from human Alzheimer's disease brains⁷. The number of differentially expressed genes (DEGs) found in both studies (Mathys et al. versus our study) was different (1,115 versus 1,726), probably owing to differences between the two studies, including in the phenotypic classification of Alzheimer's disease status (pathological only versus clinical and pathological), the different brain tissues sampled (prefrontal versus entorhinal cortex), and a different number of patients (48 versus 12). However, the DEGs that overlap between the two studies show high concordance (> 90%) of effect (upregulation or downregulation in Alzheimer's disease) across the major cell types (Extended data Figs. 3,4 and Supplementary Table 3). We also investigated the effect of other factors beyond Alzheimer's disease, such as sex, cell type, and individual (Supplementary Tables 4–6). As expected, the highest variance was due to cell type (median variance of 2.9% across all DEGs), and interindividual variability accounted for a proportion of the variance in individual cell types (median variance of 1.6% across all DEGs, Extended data Fig. 5 and Supplementary Table 6). Nevertheless, despite these other factors, Alzheimer's disease accounted for between 29.8% to 84.5% of the detected DEGs in each cell type (Supplementary Table 6).

Our analyses unveiled nine clusters of cell-type-specific and common gene expression patterns between control and Alzheimer's

disease brains (Fig. 1f,g and Extended data Fig. 6). The cell types with the most coordinated gene expression differences between control and Alzheimer's disease patients were astrocytes, endothelial cells, and microglia (further analyses reported in ref. ¹³). These coordinated changes in expression can be summarized as clusters of downregulated or upregulated genes (DEGs) in specific cell types; for example, clusters DEG1 and DEG6 in endothelial cells, DEG2 and DEG9 in astrocytes, and DEG4 and DEG8 in microglia (Fig. 1f). We also found clusters of genes that were coordinated in multiple cell types. For instance, both DEG5 and DEG7 clusters of highly and coordinately upregulated genes in Alzheimer's disease were enriched for genes involved in responses to topologically incorrect protein and cell stress, including mitochondrial, heat shock, and chaperone genes (for example, *MT-ND1–4*, *MT-CO2*, *MT-CO3*, *MT-ATP6*, *HSPA1A*, *HSP90A1*, and *DNAJA1*; Fig. 1f,g), consistent with previous reports in SH-SY5Y cells, primary mouse neurons¹⁴, and late-stage cell type-independent changes in the Alzheimer's disease patient prefrontal cortex⁷. It is possible that the cell-independent responses to topologically incorrect protein may be responses to extracellular amyloid deposition, as tau is only intraneuronal and there is evidence of plaque and oligomer interactions with multiple brain cell types¹⁵. Our analyses highlighted that, in addition to the well-studied cell-type-specific responses (that is, inflammatory responses of astrocytes and microglia and oxidative stress in neurons in Alzheimer's disease^{16,17}), there is a further coordinated response that may act to boost molecular chaperone levels to protect cells against protein misfolding and that may provide additional therapeutic avenues aimed at enhancing endogenous cell type-independent responses.

Endothelial cells in Alzheimer's disease upregulated genes involved in cytokine secretion and immune responses, including *HLA-E*, *MEF2C*, and *NFKBIA*. Endothelial cells and microglia in Alzheimer's disease were enriched in ribosomal processes and translation initiation (for example, *RPS19* and *RPS28*). The DEG2 module of genes, which includes GABA receptors (*GABRA2* and *GABRB1*), glutamate receptors (*GRIA2* and *GRID2*), and neurexin genes (*NRXN1* and *NRXN3*), was enriched in functions related to behavior, cognition, and synapse organization functions and was downregulated in Alzheimer's disease neurons, astrocytes, oligodendrocytes, and OPCs. This is consistent with a previously reported loss of synapse module connectivity in the Alzheimer's disease prefrontal cortex³. We also found that Alzheimer's disease microglia downregulated homeostatic genes (for example, *CX3CR1*, *P2RY12*, and *P2RY13* from DEG4), as has been previously described in multiple Alzheimer's disease mouse models^{13,18,19}. Furthermore, Alzheimer's disease microglia also downregulated genes related to cell–cell adhesion (for example, *CD86* and *CD83*), lipid response (*LPAR6*), and G-protein-coupled receptor pathways (*GPR183* and *LPAR6*). We observed that genes related to glial cell development and differentiation, and in particular the organization and control of myelination, were highly upregulated in Alzheimer's disease neurons, astrocytes, and oligodendrocytes (for example, *BIN1* (ref. ²⁰) and *CNTN2* (ref. ²¹)). These changes in gene expression may also reflect compensatory responses to the myelin loss observed in Alzheimer's disease (reviewed in ref. ²²), although this protective mechanism in APP/PS1 mice was previously not replicated in Alzheimer's disease post-mortem tissue by staining for protective OLIG2⁺ precursors²³. Interestingly, although there is only a 24, 32, and 22 gene overlap between the DEGs in Alzheimer's disease neurons, astrocytes, and oligodendrocytes, respectively, reported here compared to that observed by Mathys et al., that study also found a module of genes in oligodendrocytes correlated to pathology involved in oligodendrocyte differentiation and myelination⁷ (Extended data Figs. 3,4). Oligodendrocytes, astrocytes, OPCs, and endothelial cells in Alzheimer's disease were enriched for genes involved in the negative regulation of cell death, indicating a

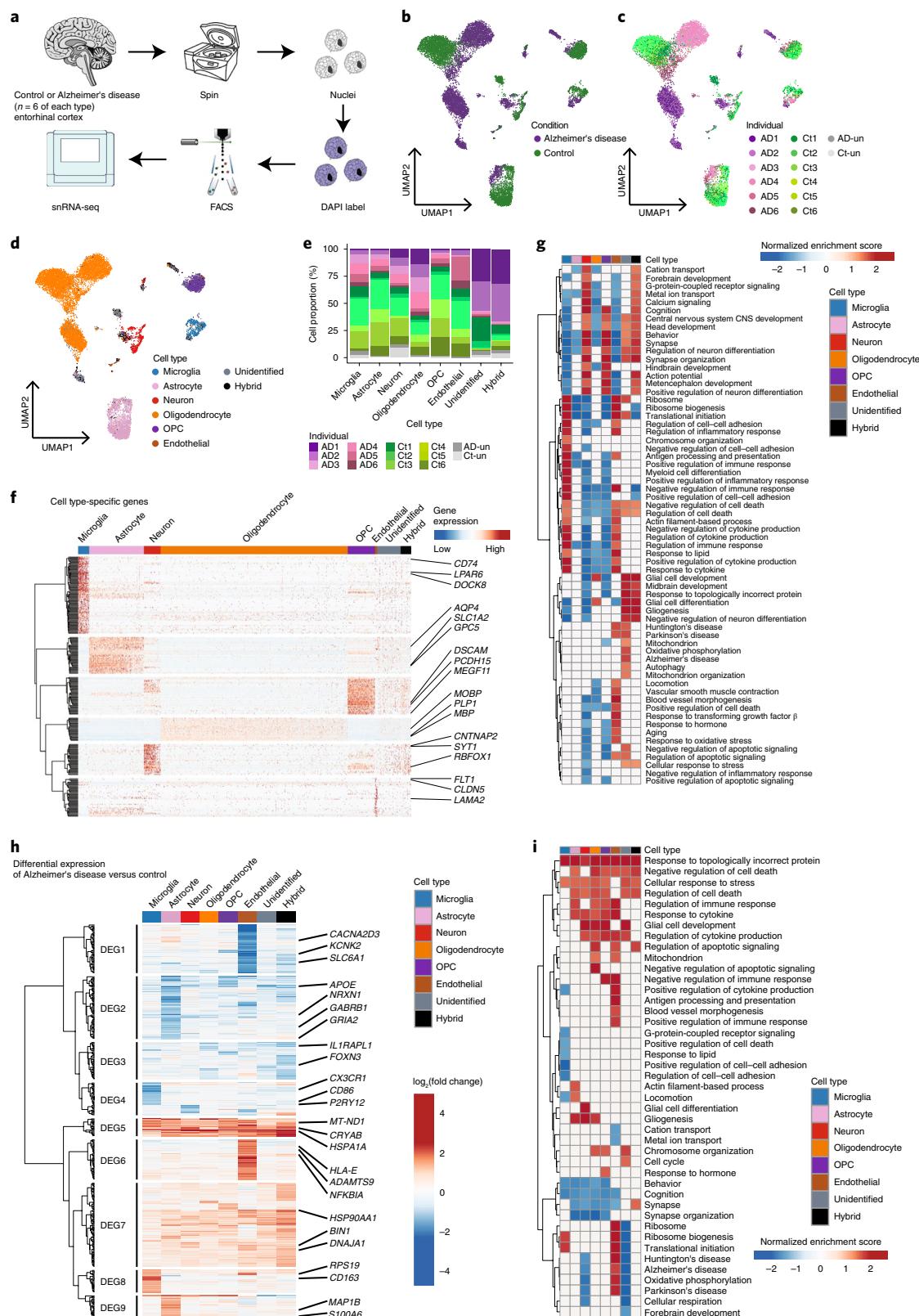


Fig. 1 | Single-nuclei sequencing of human entorhinal cortex recapitulates cell-type-specific marker genes and cell-type-specific changes in Alzheimer's disease. **a**, Schematic of nuclei isolation and RNA-seq workflow. **b-d**, UMAP visualization showing clustering of single nuclei, colored by disease diagnosis (**b**), individuals (**c**), or cell types (**d**), based on our scoring system using the gene sets determined by BRETIGEA⁵¹. **e**, Proportion of cells belonging to each individual in each cell type. **f**, Single-cell gene expression of cell-type-specific genes. **g**, Gene set enrichment analysis (GSEA) of cell-type-specific differential expression colored by statistically significant (FDR < 0.1) normalized enrichment scores. **h**, Hierarchical clustering of log fold change of DEGs (absolute log fold change > 1 and FDR < 0.01) between Alzheimer's disease and control cells for each cell type. The resulting clusters are labeled as DEG1–9. **i**, GSEA results of the differential expression between Alzheimer's disease and control for each cell type, colored by statistically significant (FDR < 0.1) normalized enrichment scores. AD, Alzheimer's disease; Ct, control; un, unidentified.

coordinated response to compensate for cell stress pathways and to protect damaged cells¹⁴.

Interestingly, although endothelial cells in Alzheimer's disease upregulated genes involved in processes related to neurodegeneration (Huntington's, Alzheimer's, and Parkinson's diseases), Alzheimer's disease neurons showed a downregulation of genes associated with neurodegeneration. Furthermore, gene set enrichment analysis indicated that the nuclear-encoded mitochondrial complex I to V genes of the *NDUF*, *SDH*, *UQCR*, *COX*, and *ATP5* families are the main contributors (60% of the genes in the leading edge, Supplementary Table 7) to the enrichment of the Alzheimer's disease gene set in the control neuron subcluster n6. Given the known specific vulnerability of cholinergic neurons in Alzheimer's disease⁵, we mapped our single cells from each neuronal subcluster onto the excitatory (Ex1–8) and inhibitory (In1–8) neuronal signatures described in ref.²⁴ (Extended data Fig. 7). Our results showed downregulation of specific excitatory neuron genes related to synaptic transmission (for example, *SNAP25* and *RIMS1*), and downregulation of ion transport and learning or memory-associated genes that was specific to inhibitory neurons (for example, *CCK*, *SST*, *RELN*, *VIP*, and *KCNIP4*; Extended data Fig. 7). Together these analyses provide a detailed snapshot of both global and cell-type-specific changes in gene expression and functional processes underlying Alzheimer's disease in the human brain.

Subcluster-specific analysis identifies homeostatic, Alzheimer's disease-specific, and shared ontological cell subclusters. As our analyses showed evidence for substantial cellular heterogeneity, we next used the Seurat algorithm to obtain subclusters within each cell type (see Methods for details). This analysis uncovered five microglial clusters, eight astrocyte clusters, six neuronal clusters, six oligodendrocyte clusters, four OPC clusters, and two endothelial clusters (Figs. 2,3a,e,i), each enriched for specific functional categories (Fig. 3d,h,l and Extended data Fig. 8). Except in neurons, in which four clusters were composed of a mixture of Alzheimer's disease and control cells, we found that Alzheimer's disease and control cells mostly segregate into different clusters (Fig. 3b,f,j), suggesting strong and penetrant disease-associated transcriptional changes across almost every cell type.

One of the Alzheimer's disease astrocyte subclusters, a1, was located closer to oligodendrocytes than to astrocytes on the global UMAP space (Fig. 2a). This a1 Alzheimer's disease astrocyte subcluster was enriched for ribosomal, mitochondrial, neuron differentiation, and heat shock responses, whereas the a2 Alzheimer's disease astrocyte subcluster showed downregulation of these processes (Fig. 3d) and was instead enriched for transforming growth factor (TGF) β signaling and immune responses (refer to our web resource for full list of gene set enrichment). Neither the a1 nor the a2 Alzheimer's disease astrocyte subcluster described here overlapped significantly with the A1 or A2 astrocyte profiles previously described by Liddelow et al. in mouse Alzheimer's disease brains, although the marker gene *C3* was upregulated in Alzheimer's disease astrocytes as previously reported¹⁷ (Extended data Fig. 9a), specifically in the presently defined a2 subcluster. In healthy patient brains, molecularly distinct astrocyte subclusters showed evidence of functional diversity, as a4 was enriched in respiratory and mitochondrial processes, and a3 and a8 were enriched in cellular responses to lipids and hormones, whereas a6 was enriched in synapse organization, action potentials, and ion channel activity²⁵ (Extended data Fig. 9a).

We next investigated subcluster-specific changes occurring in Alzheimer's disease. We observed some instances of genes implicated in both neurological and psychiatric disorders showing subcluster-specific differential expression in Alzheimer's disease. For example, *LINGO1*, which encodes a signaling protein that inhibits myelination²⁶, was strongly upregulated specifically in

several Alzheimer's disease subclusters (a1–a2, m1, n1, o1, and o3; Fig. 3c,g,k). In agreement, Mathys and collaborators recently reported an increase in *LINGO1* in Alzheimer's disease excitatory neurons and oligodendrocytes⁷ (Supplementary Table 3). Similarly, *NEAT1*, a regulatory non-coding RNA that is found increased in the serum of patients with relapsing-remitting multiple sclerosis²⁷, was highly expressed in a1, a2, and o2 Alzheimer's disease subclusters (Fig. 3c,k). Conversely, the schizophrenia-linked glutamate transporter encoded by *GRID1* (ref.²⁸) was specifically increased in o2 but downregulated in the a1 subcluster (Fig. 3k).

Beyond subcluster-specific gene expression changes in Alzheimer's disease, our data also provide insights into the molecular nature of cellular heterogeneity in Alzheimer's disease patient brains. With respect to neuronal subclusters mapped onto excitatory and inhibitory neurons (Extended data Fig. 7), we found that clusters n2–n5 contained cells from both control and Alzheimer's disease patient brains that separated according to functional excitatory or inhibitory identity. Cluster n2 comprised excitatory neurons corresponding to a mixture of Ex1–7 layer II–VI cortical neurons²⁴, whereas clusters n3–n5 each mapped to specific sets of interneuron subclusters (Extended data Fig. 9b). The n4 subcluster corresponded to PVALB⁺ and SST⁺ layer IV–VI (In6–8) cells. The n5 subcluster mapped to layer I, II, and VI VIP⁺ interneurons (In1–3), and the n3 subcluster mapped to a mixture of NDNF⁺ and CCK⁺ interneuron subsets (In1, In4–5). Conversely, clusters n1 and n6 separated according to disease status and contained a mixture of inhibitory and excitatory neurons, mostly corresponding to Ex3 layer IV excitatory neurons and CALB⁺, PVALB⁺, and SST⁺ In6–8 interneurons. The Alzheimer's disease n1 subcluster was enriched for autophagy and responses to various stimuli including hormones, lipids, and topologically incorrect protein (Fig. 3e–h). The control n6 subcluster was enriched in ribosomal (*RPS* and *RPL* families) and oxidative phosphorylation pathways (mitochondrial complex I–V genes), suggestive of high energy requirements and usage by these neurons²⁹. Together, these data indicate that the molecular identity of specific neuronal subsets is more susceptible to Alzheimer's disease-related changes, specifically responses to foreign stimuli, consistent with the reported resistance of deeper-layer neurons (layer V–VI) to the toxic effects of amyloid³⁰.

