A landscape of the genetic and cellular heterogeneity in Alzheimer disease

Logan Brase[1,2,3], Shih-Feng You[1,2,3], Ricardo D'Oliveira Albanus[1,2,3], Jorge L. Del-Aguila[12], Yaoyi Dai[11], Brenna C Novotny[1,2,3], Carolina Soriano-Tarraga[1,2,3], Taitea Dykstra[6,7], Maria Victoria Fernandez[1,2,3], John P Budde[1,2,3], Kristy Bergmann[1,2,3], John C Morris[2,4,5,10], Randall J Bateman[2,4,5,10], Richard J Perrin[2,4,5,6,10], Eric McDade[1,10], Chengjie Xiong[4,13], Alison Goate[10,14], Martin Farlow[10,15], Jasmeer P. Chhatwal[10,16], Peter Schofield[10,17,18], Helena Chui[10,19,20], Dominantly Inherited Alzheimer Network (DIAN), Greg T Sutherland[9], Jonathan Kipnis[6,7], Celeste M Karch[1,2,3,10,*], Bruno A Benitez[1,2,3,*], Carlos Cruchaga[1,2,3,10,*], and Oscar Harari[1,2,3,*,#]

Affiliations:

1. Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA
2. Hope Center for Neurological Disorders, Washington University School of Medicine, St. Louis, MO, USA
3. NeuroGenomics and Informatics, Department of Psychiatry Washington University in St. Louis
4. Knight Alzheimer Disease Research Center, Washington University School of Medicine, St. Louis, MO, USA
5. Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA
6. Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, USA
7. Center for Brain Immunology and Glia (BIG), Washington University in St. Louis, St. Louis, MO, USA
8. Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA
9. School of Medical Sciences and Charles Perkins Centre, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia
10. Dominantly Inherited Alzheimer Network (DIAN)
11. Baylor College of Medicine, Houston, TX, USA
12. Merck & Co., Inc., Boston, MA, USA
13. Division of Biostatistics, Washington University School of Medicine, St. Louis, MO, USA
14. Ronald M. Loeb Center for Alzheimer's Disease, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
15. Department of Neurology, Indiana University School of Medicine, Indianapolis, IN, USA
16. Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
17. Neuroscience Research Australia, Sydney, NSW, Australia
18. School of Medical Sciences, University of New South Wales, Sydney, NSW, Australia
19. Alzheimer's Disease Research Center, Keck School of Medicine at the University of Southern California, Los Angeles, CA, USA
20. Department of Neurology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

* These authors contributed equally

[#] To whom correspondence should be addressed:

Oscar Harari, PhD
Assistant Professor
Department of Psychiatry
Washington University School of Medicine
Campus Box 8134 | 4444 Forest Park Avenue | Office: 5587 | St. Louis, MO 63108
Tel: 314-273-1862 | Fax: 314-362-2244
Email: harario@wustl.edu

Abstract

Background

Alzheimer disease (AD) has substantial genetic, molecular, and cellular heterogeneity associated with its etiology. Much of our current understanding of the main AD molecular events associated with the amyloid hypothesis (*APP, PSEN1* and *PSEN2*) and neuroimmune modulation (*TREM2* and *MS4A*) is based on genetic studies including GWAS. However, the functional genes, downstream transcriptional ramifications, and the cell-type-specific effects of many GWAS loci remain poorly understood. Understanding these effects can point us to the cellular processes involved in AD and uncover potential therapeutic targets.

Methods

We applied a genetic-based approach to our sample selection; our cohort included carriers of AD pathogenic mutations (*APP, PSEN1),* risk variants in *TREM2*, and the resilience variant (rs1582763) in the *MS4A* cluster associated with cerebrospinal fluid (CSF) soluble TREM2 levels. We performed single-nucleus RNA-sequencing (snRNA-seq) of 1,102,459 nuclei from the human parietal cortex of these carriers. Following initial unbiased clustering and cell-type annotation, we performed deep subclustering analysis per cell type to identify unique cellular transcriptional states associated with these genetic variants. We identified differentially expressed genes between cell states and genetic variant carriers/controls, and performed differential cell proportion analyses to determine key differences among these carriers. We analyzed sequencing data from human dorsolateral prefrontal cortex and mouse models to replicate the enrichment of unique cell states in genetic variant carriers. Finally, we leveraged these cell-state differential expression results to link genes in AD GWAS loci to their functional cell types.

Findings

We identified cell-specific expression states influenced by AD genetic factors for neurons and glia. Autosomal dominant AD (ADAD) brains exhibited unique transcriptional states in all cell types. *TREM2* variant carrier brains were also enriched for specific microglia and oligodendrocyte subpopulations. Carriers of the resilience *MS4A* variant were enriched for an altered activated-microglia expression state. We mapped AD GWAS genes to their potential functional cell types, and some, including *PLCG2* and *SORL1*, were expressed in a broader range of brain cell types than previously reported.

Interpretation

AD pathogenic, risk, or resilience variants are sufficient to alter the transcriptional and cellular landscape of human brains. Overall, our results suggest that the genetic architecture contributes to the cortical cellular heterogeneity associated with disease status, which is a critical factor to consider when designing drug trials and selecting the treatment program for AD patients.

Our findings suggest that integrating genetic and single-cell molecular data facilitates our understanding of the heterogeneity of pathways, biological processes and cell types modulated by genetic risk factors for AD.

Funding

US National Institutes of Health, Hope Center, Archer foundation, Alzheimer Association, CZI.

Keywords

Single-nucleus RNA sequencing, Alzheimer disease, functional genomics, autosomal dominant Alzheimer disease

Main

Alzheimer disease (AD) is the most common form of dementia. It is characterized by the presence of amyloid (Aβ) plaques and neurofibrillary tangles (NFTs, Tau deposits) in the brain and associated with neuroinflammation, myelination changes, synaptic dysfunction, and cell death[1,2]. Genetic studies have successfully identified multiple genes and pathways associated with AD etiology including rare mutations in *amyloid precursor protein* (*APP*)[3,4], *presenilin 1* (*PSEN1*)[5], and *presenilin 2* (*PSEN2*)[6] which cause familial autosomal dominant AD (ADAD), mainly due to the neuronal amyloid cascade hypothesis[7]. The strongest risk factor for sporadic AD (sAD), *apolipoprotein E* (*APOE)* ε4, implicates cholesterol metabolism and Aβ clearance mediated by astrocytes and microglia[8]. Integration of epigenetics and genome-wide association studies (GWAS) signals have implicated immune system dysfunction in AD[9]. In particular, rare genetic variants in t*riggering receptor expressed on myeloid cells 2* (*TREM2*), a gene that drives microglia activation, increase AD risk[10]. Recent GWAS have identified additional common genetic risk factors for AD, including rs1582763, an intergenic variant in the *MS4A* locus that confers resilience to AD and is associated with higher cerebrospinal fluid (CSF) soluble TREM2 levels[11-14]. However, linkage and association studies do not reveal the downstream transcriptional ramifications nor the cell types these variants affect.

Single-nucleus RNA-sequencing (snRNA-seq) has emerged as a powerful approach to interrogate the underlying transcriptional landscape of the cellularly-complex human brain. SnRNA-seq studies of sAD and *TREM2* p.R47H variant carriers to date have profiled the dorsolateral prefrontal cortex (DLPFC)[15,16], entorhinal cortex[17], and temporal neocortex[17], uncovering significant transcriptional changes in microglia, astrocytes, and oligodendrocytes associated with AD. A recent single-cell RNA-sequencing study found that within neurons, increased *APOE* expression leads to increased *MHC-I* expression and eventual cell death[18]. Previously, we generated snRNA-seq data from the parietal cortex from a *PSEN1* p.A79V mutation carrier and found a reduced proportion of excitatory neurons compared to two sAD samples, but a similar distribution of inhibitory neurons[19]. These studies highlight the potential for snRNA-seq to provide highly detailed expression profiles at the cellular level in a wide array of genetically defined individuals, uncovering novel cell-specific effects of AD risk genes.

In this study, we selected brains from the Dominantly Inherited Alzheimer Network (DIAN) and Charles F. And Joanne Knight Alzheimer Disease Research Center (Knight ADRC) biobanks to enrich our cohort with carriers of pathogenic, risk, and resilience genetic variants to leverage the naturally occurring perturbations of these genes and elucidate their role in AD pathogenesis. Specifically, we performed snRNA-seq utilizing the parietal cortex (Brodmann area 1-3,7) to investigate the cellular and transcriptional profiles in carriers of AD-risk-modifying variants. We focused on the parietal cortex: an understudied brain region with high level of plaques and tangle burden but relatively low atrophy in the early stages of AD pathology progression[20]. After stringent filtering, we obtained 294,114 high-quality nuclei from 67 individuals. We investigated the cellular diversity and proportions in AD by iteratively extracting nuclei by the cell type from the dataset (digital sorting), subclustering these nuclei into cell-type transcription states (cell states), and identifying differentially expressed genes (DEGs) among these cell states. We analyzed the differences in cell-type composition associated with ADAD, *APOE*ε4, *TREM2,* and *MS4A* carriers, and identified DEGs between the genetic factors and control samples. Finally, we created an online browser (http://ngi.pub/SNARE/) for convenient access to the snRNA-seq expression data.

To further investigate and validate our initial findings, we integrated brain snRNA-seq data from the dorsolateral prefrontal cortex of human donors from the ROSMAP cohort and 5xFAD mouse model (see details in Methods). We confirmed similar upregulated gene signatures between our cohort and the replication data in specific cell states of interest. Finally, we performed cell state proportional analyses to validate altered proportions associated with genetic factors.

Our findings highlight the power of leveraging genetic and single-cell molecular data to understand the heterogeneity of pathways, biological processes and cell types modulated by genetic risk factors for AD.

Results

Diverse transcriptional profiles within each cell type in cortical samples from genetically defined AD patients. To study the cellular and genetic heterogeneity of cortex from AD patients, we selected 74 participants from the DIAN and Knight ADRC brain banks, which included detailed clinical data, post-mortem pathological data, and genetic characterization. We generated whole transcriptomes of the parietal cortices at a single-cell resolution using 10x Genomics Next GEM technology. After data cleaning and quality control (QC; see Methods), we retained 67 samples and obtained molecular data for 16 carriers of pathogenic mutations in *APP* and *PSEN1* (autosomal dominant AD; ADAD), 19 carriers of AD risk variants in *TREM2*, 16 sAD non-carriers of any of these variants, three presymptomatic samples that meet AD pathological criteria but without cognitive impairment at age-of-death (presymptomatic), eight samples with non-AD neurodegenerative pathologies (others), and nine samples without neurodegenerative pathology nor deficits in their cognition (controls) (Table 1). Within the cohort, 41 samples carried the minor allele (A) for rs1582763, a SNP in the *MS4A* gene cluster[11,13,14], and 24 samples carried the *APOE*ε4 allele (Table 1, Fig. 1a).