In the case of oligodendrocytes, our data revealed that, although the control oligodendrocyte subclusters (o5 and o6) are composed of a mixture of cells originating from all samples (Fig. 3i,j), the Alzheimer's disease-specific o1–3 subclusters distinctly cluster by sample (Fig. 3i–l).

Cell subcluster-specific activity and regulation of GWAS candidate genes in the human Alzheimer's disease brain. Large-scale genetic screening studies of Alzheimer's disease susceptibility, such as GWAS and exome sequencing, have provided a large repertoire of pathological pathways and candidate gene targets^{2,31}, and have highlighted the microglial genetic vulnerability in Alzheimer's disease^{1,32}, spurring an entire field of inquiry. However, how they relate to specific cell type activity has remained unknown. Here, we leveraged our single-cell data resource to examine the cell type specificity of the expression of about 1,000 GWAS candidate genes for Alzheimer's disease and Alzheimer's disease-related traits (Supplementary Table 8). To assess cell type specificity, we calculated the specificity score of each GWAS gene associated with Alzheimer's disease, LOAD, Alzheimer's disease biomarkers, and Alzheimer's disease neuropathological change. Briefly, the specificity score is a measure of the proportion of the expression of a particular gene across all cell types. Therefore, a high gene specificity score for a cell type indicates a high proportion of the expression of that gene being captured in that particular cell type. First, corroborating earlier evidence, several previously described microglia-specific GWAS genes, including *INPP5D*, *HLA-DRB5*, *PLCG2*,

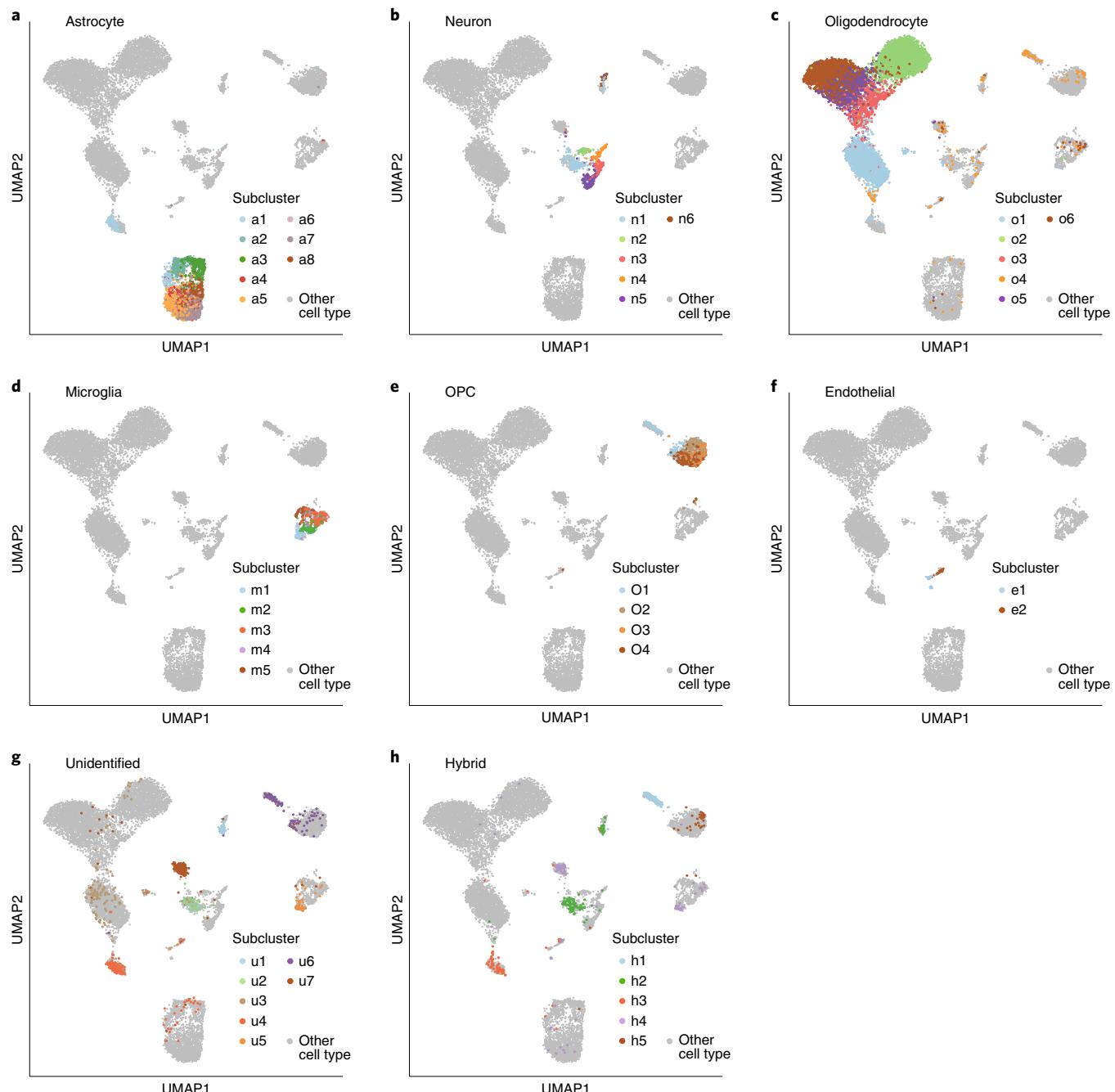


Fig. 2 | Single-nuclei sequencing of human entorhinal cortex uncovers high cellular heterogeneity with each cell type. Global UMAP visualization showing the location of cells within each cell subcluster of astrocytes (a), neurons (b), oligodendrocytes (c), microglia (d), OPCs (e), endothelial cells (f), unidentified cells (g), and hybrid cells (h). The cell types were determined previously in Fig. 1d.

HLA-DRB1, *CSF3R*, and *MS4A6A*, showed highly specific expression in microglia (Fig. 4a, left panel). In addition, we detected two microglia-specific GWAS genes not previously associated with microglia, *RIN3* and *TBXAS1* (Fig. 4a, left panel). *RIN3* is involved in microglial endocytosis and interacts with *BIN1* and *CD2AP*, indicating possible involvement in APP trafficking³³. *TBXAS1* is a potent vasoconstrictor in cerebral circulation and has been linked to ischemic stroke³⁴. Notably, *TBXAS1* is also the target of two FDA-approved drugs for stroke and peripheral vascular disease, including dipyridamole³⁵, which was shown to inhibit amyloid-induced microglial inflammation³⁶. Therefore, our data prompt further detailed functional investigations in a microglial model to study the

contribution of the gene to Alzheimer's disease, potentially in the context of cardiovascular risk³⁷.

Functionally, LOAD-associated GWAS genes were enriched for inflammatory, phagocytic, or endocytic pathways, lipid metabolism, and synaptic and axonal function². A number of neuron-specific GWAS genes have been linked to neuronal migration (for example, *NRG1*) and synaptic transmission or adhesion (for example, *GABRG3* and *PCDH8*, Fig. 4a). Interestingly, two of the neuronal-associated GWAS genes are involved in RNA splicing pathways (*SRRM4* and *ELAVL2*, Fig. 4a, left panel), as aberrant splicing was recently shown to be associated with Alzheimer's disease in a combined aging cohort of 450 patients³⁸.

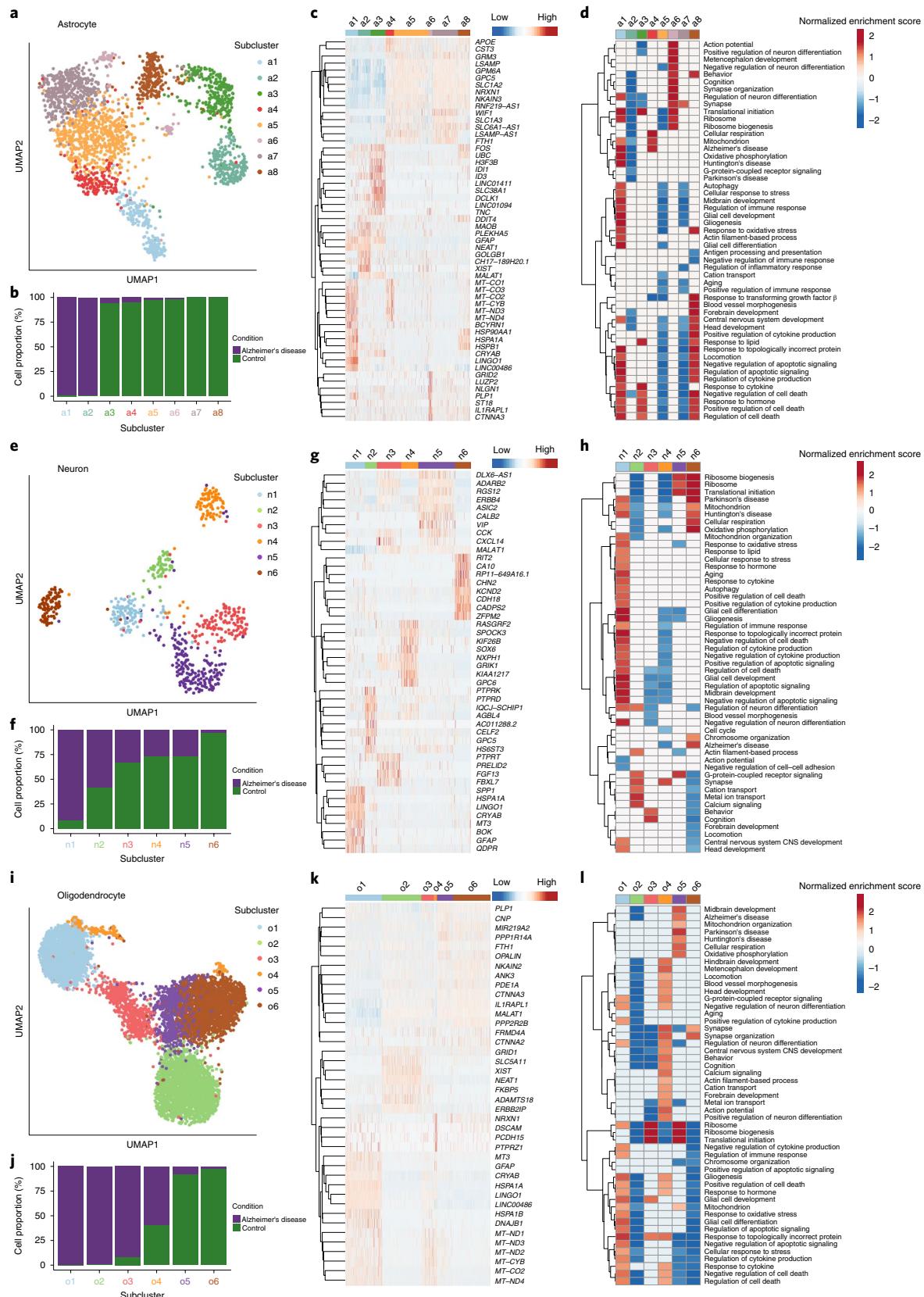


Fig. 3 | Single-nuclei sequencing of human Alzheimer's disease and control entorhinal cortex reveals homeostatic, Alzheimer's disease-specific, and shared ontological cell subclusters. **a,e,i, UMAP visualization of subclusters of astrocytes (**a**), neurons (**e**), and oligodendrocytes (**i**), showing the composition of cells in subclusters by disease state (**b,f,j**); hierarchical clustering and heatmap colored by single-cell gene expression of subcluster-specific genes (top eight genes ordered by their log fold change shown per cluster) (**c,g,k**). **d,h,l**, GSEA of subcluster-specific differential expression, colored by statistically significant (FDR < 0.1) normalized enrichment scores for selected gene ontologies shown in each cell subcluster.**

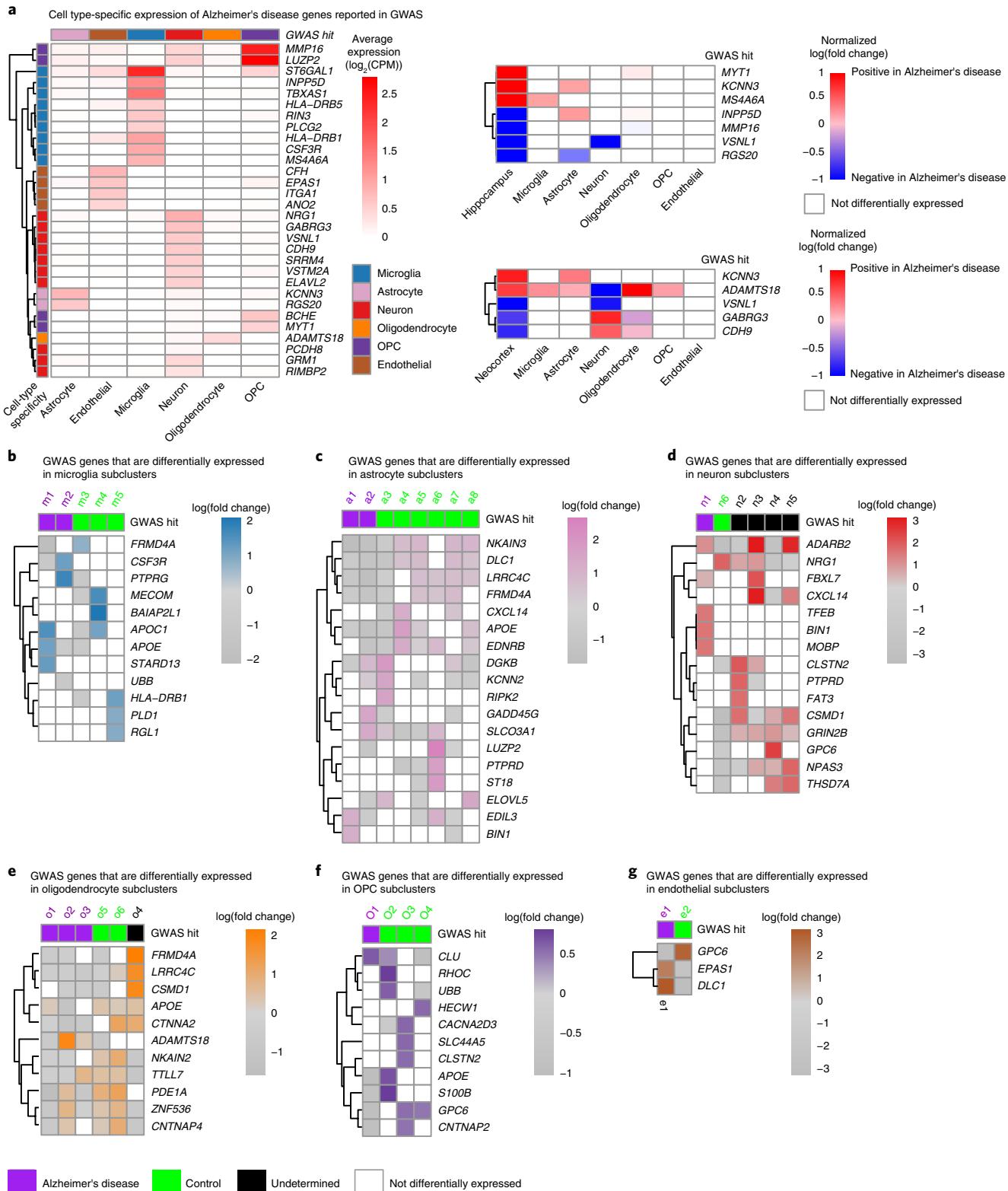


Fig. 4 | Alzheimer's disease genes identified by GWAS show specific gene expression patterns across cell types and cell-type subclusters. **a**, Left: expression of cell-type-specific GWAS hits ($n=30$ genes, >99 th percentile for specificity). Right four panels: comparison of cell-type specificity of GWAS DEGs identified by previous bulk tissue microarray studies between control and Alzheimer's disease brains. **b-g**, Heatmaps showing the log fold change of significant ($\text{FDR} < 0.05$) differential expression between Alzheimer's disease ($n=6$) and control individuals ($n=6$) for GWAS genes within subclusters with respect to each cell type; for each subcluster, the top three GWAS genes based on absolute log fold change were chosen for visualization; microglia (**b**), astrocytes (**c**), neurons (**d**), oligodendrocytes (**e**), OPCs (**f**), or endothelial cells (**g**). Subclusters were considered as Alzheimer's disease or control subclusters if more than 80% of cells in that subcluster were from either Alzheimer's disease or control brains, respectively; otherwise they were considered undetermined. For each cell type, differential expression was performed using the empirical Bayes QLF test and DEGs were determined by $\text{FDR} < 0.05$. CPM, counts per million.