In total, we sequenced 1,102,459 nuclei. After data cleaning and QC (Table 2), we retained and clustered 294,114 nuclei (Table 3), identifying 15 clusters covering major cell types found in the brain (Fig. 1b,c). These clusters were annotated based on the expression of well-known cell-type markers[19] (Fig. 1c, and Supplementary Table 1). Clusters 0, 1, 2, and 9 were identified as oligodendrocytes, clusters 3, 8, 10, 11, and 13 as neurons, clusters 4, 6, and 14 as astrocytes. Microglia, oligodendrocyte precursor cells (OPCs), and endothelial cells were identified in clusters 7, 5, and 12, respectively (Fig. 1b,c). All samples were equally distributed among the different clusters (Supplementary Table 2). The distribution of nuclei by clusters (Table 3) showed that glial cells accounted for 81.9% of the nuclei, whereas neuronal nuclei accounted for 18.1%. We observed a significant increase of OPCs (β=0.15; P=2.99×10^{-2}) and endothelial cells (β=0.07; P=1.25×10^{-2}) in ADAD participants compared to controls.

After extracting each cell type and subclustering, we identified 5-9 subclusters or cell-type transcription states (cell states), within each cell type (Table 3). We observed an average of 2,696 upregulated genes (average of 16,223 tested) that passed multiple test correction in each cell state compared to all other cell states in the same cell type (Supplementary Table 3, Supplementary

Material). This highlights the incredible multiplicity within cell types. As expected due to statistical power, there was a positive correlation between the number of detectable DEGs and the number of nuclei in the cell type. However, the highest average number of DEGs was found in inhibitory and excitatory neurons, which had three times less nuclei than that of oligodendrocytes, most likely due to the neuronal differences between cortical layers[21]. We identified a comparatively small number of endothelial nuclei, which were not further subclustered (Table 3).

Pleiotropic effects of *APP* and *PSEN1* mutations across brain cell-types. The core of the amyloid hypothesis is that ADAD mutations in *APP* and *PSEN1* lead to dysfunctional neurons, the main source of aggregation-prone Aβ. Strikingly, every cell type we analyzed exhibited cell states enriched within ADAD carriers (Fig. 2a-e and Supplementary Table 4). Two astrocyte cell states were specific to ADAD subjects. One in particular, Astro.4 (β=0.15, P=4.39×10^{-2}), had increased expression of *OSMR* (log2FC=1.48, Adj.P=4.37×10^{-4}), *VIM* (log2FC=1.80, Adj.P=8.92×10^{-13}), and *CTSB* (log2FC=1.56, Adj.P=4.03×10^{-5}) compared to the other astrocyte cell states, similar to the disease-associated astrocytes (DAA) identified in the 5xFAD mouse model[22] (Extended Fig. 1a).

Microglia contained a prominent cell state, Mic.4 (β=0.40, P=2.50×10^{-3}), composed of nuclei from 11 *PSEN1* carriers and one *TREM2* p.R136W carrier (Fig. 2b). This cell state did not show a significant overlap of DEGs with genes featured in established microglia signatures including 'disease-associated microglia' (DAM)[23], 'microglial neurodegenerative' (MGnD)[24], and 'human AD microglia' (HAM)[25] (Supplementary Table 5). Interestingly, Mic.4 showed a significant increase in the expression of *MECP2* (β=0.67, Adj.P=3.20×10^{-10}), which has previously shown differential expression in AD brains and when knocked down in microglia, caused NMDA receptor-dependent excitotoxic neuronal cell death in a mouse model of Rett syndrome[26]. Pathway analysis revealed that the 491 upregulated genes in Mic.4 were associated with MAPK signaling (Adj.P=1.37×10^{-2}), estrogen signaling (Adj.P=1.37×10^{-2}), NOD-like receptor signaling (Adj.P=1.90×10^{-2}), and necroptosis pathways (Adj.P=2.42×10^{-2}), indicating an acute cellular response leading to cell death (Supplemental Table 6). Furthermore, we also observed this microglia cell state in the cortex of seven-month-old 5xFAD mice (high amyloid plaque load at this age), with the same upregulated gene signature (upregulation signature score P=1.92×10^{-374}; Fig. 2f).

Oligo.3 was also significantly associated with ADAD subjects (β=0.63, P=1.48×10^{-6}). Pathway analysis of the oligo.3 upregulated genes revealed an enrichment in spliceosome genes (Adj.P=4.64×10^{-24}) many of which were from the family of heterogeneous nuclear ribonucleoproteins (HNRNP), including *HNRNPA1, HNRNPA2B1, HNRNPA3, HNRNPC, HNRNPD, HNRNPH3, HNRNPK, HNRNPM,* and *HNRNPU,* which have previously been linked to late-onset AD[27] (Extended Fig. 1b, Supplementary Table 6h). The AD risk genes, *PICALM, CLU, APP,* and *MAP1B*, that have intronic excision levels correlated with the expression of HNRNP genes[27], were also overexpressed in Oligo.3 (Extended Fig. 1b). This suggests that *PICALM*, *CLU*, *APP*, and *MAP1B* could be alternatively spliced through HNRNP splicing repression within oligodendrocytes. HNRNP genes also play a role in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)[28] and promote translation of APP[29]. Similarly, additional clusters in oligodendrocytes (Oligo.1), OPCs (clusters 4, 5, 6), and Neurons (Neuro.6) were enriched within ADAD carriers (Fig. 2c-e).

***TREM2* variants modulate microglial and oligodendrocytic transcription states.** Individuals with specific *TREM2* AD risk variants (p.R47H, p.R62H and p.H157Y) showed increased proportions of nuclei in Mic.2 (Fig. 3a; β=0.23, P=3.29×10^{-2}). Mic.2 showed high expression of resting-state-microglia marker genes (*TMEM119, P2RY13, MED12L* and *SELPLG*) and minimal elevation of activated (*ABCA1*, *C5AR1*, *TNFAIP3,* and *CD83*) marker genes compared to our resting (Mic.0) and activated (Mic.1) microglia cell states (Fig. 3b). This is consistent with the reported association of these *TREM2* risk variants with reduced microglial activation[16,30]. To replicate the enrichment of this cluster within *TREM2* risk variant carriers, we analyzed microglia from 32 ROSMAP snRNA-seq samples (DLPFC) that included 11 *TREM2* p.R62H carriers (Synapse ID: syn21125841)[16]. We found that 10.2% of the ROSMAP microglia recapitulated the Mic.2 transcriptional state (signature score: β=0.22 P=7.4×10^{-107}), and that on average, carriers of *TREM2* p.R62H had a higher proportion of their microglia in this unique state than non-carriers (Fig. 3a, Discovery: β=0.20, P=2.58×10^{-2}; Meta-analysis: P=2.26×10^{-2}).

These *TREM2* risk variant carriers were also enriched for Oligo.5 (β=0.13, P=4.66×10^{-2}; Fig. 3c; Supplementary Table 4b), which exhibited upregulation of 1,124 genes including *TFEB* (Log2FC=0.15; Adj.P=8.69×10^{-6}) compared to other oligodendrocyte cell states (Fig. 3d). Altered *TFEB* expression may be driven by the interaction of *TREM2* with *mTOR* which is upstream of *TFEB*[31,32]. *TFEB* is a central regulator of lysosomal biogenesis[33,34] and represses myelination at

different developmental stages[35]. Altered *TFEB* signaling has been implicated in a number of neurodegenerative diseases[33,36]. Oligo.5 also had increased proportions among sAD subjects (Supplementary Table 4d). This cell state was also present in 7.1% of oligodendrocytes from the ROSMAP cohort (upregulation signature score of Oligo.5 $p=9.71\times10^{-96}$) with an increased proportion in p.R62H carriers (Discovery: $\beta=0.13$, $P=2.48\times10^{-2}$; Meta-analysis: $P=6.11\times10^{-3}$) (Fig 3c). Changes in autophagy markers were reported previously in microglia of p.R62H carriers[31] and our data suggest that autophagy dysregulation may also be a feature of oligodendrocytes in p.R62H carriers.

An inspection of the *TREM2* p.R136W variant[10] carrier revealed a divergent signature compared to other *TREM2* carriers. Cellular studies showed that *TREM2* p.R136W severely altered both cell surface and overall TREM2 expression[37]. Three *TREM2* variant carriers, including p.R136W, were diagnosed with AD prior to age 65, but only the p.R136W carrier had a Braak stage of VI (6). In each cell type, this individual's nuclei tended to cluster with the ADAD nuclei (Fig 3e), suggesting that this mutation could altered expression networks more similarly to *PSEN1* or *APP* mutations than to other *TREM2* variants. Due to the rarity of this variant, only a single brain was analyzed; thus, additional replication is required to validate these observations.

MS4A resilience variant carriers show a specific inflammatory microglial activation state. Carriers of rs1582763-A, an intergenic allele associated with reduced risk for AD and higher CSF sTREM2 levels[14], exhibited a trend for decreased proportions of nuclei in Mic.1 (Supplementary Table 4c), an activated microglia state. Mic.1 had an upregulation of activated-microglia genes *CD68* (log2FC=0.39, $P=1.06\times10^{-4}$), *CD83* (log2FC=1.34, $P=9.44\times10^{-37}$), *TNFAIP3* (log2FC=0.57, $P=2.16\times10^{-7}$), *C5AR1* (log2FC=1.23, $P=4.03\times10^{-46}$), *GPNMB* (log2FC=1.73, $P=6.99\times10^{-102}$), and *ABCA1* (log2FC=0.82, $P=1.78\times10^{-45}$; Fig. 3b), and a significant overlap with genes also upregulated in the DAM[23], MGnD[24], 'human Alzheimer's microglia' (HAM)[25], and aging[38] signatures (Supplementary Table 5). In contrast, rs1582763-A carriers showed increased proportions of nuclei in Mic.3 compared to non-carriers ($\beta=0.15$, $P=1.67\times10^{-3}$; Fig 4a,b; Supplementary Table 4c). The Mic.3 state displayed a proinflammatory profile, including the upregulation of *ILB1* (log2FC=3.45, $Adj.P=2.07\times10^{-36}$), *CD14* (log2FC=0.63, $P=1.87\times10^{-3}$), *FCGR3A* (log2FC=0.50, $P=3.31\times10^{-3}$), and *CD40* (log2FC=0.91, $P=9.84\times10^{-3}$; Fig. 4c). Although this rs1582763-associated microglia cell state was present in the ROSMAP cohort (upregulation signature score of Mic.3 $P=8.64\times10^{-152}$), it was not significantly associated with rs1582763,

possibly because the ROSMAP cohort contained few homozygous recessive carriers (n=3) which were enriched by the design of our study.

We then performed differential expression analyses to characterize the genes that differentiated the main activated microglia (Mic.1) from the rs1582763-A activated microglia (Mic.3). Mic.1 showed increased expression of *C5* (log2FC=1.79, Adj.P=3.46×10^{-3}), whereas Mic.3 had increased expression of *C3* (log2FC =-1.24, Adj.P=4.98×10^{-26}; Fig. 4d; Supplementary Table 7a). Previous analyses have indicated a protective role for *C3* in AD and a potentially detrimental role for *C5*[39]. This is consistent with the protective effect of rs1582763-A. Mic.1 also had increased *ACVR1* (log2FC =0.97, Adj.P=2.92×10^{-3}) and *BMPR2* (log2FC =0.39, Adj.P=3.54×10^{-2}), indicating increased BMP signaling, whereas Mic.3 had high *TGFBR1* (log2FC =-0.98, Adj.P=4.82×10^{-17}) and *TGFBR2* (log2FC =-0.35, Adj.P=4.50×10^{-2}), indicating increased TGF-β signaling (Fig. 4d, Supplementary Table 7b). These genes encode for related receptors in the TGF-β superfamily which is implicated in multiple neurological disorders[40]. A gene ontology analysis also showed an upregulation of genes related to cytokine response/production (Adj.P=2.89×10^{-9}) and response to lipopolysaccharides (Adj.P=1.78×10^{-4}). In addition, carriers of rs1582763-A had decreased proportions of resting astrocytes (Astro.0) and a trend towards increased proportions of activated astrocytes (Astro.1) (Fig. 4e, Extended Fig. 1a, Supplementary Table 4). The MS4A genes are not expressed in astrocytes which suggests a possible cellular cross-talk synchronizing the activation of microglia and astrocytes.

Coexpression of *APOE* and *MHC-I* highlights inhibitory neuron vulnerability to neurodegeneration. *APOE*ε4 carriers did not exhibit a unique cell state signature within any cell type likely because we did not enrich the data set for this allele (Table 1, Supplementary Table 4e). However, *APOE*ε4 carriers showed an upregulation of genes involved in ribosome biogenesis (Adj.P=1.42×10^{-2}) and mitotic G1/S transition checkpoint signaling (Adj.P=1.42×10^{-2}) in microglia, maintenance of protein localization in ER (Adj.P=7.75×10^{-3}) and COPI coating of Golgi vesicles (Adj.P=7.75×10^{-3}) in astrocytes, and cytoplasmic translation in both neurons (Adj.P=3.24×10^{-9}) and OPCs (Adj.P=4.34×10^{-5}; Supplementary Table 8). We also explored the role of *APOE* expression in neurons. As was previously described[18], we found a moderate proportion of *APOE*-high expression neurons in controls (20%) with significantly higher proportions in presymptomatic brains (69%, P=6.52×10^{-4}) and considerably lower proportions in both sAD (5.7%, P=2.91×10^{-2}) and ADAD (3.4%, P=2.20×10^{-2}) brains (Fig. 5a,b). We also

observed a correlation between *APOE* and *MHC-I* within specific populations of neurons (Fig. 5c-h). Only inhibitory neurons (Neuron.3 and Neuron.5) showed significant correlations in almost all sub-groups (ADAD, sAD, presymptomatic, and controls), but all inhibitory and excitatory cell states had substantial correlations in ADAD participants (Fig. 5d-h, Supplementary Table 9). An increased expression of *APOE* and *MHC-I* within a neuron seems to tag it for removal[18], which agrees with the sharp increase in *APOE*-high neurons prior to clinical manifestations (presymptomatic) of AD followed by a sharp decrease after (sAD and ADAD). The correlation in Neuron.3 and Neuron.5 across sub-groups supports the premise that these inhibitory neurons are the first to succumb to neurodegeneration[41] and that as pathology progresses, excitatory neurons are also targeted, as seen in the ADAD participants. *APOE*-high nuclei in Neuron.3 and Neuron.5 had upregulated genes in 'cellular response to cytokine stimulus' (Adj.P=3.24×10^{-10}), suggesting cytokines could trigger the increased *APOE* and *MHC-I* expression in these cells[42] (Supplementary Table 10a). Dysregulated genes were involved in circadian entrainment, cAMP signaling, and NMDA receptor activity as previously reported and highlight sleep and circadian rhythm disruption in AD[43] (Supplementary Table 10b).

Genes within risk AD GWAS loci show variable expression between cell states. GWAS meta-analyses have successfully identified genomic loci associated with sAD[11-13]. However, several major hurdles have prevented the field from understanding the contribution of these genomic loci to AD risk. For example, the cell types in which the genomic loci are expressed and their function/dysfunction in these cell types is poorly understood. We sought to leverage our glial and neuronal cell states to map genomic loci implicated in the general AD population onto genes and cell types. We curated a list of genes in loci identified in recent AD GWAS, starting with 89 genes previously prioritized[13,44,45] (77 measured in this study) from 39 different loci (Supplementary Table 11a); we extended the list by adding the 'non-prioritized' genes within all AD GWAS associated loci also present in this data set (540 genes, 46 loci; Supplementary Table 11b,c).

We then inspected the snRNA-seq data to identify which genes were differentially expressed among cell states. Out of the 77 prioritized genes, 39 (from 29 loci) showed differential expression between the cell states of at least one cell type (Fig. 6). We also found that for 13 genes, the cell type with the greatest expression level variability (log fold change between cell states) did not match the cell type with the highest average expression. For example, *BIN1* showed the highest average expression in oligodendrocytes, but OPCs had the largest log2 fold change in expression

(log2FC) between cell states. This suggests that transcriptional states can provide additional independent information to map GWAS genes to cell types. Rare coding variants in *PLCG2*, implicated in *TREM2*-dependent microglial function in AD[46], have been recently reported to protect against AD[47]. We found differential expression of *PLCG2* between cell states within all cell types, suggesting an extended functionality beyond that of the *TREM2* signaling pathway in microglia (Fig 6)[48,49]. *PLCG2* expression was significantly increased in ADAD brains across all cell types except microglia, which was only nominally significant (Supplementary Table 12). *SORL1,* predominantly studied in neurons in the context of AD[50], showed higher average expression in microglia than in neurons and greater log2FC within oligodendrocytes than in neurons. Finally, we report that the three prioritized genes in the *EPHA1* locus (*EPHA1, ZYX,* and *FAM131B*) showed low average expression and minimal log2FC in all cell types (Fig 6). However, when we included all genes in the locus, we found moderate average expression and significant log2FC of *TCAF1* in all cell types except microglia suggesting that this gene could be the lead gene at this locus (Supplementary Material). *TCAF1* helps traffic the receptor *TRPM8* to the plasma membrane, where it regulates inflammation[51,52]. An alternative explanation is that the effector cell type was not well represented in the present study (e.g. endothelial cells, pericytes). These results will be valuable in identifying the relevant genes and cell types linked to GWAS loci.

Discussion

We performed snRNA-seq on over one million nuclei from the parietal cortex (area B1-3,7) of 74 brains with AD genetic variants in *APP, PSEN1, TREM2*, and rs1582763-A, a resilience intergenic variant in the *MS4A* locus associated with increased levels of soluble TREM2 in CSF and reduced AD risk[14]. After a stringent QC process, we retained 67 brains and 294,114 high-quality nuclei representing most major cortical cell types. Our high numbers of nuclei enabled re-clustering of the individual cell types to identify unique biological cell states.

One of the premises of the amyloid hypothesis is that the molecular changes in ADAD could be extrapolated to sAD. Interestingly, we observed cell states that were either unique to or enriched within ADAD subjects for all cell types. Thus, one could conclude that ADAD has altered cellular architecture compared with sAD invalidating extrapolation. However, upon integration of our microglia and oligodendrocyte data from the parietal cortex with snRNA-seq data from the DLPFC of ROSMAP participants, we observed ROSMAP (sAD) nuclei clustering with our ADAD enriched clusters despite the absence of ADAD subjects in the ROSMAP cohort (Extended Fig. 2b,e). The parietal cortex is affected in the later stages of AD progression, whereas the DLPFC is affected earlier in disease progression[53,54]. Therefore, the cell states identified in the parietal cortex of ADAD participants could represent advanced/accelerated stages of pathology which can also be detected in brain regions with high pathology in LOAD like the DLPFC. This would be consistent with the elevated tau PET found in the parietal region of ADAD compared to sAD patients[55]. The differences observed in the parietal cortex could also be due to younger ages in ADAD (average of 32.9 years younger). Additional studies capturing multiple brain regions with varying degrees of pathology and ages are required to understand these observations in more detail.

We report microglia and oligodendrocyte cell states associated with *TREM2* risk variant carriers. *TREM2* is mainly expressed in microglia. However, loss of function mutations in TREM2 cause Nasu-Hakola disease, which is characterized by white matter changes, including loss of myelination, suggesting an oligodendrocyte-microglia crosstalk[56]. Therefore, it is likely that AD patients with *TREM2* variants had altered microglia behavior which changed the microenvironment, driving the oligodendrocyte cell state found in *TREM2* variant carriers. The microglial cluster associated with these reduced activation variant carriers displayed a resting-like state, suggesting they might benefit from antibody treatments targeting *TREM2* to prevent its

cleavage, increasing microglial activation[57]. However, not all *TREM2* risk variants are functionally equivalent[14,58,59]. *TREM2* p.R47H, p.R62H, and p.H157Y show reduced cell activation *in vitro*[16,30] and we demonstrate that brains from these variant carriers exhibit altered microglia and oligodendrocytes populations. Interestingly, a single *TREM2* p.R136W carrier clustered closer with the ADAD subjects than the other *TREM2* carriers in astrocytes, microglia, oligodendrocytes, and OPCs.

We also observed a microglia state expressing activation markers enriched for homozygous carriers of the *MS4A* resilient variant. In contrast to the major activated microglia state, this cell state showed upregulation of proinflammatory genes and other genes related to cytokine signaling. We also replicated the finding that the number of *APOE*-high neurons increases before the clinical onset of AD and decreases during disease progression[18]. We also found a correlation between *APOE* expression and *MHC-I* expression. This correlation was predominantly seen in inhibitory neurons, which supports the current view that inhibitory neurons are affected before their excitatory counterparts[41]. ADAD participants showed significant coexpression of *APOE/MHC-I* in all neuronal cell states.