Comparison of our data with previous bulk microarray data from the hippocampus and neocortex of Alzheimer's disease patients^{39,40} (Fig. 4a, right panels) revealed that *ADAMTS18*, a GWAS hit previously identified as specific to oligodendrocytes, was upregulated in Alzheimer's disease microglia, astrocytes, oligodendrocytes, and OPCs, but downregulated in neurons. Other concordant changes of GWAS genes were dominated by a single cell type; for instance, upregulation of *MYT1* in Alzheimer's disease oligodendrocytes, *MS4A6A* in Alzheimer's disease microglia, and *KCNN3* in Alzheimer's disease astrocytes, as well as downregulation of *VSNL1* in Alzheimer's disease neurons and *RGS20* in Alzheimer's disease astrocytes. Not unexpectedly, a large majority of previously identified GWAS DEGs between Alzheimer's disease and control patient brains displayed cell-type-specific expression patterns in our data, further highlighting the value of single-cell analyses in human Alzheimer's disease.

We further explored the functional relevance of GWAS genes using two paradigms: we first examined subcluster-specific changes in the expression of top Alzheimer's disease GWAS genes (Fig. 4b–g and Supplementary Table 9), and we integrated GWAS genes into transcription factor-driven regulatory modules to understand the drivers of cell transitions from a healthy state into an Alzheimer's disease state (Fig. 5 and Extended data Fig. 10). Importantly, our single-cell data allowed us to uncover sets of GWAS genes with divergent cell-type-specific expression in Alzheimer's disease cell subclusters. For instance, we observed that *APOE* was downregulated in Alzheimer's disease OPC (O1), oligodendrocyte (o2), and astrocyte subclusters (a1 and a2), whereas it was upregulated in a microglial Alzheimer's disease subcluster (m1), consistent with increased microglial *APOE* and reduced astrocytic *APOE* expression in the Alzheimer's disease prefrontal cortex⁷ (Supplementary Table 3). Our results are also in line with recent reports that reduced *APOE* expression in *APOE4* isogenic induced pluripotent stem cells (iPSC)-derived astrocytes resulted in impaired lysosomal functions and amyloid clearance⁴¹, while increased *Apoe* expression in microglia has been associated with Alzheimer's disease phenotypes^{13,18,19}. To our knowledge, no previous study has examined the effects of *APOE* expression or genotype on oligodendrocyte or OPC function. However, given the important role of oligodendrocytes in the synthesis and secretion of cholesterol, it is possible that loss of endogenous *APOE* expression in OPC or oligodendrocyte subclusters may impair myelination. Indeed MRI analyses in healthy patients showed an allele-dependent effect of *APOE* on myelin breakdown⁴². Together, our data uncover cell subcluster-specific expression patterns of *APOE* in human Alzheimer's disease brains, the combined and cell-specific consequences of which warrant further examination in iPSC-derived and mouse models of Alzheimer's disease pathogenesis.

We also detected instances in which GWAS genes showed coordinated upregulation or downregulation in multiple Alzheimer's disease cell subclusters. After *APOE*, the most strongly associated genetic risk factor for LOAD is *BIN1*, which was previously reported to be increased in Alzheimer's disease brains¹³. We found that *BIN1* was specifically increased in Alzheimer's disease subclusters a1 and n1 (Fig. 4c,d), therefore identifying the exact cell subpopulations. Both clusters were enriched for genes relating to autophagy and responses to incorrectly folded proteins (Fig. 3d,h), in line with the reported direct interaction with tau⁴³. However, *BIN1* has ten transcripts, the relative abundances of which cannot be resolved from 3' biased 10x data. Different *BIN1* isoforms exhibit specific functions in endocytosis, calcium signaling, and apoptosis, and distinct expression patterns of these transcripts were associated with greater or reduced amyloid load in Alzheimer's disease patients⁴⁴. *FRMD4A*, a LOAD GWAS gene⁴⁵ that we found to be expressed in all cell types, showed concordant downregulation in subclusters of

Alzheimer's disease patient cells in microglia, astrocytes, and oligodendrocytes (Fig. 4b,c,e).

An intriguing observation is that control subclusters also displayed higher expression levels of Alzheimer's disease GWAS genes (that is, *APOE* was upregulated in a4/a8, *APOC1* was upregulated in m4, and *HLA-DRB1* was upregulated in m5), which may represent differential susceptibility of particular functional subpopulations to gene variants identified by GWAS (Fig. 4b,c). Together, these data highlight the advantage of studying single-cell data to understand the effect of disease gene variants on cell subtype-specific genetic susceptibility, and may explain why conventional (whole-body) rather than conditional or cell-type-specific gene knockouts in Alzheimer's disease models have often yielded discrepant results.

The master lysosomal regulator TFEB drives a network of GWAS genes that controls cell state transition to Alzheimer's disease in astrocytes. We next systematically examined the transcription factor dynamics underlying subcluster-specific cell state transitions towards Alzheimer's disease. Using CellRouter we predicted transcription factors that drive transitions from control to Alzheimer's disease subclusters, building gene regulatory networks (GRNs). Subsequently, GRN scores were calculated to assess how well a transcription factor and its downstream targets correlate with the identified trajectory. Our analysis unveiled common transcription factors predicted to control multiple transitions within a cell type, such as *AEBP1* (Fig. 5a), with previously reported upregulation in Alzheimer's disease brains⁴⁶ and association with amyloid pathology⁴⁷. *AEBP1* displayed a high GRN score for transitions from every control astrocyte subcluster towards a1 Alzheimer's disease astrocytes. Our analysis identified several other transcription factors with high GRN scores for multiple lineage conversions to Alzheimer's disease subclusters, including *SOX10*, *MYRF*, and *NKX6-2*, predicted to regulate the generation of n1, o1/o2, and a1 subclusters (Fig. 5a), with functional enrichment of glial and neuronal development, myelination, and the unfolded protein response, respectively (Extended data Fig. 9c). CellRouter did not detect any transcription factors for the endothelial subpopulation.

We set out to investigate the relationships between transcription factors and their regulation of GWAS genes associated with Alzheimer's disease, highlighting relevant GRNs containing GWAS genes as downstream targets (Extended data Fig. 10). These GRNs had the highest average log fold change across their GWAS-associated downstream targets. Integration of transcription factor-driven regulatory modules with downstream GWAS gene targets revealed subcluster-specific regulation of Alzheimer's disease genetic susceptibility. For example, we discovered that *HIF3A*, a transcription factor that inhibits hypoxia-induced gene expression⁴⁸, regulates multiple Alzheimer's disease GWAS genes (*NPAS3*, *BIN1*, *MOBP*, *CLDN11*, and *ADARB2*) and drives cell state transitions from n6 to n1. In addition, *HIF3A* was also involved in the transitions towards o1, a1–a2, m1, and O1 subclusters (Extended data Fig. 10). However, different functional gene networks downstream of *HIF3A* were predicted for other transitions (Extended data Fig. 9c). Our analyses also showed that the *TFEB* gene, which was upregulated in diseased astrocytes⁴⁹, acts upstream of ten GWAS loci for Alzheimer's disease (*BIN1*, *CLDN11*, *POLN*, *STK32B*, *EDIL3*, *AKAP12*, *HECW1*, *WDR5*, *LEMD2*, and *DLC1*), which are also dysregulated in Alzheimer's disease astrocytes (Fig. 5b). Notably, this upregulated regulatory module, which is enriched in chaperone-mediated responses (Fig. 5c), underlies the cellular transitions from control to Alzheimer's disease states in specific astrocyte subpopulations, which had otherwise been missed in bulk tissue expression data alone. The full list of transcription factors and downstream target genes is accessible via <http://adns.ddnetbio.com>. These results establish a functional link between specific astrocyte subpopulations and Alzheimer's disease, and

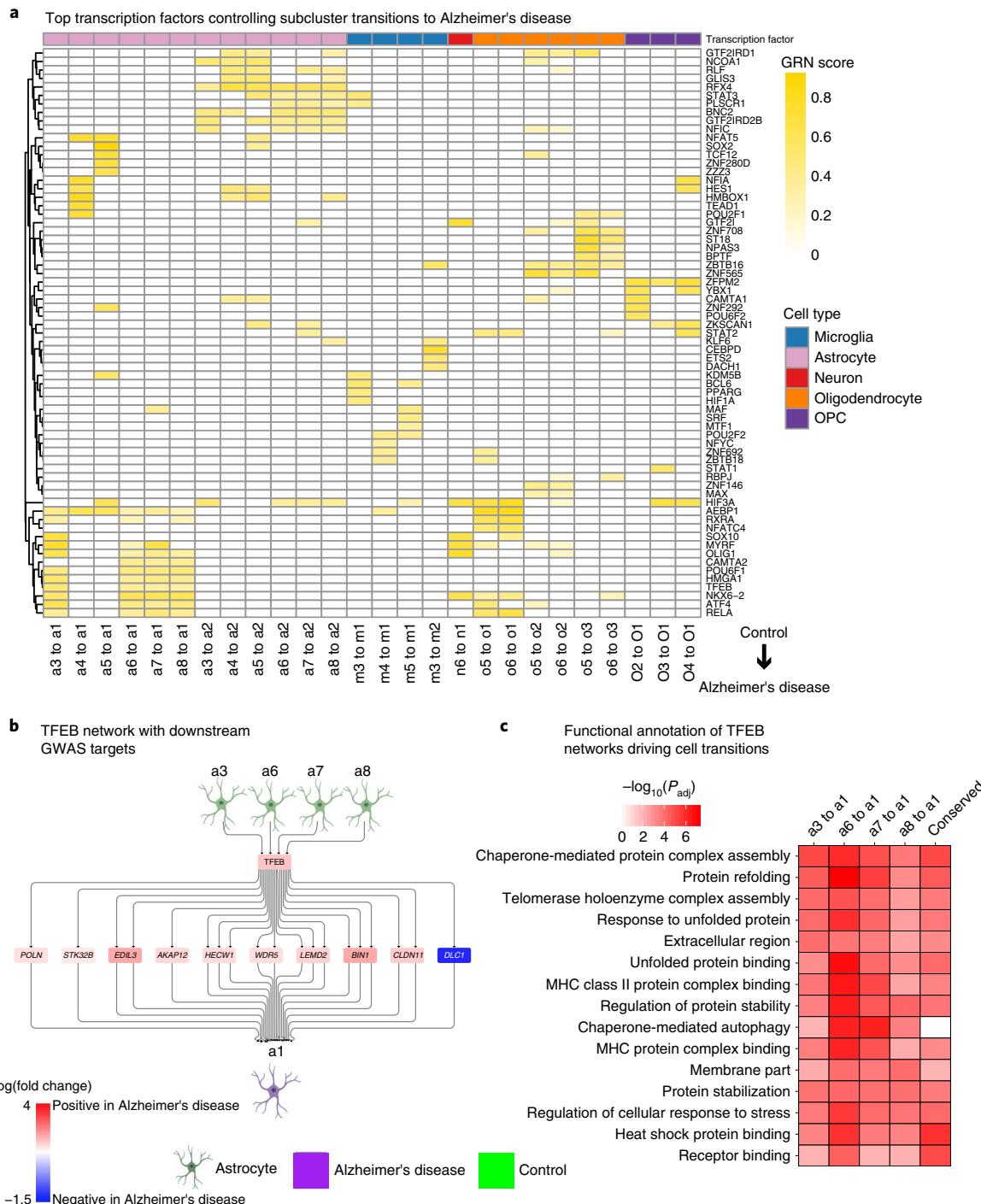


Fig. 5 | GRN analysis predicts transcription factors for conversion of control to Alzheimer's disease subcluster signatures. **a**, Heatmap of scores from top transcription factor regulons with respect to each cell transition from control subclusters to Alzheimer's disease subclusters. Note that some transitions, such as m4 to m2 and m5 to m2, are not shown, as no significant transcription factor regulons were identified after post-processing of CellRouter outputs (see Methods). **b**, TFEB network with downstream GWAS targets. **c**, Top functional enrichment terms of TFEB networks driving cell transitions for transitions a3 to a1 ($n=139$ genes), a6 to a1 ($n=57$ genes), a7 to a1 ($n=166$ genes), and a8 to a1 ($n=126$ genes), as well as genes that conserved all of the four aforementioned trajectories ($n=37$ genes). Multiple testing correction was done using the Benjamini-Hochberg method; P_{adj} , adjusted P value. Enrichment analysis was performed using the hypergeometric test.

suggest that the contribution of multiple GWAS loci to Alzheimer's disease susceptibility is coordinated and mediated by TFEB activity in these astrocytes.

An online single-cell atlas of human Alzheimer's disease brain transcriptomes. To facilitate Alzheimer's disease researchers to

access and mine our data resource beyond the standard Gene Expression Omnibus (GEO) data repository, we provide an interactive online tool (<http://adsn.ddnetbio.com>) to visualize and interrogate our dataset with numerous explorative analyses, including the ability to display the expression of any gene of interest in all individual cells, to examine cell identity, cell heterogeneity, and the

differential gene expression and ontologies between subclusters, and to find novel markers of cell types and subclusters.

Discussion

We generated a comprehensive data resource for single-cell transcriptomic analysis of Alzheimer's disease patient brains, providing a unique resource for future studies seeking to understand cellular heterogeneity and define functional changes at single-cell resolution in the human Alzheimer's disease brain. To demonstrate the utility of this resource in helping to uncover the genetic mechanism underlying cellular heterogeneity and function in Alzheimer's disease, we have provided several examples of applications and data analyses, including integration with external datasets related to genetic susceptibility to Alzheimer's disease. Importantly, by doing this, we have been able to gain several novel insights into cell subtype-specific regulation and cell identity changes in Alzheimer's disease.

The identification of transcription factors that orchestrate the conversion of control to Alzheimer's disease cell signatures in the different cell populations of the human brain can pinpoint specific molecular processes and mechanisms for therapeutic intervention. Given the cellular complexity observed in Alzheimer's disease pathophysiology, knowledge of the specific cell population (or populations) in which a given regulatory program is operating to drive disease will aid the design of more robust and effective (that is, cell-type-specific) cellular studies and therapeutic approaches.

The many disease-associated genes identified by GWAS remain largely orphan with regards to a specific functional context in human Alzheimer's disease brain. Our data and analysis allowed us to provide direct insights into the cells and ontological cell subclusters in which these disease genes operate in the human Alzheimer's disease brain. Notably, our data reveal complex patterns of expression changes for multifold Alzheimer's disease genes within or across specific cell populations. This complexity should be taken into account to enhance our interpretation of genetic discoveries in Alzheimer's disease. For example, our data on cell-type-specific expression of GWAS genes will prompt expression quantitative trait loci (eQTL)¹ analyses tailored to specific cell populations, to inform functional specialization of Alzheimer's disease gene variants and regional (that is, cell-type-specific) susceptibility to disease.

Finally, our data resource allowed us to explore genetic contributors to Alzheimer's disease beyond analysis of single gene effects in specific cell types or subpopulations. Specifically, integrated gene networks and GWAS data at the single-cell level to detail the coordinated (transcription factor-driven) contribution of multiple GWAS genes to Alzheimer's disease susceptibility in specific cellular transitions from control to Alzheimer's disease states. For example, we show how the TFEB transcription factor, which is upregulated in diseased astrocytes, acts upstream of ten GWAS loci for Alzheimer's disease that are also dysregulated in astrocytes from the brains of Alzheimer's disease patients. These results provide insights into disease networks in the human brain, and uncover molecular interactors that relate to driving nodes (for example, transcription factors) connecting pathways to genetic susceptibility (for example, GWAS genes) to disease in specific cell populations. Therefore, the specific differential expression patterns shown here for Alzheimer's disease risk genes help to elucidate the underlying cellular context and regulatory mechanisms, and can inform targeted functional studies to regulate these genes in specific subcellular compartments in the Alzheimer's disease brain.