The majority of the AD GWAS hits are non-coding variants and potentially cause altered gene expression. Thus, we inspected our cell state differential expression results to identify genes within AD GWAS loci that showed variable expression within cell types. The results highlight the specific cell types influenced by the genes and show that many genes play essential roles across cell types, including *PLCG2* and *SORL1*. In the future, cell-type-specific eQTL analyses using the lead GWAS hits could confirm if any of the expression variability is due to genotypic status.

In conclusion, a single AD risk variant can alter the transcriptional landscape of multiple brain cell types. Pathogenic mutations in *APP* and *PSEN1* influenced profiles of neurons but more especially glia when compared to controls and sAD parietal cortices. We found *TREM2* risk variants that shifted microglial and oligodendrocytic profiles and that the resilience *MS4A* variant altered the 'activated' microglia profile. Each of these changes has the potential to modify the pathological progression and the clinical manifestations of AD. Overall, this suggests the critical need to incorporate genetics when screening participants for drug trials and selecting the optimal treatment program for AD patients.

Materials and Methods

Processing of brain tissue samples

The Neuropathology Core of the Knight-Alzheimer Disease Research Center (Knight-ADRC) and the Dominantly Inherited Alzheimer Network (DIAN) provided the parietal lobe tissue of postmortem brains for each sample. These samples were obtained with informed consent for research use and were approved by the review board of Washington University in St. Louis. AD neuropathological changes were assessed according to the National Institute on Aging-Alzheimer's Association (NIA-AA) criteria. Their demographic, clinical severity and neuropathological information are presented in Table 1.

Nuclei isolation and snRNA-seq on the 10X Genomics platform

The 74 frozen human parietal cortices were processed according to the 'Nuclei extraction and library preparation' protocol described in [19]. Briefly, the tissue was homogenized and the nuclei were isolated using a density gradient. The nuclei were then sequenced using the 10X Chromium single cell Reagent Kit v3, with 10,000 cells per sample and 50,000 reads per cell for each of the 74 samples.

SnRNA-seq data processing with 10X Genomics CellRanger software and data filtering

The CellRanger (v3.0.2 10XGenomics) software was employed to align the sequences and quantify gene expression. We used the GRCH38 (3.0.0) reference to prepare a pre-mRNA reference according to the steps detailed by 10X Genomics. The software was packaged into a Docker container (https://hub.docker.com/r/ngicenter/cellranger3.0.2), allowing us to launch it within the McDonnell Genome Institute (MGI) infrastructure, reducing the computing time for generating the BAM files.

Filtering and QC were done using the Seurat package (3.0.1) on each subject individually. Each raw gene expression matrix for each sample was plotted, using *BarcodeInflectionsPlot* to calculate the inflection points derived from the barcode-rank distribution. Once the thresholds were determined, a subset of the data was isolated. We removed nuclei with high mitochondria gene expression following the dynamic model proposed by Tsai et al.[15]. Briefly, the nuclei were grouped by their percentage of mitochondria values using k=2 clustering, and the group with the higher percentage values was removed. Genes that were not expressed in at least ten nuclei were removed from the final matrix. To detect and discard doublets, we used *Doublet Finder*[60], which removes nuclei with expression profiles that resemble synthetically mixed nuclei from the dataset. The gene

expression matrices from all samples were combined in R independently for further processing using the Seurat protocol.

Dimensionality reduction, clustering, and UMAP

The merged expression matrix was normalized using the *SCTransform* protocol by Seurat. This function calculates a model of technical noise in scRNA-seq data using 'regularized negative binomial regression' as described previously[61]. We regressed out, during the normalization, the number of genes, the number of UMIs, and the percentage of mitochondria. The principal components were calculated using the first 3000 variable genes, and the Uniform Manifold Approximation and Projection (UMAP) analysis was performed with the top 14 PCs. The clustering was done using a resolution of 0.2.

Cluster annotation and quantification of regional and individual contributions to cell types.

We employed a list of marker genes that we had previously curated[19] to annotate brain single-nuclei RNA-seq data. We used the *DotPlot* function (Seurat package) to visualize the average expression of genes related to specific cell types. This approach enabled the labeling of cell types based on the overall expression profile of the nuclei, regardless of dropout events. In addition, we employ a supervised method termed Garnett[62] that leverages machine learning to classify each of the nuclei and estimate cluster homogeneity. This method also provides a metric of gene ambiguity, which enables further optimization of the marker genes to be included in the classification process. For this method, we employed *SYT1, SNAP25, GRIN1* to classify neurons, *NRGN, SLC17A7, CAMK2A* for excitatory neurons and *GAD1, GAD2* for inhibitory neurons; *AQP4, GFAP* for astrocytes; *CSF1R, CD74, C3* for microglia; *MOBP, PLP1, MBP* for oligodendrocytes; *PDGFRA, CSPG4, VCAN* for oligodendrocyte precursor cells (OPCs); *CLDN5, TM4SF1, CDH5* for endothelial cells and ANPEP for pericytes. We employed the function *check_markers* (Garnett package) to evaluate the ambiguity score and the relative number of cells for each cell type. A classifier was then trained using the marker file, with “num_unknown” set to 50. This classifier annotates cells with cell-type assignments extended to nearby cells using the “clustering-extended type” labeling option. At this stage, 1 ambiguous cluster and 1 subject-specific cluster were dropped.

Identification of alternative cell-type transcription states

The nuclei within the primary cell-type clusters were each isolated from the full dataset and re-clustered. We re-normalized the data subset using the same protocol as explained in section 4 in

methods. The number of PCs used for UMAP dimensionality reduction was different for each cell type, 4, 8, 10, 6, 5, for Neuron, Oligodendrocytes, Microglia, Astrocytes, and OPC, respectively. We then employed Seurat's *FindNeighbors* and *FindClusters* functions to identify unique cell states or subclusters (resolution=0.1, 0.2, 0.2, 0.05, 0.15). Additionally, we used the Garnett protocol to examine nuclei in each expression state within each cell type to detect and remove those nuclei that did not resemble a trustworthy expression profile from downstream analyses. We dropped 1 subject-specific subcluster in Neurons. After this final stage of QC, we ended with 67 of the 74 brains and 294,114 of the 1,102,459 nuclei.

Differential proportion analysis

To identify associations between cell-type transcriptional state and disease status or genetic strata (ADAD, sAD, TREM2, TREM2_reduced, rs158276), we employed linear regression models to test each subject's cell state compositions. More explicitly, the number of nuclei a subject had in a specific cell state was divided by the subject's total nuclei count for that cell type creating a proportion. The proportions were normalized using a cube root transformation and were corrected

$$Proportion = \frac{numberNuclei(state_i)}{totalNuclei(cell\text{-}type)}$$

by sex, age of death (AOD), and disease status depending on the variable of interest. We removed subjects that contributed fewer than 60 nuclei to the cell type cluster. The TREM2 analyses only included the 31 sAD subjects. We utilized glm, a standard function in R, to implement the model.

rs1582763 (Additive):	$(Proportion)^{1/3} \sim Genotype + Sex + ADstatus$
APOEε4+:	$(Proportion)^{1/3} \sim APOEstatus + Sex + ADstatus$
TREM2, TREM2_reduced:	$(Proportion)^{1/3} \sim TREM2status + Sex + AOD$
ADAD:	$(Proportion)^{1/3} \sim Sex + ADstatus$
sAD:	$(Proportion)^{1/3} \sim Sex + ADstatus$

Cell state differential expression

To determine if there was unique functionality or potentially altered cell states due to disease, we fitted a linear mixed model that predicted the expression level of each gene for the individual nuclei by cell state and corrected for the subject of origin and sex[63]. Control, sAD, and ADAD subjects were used to calculate cell state differentially expressed genes. Expression levels were extracted from the Seurat objects using *GetAssayData* with the 'slot' parameter set to 'counts'. Age of death (AOD) was not included in the model because AOD is correlated with ADAD status. R package nebula version 1.1.5 was used to implement the model, including parameters for a zero-inflated

negative binomial distribution and the random effect of the subject of origin[64]. The number of UMI's per nuclei was already corrected for during *SCTransformation*, so the model did not need to account for the number of UMI's.

Expression ~ CellStateStatus + Sex + (1|Subject); ZI model= ~1; ZINB distribution

Within each cell type, differential expression (DE) was calculated on each cell state versus all other states and each state against each other state individually.

Genetic factor differential expression

Differentially expressed genes were identified within each cell type by genetic status, namely ADAD, TREM2, or sAD vs. Controls and *APOE*ε4- vs. *APOE*ε4+. The nuclei for each status (ADAD, TREM2 variant carriers, sAD, and APOEe4+) were isolated by groups. Each group was individually compared to controls (neuropath free controls or APOEe4-) using linear mixed models as explained above. The following model was used:

Expression ~ GeneticStatus + AOD + Sex + (1|Subject); ZI model= ~1; ZINB distribution

Age of Death (AOD) was not included when analyzing ADAD and *APOE*ε4+. This model was run on the entire cell type as a whole and each individual cell state within the cell types.

Pathway analysis

The upregulated genes identified for each cell state were used in a subsequent pathway analysis. We used the R-based application Enrichr hosted by the Ma'ayan Laboratory[65,66]. We used gene sets to determine pathway enrichment by using the "KEGG 2021 Human" or "GO Biological Process 2021" gene sets. Downregulated genes were also run in the APOE-high neuron analysis.

Microglia expression states

We collected 12 gene sets (Aging[38], Homeostatic[67], Lipid-droplet-accumulating[68], Neurodegenerative[24], Proliferative-region-associated[69], Injury-responsive[70], Activated-response[71], Interferon-response[71], Human-Alzheimer's[25], Disease-associated[23], LRRK2, and Granulin microglia) associated with different microglia functional states that had been described in the literature. Each set was split into their up and down-regulated gene lists. A hypergeometric test was performed to identify which microglia transcriptional states previously reported were recapitulated in this dataset.

Neuronal *APOE* and *MHC-I* coexpression

We followed the methods outlined by Zalocusky *et al.*[18]. Briefly, MAGIC[72] was used to impute gene expression, *APOE* expression greater than two standard deviations marked *APOE*-high

expression, and the genes *HLA-A, HLA-B, HLA-C, HLA-E, HLA-F*, and *B2M* were summed to represent *MHC-I* expression. Correlation between *APOE* and *MCH-I* expression was calculated using the native *cor.test* function in R. Genes differentially expressed between *APOE*-high and *APOE*-low neurons were identified using linear mixed models as previously described (*nebula* R package; model: expression ~ apoeHigh + Sex). These DE analyses were performed only on the neuro.3 and neuro.5 cell states split by AD condition. Enriched pathways in the significantly upregulated genes were identified using EnrichR as previously described.

Characterization of loci identified in AD risk GWAS

A list of genes identified through AD GWAS was collected. We started with 89 genes from 39 different loci that were previously prioritized[13,44,45]. We then added in all 878 genes from all the significant loci. This list was filtered by genes in our snRNA-seq data set, leaving 540 genes in 46 different loci. The significant cell state DE results were queried for the genes within our curated lists and extracted by cell type. We did not include cell states that contained less than 5% of that cell type's nuclei. For DE analyses that were a cell state against all others (e.g. mic.1 vs all other mic cell states), we removed gene hits that had a negative estimate. We calculated the log2 fold change from the estimates provided by nebula by using this equation:

$\log_2 FC = \log_2(e^{estimate})$

If a gene had multiple hits within a single cell type, we selected the maximum log fold change in our plots. The log2FC values greater than three were capped at three to preserve visual variability. For each gene in our gene list, we calculated the mean expression in each cell type.

ADAD-specific microglia cluster validation

To confirm the biological existence of Mic.4, a microglia expression state composed of ADAD subjects, single microglia cells (n=1551) from 5xFAD mice were collected as described in [73]. Protein encoding mouse genes were converted to their human orthologs using biomaRt. The data were normalized using *SCTransform*, regressing out numbers of genes and UMIs. Seurat's best practices workflow was followed to integrate the mouse and human microglia using the human microglia as the reference dataset. Seven clusters were assigned using *FindNeighbors* and *FindClusters* with the first eight principal components and a resolution of 0.2 as input. A total of 297 mouse cells and 1413 of the 1429 human Mic.4 nuclei were recaptured in the post-integration cluster 2. The mouse cells were then re-isolated. A human Mic.4 signature score was calculated for each mouse cell by running the 412 upregulated genes that passed multiple test corrections into

Seurat's *AddModuleScore.* The significant pairwise differences in cluster scores were calculated using the linear mixed model *moduleScore ~ cluster, ziformula=~0, family='gaussian'* executed using *glmmTMB.*

TREM2-enriched microglia cluster validation

Human single nucleus data from the ROSMAP samples (DLPFC) generated as described in[16] were used to confirm the Mic.2 cell state, enriched for TREM2 p.R47H, p.R62H and p.H157Y. The ROSMAP data has 11 sAD, 11 TREM2 R62H, and 10 control subjects with 3,986 microglia. The microglia were isolated and normalized using Seurat's *SCTransform* function with "return.only.var.genes" set to FALSE and regressing out "nCount_RNA" and "nFeature_RNA". The microglia were then integrated using 3000 features in *SelectIntegrationFeatures*, *PrepSCTintegration*, our data as a reference in *FindIntegrationAnchors*, and *IntegrateData*. To identify which nuclei fell into our original clusters, the integrated data was clustered using the first ten principal components as input for *FindNeighbors* and a resolution of 15 in *FindClusters*. This shattered the data finding 147 clusters. We assigned each of these 147 clusters an 'original' cluster identity by isolating our cohort of nuclei from the individual clusters and identifying the most common original ID. This ID was then transferred to the ROSMAP nuclei, similar to a nearest neighbor classifier[74]. These cluster identities were mapped to the pre-integrated normalized ROSMAP data. To measure the accuracy of our label transfer, a mic.2 signature score was calculated for each nucleus by running the mic.2 significantly upregulated genes into Seurat's A*ddModuleScore.* Out of the 125 upregulated genes, 114 were present in the ROSMAP data. The significant pairwise differences between clusters were calculated using the linear mixed model *moduleScore ~ cluster + AOD + Sex + (1|subject), ziformula=~0, family='gaussian'* executed using *glmmTMB.* The previously described 'Differential Proportion Analysis' methods were followed to verify the enrichment of TREM2 R62H nuclei in the ROSMAP Mic.2 cluster.

TREM2-enriched oligodendrocyte cluster validation

The two oligodendrocyte clusters (29,478 nuclei) in the ROSMAP data were integrated with our oligodendrocytes and labeled with our original cluster identities as described above. A total of 264 clusters were identified using eight PCA dimensions at a resolution of 15. Because of the large number of nuclei and the reduced variability of our Oligodendrocytes, many of the Oligodendrocyte cell states shared upregulated genes. To find a clear signal for Oligo.5, we removed all upregulated genes from Oligo.5 that had a larger β (estimate) in another cell type. This

left 77 of the 638 upregulated genes. 73 of those 77 were present in the ROSMAP data and used to calculate the Oligo.5 signature score. Pairwise significant differences in signature score and enrichment of TREM2 R62H variant carriers were calculated as described above. The previously described 'Differential Compositional Analysis' methods were followed to verify the enrichment of TREM2 R62H nuclei in the ROSMAP Oligo.5 cluster.

Rs1582763-enriched microglia cluster validation

Using the same cluster-labelled ROSMAP microglia as described above, we calculated a mic.3 signature score. 192 of the 207 upregulated Mic.3 genes were present in the ROSMAP data. Pairwise significant differences in signature score and enrichment of rs1582763 carriers were calculated as described above. The previously described 'Differential Compositional Analysis' methods were followed to verify the enrichment of rs1582763 nuclei in the ROSMAP Mic.3 cluster.

Data visualization browser

We developed the Single Nucleus Alzheimer disease RNA-seq Explorer (SNARE) to host the single nucleus expression data using the cellxgene[75] platform. It can be publically accessed at http://ngi.pub/SNARE. We include the UMAP representations for the full data set (includes all cell types) and the individual cell type subclustering.

Data availability

The single nucleus data from the Knight ADRC is publicly available by request from the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) with accession number NG00108 (https://www.niagads.org/datasets/ng00108). To access the data from the DIAN brain bank, special request must be made using this URL: https://dian.wustl.edu/our-research/for-investigators/. The 5xFAD mouse microglia data are located in the Gene Expression Omnibus (GEO database) under the accession number GSE141917 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141917). The ROSMAP single nucleus RNA sequencing data is available at Synapse under Synapse ID syn21125841 (https://www.synapse.org/#!Synapse:syn21125841/wiki/597278).

Code availability

Custom code used to analyze the snRNA-seq data and datasets generated and/or analyzed in the current study are available from the corresponding authors upon request.

Supplementary Material

Locus_Plots – pdf file

DEG_by_cell_state – zip folder of xlsx files

Funding

This work was possible thanks to the following governmental grants from the National institute of Health: NIA R01AG057777 (OH), R56AG067764 (OH) P30AG066444 (JCM), P01AGO26276 (JCM), U19AG032438 (RJB), R01AG044546 (CC), P01AG003991 (JCM), RF1AG053303 (CC), RF1AG058501 (CC), U01AG058922 (CC), NINDS R01NS118146 (BAB), RFNS110809 (CMK), R01AG062734 (CMK), U01AG072464 (OH, CMK), the BrightFocus Foundation (CMK), and the Chan Zuckerberg Initiative (CZI). O.H. is an Archer Foundation Research Scientist.

This work was supported by access to equipment made possible by the Hope Center for Neurological Disorders, and the Departments of Neurology and Psychiatry at Washington University School of Medicine.

Acknowledgments

Data collection and sharing for this project was supported by The Dominantly Inherited Alzheimer Network (DIAN, U19AG032438) funded by the National Institute on Aging (NIA),the

Alzheimer's Association (SG-20-690363-DIAN), the German Center for Neurodegenerative Diseases (DZNE), Raul Carrea Institute for Neurological Research (FLENI), Partial support by the Research and Development Grants for Dementia from Japan Agency for Medical Research and Development, AMED, and the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), Spanish Institute of Health Carlos III (ISCIII), Canadian Institutes of Health Research (CIHR), Canadian Consortium of Neurodegeneration and Aging, Brain Canada Foundation, and Fonds de Recherche du Québec – Santé. This manuscript has been reviewed by DIAN Study investigators for scientific content and consistency of data interpretation with previous DIAN Study publications. We acknowledge the altruism of the participants and their families and contributions of the DIAN research and support staff at each of the participating sites for their contributions to this study

The results published here are in part based on data obtained from the AD Knowledge Portal (https://adknowledgeportal.org). Study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by NIA grants P30AG10161 (ROS), R01AG15819 (ROSMAP; genomics and RNAseq), R01AG17917 (MAP), R01AG30146, R01AG36042 (5hC methylation, ATACseq), RC2AG036547 (H3K9Ac), R01AG36836 (RNAseq), R01AG48015 (monocyte RNAseq) RF1AG57473 (single nucleus RNAseq), U01AG32984 (genomic and whole exome sequencing), U01AG46152 (ROSMAP AMP-AD, targeted proteomics), U01AG46161(TMT proteomics), U01AG61356 (whole genome sequencing, targeted proteomics, ROSMAP AMP-AD), the Illinois Department of Public Health (ROSMAP), and the Translational Genomics Research Institute (genomic). Additional phenotypic data can be requested at www.radc.rush.edu.

Figure 1 was created with BioRender.com

Competing Interests

CC receives research support from: Biogen, EISAI, Alector and Parabon. The funders of the study had no role in the collection, analysis, or interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. CC is a member of the advisory board of Vivid genetics, Halia Therapeutics and ADx Healthcare.

Reference

1 LaFerla, F. M. & Oddo, S. Alzheimer's disease: Abeta, tau and synaptic dysfunction. *Trends Mol Med* **11**, 170-176, doi:10.1016/j.molmed.2005.02.009 (2005).

2 Long, J. M. & Holtzman, D. M. Alzheimer Disease: An Update on Pathobiology and Treatment Strategies. *Cell* **179**, 312-339, doi:10.1016/j.cell.2019.09.001 (2019).

3 Goate, A. *et al.* Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349**, 704-706, doi:10.1038/349704a0 (1991).

4 Murrell, J., Farlow, M., Ghetti, B. & Benson, M. D. A mutation in the amyloid precursor protein associated with hereditary Alzheimer's disease. *Science* **254**, 97-99, doi:10.1126/science.1925564 (1991).

5 St George-Hyslop, P. *et al.* Genetic evidence for a novel familial Alzheimer's disease locus on chromosome 14. *Nat Genet* **2**, 330-334, doi:10.1038/ng1292-330 (1992).

6 Levy-Lahad, E. *et al.* Candidate gene for the chromosome 1 familial Alzheimer's disease locus. *Science* **269**, 973-977, doi:10.1126/science.7638622 (1995).

7 Karran, E., Mercken, M. & De Strooper, B. The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics. *Nat Rev Drug Discov* **10**, 698-712, doi:10.1038/nrd3505 (2011).