To quantify the effect of interindividual genetic variation and other factors (for example, age), larger population cohorts than those used in our study and that by Mathys et al. need to be considered. Despite this, comparing our dataset with that of Mathys et al. supports the replicability of single nuclear RNA sequencing (snRNA-seq) experiments across studies and different human brain tissues. Intriguingly, our data indicate the possible conservation

of subcellular changes in Alzheimer's disease in both the earliest affected entorhinal³⁰ and later affected prefrontal cortices. In addition, corroborating the finding by Mathys et al of the disruption of myelin processes in Alzheimer's disease, we also found that, out of all major brain cell types, oligodendrocytes have the greatest inter-individual differences between Alzheimer's disease patients, further implicating myelination in Alzheimer's disease pathogenesis.

Our assessment of interindividual variability showed that, within each cell type, detection of some DEGs was also explained by individual or sex differences (Extended data Fig. 5). Therefore, although our study contributes significantly to our understanding of the transcriptional changes underpinning cellular heterogeneity and changes in Alzheimer's disease, examining larger patient cohorts in the future will enable assessment of the effects and relative contribution of underlying genetic factors to the described cellular and transcriptomic changes in disease. We anticipate that our resource will stimulate and allow further discoveries in many different areas, as exemplified by our work.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41593-019-0539-4>.

Published online: 25 November 2019

References

- Huang, K.-L. et al. A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease. *Nat. Neurosci.* **20**, 1052–1061 (2017).
- Lambert, J. C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
- Zhang, B. et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**, 707–720 (2013).
- Füger, P. et al. Microglia turnover with aging and in an Alzheimer's model via long-term *in vivo* single-cell imaging. *Nat. Neurosci.* **20**, 1371–1376 (2017).
- Whitehouse, P. J. et al. Alzheimer's disease and senile dementia: loss of neurons in the basal forebrain. *Science* **215**, 1237–1239 (1982).
- Zhong, S. et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).
- Mathys, H. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
- Wilhelmsen, U. et al. Injury leads to the appearance of cells with characteristics of both microglia and astrocytes in mouse and human brain. *Cereb. Cortex* **27**, 3360–3377 (2017).
- Zeisel, A. et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
- Darmanis, S. et al. A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl Acad. Sci. USA* **112**, 7285–7290 (2015).
- Miceli, F. et al. KCNQ3-Related Disorders. In *Gene Reviews* (ed. Adam, M. P.) <https://www.ncbi.nlm.nih.gov/books/NBK201978/> (Univ. Washington, 2014).
- Ebermann, I. et al. GPR98 mutations cause Usher syndrome type 2 in males. *J. Med. Genet.* **46**, 277–280 (2009).
- Grubman, A., Choo, X. Y., Chew, G., Ouyang, J. F. & Sun, G. Mouse and human microglial phenotypes in Alzheimer's disease are controlled by amyloid plaque phagocytosis through Hif1α. Preprint at *biorXiv* <https://www.biorxiv.org/content/10.1101/639054v1> (2019).
- Veereshwarayya, V., Kumar, P., Rosen, K. M., Mestril, R. & Querfurth, H. W. Differential effects of mitochondrial heat shock protein 60 and related molecular chaperones to prevent intracellular β-amyloid-induced inhibition of complex IV and limit apoptosis. *J. Biol. Chem.* **281**, 29468–29478 (2006).
- De Strooper, B. & Karan, E. The cellular phase of Alzheimer's disease. *Cell* **164**, 603–615 (2016).
- Xie, H. et al. Rapid cell death is preceded by amyloid plaque-mediated oxidative stress. *Proc. Natl Acad. Sci. USA* **110**, 7904–7909 (2013).
- Liddelow, S. A. et al. Neurotoxic reactive astrocytes are induced by activated microglia. *Nature* **541**, 481–487 (2017).
- Krasemann, S. et al. The TREM2-APOE pathway drives the transcriptional phenotype of dysfunctional microglia in neurodegenerative diseases. *Immunity* **47**, 566–581.e9 (2017).

19. Keren-Shaul, H. et al. A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**, 1276–1290.e17 (2017).
20. De Rossi, P. et al. Predominant expression of Alzheimer's disease-associated BIN1 in mature oligodendrocytes and localization to white matter tracts. *Mol. Neurodegener.* **11**, 59 (2016).
21. Savvaki, M. et al. The expression of TAG-1 in glial cells is sufficient for the formation of the juxtaparanodal complex and the phenotypic rescue of tag-1 homozygous mutants in the CNS. *J. Neurosci.* **30**, 13943–13954 (2010).
22. Bartzikis, G. Age-related myelin breakdown: a developmental model of cognitive decline and Alzheimer's disease. *Neurobiol. Aging* **25**, 5–18 (2004).
23. Behrendt, G. et al. Dynamic changes in myelin aberrations and oligodendrocyte generation in chronic amyloidosis in mice and men. *Glia* **61**, 273–286 (2013).
24. Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
25. John Lin, C.-C. et al. Identification of diverse astrocyte populations and their malignant analogs. *Nat. Neurosci.* **20**, 396–405 (2017).
26. Mi, S. et al. LINGO-1 negatively regulates myelination by oligodendrocytes. *Nat. Neurosci.* **8**, 745–751 (2005).
27. Santoro, M. et al. Expression profile of long non-coding RNAs in serum of patients with multiple sclerosis. *J. Mol. Neurosci.* **59**, 18–23 (2016).
28. Fallin, M. D. et al. Bipolar I disorder and schizophrenia: a 440-single-nucleotide polymorphism screen of 64 candidate genes among Ashkenazi Jewish case-parent trios. *Am. J. Hum. Genet.* **77**, 918–936 (2005).
29. Hall, C. N., Klein-Flügge, M. C., Howarth, C. & Attwell, D. Oxidative phosphorylation, not glycolysis, powers presynaptic and postsynaptic mechanisms underlying brain information processing. *J. Neurosci.* **32**, 8940–8951 (2012).
30. Romito-DiGiacomo, R. R., Menegay, H., Cicero, S. A. & Herrup, K. Effects of Alzheimer's disease on different cortical layers: the role of intrinsic differences in Abeta susceptibility. *J. Neurosci.* **27**, 8496–8504 (2007).
31. Bis, J. C. et al. Whole exome sequencing study identifies novel rare and common Alzheimer's-associated variants involved in immune response and transcriptional regulation. *Mol. Psychiatry* <https://doi.org/10.1038/s41380-018-0112-7> (2018).
32. Sims, R. et al. Rare coding variants in *PLCG2*, *ABI3*, and *TREM2* implicate microglial-mediated innate immunity in Alzheimer's disease. *Nat. Genet.* **49**, 1373–1384 (2017).
33. Kajihara, H. et al. Characterization of RIN3 as a guanine nucleotide exchange factor for the Rab5 subfamily GTPase Rab31. *J. Biol. Chem.* **286**, 24364–24373 (2011).
34. He, Z.-Y., Li, L., Wang, Y.-Z., Liu, X. & Yuan, L.-Y. Associations between thromboxane A synthase 1 gene polymorphisms and the risk of ischemic stroke in a Chinese Han population. *Neural Regen. Res.* **13**, 463 (2018).
35. Diener, H. C. et al. European Stroke Prevention Study 2. Dipyridamole and acetylsalicylic acid in the secondary prevention of stroke. *J. Neurol. Sci.* **143**, 1–13 (1996).
36. Paris, D. et al. Inhibition of Alzheimer's beta-amyloid induced vasoactivity and proinflammatory response in microglia by a cGMP-dependent mechanism. *Exp. Neurol.* **157**, 211–221 (1999).
37. Alzheimer's Association. 2018 Alzheimer's disease facts and figures. *Alzheimers Dement.* **14**, 367–429 (2018).
38. Raj, T. et al. Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nat. Genet.* **50**, 1584–1592 (2018).
39. Tan, M. G. et al. Genome wide profiling of altered gene expression in the neocortex of Alzheimer's disease. *J. Neurosci. Res.* **88**, 1157–1169 (2010).
40. Blalock, E. M., Buechel, H. M., Popovic, J., Geddes, J. W. & Landfield, P. W. Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease. *J. Chem. Neuroanat.* **42**, 118–126 (2011).
41. Lin, Y.-T. et al. APOE4 causes widespread molecular and cellular alterations associated with Alzheimer's disease phenotypes in human iPSC-derived brain cell types. *Neuron* **98**, 1294 (2018).
42. Bartzikis, G. et al. Apolipoprotein E genotype and age-related myelin breakdown in healthy individuals: implications for cognitive decline and dementia. *Arch. Gen. Psychiatry* **63**, 63–72 (2006).
43. Chapuis, J. et al. Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. *Mol. Psychiatry* **18**, 1225–1234 (2013).
44. Yu, L. et al. Association of brain DNA methylation in *SORL1*, *ABCA7*, *HLA-DRB5*, *SLC24A4*, and *BIN1* with pathological diagnosis of Alzheimer disease. *JAMA Neurol.* **72**, 15–24 (2015).
45. Lambert, J.-C. et al. Genome-wide haplotype association study identifies the *FRMD4A* gene as a risk locus for Alzheimer's disease. *Mol. Psychiatry* **18**, 461–470 (2013).
46. Hokama, M. et al. Altered expression of diabetes-related genes in Alzheimer's disease brains: the Hisayama study. *Cereb. Cortex* **24**, 2476–2488 (2014).
47. Shijo, M. et al. Association of adipocyte enhancer-binding protein 1 with Alzheimer's disease pathology in human hippocampi. *Brain Pathol.* **28**, 58–71 (2018).
48. Maynard, M. A. et al. Human HIF-3α is a dominant-negative regulator of HIF-1 and is down-regulated in renal cell carcinoma. *FASEB J.* **19**, 1396–1406 (2005).
49. Martini-Stoica, H. et al. TFEB enhances astroglial uptake of extracellular tau species and reduces tau spreading. *J. Exp. Med.* **215**, 2355–2377 (2018).
50. de Toledo-Morrell, L., Goncharova, I., Dickerson, B., Wilson, R. S. & Bennett, D. A. From healthy aging to early Alzheimer's disease: in vivo detection of entorhinal cortex atrophy. *Ann. N. Y. Acad. Sci.* **911**, 240–253 (2000).
51. McKenzie, A. T. et al. Brain cell type specific gene expression and co-expression network architectures. *Sci. Rep.* **8**, 8868 (2018).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

Methods

Nuclei isolation sorting from human Alzheimer's disease brain tissue.

Entorhinal cortex tissue from post-mortem Alzheimer's disease and non-diseased age-matched individuals was obtained from the Victorian Brain Bank (ethics approval for patient tissue banking and consent, University of Melbourne HREC Approval No: 1545740; approval for transcriptomic analysis using banked tissue, Monash University MUHREC 2016–0554; patient demographics in Supplementary Table 1). The study participants were allocated into disease or control groups based on overall amyloid and tau pathology. Nuclei isolation was carried out using the Nuclei Isolation Kit: Nuclei EZ Prep (Sigma, NUC101) as described in ref.⁵². Briefly, tissue samples were homogenized using a glass dounce grinder in 2 ml of ice-cold EZ PREP buffer (Sigma, N3408) and incubated on ice for 5 min. Centrifuged nuclei (500 g, for 5 min at 4 °C) were washed in ice-cold EZ PREP buffer, and nuclei suspension buffer (NSB, consisting of 1× PBS (Gibco, 14190-144), 1% (w/v) BSA (Sigma, A1470), and 0.2 U μ l⁻¹ RNase inhibitor (Clontech, 2313A)). Isolated nuclei were resuspended in NSB to 10⁶ nuclei per 400 μ l, filtered through a 40 μ m cell strainer, and counted with Trypan blue (Nano EnTek, EBT-001). Nuclei enriched in NSB were stained with DAPI (1:1,000) (ThermoFisher Scientific, D1306) for nuclei isolation using the Influx cell sorter (BD Biosciences, 70 μ m nozzle, 21–22 p.s.i.). Nuclei were defined as DAPI-positive singlets. Sorted nuclei were counted twice before loading onto the 10x Chromium (10x Genomics). Library construction was performed using the Chromium Single Cell 3' Library & Gel Bead Kit v2 (10x Genomics, PN-120237) with 18 complementary DNA pre-amplification cycles and sequencing on one high-output lane of the NextSeq 500 (Illumina). Data collection and analysis were not performed blind to the conditions of the experiments. Our sequencing saturation was high, ranging from 83.2% to 97.4%. Sequencing was done using samples from eight Alzheimer's disease and eight control individuals. Our single-nuclei data initially comprised eight 10x runs consisting of four Alzheimer's disease runs and four control runs. Each run had two individuals (Supplementary Table 1) that were both randomly chosen from within the same group (either control or disease). Two runs (one Alzheimer's disease run and one control run) were discarded because of high neuronal enrichment (see Extended data Fig. 2a), possibly indicating neuronal contamination or technical artefacts. In all, this resulted in a final dataset of six 10x runs consisting of three Alzheimer's disease runs (six Alzheimer's disease individuals) and three control runs (six control individuals), giving us a total sample size of 12 individuals. We did not use statistical methods to pre-determine sample sizes but our sample sizes are similar to, or greater than, those reported in previous publications^{6,24,53}.

Mapping single-nuclei reads to the genome and disentangling individuals from mixed-individual libraries. Using the Grch38 (1.2.0) reference from 10x Genomics, we made a pre-mRNA reference according to the steps detailed by 10x Genomics (<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/advanced/references>). Cellranger count was used to obtain raw counts. To disentangle the individual donor identity of every cell, we used the Bayesian demultiplexing tool vireo (Version 0.1.2, <https://doi.org/10.1101/598748>) and its associated pipeline (<https://github.com/huangyh09/vireo-manual>). Note that vireo is only able to distinguish the cells with respect to the two donors, but it is unable to assign the actual donor identity to each cell. Briefly, for each 10x library, we compiled a list of single nucleotide polymorphisms (SNPs) using cellSNP (Version 0.1.6) with a minimum allele frequency of 0.1 and minimum unique molecular identifiers (UMIs) of 20. cellSNP requires a reference list of human variants on which SNPs are called. To this end, we used a pre-compiled list of SNPs from the 1,000 genome project provided by the authors of vireo (<https://sourceforge.net/projects/cellsnp/files/SNPlist/>). Specifically, we used the list of variants based on the Grch38 build, ranging from chromosome 1 to 22 and X with a minimum allele frequency of 0.05. Subsequently, vireo was performed to demultiplex each 10x library by separating the cells into two donor populations. Cells that did not pass the vireo algorithm criteria were deemed unassigned with respect to donor identity. In addition, cells with features from both donors were termed doublets. Overall, the percentages of cells excluding doublets and unassigned cells were high, ranging from 88.8% to 99.2% in the three Alzheimer's disease libraries, and from 94.0% to 97.6% in the three control libraries.

Quality control for the expression matrix. The raw expression matrix was composed of 33,694 genes and 14,876 cells. Genes without any counts in any cells were filtered out. A gene was defined as detected if two or more transcripts were present in at least ten cells. The 100 postmortem interval (PMI)-associated genes, as defined by Zhu et al. in the cerebral cortex, that were detected in our data were removed⁵⁴.

For cell filtering, cells outside the 5th and 95th percentile with respect to the number of genes detected and the number of UMIs were discarded. In addition, cells with more than 10% of their UMIs assigned to mitochondrial genes were filtered out. The matrix was normalized with a scale factor of 10,000, as recommended by the Seurat pipeline⁵⁵ before FindVariableGenes was used to define variable genes with the parameters x.low.cutoff = 0.0125, x.high.cutoff = 3, and y.cutoff = 0.5. ScaleData was used to center the gene expression. Overall, the resulting filtered matrix consisted of 10,850 genes and 13,214 cells.