8 Martins, I. J. *et al.* Apolipoprotein E, cholesterol metabolism, diabetes, and the convergence of risk factors for Alzheimer's disease and cardiovascular disease. *Mol Psychiatry* **11**, 721-736, doi:10.1038/sj.mp.4001854 (2006).

9 Podlesny-Drabiniok, A., Marcora, E. & Goate, A. M. Microglial Phagocytosis: A Disease-Associated Process Emerging from Alzheimer's Disease Genetics. *Trends Neurosci* **43**, 965-979, doi:10.1016/j.tins.2020.10.002 (2020).

10 Jin, S. C. *et al.* Coding variants in TREM2 increase risk for Alzheimer's disease. *Hum Mol Genet* **23**, 5838-5846, doi:10.1093/hmg/ddu277 (2014).

11 Hollingworth, P. *et al.* Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nature Genetics* **43**, 429-+, doi:10.1038/ng.803 (2011).

12 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, doi:10.1038/ng.2802 (2013).

13 Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* **51**, 414-430, doi:10.1038/s41588-019-0358-2 (2019).

14 Deming, Y. *et al.* The MS4A gene cluster is a key modulator of soluble TREM2 and Alzheimer's disease risk. *Sci Transl Med* **11**, doi:10.1126/scitranslmed.aau2291 (2019).

15 Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332-337, doi:10.1038/s41586-019-1195-2 (2019).

16 Zhou, Y. *et al.* Human and mouse single-nucleus transcriptomics reveal TREM2-dependent and TREM2-independent cellular responses in Alzheimer's disease. *Nat Med* **26**, 131-142, doi:10.1038/s41591-019-0695-9 (2020).

17 Grubman, A. *et al.* A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat Neurosci* **22**, 2087-2097, doi:10.1038/s41593-019-0539-4 (2019).

18 Zalocusky, K. A. *et al.* Neuronal ApoE upregulates MHC-I expression to drive selective neurodegeneration in Alzheimer's disease. *Nat Neurosci* **24**, 786-798, doi:10.1038/s41593-021-00851-3 (2021).

19 Del-Aguila, J. L. *et al.* A single-nuclei RNA sequencing study of Mendelian and sporadic AD in the human brain. *Alzheimers Res Ther* **11**, 71, doi:10.1186/s13195-019-0524-x (2019).
20 Braak, H. & Braak, E. Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiol Aging* **16**, 271-278; discussion 278-284, doi:10.1016/0197-4580(95)00021-6 (1995).
21 Lake, B. B. *et al.* Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586-1590, doi:10.1126/science.aaf1204 (2016).
22 Habib, N. *et al.* Disease-associated astrocytes in Alzheimer's disease and aging. *Nat Neurosci* **23**, 701-706, doi:10.1038/s41593-020-0624-8 (2020).
23 Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276-1290 e1217, doi:10.1016/j.cell.2017.05.018 (2017).
24 Krasemann, S. *et al.* The TREM2-APOE Pathway Drives the Transcriptional Phenotype of Dysfunctional Microglia in Neurodegenerative Diseases. *Immunity* **47**, 566-581 e569, doi:10.1016/j.immuni.2017.08.008 (2017).
25 Srinivasan, K. *et al.* Alzheimer's patient brain myeloid cells exhibit enhanced aging and unique transcriptional activation. *bioRxiv*, 610345, doi:10.1101/610345 (2019).
26 Kim, B., Choi, Y., Kim, H. S. & Im, H. I. Methyl-CpG Binding Protein 2 in Alzheimer Dementia. *Int Neurourol J* **23**, S72-81, doi:10.5213/inj.1938196.098 (2019).
27 Raj, T. *et al.* Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nat Genet* **50**, 1584-1592, doi:10.1038/s41588-018-0238-1 (2018).
28 Bampton, A., Gittings, L. M., Fratta, P., Lashley, T. & Gatt, A. The role of hnRNPs in frontotemporal dementia and amyotrophic lateral sclerosis. *Acta Neuropathol* **140**, 599-623, doi:10.1007/s00401-020-02203-0 (2020).
29 Lee, E. K. *et al.* hnRNP C promotes APP translation by competing with FMRP for APP mRNA recruitment to P bodies. *Nat Struct Mol Biol* **17**, 732-739, doi:10.1038/nsmb.1815 (2010).
30 Song, W. *et al.* Alzheimer's disease-associated TREM2 variants exhibit either decreased or increased ligand-dependent activation. *Alzheimers Dement* **13**, 381-387, doi:10.1016/j.jalz.2016.07.004 (2017).
31 Ulland, T. K. *et al.* TREM2 Maintains Microglial Metabolic Fitness in Alzheimer's Disease. *Cell* **170**, 649-663 e613, doi:10.1016/j.cell.2017.07.023 (2017).
32 Napolitano, G. *et al.* mTOR-dependent phosphorylation controls TFEB nuclear export. *Nat Commun* **9**, 3312, doi:10.1038/s41467-018-05862-6 (2018).
33 Liu, J. & Li, L. Targeting Autophagy for the Treatment of Alzheimer's Disease: Challenges and Opportunities. *Front Mol Neurosci* **12**, 203, doi:10.3389/fnmol.2019.00203 (2019).
34 Sardiello, M. *et al.* A Gene Network Regulating Lysosomal Biogenesis and Function. *Science* **325**, 473-477, doi:10.1126/science.1174447 (2009).
35 Meireles, A. M. *et al.* The Lysosomal Transcription Factor TFEB Represses Myelination Downstream of the Rag-Ragulator Complex. *Dev Cell* **47**, 319-330 e315, doi:10.1016/j.devcel.2018.10.003 (2018).
36 Xu, Y. *et al.* TFEB regulates lysosomal exocytosis of tau and its loss of function exacerbates tau pathology and spreading. *Mol Psychiatr*, doi:10.1038/s41380-020-0738-0 (2020).
37 Sirkis, D. W. *et al.* Rare TREM2 variants associated with Alzheimer's disease display reduced cell surface expression. *Acta Neuropathol Commun* **4**, 98, doi:10.1186/s40478-016-0367-7 (2016).
38 Galatro, T. F. *et al.* Transcriptomic analysis of purified human cortical microglia reveals age-associated changes. *Nat Neurosci* **20**, 1162-1171, doi:10.1038/nn.4597 (2017).
39 Veerhuis, R., Nielsen, H. M. & Tenner, A. J. Complement in the brain. *Mol Immunol* **48**, 1592-1603, doi:10.1016/j.molimm.2011.04.003 (2011).

40 Kashima, R. & Hata, A. The role of TGF-beta superfamily signaling in neurological disorders. *Acta Bioch Bioph Sin* **50**, 106-120, doi:10.1093/abbs/gmx124 (2018).
41 Shimojo, M. *et al.* Selective Disruption of Inhibitory Synapses Leading to Neuronal Hyperexcitability at an Early Stage of Tau Pathogenesis in a Mouse Model. *Journal of Neuroscience* **40**, 3491-3501, doi:10.1523/Jneurosci.2880-19.2020 (2020).
42 Zhang, H. L., Wu, L. M. & Wu, J. Cross-Talk between Apolipoprotein E and Cytokines. *Mediat Inflamm* **2011**, doi:10.1155/2011/949072 (2011).
43 Phan, T. X. & Malkani, R. G. Sleep and circadian rhythm disruption and stress intersect in Alzheimer's disease. *Neurobiol Stress* **10**, doi:10.1016/j.ynstr.2018.10.001 (2019).
44 Andrews, S. J., Fulton-Howard, B. & Goate, A. Interpretation of risk loci from genome-wide association studies of Alzheimer's disease. *Lancet Neurol* **19**, 326-335, doi:10.1016/S1474-4422(19)30435-1 (2020).
45 Schwartzentruber, J. *et al.* Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat Genet* **53**, 392-402, doi:10.1038/s41588-020-00776-w (2021).
46 Andreone, B. J. *et al.* Alzheimer's-associated PLC gamma 2 is a signaling node required for both TREM2 function and the inflammatory response in human microglia. *Nature Neuroscience* **23**, 927-+, doi:10.1038/s41593-020-0650-6 (2020).
47 Takalo, M. *et al.* The Alzheimer's disease-associated protective Plc gamma 2-P522R variant promotes immune functions. *Mol Neurodegener* **15**, doi:10.1186/s13024-020-00402-7 (2020).
48 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, doi:10.1038/ng.3916 (2017).
49 Magno, L. *et al.* Alzheimer's disease phospholipase C-gamma-2 (PLCG2) protective variant is a functional hypermorph. *Alzheimers Res Ther* **11**, 16, doi:10.1186/s13195-019-0469-0 (2019).
50 Yin, R. H., Yu, J. T. & Tan, L. The Role of SORL1 in Alzheimer's Disease. *Mol Neurobiol* **51**, 909-918, doi:10.1007/s12035-014-8742-5 (2015).
51 Gkika, D. *et al.* TRP channel-associated factors are a novel protein family that regulates TRPM8 trafficking and activity. *J Cell Biol* **208**, 89-107, doi:10.1083/jcb.201402076 (2015).
52 Silverman, H. A., Chen, A. D., Kravatz, N. L., Chavan, S. S. & Chang, E. H. Involvement of Neural Transient Receptor Potential Channels in Peripheral Inflammation. *Front Immunol* **11**, doi:10.3389/fimmu.2020.590261 (2020).
53 Elahi, F. M. & Miller, B. L. A clinicopathological approach to the diagnosis of dementia. *Nat Rev Neurol* **13**, 457-476, doi:10.1038/nrneurol.2017.96 (2017).
54 Smith, A. D. Imaging the progression of Alzheimer pathology through the brain. *Proc Natl Acad Sci U S A* **99**, 4135-4137, doi:10.1073/pnas.082107399 (2002).
55 Gordon, B. A. *et al.* Tau PET in autosomal dominant Alzheimer's disease: relationship with cognition, dementia and other biomarkers. *Brain* **142**, 1063-1076, doi:10.1093/brain/awz019 (2019).
56 Bianchin, M. M., Martin, K. C., de Souza, A. C., de Oliveira, M. A. & Rieder, C. R. Nasu-Hakola disease and primary microglial dysfunction. *Nat Rev Neurol* **6**, 2 p following 523, doi:10.1038/nrneurol.2010.17-c1 (2010).
57 Schlepckow, K. *et al.* Enhancing protective microglial activities with a dual function TREM2 antibody to the stalk region. *EMBO Mol Med* **12**, e11227, doi:10.15252/emmm.201911227 (2020).
58 McQuade, A. *et al.* Gene expression and functional deficits underlie TREM2-knockout microglia responses in human models of Alzheimer's disease. *Nature Communications* **11**, doi:10.1038/s41467-020-19227-5 (2020).

59 Piccio, L. *et al.* Cerebrospinal fluid soluble TREM2 is higher in Alzheimer disease and associated with mutation status. *Acta Neuropathologica* **131**, 925-933, doi:10.1007/s00401-016-1533-5 (2016).
60 McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337 e324, doi:10.1016/j.cels.2019.03.003 (2019).
61 Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, 296, doi:10.1186/s13059-019-1874-1 (2019).
62 Pliner, H. A., Shendure, J. & Trapnell, C. Supervised classification enables rapid annotation of cell atlases. *Nat Methods* **16**, 983-986, doi:10.1038/s41592-019-0535-3 (2019).
63 Zimmerman, K. D., Espeland, M. A. & Langefeld, C. D. A practical solution to pseudoreplication bias in single-cell studies. *Nat Commun* **12**, 738, doi:10.1038/s41467-021-21038-1 (2021).
64 He, L. *et al.* NEBULA is a fast negative binomial mixed model for differential or co-expression analysis of large-scale multi-subject single-cell data. *Commun Biol* **4**, 629, doi:10.1038/s42003-021-02146-6 (2021).
65 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128, doi:10.1186/1471-2105-14-128 (2013).
66 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).
67 Friedman, B. A. *et al.* Diverse Brain Myeloid Expression Profiles Reveal Distinct Microglial Activation States and Aspects of Alzheimer's Disease Not Evident in Mouse Models. *Cell Rep* **22**, 832-847, doi:10.1016/j.celrep.2017.12.066 (2018).
68 Marschallinger, J. *et al.* Lipid-droplet-accumulating microglia represent a dysfunctional and proinflammatory state in the aging brain. *Nat Neurosci* **23**, 194-208, doi:10.1038/s41593-019-0566-1 (2020).
69 Li, Q. *et al.* Developmental Heterogeneity of Microglia and Brain Myeloid Cells Revealed by Deep Single-Cell RNA Sequencing. *Neuron* **101**, 207-223 e210, doi:10.1016/j.neuron.2018.12.006 (2019).
70 Hammond, T. R. *et al.* Single-Cell RNA Sequencing of Microglia throughout the Mouse Lifespan and in the Injured Brain Reveals Complex Cell-State Changes. *Immunity* **50**, 253-271 e256, doi:10.1016/j.immuni.2018.11.004 (2019).
71 Sala Frigerio, C. *et al.* The Major Risk Factors for Alzheimer's Disease: Age, Sex, and Genes Modulate the Microglia Response to Abeta Plaques. *Cell Rep* **27**, 1293-1306 e1296, doi:10.1016/j.celrep.2019.03.099 (2019).
72 van Dijk, D. *et al.* Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174**, 716-+, doi:10.1016/j.cell.2018.05.061 (2018).
73 Da Mesquita, S. *et al.* Meningeal lymphatics affect microglia responses and anti-Abeta immunotherapy. *Nature* **593**, 255-260, doi:10.1038/s41586-021-03489-0 (2021).
74 Cover, T. M. & Hart, P. E. Nearest Neighbor Pattern Classification. *Ieee T Inform Theory* **13**, 21-+, doi:Doi 10.1109/Tit.1967.1053964 (1967).
75 Megill, C. *et al.* cellxgene: a performant, scalable exploration platform for high dimensional sparse matrices. *bioRxiv*, 2021.2004.2005.438318, doi:10.1101/2021.04.05.438318 (2021).

Table 1: Demographic Characteristics of Samples

Samples	Control	ADAD	sAD	Presym	Other
Total	9	16	31	3	8
MS4A (AG %)*	55.6	46.7	45.2	33.3	12.5
TREM2	-	-	15	-	4
PSEN1	-	13	-	-	-
APP	-	3	-	-	-
Gender (male)%	33.3	56.3	45.2	33.3	50.0
AOD (mean,sd)y	90.1(9.6)	51.0(6.9)	81.5(6.4)	77.3(15.3)	88.8(6.1)
APOE Ɛ4+ %$	11.1	25.0	54.8	33.3	12.5
PMI (mean,sd)y	10.9(5.5)	14.2(7.7)	11.9(6.3)	12.4(1.9)	11.3(9.1)

Abbreviations:

ADAD: Autosomal Dominant Alzheimer's Disease

sAD: Sporadic Alzheimer's Disease

Presym : Presymptomatic

Other: (1:Dementia with Lewy bodies, , 4:Argyrophilic grain disease, 1:Encephalopathy, 1:Neurofibrillary tangle-predominant AD, 1:Cerebrovascular disease)

*: MS4A is referring to SNP rs1582763 (**AG**:28, **AA**:13 and **GG**:25)

$: the total number of APOE Ɛ4+ were 24 (**23**:4, **24**:2, **33**:39, **34**:19, **44**:3)

Table 2: Quality Control Steps

	Cell ranger	Sigmoid correction	Distrib UMI	Distrib Genes	% mit.	>10 cells*	Double Finder
Total #Nuclei	1,102,459	566,486	423,937	414,303	380,625	380,625	346,264
Mean #Genes	3,154	2,418	2,242	2,228	2,234	2,234	2,221
Mean #UMI	10,279	5,646	4,969	4,899	4,897	4,897	4,888
Avg %Mitochondria	2.35	2.34	1.98	1.94	1.22	1.22	1.21

Abbreviations:
#: number
mit: mitochondria
*: genes were removed if they were not expressed in at least 10 nuclei

Table 3: Cell Type Details

Cell-type	Subclusters	Nuclei	Avg DEG
Oligodendrocytes	9	164,437	3,329
Neurons (Ex/Inh)	4/3	38,164/15,411	4,713/3,881
Astrocytes	5	38,665	3,326
OPC	7	17,363	1,962
Microglia	9	16,941	1,172
Endothelial	--	3,126	--
Total		**294,114**	

Ex: Excitatory
Inh: Inhibitory

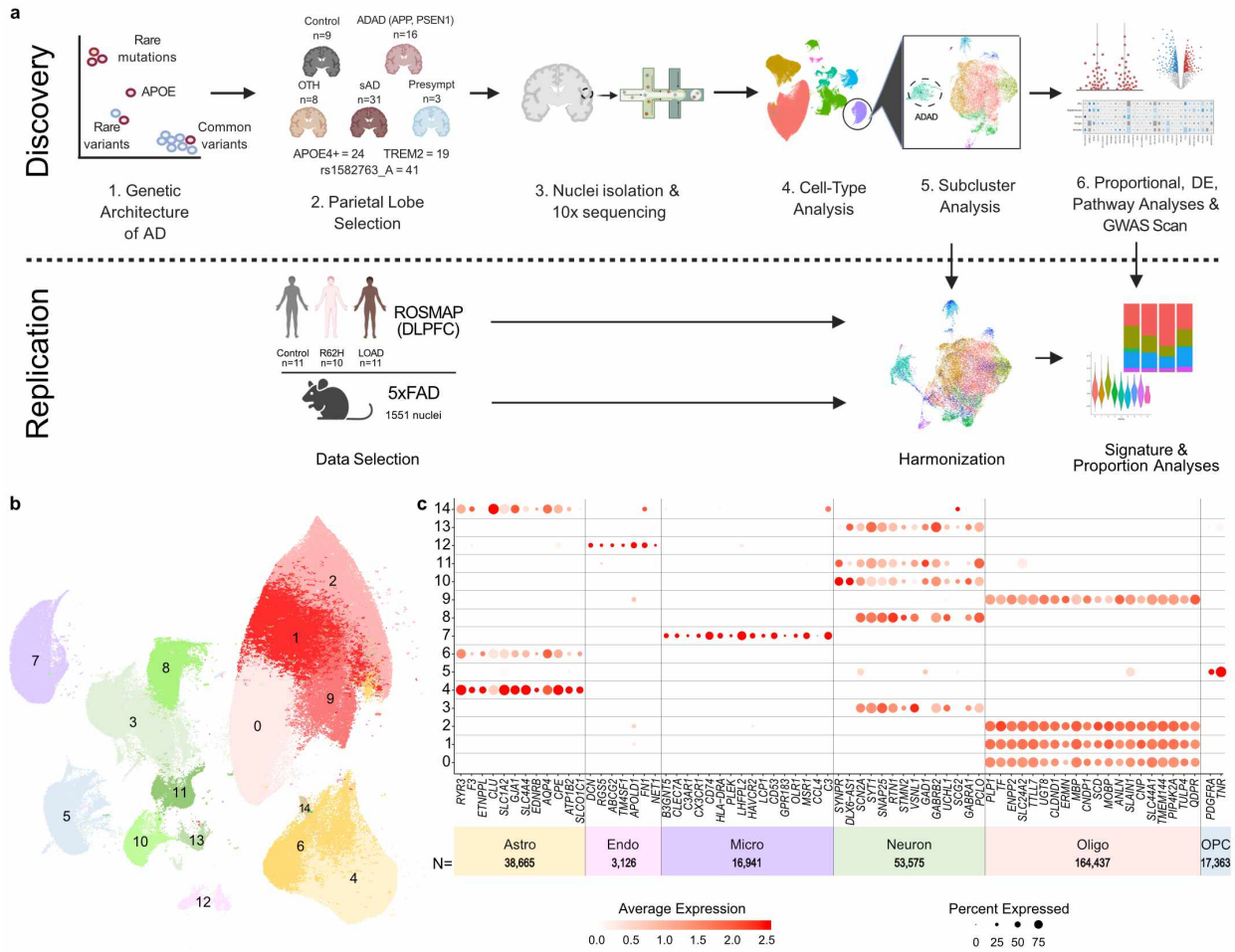


Figure 1. SnRNA-seq distinguishes major cell types using 67 Human brains.
a) Diagram of the study design. **b**) UMAP plot showing 15 distinguished clusters, 0–14, with 294,114 total cells. **c**) DotPlot depicting expression of cell type specific markers genes to identify each cluster in **b**.