Cell type identification. BRETİGEA⁵¹ is an R package that uses brain cell-type marker gene sets curated from independent human and mouse single-cell RNA datasets for cell type proportion estimation in bulk RNA datasets. It can also be used in single-cell RNA datasets for the identification of cell types. For the reference datasets, BRETİGEA uses well-annotated and well-referenced datasets from ref.⁹ and ref.¹⁰. The six main cell type lineages identified by BRETİGEA are neurons, astrocytes, oligodendrocytes, microglia, OPCs, and endothelial cells. We obtained the gene sets from BRETİGEA for all six cell types and calculated a module score for each cell type using Seurat's *AddModuleScore* function. This function calculates the average expression levels of each cell type gene set subtracted by the aggregated expression of a background gene set. Cell type identification was then performed in two steps.

Firstly, each cell was assigned a cell type based on the highest cell type score across all six cell types. Furthermore, we defined a cell as a hybrid cell if the difference between the first and second highest cell type scores were within 20% of the highest cell type score:

$$(x_1 - x_2)/x_1 < 0.2$$

where x_1 is the highest cell type score and x_2 is the second highest cell type score.

Secondly, for each cell type, we applied z-score transformation to the gene score distribution. Data distribution was assumed to be normal but this was not formally tested. Gene score distributions are shown in Supplementary Fig. 1. Subsequently, to consider poorly identified cells (that is, cells with low cell type scores), for each cell type, cells with low cell type score (5th percentile and below) were relabeled as unidentified cells.

Alternatively, these cells may represent cells in which a true cell signature is more difficult to determine owing to a large number of gene dropouts: indeed, a large proportion of hybrid and unidentified cells were from the AD1–AD2 library corresponding to the two oldest Alzheimer's disease patients, in which we detected a median of only 473 genes per cell.

Overall, we obtained 449 microglia, 2,171 astrocytes, 656 neurons, 7,432 oligodendrocytes, 1,078 OPCs, 98 endothelial cells, 925 unidentified cells, and 405 hybrid cells. From UMAP visualization, the clusters separate well based on cell type, supporting the accuracy of our cell type identification method.

Human single-nuclei UMAP and clustering analysis. Seurat was used for normalization, scaling, and finding variable genes in the same manner as described in the 'Quality control for expression matrix' section. For each cell type, Seurat's *SubsetData* was used to subset the data. Subsequently, each cell type underwent the same procedure of normalization, scaling, and finding of variable genes. Before UMAP calculation, principal component analysis (PCA) was first performed to obtain a small number of principal components as input to the UMAP algorithm. The number of principal components to be included varies across the individual cell types and the global dataset and was determined using the Elbow method. The percentage variance explained by each principal component was plotted and the number of principal components was then chosen at the 'elbow' of the plot where a substantial drop was observed in the proportion of variance explained. Seurat's *FindClusters* was used to derive subclusters within each cell type with a resolution of 0.8. The number of principal components for each cell type used are as follows: all cell types, 19; microglia, 15; astrocytes, 16; neurons, 15; oligodendrocytes, 15; OPCs, 16; endothelial cells, 18; unidentified cells, seven; and hybrid cells, nine.

Identification of individual- and sex-specific genes. To identify the individual- and sex-specific genes within each cell type, we first separated the entire dataset into each of the identified major brain cell types. Notably, endothelial cells were omitted as some individuals had insufficient numbers of endothelial cells sequenced for reliable differential expression analysis. As the individual and sex covariates are confounded by disease (Alzheimer's disease versus control), we further separated the cells of each cell type into cells belonging to individuals with Alzheimer's disease and control individuals. For each group of cells, we then performed differential expression using edgeR (v3.20.8) between cells from male and female individuals to identify sex-specific genes (threshold of absolute log fold change > 0.5 and false discovery rate (FDR) < 0.01). To identify individual-specific genes, we performed differential expression using edgeR (v3.20.8) between the individual of interest and the average expression of the remaining individuals (threshold of absolute log fold change > 0.5 and FDR < 0.01). Overall, for either the individual or sex covariate, this procedure generated ten sets of genes (from each of the five cell types and from either Alzheimer's disease or control cells—that is, microglia–Alzheimer's disease, microglia–control, astrocyte–Alzheimer's disease, astrocyte–control, and so on).

Human single-nuclei differential expression and gene set enrichment analysis. Differential expression was performed using the empirical Bayes quasi-likelihood F-tests (QLF) in the edgeR package (v3.20.8). Genes were deemed significantly differentially expressed if they had an absolute log fold change greater than 0.5 and FDR less than 0.01. Several differential expression comparisons were performed. First, to determine cell-type-specific genes, differential expression was performed between the cell type of interest and the average of the remaining five other cell types. For example, to compare the first cell type with the average

of the seven remaining cell types (including unidentified and hybrid cells), the argument `contrast=c(1,-1/7,-1/7,-1/7,-1/7,-1/7,-1/7)` was supplied to the `glmQLFTest` function in edgeR. Second, to identify transcriptomic differences between individuals with Alzheimer's disease and control individuals at the cell type level, differential expression was performed between cells from Alzheimer's disease patient libraries and control patient libraries for each cell type. Third, we designated subclusters to be Alzheimer's disease, control, or undetermined based on the cell composition. A subcluster was designated to be Alzheimer's disease if more than 80% of the cells originated from an Alzheimer's disease patient and the same applied for assigning control subclusters. Otherwise, a subcluster was designated to be undetermined. Differential expression was then performed between cells in the Alzheimer's disease subclusters and control subclusters for each cell type. Fourth, to identify subcluster-specific genes within each cell type, differential expression was performed between the subcluster of interest and the average of the remaining subclusters of the same cell type. Fifth, differential expression was also performed between pairs of subclusters of the same cell type. For all of the above-mentioned differential expression analyses, gene set enrichment analysis was also performed using the fgsea package (v1.4.1) with 100,000 permutations.

Obtaining novel cell type markers. We first obtained DEGs between each cell type and all seven other cell types (including unidentified and hybrid cells). For each cell type, we overlapped DEGs with FDR less than 0.05 and log fold change greater than 2 with the human marker genes obtained from BRETIGEA. The non-intersected genes were deemed the novel cell type markers.

Variance explained by each covariate. To quantify the variance explained by each covariate (namely cell type, individual, Alzheimer's disease-versus-control, number of detected genes, number of total UMIs, and sex), the `getVarianceExplained` function in the scater package (v1.10.1) was performed on all of the genes that were significantly differentially expressed between Alzheimer's disease and control individuals in any of the major brain cell types.

CellRouter analyses. *GRN analysis.* CellRouter is a single-cell RNA sequencing algorithm used to identify gene regulatory changes along transitions between user-defined cell states³⁶. CellRouter analysis was run using the pipeline provided by the original authors (<https://github.com/edroaldo/CellRouter>) as of 14 December 2018. For each cell type, we input raw counts into CellRouter. This was followed by built-in normalization in CellRouter, scaling, PCA calculation, and t-distributed stochastic neighbor embedding plot (t-SNE) visualization³⁷. t-SNE was performed with a perplexity of 20 and 1,000 iterations. For each cell type, a k-nearest neighbor (KNN) network was built with `buildKNN` in CellRouter using jaccard similarity with parameters $k = 10$ and the number of principal components derived from the section "Human single-nuclei UMAP and clustering analysis". In CellRouter, `findPaths` was used to find the trajectory from one subcluster to another. Paths were processed with `processTrajectories` in CellRouter by minimizing path cost and setting the minimum number of cells in a trajectory and neighbors to 2 and 3, respectively. To identify genes regulated along each trajectory, `correlationPseudotime` was performed in CellRouter using Spearman correlation and using only the top 75th (positively correlated) and bottom 25th (negatively correlated) percentiles to define our gene networks in the next step. `smoothDynamics` and `clusterGenesPseudotime` in CellRouter were subsequently performed. GRNs were constructed using `buildGRN` with a z-score threshold of 1.65. With respect to the GRN scores calculated by CellRouter, only the top 90th and bottom 10th percentile GRNs were taken as the final outputs from the CellRouter pipeline. 'Hs' was used as the species input. Igraph R package was used within the CellRouter pipeline for processing networks (<http://igraph.org>).

Pruning of GRNs. Importantly, for each GRN, we noticed that the downstream targets identified by CellRouter might not be differentially expressed. Therefore, we pruned each network by removing non-differentially expressed downstream targets ($FDR > 0.05$). Next, for each transition, we selected GRNs containing the DEGs that were the most highly differentially expressed. We first calculated a gene score for each downstream target by multiplying absolute log₂(fold change) and $-\log_{10}(FDR)$. For each transition, we removed a GRN from our analysis if its gene scores were not significantly higher than the gene scores of all the other DEGs for that respective transition via a one-sided Wilcoxon non-parametric test, which assumes that data are paired and derived from the same population, that each pair is randomly and independently selected, and that data are measured using an interval scale. This ensured that the remaining GRNs had downstream targets with gene scores significantly higher than the gene scores of the DEGs that are not part of each respective GRN. In other words, for each trajectory, the remaining GRNs have downstream targets representing the most highly DEGs. Last, we ensured that the transcription factors identified were indeed human transcription factors by overlapping them with the list from Lambert et al.³⁸. It is important to note that, after these post-processing and pruning steps, some transitions did not have any significant transcription factors, perhaps suggesting that the transition was between subclusters that have similar transcriptional landscapes.

Visualization of GRN scores and networks. For Fig. 4a, pheatmap³⁹ was used to plot up to the top five GRN scores for each trajectory. For this visualization, we only considered transcription factors that are inducers (that is, upregulation of the transcription factor leads to mostly upregulation of downstream targets). ggplot2 (ref. ⁶⁰) was used for visualizing the GRN scores for each trajectory in the R shiny app. Cytoscape⁶¹ was used to construct the network, and yFiles was used to construct the hierarchical network layout. Cell symbols were created with BioRender (<https://app.biorender.com>). For each cell type, the networks plotted are the networks with the highest average log fold change of GWAS hits taken from the differential expression between the source subcluster and each of its target subclusters. These GWAS hits are colored by the average log fold change.

Single cell-level annotation of GWAS candidate genes. To calculate GWAS specificity scores, GWAS data were downloaded from the NHGRI-EBI catalog⁴² on 25 February 2019. Specifically, data for four traits were downloaded: 'Alzheimer's Disease' (EFO-0000249), 'Alzheimer's Disease Biomarker Measurements' (EFO-0006514), 'Late-Onset Alzheimer's Disease' (EFO-1001870), and 'Alzheimer's Disease Neuropathologic Change' (EFO-0006801); see Supplementary Table 8 for list of genes. We first removed gene duplicates and GWAS loci in intergenic regions, and used a *P* value of 9×10^{-6} or lower to identify significant associations. Then, since GWAS signals can point to multiple candidate genes within the same associated locus, for each Alzheimer's disease-related phenotype we chose to focus on the 'Reported Gene(s)' (that is, genes reported as associated by the authors of each GWAS study). Overall, we retrieved 980 GWAS-reported genes for downstream analysis. For calculation of cell type specificity scores, we followed the expression weighted cell type enrichment (EWCE) method described by Skene et al.⁶³. Briefly, this approach involves the following steps. First, an analysis of variance (ANOVA) was performed to remove uninformative genes (genes with expression that did not vary significantly across cell types). We implemented a *P* value threshold of 0.00001 for the F-statistic, which is the default value used in the EWCE R package. Second, specificity scores were calculated using `generate.celltype.data`. Finally, we defined a GWAS-associated gene as cell-type-specific if its specificity score was at least at the top 99th percentile of the specificity score distribution across all cells with respect to all GWAS-associated genes. Using this criterion, we identified 30 cell-type-specific genes, the average expression of which across cell types is shown in Fig. 4a.

Comparison of single-cell data to bulk Alzheimer's disease gene expression profiles. *Comparing with public datasets.* Blalock et al.⁴⁰ performed microarray analyses for laser-captured hippocampal tissue from control and Alzheimer's disease patients and generated a list of genes and corresponding directionality of expression in Alzheimer's disease. 'UP' genes were given a log fold change of +1 and 'DOWN' genes were given a log fold change of -1. We overlapped this list with the 30 cell-type-specific GWAS genes provided that these genes were differentially expressed between subcluster-driven Alzheimer's disease and control clusters ($FDR < 0.10$). For both the microarray and our datasets, we normalized the log fold change by the respective maximum absolute log fold change before visualization. We also compared our data with that of Tan et al.³⁹, who performed microarray analysis in the neocortex of Alzheimer's disease patients. We overlapped the DEGs identified by Tan et al.³⁹ ($FDR < 0.10$) with our 30 cell-type-specific GWAS genes. Once again, genes were considered differentially expressed between subcluster-driven Alzheimer's disease and control clusters if the *FDR* was less than 0.10. Log fold changes were normalized using the same methodology described earlier in this section.

Finding subcluster-specific GWAS hits. After obtaining the DEGs ($FDR < 0.05$) between each subcluster and the other subclusters for each cell type, we overlapped these genes with the 980 GWAS-reported genes (Supplementary Table 9). For visualization in Fig. 4b, up to the top three genes with the highest absolute log fold change for each subcluster were plotted for each cell type.

Functional annotation of GRN. *Functional annotation of TFEB networks.* Enrichment of the pruned TFEB networks for specified transitions—a3 to a1, a6 to a1, a7 to a1, and a8 to a1—was performed using `moduleGO` from the R packages DGCA⁶⁴ and GOstats⁶⁵. The top five gene ontology terms for each transition were plotted in Fig. 5c. The conserved TFEB network represents the genes that are common across all four transitions. Reported *P* values are adjusted for multiple testing by the Benjamini–Hochberg method⁶⁶.

Functional annotation of top GRN networks. Enrichment of the pruned GRN networks was performed using the same procedure described above. For each GRN, the total gene set is the union of all specific gene sets with respect to the trajectories as shown in Extended data Fig. 10. The top five gene ontology terms for each GRN network were plotted in Extended data Fig. 9c.

Comparison with neuronal data from Lake et al. Processed neuronal expression data from Lake et al.²⁴ were downloaded from <https://hemberg-lab.github.io/scRNA.seq.datasets/human/brain/>. Briefly, Lake et al. used single-nucleus sequencing on post-mortem human brains from six different regions and identified

16 neuronal subtypes. The original log(CPM) expression matrix consists of 25,051 genes and 3,042 cells. For each of the 16 neuronal subtypes, we calculated the average log(CPM) for each gene to construct a 25,051 by 16 reference expression matrix. We projected our neuronal data onto this reference expression matrix using methods as described in the reference component analysis (RCA)⁶⁷. In short, the RCA pipeline involves the calculation of the Pearson correlation coefficient to find out how closely the projected data are associated to the reference data. The visualized outputs are the z-score transformations of the correlation coefficient raised to the fourth power.

Comparison with astrocyte data from Lin et al. and Liddelow et al. Astrocyte data were downloaded from Lin et al.²⁵. Briefly, Lin et al. identified five astrocyte subpopulations in the human adult brain. Each of these five reported astrocyte subclusters has its own module or set of marker genes (FDR < 0.10). For each module, we converted mice gene names to human orthologs using *biomaRt* (v2.34.2). In addition, for each module, we removed genes that appeared in other modules, thereby giving us specific genes for each module. We plotted the average expression of each module in our eight astrocytic subpopulations. For visualization, we normalized the expression of each module by the maximum module expression across all eight astrocytic subpopulations. Similarly, we plotted the average expression of the pan-reactive, A1-specific, and A2-specific gene modules as identified by Liddelow et al.¹⁷ across all eight astrocytic subpopulations.

Comparison with Mathys et al. Differential gene expression results between 'Alzheimer pathology' and 'no Alzheimer pathology' were obtained from Supplementary Table 2 (ref. 7). For our own Alzheimer's disease versus control (a priori) differential expression analysis, DEGs are defined as having FDR less than 0.01 and absolute log fold change greater than 0.5. Subsequently, for each cell type, we calculated the significance of overlap between DEGs from Mathys et al. (Alzheimer's disease pathology versus no Alzheimer's disease pathology) and Grubman et al. (Alzheimer's disease versus control) using the hypergeometric test (*hyper*). *VennDiagram* (v1.6.20) was used to visualize the overlap. Among the overlapping genes, we calculated the number of concordant and discordant genes. Concordant genes refer to genes with the same directionality in both Mathys et al. and Grubman et al. *clusterProfiler* (v3.6.0)⁸ was used to calculate gene ontology adjusted P values (biological process) for three gene sets: overlapping DEGs, Mathys et al. specific DEGs, and Grubman et al. specific DEGs. To remove redundant gene ontology terms, the *simplify* function from *clusterProfiler* was used with inputs of 'p.adjust' and 'min' for the arguments 'select_fun' and 'by', respectively. The similarity cutoff was set at 0.5. The *bitr* function was used to map gene symbols to gene entrez identifiers using *org.Hs.eg.db* (3.5.0). For visualization onto heatmap, *pheatmap* was used to plot up to the ten most significant gene ontology terms with respect to each gene set. Gene ontology terms not meeting the adjusted P value cutoff of 0.05 were removed. Note that our analyses of neuron (excitatory) and neuron (inhibitory) DEGs from Mathys et al. were performed by comparing them to our neuron (Alzheimer's disease versus control) DEGs. Multiple testing correction was done using the Benjamini–Hochberg method.

Shiny app development. *R shiny web application.* To allow for easy visualization of the extensive dataset and analysis involved in this work, a web interface (<http://adsn.ddnetbio.com>) was created using the shiny package (v1.1.0).

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

Data availability

All single-cell RNA sequencing data are available from the Gene Expression Omnibus (GEO) under the accession number GSE138852. Data can also be visualized via the interactive web application at adsn.ddnetbio.com. Single-cell gene expression data and metadata can also be downloaded directly via adsn.ddnetbio.com.

Code availability

Code is available from the authors by reasonable request.

References

52. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
53. Tirosh, I. et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendrogloma. *Nature* **539**, 309–313 (2016).
54. Zhu, Y., Wang, L., Yin, Y. & Yang, E. Systematic analysis of gene expression patterns associated with postmortem interval in human tissues. *Sci. Rep.* **7**, 5435 (2017).
55. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
56. Lummertz da Rocha, E. et al. Reconstruction of complex single-cell trajectories using CellRouter. *Nat. Commun.* **9**, 892 (2018).
57. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
58. Lambert, S. A. et al. The human transcription factors. *Cell* **175**, 598–599 (2018).
59. Kolde, R. *pheatmap*: pretty heatmaps. *R package version 61*, 926 (2012).
60. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer: 2016)..
61. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
62. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
63. Skene, N. G. & Grant, S. G. N. Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment. *Front. Neurosci.* **10**, 16 (2016).
64. McKenzie, A. T., Katsiyi, I., Song, W.-M., Wang, M. & Zhang, B. DGCA: a comprehensive R package for differential gene correlation analysis. *BMC Syst. Biol.* **10**, 106 (2016).
65. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
66. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
67. Li, H. et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat. Genet.* **49**, 708–718 (2017).
68. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
69. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).

Acknowledgements

The authors acknowledge Flowcore and Micromon, Monash University, for the provision of instrumentation, training, and technical support. Tissues were received from the Victorian Brain Bank, supported by The Florey Institute of Neuroscience and Mental Health, The Alfred and the Victorian Institute of Forensic Medicine, and funded in part by Parkinson's Victoria and MND Victoria. The Australian Regenerative Medicine Institute is supported by grants from the State Government of Victoria and the Australian Government. This work was supported by a Dementia Australia Research Foundation Grant to A.G., J.M.P. and E.P., a Monash Network of Excellence grant to J.M.P., E.P., and O.J.L.R., and a Yulgilbar Foundation grant to A.G. and G.S. A.G. was supported by a National Health and Medical Research Council (NHMRC) and Australian Research Council (ARC) Dementia Research Development Fellowship (GNT1097461). J.M.P. was supported by a Sylvia-Charles Viertel Fellowship. O.J.L.R. and J.F.O. were supported by a Singapore National Research Foundation Competitive Research Programme (NRF-CRP20-2017-0002). S.B. was supported by a National Health and Medical Research Council (NHMRC) and Australian Research Council (ARC) Dementia Research Development Fellowship (GNT1111206). This work was supported by a Sylvia and Charles Viertel Senior Medical Research Fellowship to R.L. and a Howard Hughes Medical Institute International Research Scholarship to R.L. We acknowledge S. Freytag (University of Western Australia) for her advice regarding disentangling individuals from mixed-individual libraries.

Author contributions

A.G. and J.M.P. conceived the study and designed experiments, and together with E.P. and O.J.L.R., designed the bioinformatics analyses. A.G., G.S., and X.Y.C. performed nuclei isolation and fluorescence-activated cell sorting (FACS). C.M. performed the pathological assessment of human control and Alzheimer's disease cases. G.C. and E.P. performed GWAS integration, and CellRouter and network analyses. J.F.O. and O.J.L.R. performed cell-type and cell-subcluster identification and performed differential gene expression and GSEA analyses. J.F.O. and O.J.L.R. developed the shiny web interface. J.P., R.S., S.B., D.V.L., D.P., and R.L. worked up the protocol for single-nuclei sequencing from the human brain. A.G., G.C., J.F.O., O.J.L.R., E.P., and J.M.P. wrote the manuscript. All authors approved of, and contributed to, the final version of the manuscript.

Competing interests

O.J.L.R. and J.M.P. are co-inventors of the patent (WO/2017/106932) and are co-founders, shareholders, and directors of Mogrify Ltd., a cell therapy company. All other authors declare no competing interests.

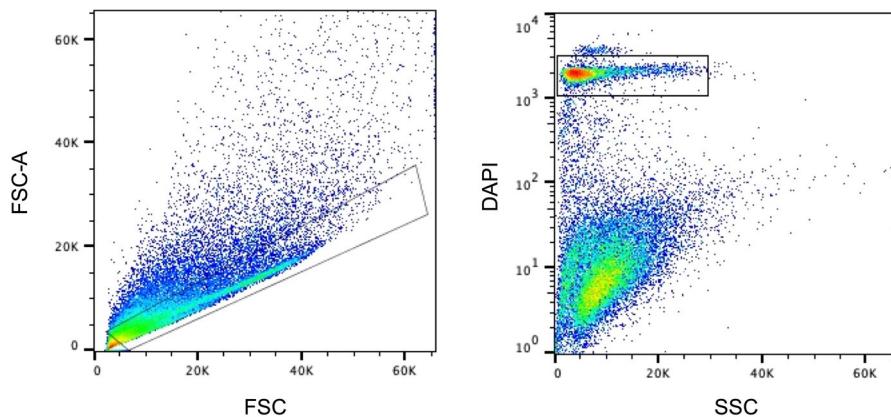
Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41593-019-0539-4>.

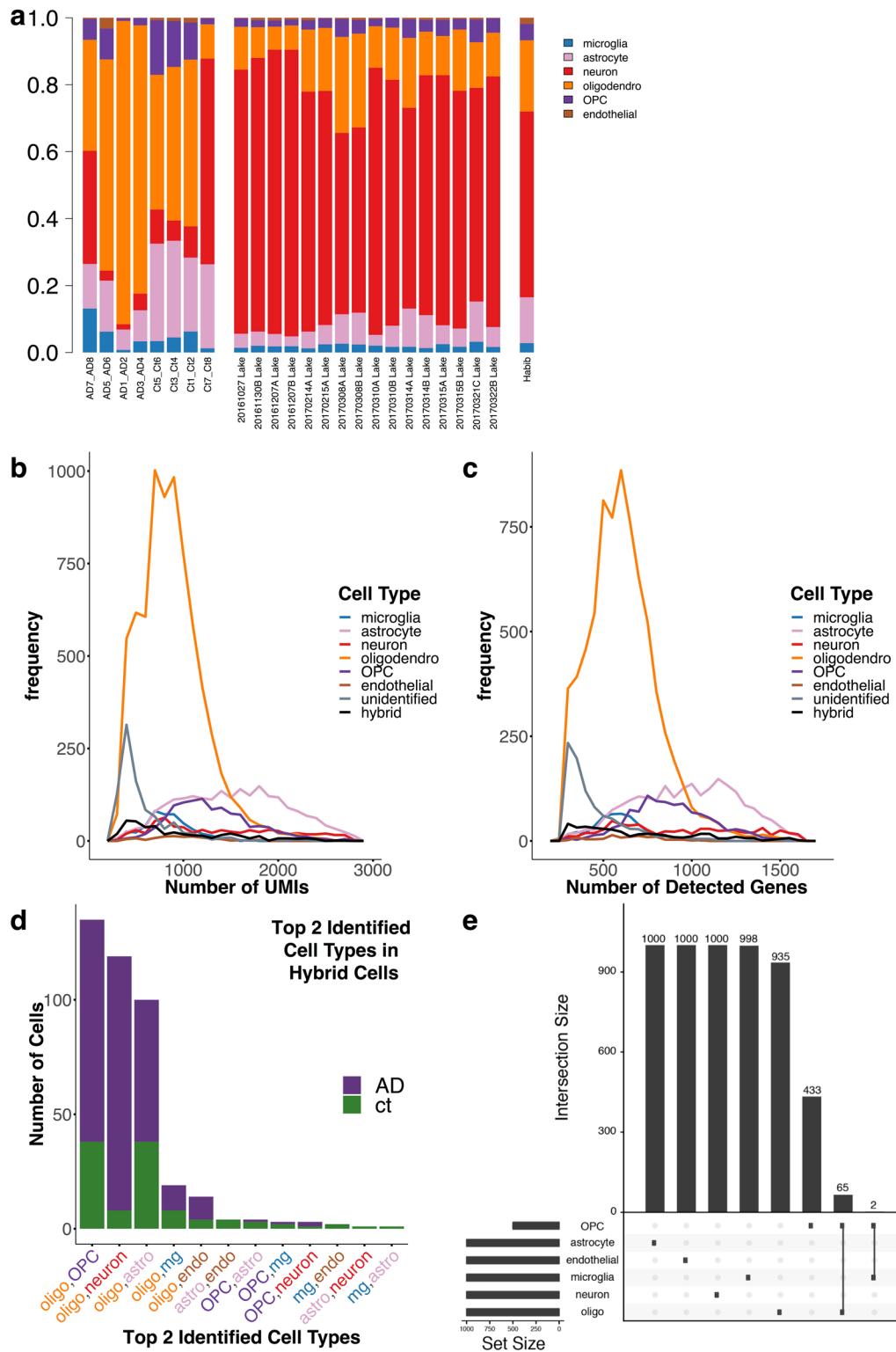
Supplementary information is available for this paper at <https://doi.org/10.1038/s41593-019-0539-4>.

Correspondence and requests for materials should be addressed to O.J.L.R., E.P. or J.M.P.

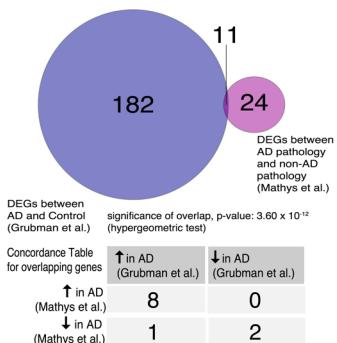
Reprints and permissions information is available at www.nature.com/reprints.



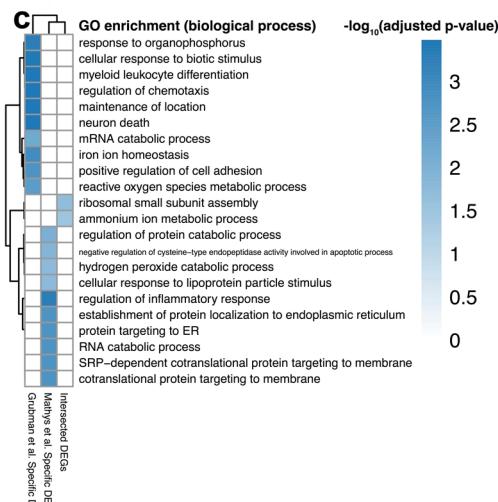
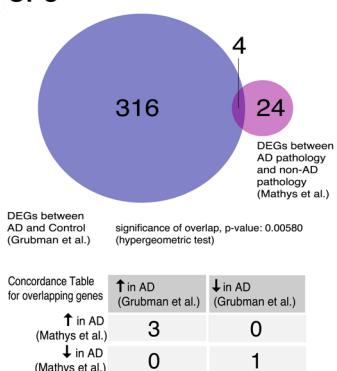
Extended data Fig. 1 | Gating strategy for FACS isolation of single nuclei. Nuclei isolated using the EZ-Prep kit from entorhinal cortex of AD and control patients were FACS sorted on BD Influx using 70 μ m nozzle, 21–22 psi. Single DAPI⁺ events were considered nuclei.



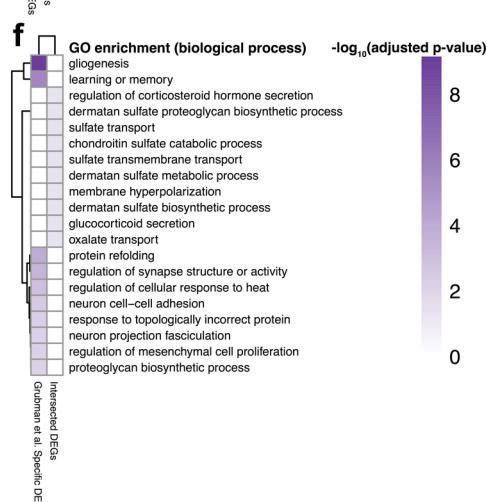
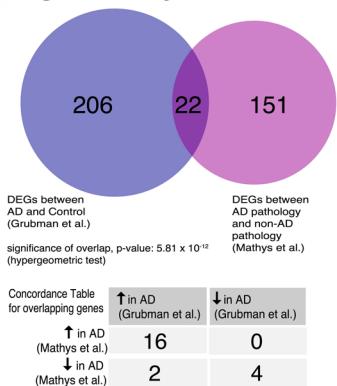
Extended data Fig. 2 | Single nuclei metadata and analysis of hybrid cells. **a**, cell proportion plots by library for eight libraries showing exclusion of two libraries due to high neuronal count and comparison of cell type proportions recovered in this study compared to cell proportions in other single nuclei studies²⁴ and⁵². Cell proportions are shown by sequencing library, where this information is available. **b**, Frequency histogram of number of Unique Molecular Identifiers (UMIs) across all 8 cell type groups. **c**, Frequency histogram of number of detected genes across all 8 cell type groups. **d**, Barplot showing number of top two identified cell types in hybrid cell types in AD and control libraries as proportion. **e**, UpSetR⁶⁹ plot showing the BRETAGE number of distinct and overlapping cell type markers across the six major cell types - microglia, astrocyte, neuron, oligodendrocyte, OPC, and endothelial cells.

a Microglia**b****Overlapping genes (n=11)****Concordant fold change in AD****Discordant fold change in AD**

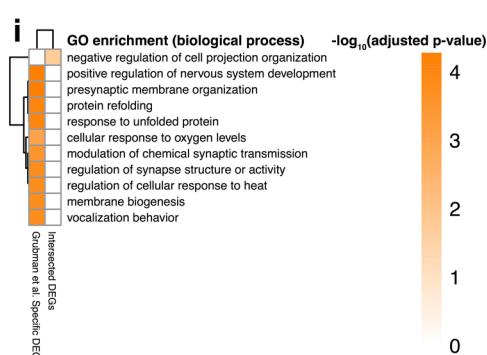
*PTPRG NR4A2
SPP1 APOC1
VSIG4 RPS28
RPS28 MYO1E
SAT1 FCGBP
RNF149*

**d OPC****e****Overlapping genes (n=4)****Concordant fold change in AD****Discordant fold change in AD**

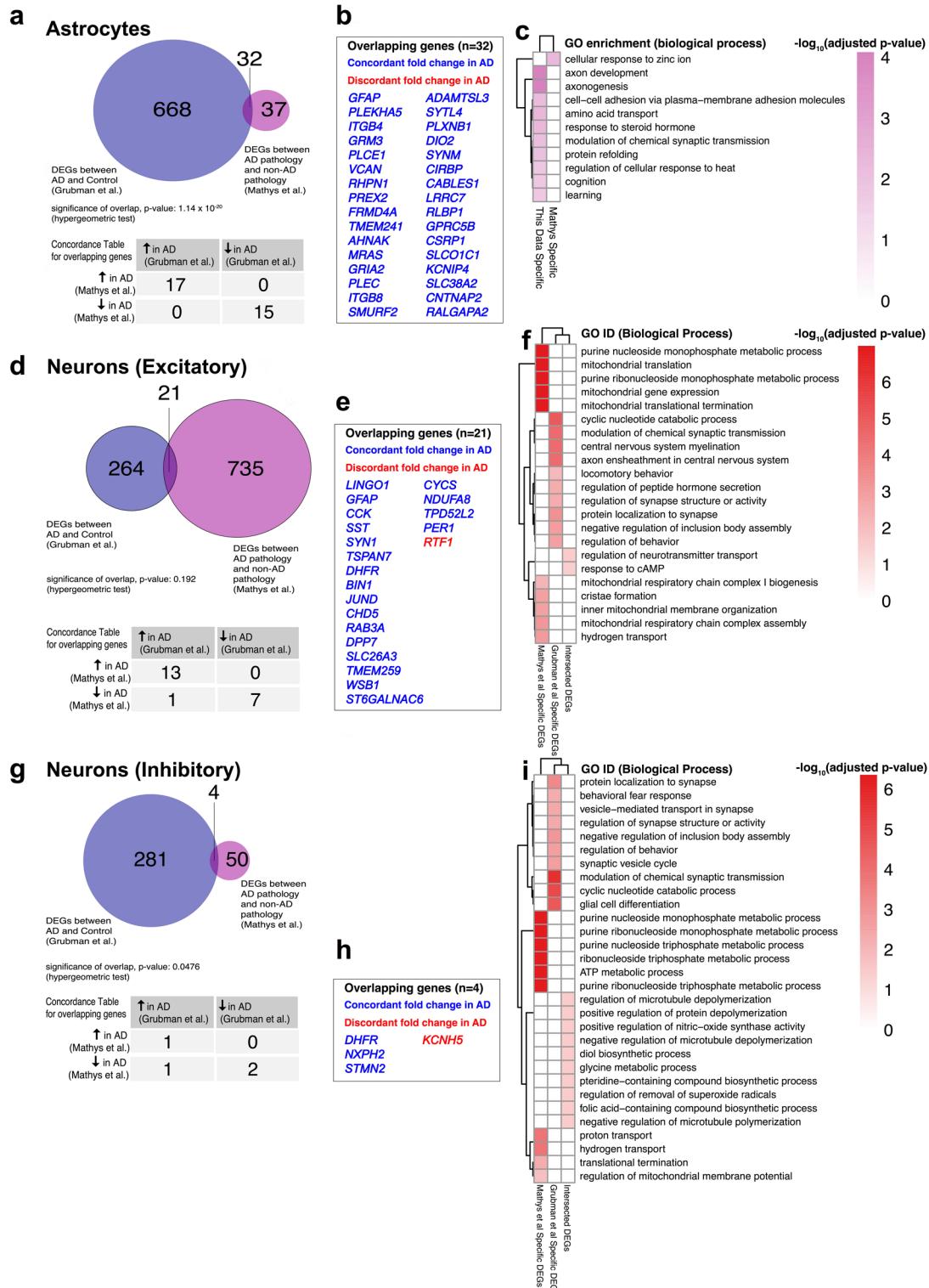
*SLC26A3 CD163L1
VCAN GALR1*

**g Oligodendrocytes****h****Overlapping genes (n=22)****Concordant fold change in AD****Discordant fold change in AD**

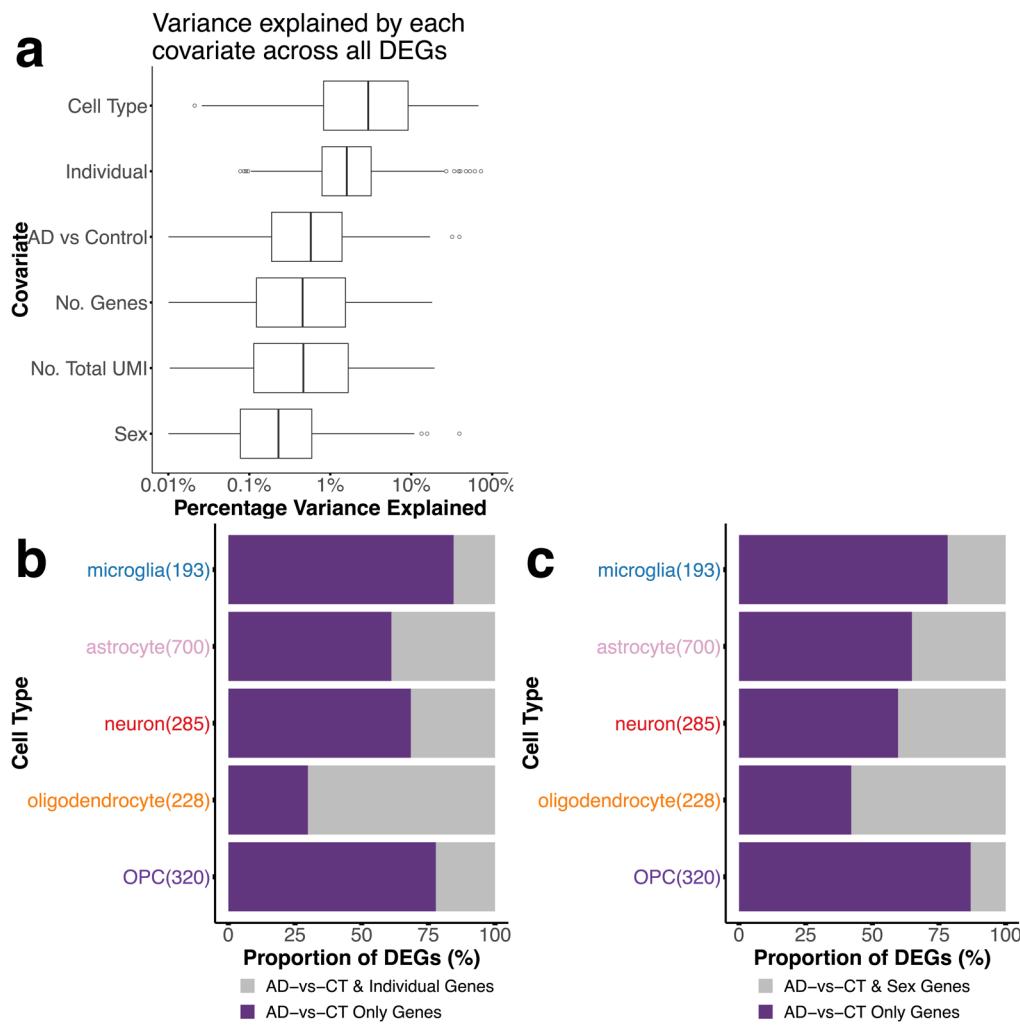
*LINGO1 METTL7A
ERBB2IP CARS1
SPP1 LDLRAD4
ITPKB ATF4
SLC38A2 RPS27
CCP110 PPA1
KCNH8 GPM6A
MID1IP1 HP1BP3
FKBP4 PIM3
QDPR SLC5A11
HTRA1 EVA1C*

**Extended data Fig. 3 | Comparison of differentially expressed genes (DEGs) within microglia, OPCs and oligodendrocytes with Mathys et al. 2019.**

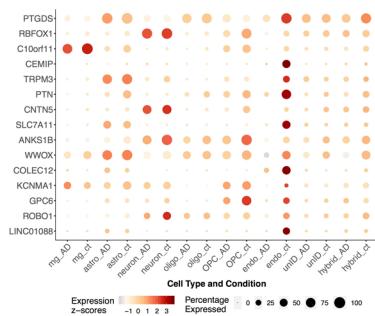
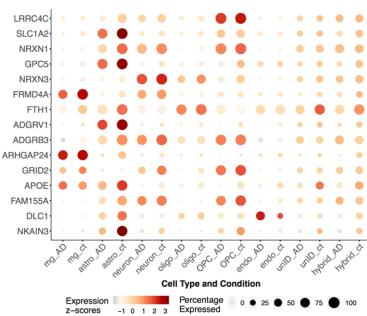
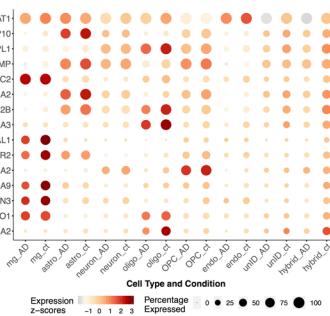
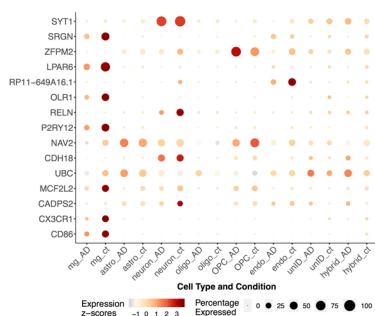
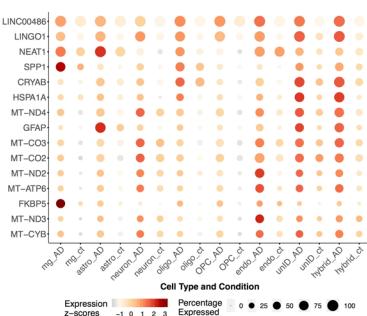
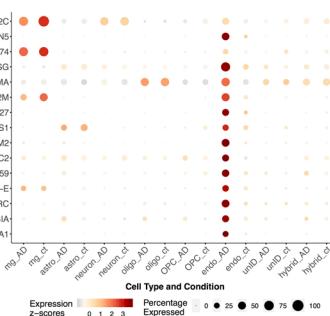
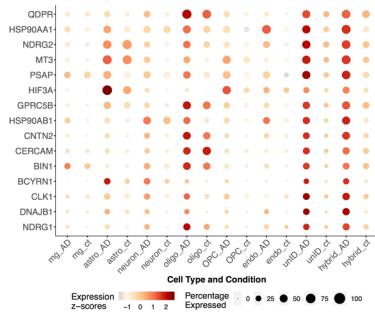
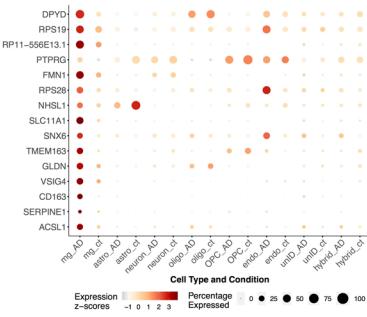
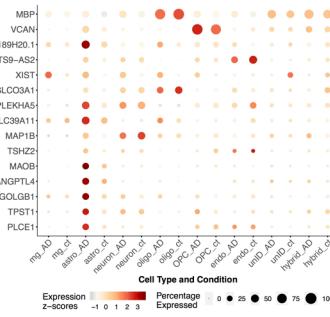
a, d, g, Venn diagram showing overlap of DEGs detected in Mathys et al. ($n = 24$ AD-pathology vs $n = 24$ no-pathology individuals) and those identified in the present study ($n = 6$ AD vs $n = 6$ control individuals) in microglia (a), OPCs (d) and oligodendrocytes (g), and the table shows the number of and concordance of dysregulation in AD of genes within the overlap of DEGs in both studies. Hypergeometric test was used to test for significance of overlap. **b, e, h**, List of concordant and discordant overlapping DEGs in microglia (b), OPCs (e) and oligodendrocytes (h). **c, f, i**, Significant gene ontology (GO) terms for each gene set comparison in microglia (c), OPCs (f) and oligodendrocytes (i). Multiple testing correction was done using the Benjamini-Hochberg method. Enrichment analysis was performed using hypergeometric test.



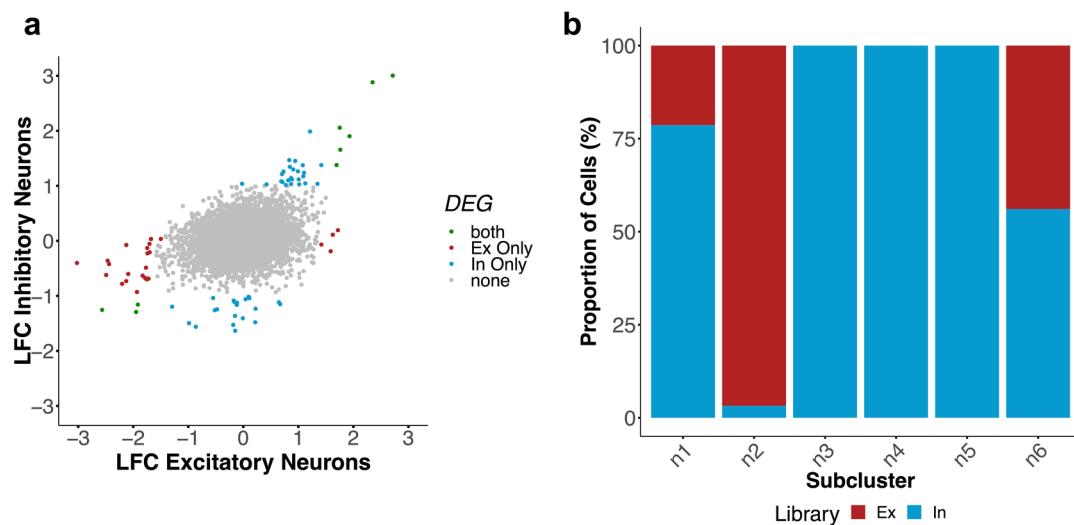
Extended data Fig. 4 | Comparison of differentially expressed genes (DEGs) within astrocytes, excitatory neurons and inhibitory neurons with Mathys et al. 2019. **a, d, g,** Venn diagram showing overlap of DEGs detected in Mathys et al. ($n = 24$ AD-pathology vs $n = 24$ no-pathology individuals) and those identified in the present study ($n = 6$ AD vs $n = 6$ control individuals) in astrocytes (**a**), excitatory neurons (**d**), and inhibitory neurons (**g**), and the table shows the number of and concordance of dysregulation in AD of genes within the overlap of DEGs in both studies. Hypergeometric test was used to test for significance of overlap. **b, e, h,** List of concordant and discordant overlapping DEGs in astrocytes (**b**), excitatory neurons (**e**), and inhibitory neurons (**h**). **c, f, i,** Significant gene ontology (GO) terms for each gene set comparison in astrocytes (**c**), excitatory neurons (**f**), and inhibitory neurons (**i**). Multiple testing correction was done using the Benjamini–Hochberg method. Enrichment analysis was performed using hypergeometric test.



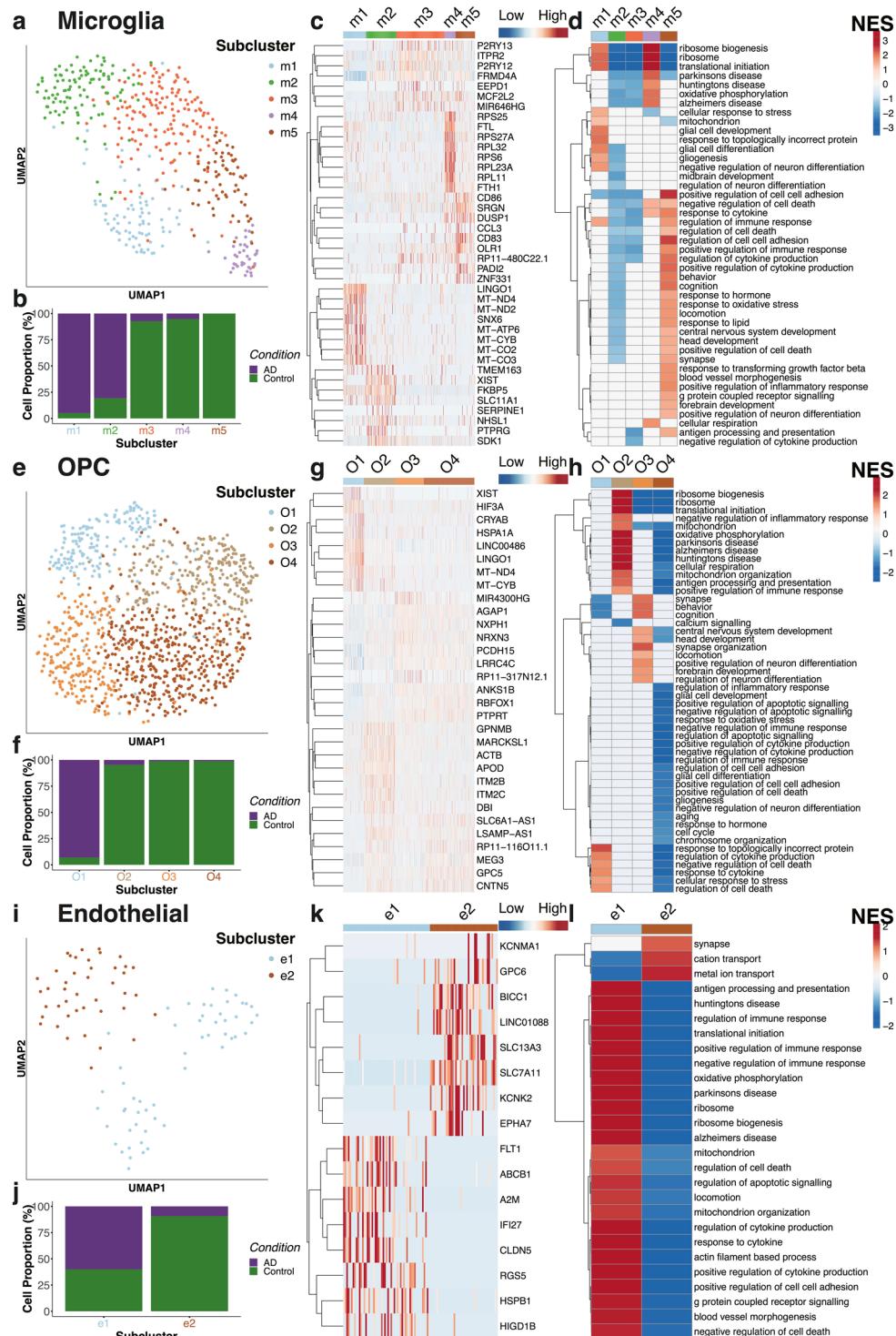
Extended data Fig. 5 | Analysis of sources of variation. **a**, Box plot (center line: median, box limits: upper and lower quartiles, whiskers: 1.5x interquartile range, points outside whiskers: outliers) showing percentage variance explained by each covariate for genes that are differentially expressed between AD and Control groups in any of the identified cell types. **b**, Proportion of AD-vs-Control differentially expressed genes (DEGs) that are also individual-associated genes and genes that are only DE between AD and Control. **c**, Proportion of AD-vs-Control DEGs that are also sex-associated genes and genes that are only DE between AD and Control. The number of DEGs are given in parenthesis.

a DEG1**b DEG2****c DEG3****d DEG4****e DEG5****f DEG6****g DEG7****h DEG8****i DEG9**

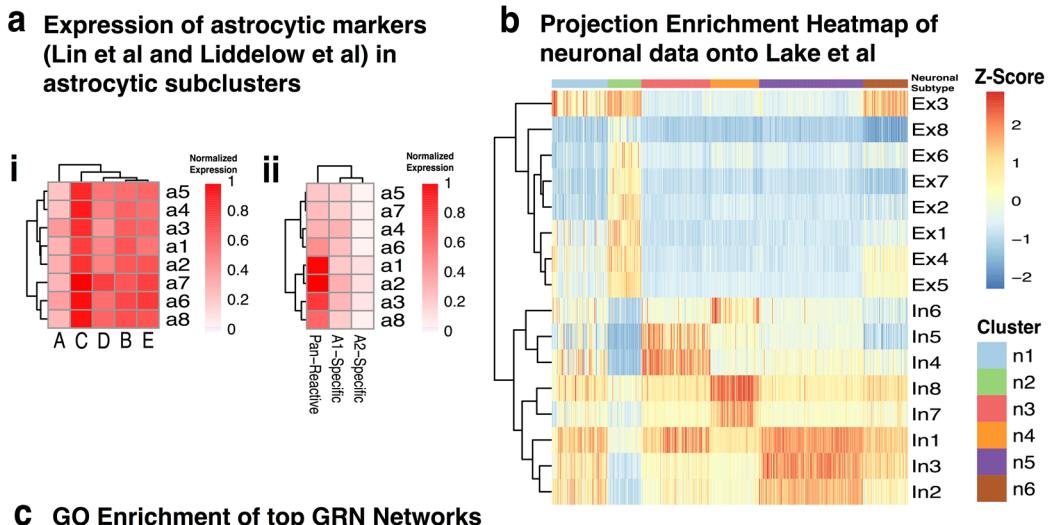
Extended data Fig. 6 | Cell-type-specific and shared AD related changes. **a-i**, Bubble plot showing the top 15 differentially expressed genes (DEGs) within the DEG1-DEG9 groups (see Fig. 1f), where the colour represents the relative gene expression and the bubble size is the proportion of cells in the group expressing the gene.



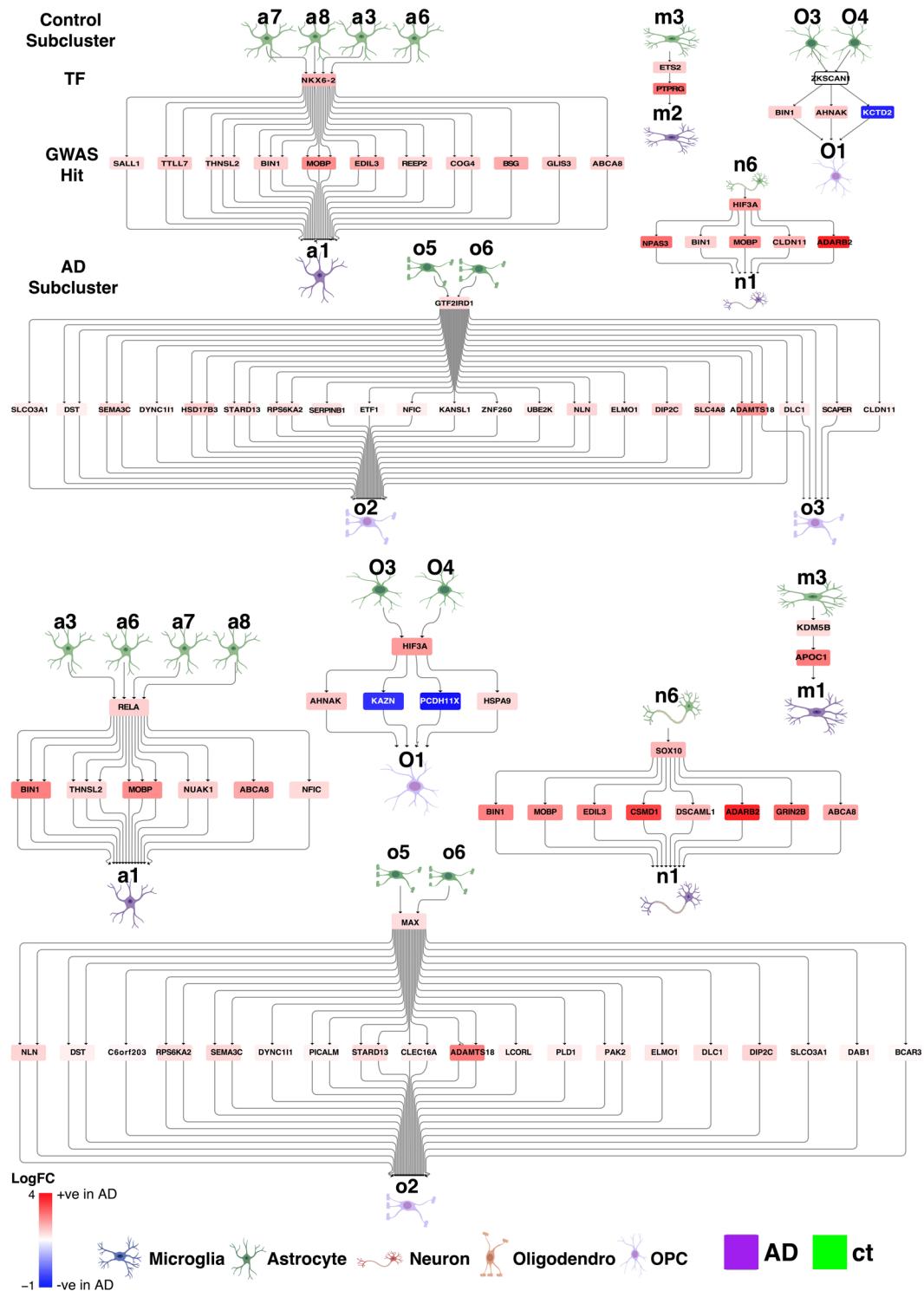
Extended data Fig. 7 | AD related changes in excitatory and inhibitory neurons. Scatter plot showing the relationship between cell-type-specific and common differentially expressed genes (DEGs) in control and AD, excitatory and inhibitory neurons. Cutoff for significance is $\text{abs}(\log \text{fold change}) > 0.5$ and false discovery rate (FDR) < 0.01 .



Extended data Fig. 8 | Single nuclei sequencing of human AD and control entorhinal cortex reveals homeostatic, AD-specific and shared ontological cell subclusters. **a,e,i**, Uniform Manifold Approximation and Projection (UMAP) visualization of subclusters of microglia (**a**), OPCs (**e**) and endothelial cells (**i**) showing **b,f,j**, the composition of cells in subclusters by disease state, **c,g,k**, hierarchical clustering and heatmap colored by single-cell gene expression of subcluster-specific genes (top eight genes were shown per cluster). **d,h,l**, Gene Set Enrichment Analysis (GSEA) results of subcluster specific genes coloured by normalised gene set enrichment scores for the gene ontologies shown in each cell subcluster.



Extended data Fig. 9 | Comparison of signatures across subclusters. Comparison of astrocyte signatures to **aⁱ**,¹⁷ **aⁱⁱ**,²⁵ **b**, Comparison of neuronal subcluster signatures to²⁴ (Ex1–8, In1–8). **c**, Functional annotation of gene regulatory networks (GRNs) controlled by TFs in Fig. 5a. GRNs are shown in Extended data Fig. 10, *ETS2* ($n = 9$ genes), *GTF2IRD1* ($n = 339$ genes), *HIF3A*-neuron ($n = 112$ genes), *HIF3A*-OPC ($n = 96$ genes), *KDM5B* ($n = 44$ genes), *MAX* ($n = 273$ genes), *NKX6-2* ($n = 332$ genes), *RELA* ($n = 224$ genes), *SOX10* ($n = 114$ genes), and *ZKSCAN1* ($n = 37$ genes). Multiple correction testing was done using the Benjamini-Hochberg method. Enrichment analysis was performed using the hypergeometric test in R package GOSTats.



Extended data Fig. 10 | Gene regulatory network analysis predicts transcription factors regulating GWAS genes for conversion of control to AD subcluster signatures. Trajectories of the top transcription factors (TFs): NFKX6-2, ETS2, ZKSCAN1, HIF3A, GTF2IRD1, RELA, SOX10, KDM5B and MAX, with downstream GWAS gene targets. For each cell type, the reported TFs-driven gene regulatory networks include GWAS gene hits, which are direct target genes of the TF and have the highest average log fold change in gene expression between the source subcluster/s and target subcluster/s. Each gene is colored according to its average log fold change between the source subcluster/s and target subcluster/s. Only significantly differentially expressed GWAS gene hits are reported (false discovery rate < 0.05).

Reporting Summary

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

Statistics

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

n/a Confirmed

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

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

Software and code

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

Data collection

No software was used for data collection.

Data analysis

Data was analyzed using: Cellranger (2.1.1), Seurat (2.3.4), reshape2 (1.4.3), ggplot2 (3.0.0), ggbeeswarm (0.6.0), igraph (1.2.2), magrittr (1.5), dplyr (0.7.8), Cellrouter (as of 14th December 2018) (<https://github.com/edroaldo/cellrouter>), RETIGEA (1.0.0), RColorBrewer (1.1.2), pheatmap (1.0.10), GOstats (2.44.0), org.Hs.eg.db (3.5.0), DGCA (1.0.1), HGNChelper (0.3.6), Cytoscape (3.5.1), yFiles (1.0.1), R.utils (2.6.0), edgeR (3.20.8), fgsea (1.4.1), WGCNA (1.6.1), SingleCellExperiment (1.0.0), scater (1.10.1), gplots (3.0.1), preprocessCore (1.40.0), flashClust (1.1.2), RCA (2.0), VennDiagram (v1.6.20), clusterProfiler (v3.6.0), virore (v0.1.2), cellSNP (v0.1.6), biomart (v2.34.2), shiny (v1.1.0)

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

Data

Policy information about [availability of data](#)

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

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

All raw sequencing data is available on GEO under the accession XXXX. Processed data is available at <http://adsn.ddnetbio.com>.

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	No statistical methods were used to pre-determine sample sizes but our sample sizes are similar or greater than those reported in previous publications (Tirosh et al 2016, Nature; Lake et al 2016, Science; Zhong et al, 2018, Nature)
Data exclusions	Two single-cell libraries were discarded due to high neuronal enrichment possibly indicating neuronal contamination or technical artifacts. Exclusion criteria for individual cells are described in the manuscript as follows: For cell filtering, cells outside the 5th and 95th percentile with respect to number of genes detected and number of unique molecular identifiers (UMI) were discarded. In addition, cells with more than 10% of their UMIs assigned to mitochondrial genes were filtered out. Overall, the resulting filtered matrix consisted of 10,850 genes and 13,214 cells.
Replication	Experimental replication was not attempted.
Randomization	The study participants were allocated into disease or control groups based on the overall amyloid and tau pathology.
Blinding	Data collection and analysis were not performed blind to the conditions of the experiments.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Human research participants

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

Population characteristics	8 AD and 8 control patients were selected based on pathological analysis of Amyloid and Braak clinical staging. Patients were age and sex matched, with mean age 77.6, range 67.3-91 years and with a range of ApoE genotypes.
Recruitment	No donors were recruited, tissue has been obtained from brain donors to the Victorian Brain Bank. Samples with the lowest PMI available, age and sex matched, with a range of ApoE genotypes and with entorhinal cortex tissue available were chosen.
Ethics oversight	Ethics Approval for this study: Monash University MUHREC 2016-0554. Ethics Approval for Victorian Brain Bank patient tissue banking and consent : University of Melbourne HREC Approval No.: 1545740

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

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Entorhinal cortex tissue from post-mortem Alzheimer's disease and non-disease aged matched individuals was obtained from the Victorian Brain Bank (Ethics Approval: MUHREC 2016-0554). Nuclei isolation was carried out using the Nuclei Isolation Kit: Nuclei EZ Prep (Sigma, #NUC101) as described(Habib et al. 2017). Briefly, tissue samples were homogenized using a glass dounce grinder in 2 ml of ice-cold EZ PREP and incubated on ice for 5 min. Centrifuged nuclei (500 × g, 5 min and 4 °C) were washed in ice-cold EZ PREP buffer, and Nuclei Suspension Buffer (NSB; consisting of 1× PBS, 1% (w/v) BSA and 0.2 U/µl RNase inhibitor (Clontech, #2313A). Isolated nuclei were resuspended in NSB to 106 nuclei per 400 µl), filtered through a 40 µm cell strainer and counted with Trypan blue. Nuclei enriched in Nuclei Suspension Buffer were stained with DAPI (1:1000) for nuclei isolation using the FACSaria™ III cell sorter (BD Biosciences, Franklin Lakes, NJ; 70 µm nozzle, 21–22 psi). Nuclei were defined as DAPI+ singlets. Sorted nuclei were counted twice prior to loading onto the 10X Chromium (10X Genomics).

Instrument

BD Influx, 5 Laser

Software

FlowJo was used to analyse data

Cell population abundance

Sorted nuclei were counted twice with the Countess Cell counter, and live/dead proportions were analysed based on Trypan Blue Staining. Samples were >99% pure post sorting.

Gating strategy

Nuclei were gated on FSC/SSC, singlets and on DAPI fluorescence.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.