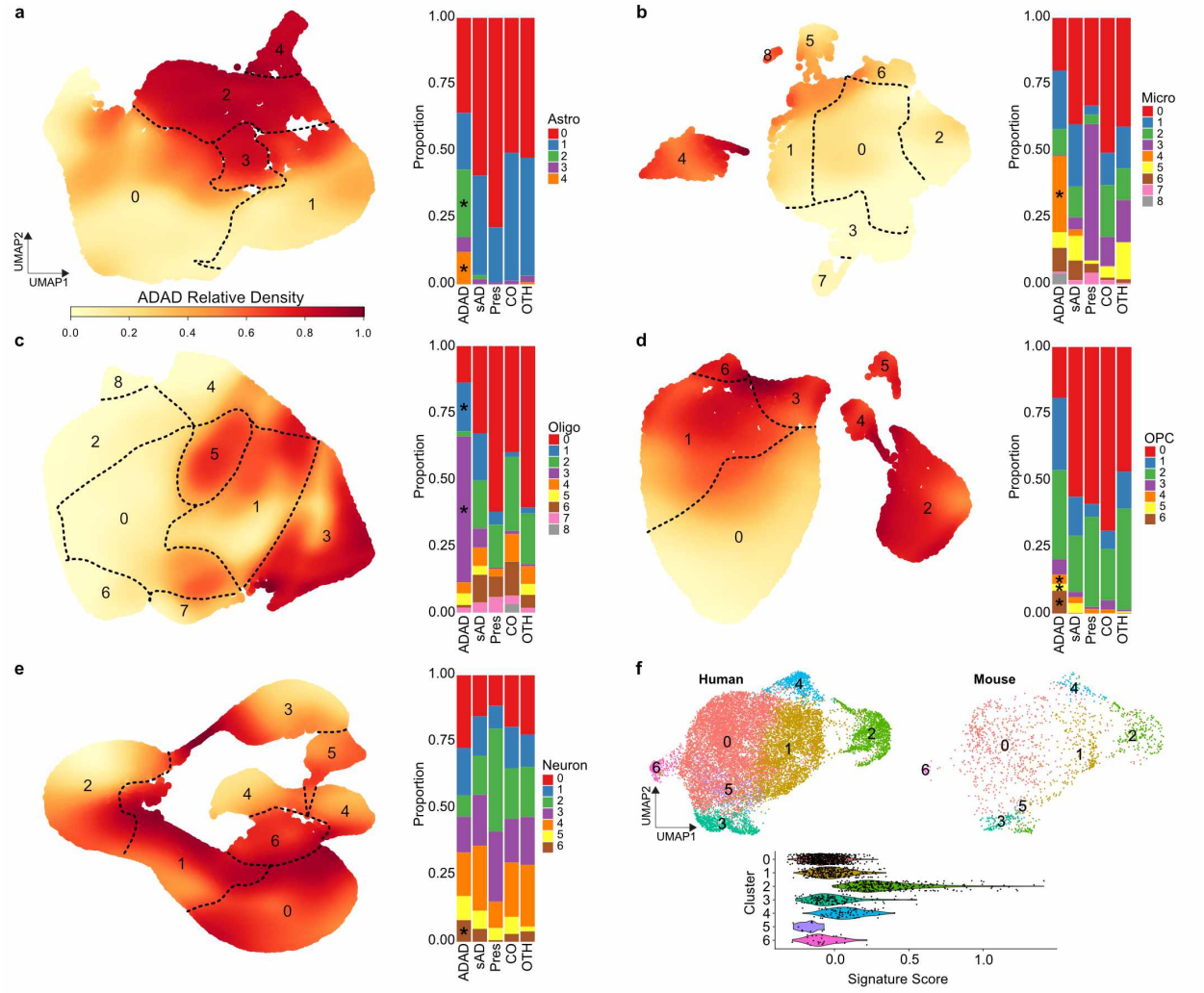


Figure 2. **ADAD participants exhibit distinct signatures in Astrocytes, Microglia, OPCs, Oligodendrocytes, and Neurons**.

a - e) Left panel, UMAP Enrichment plots show the normalized density of ADAD nuclei by cell type. 0 and 1 represent the lowest and highest measured normalized densities respectively. Right panel, a proportion plot shows the enrichment of certain cell states in ADAD participants. (*) represent a significant (P<0.05) enrichment. ADAD = autosomal dominant AD; sAD = sporadic AD; Pres = presymptomatic; CO = neuropath free; OTH = non-AD neurodegenerative. **a)** Astrocytes. **b)** Microglia. **c)** Oligodendrocytes. **d)** OPCs. **e)** Neurons. **f)** 5xFAD mouse validation of microglia ADAD cluster (cluster 4 in **b**, cluster 2 here). On the top, a UMAP of integrated microglia split by species. On the bottom, a violin plot showing that mouse cells in the ADAD cluster have a higher human microglia ADAD cluster signature score than cells in other clusters.

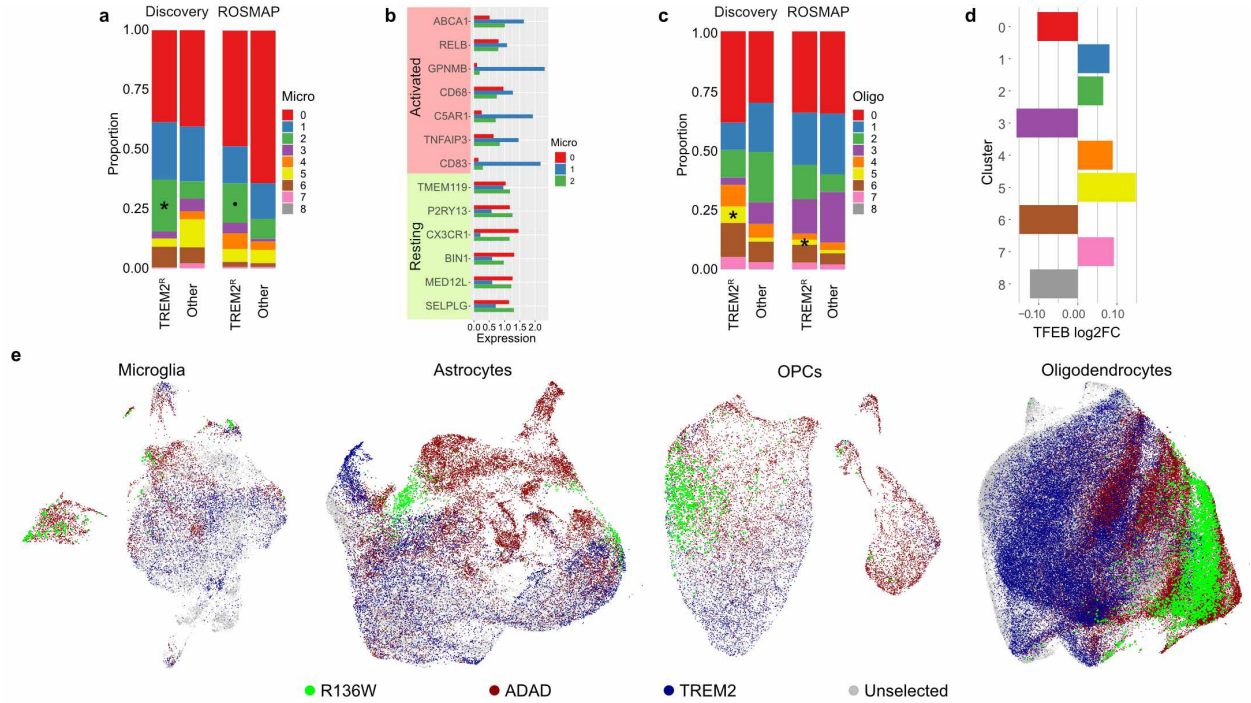


Figure 3. *TREM2* reduced-activation variant carriers (p.R47H, p.R62H, and p.H157Y) have distinct microglia and oligodendrocyte profiles.

a) Proportion plot shows a significant (P<0.05) enrichment (*) of mic.2 in *TREM2* reduced activation carriers ($TREM2^R$) and a trend (·) in the ROSMAP replication. 'Other' indicates all sAD non-*TREM2* reduced activation carriers including carriers of other *TREM2* variants. **b)** Bar plot shows the expression of both activated and resting microglia marker genes in mic.2 compared to the resting (mic.0) and activated cell states (mic.1). **c)** Proportion plot shows a significant (P<0.05) enrichment (*) of oligo.5 in $TREM2^R$ carriers in both the discovery and ROSMAP cohorts (P<0.05). **d)** Bar plot shows the log2 fold change of *TFEB* by oligodendrocyte subcluster. **e)** *TREM2* p.R136W carrier (green) clusters more closely with ADAD (red) than other *TREM2* carriers (blue) in Microglia, Astrocytes, OPCs, and Oligodendrocytes (from left to right).

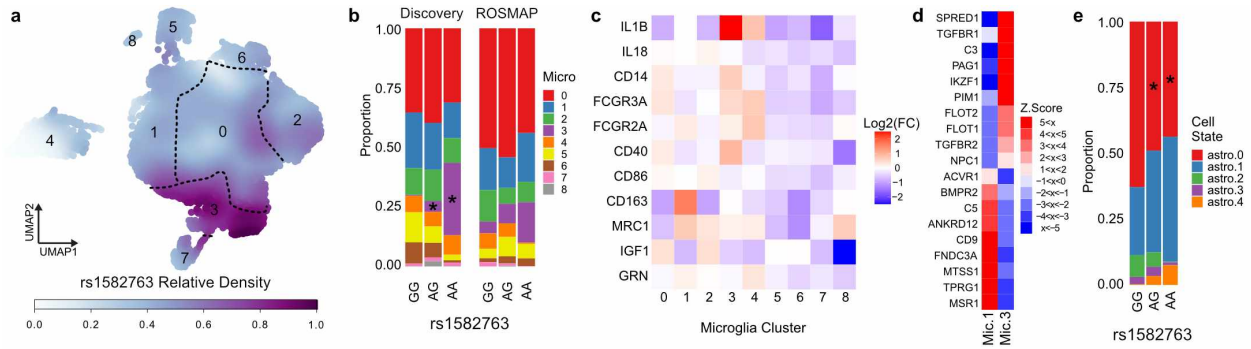


Figure 4. Unique microglia signatures for *MS4A* rs1582763 carriers.
a) UMAP Enrichment plots show the normalized density of rs1582763_A alleles in microglia. 0 and 1 represent the lowest and highest measured normalized densities respectively. **b)** A proportion plot shows a dose dependent enrichment of microglia expression state 3 in rs1582763 carriers. (*) represents a significant (P<0.05) enrichment and (·) represents a trend. **c)** A heatmap of proinflammatory (*IL1B – CD86*) and anti-inflammatory (*CD163 – GRN*) microglia marker gene log2 fold change (Log2(FC)) within the microglia subclusters. **d)** A heatmap of linear regression z scores shows the differences in gene expression between the 'main' (Mic.1) and 'MS4A' (Mic.3) activated microglia clusters. **e)** A proportion plot depicts a significant (P<0.05) depletion (*) of astro.0 (non-activated) in rs1582763 carriers and a trend for enrichment of astro.1 (activated).

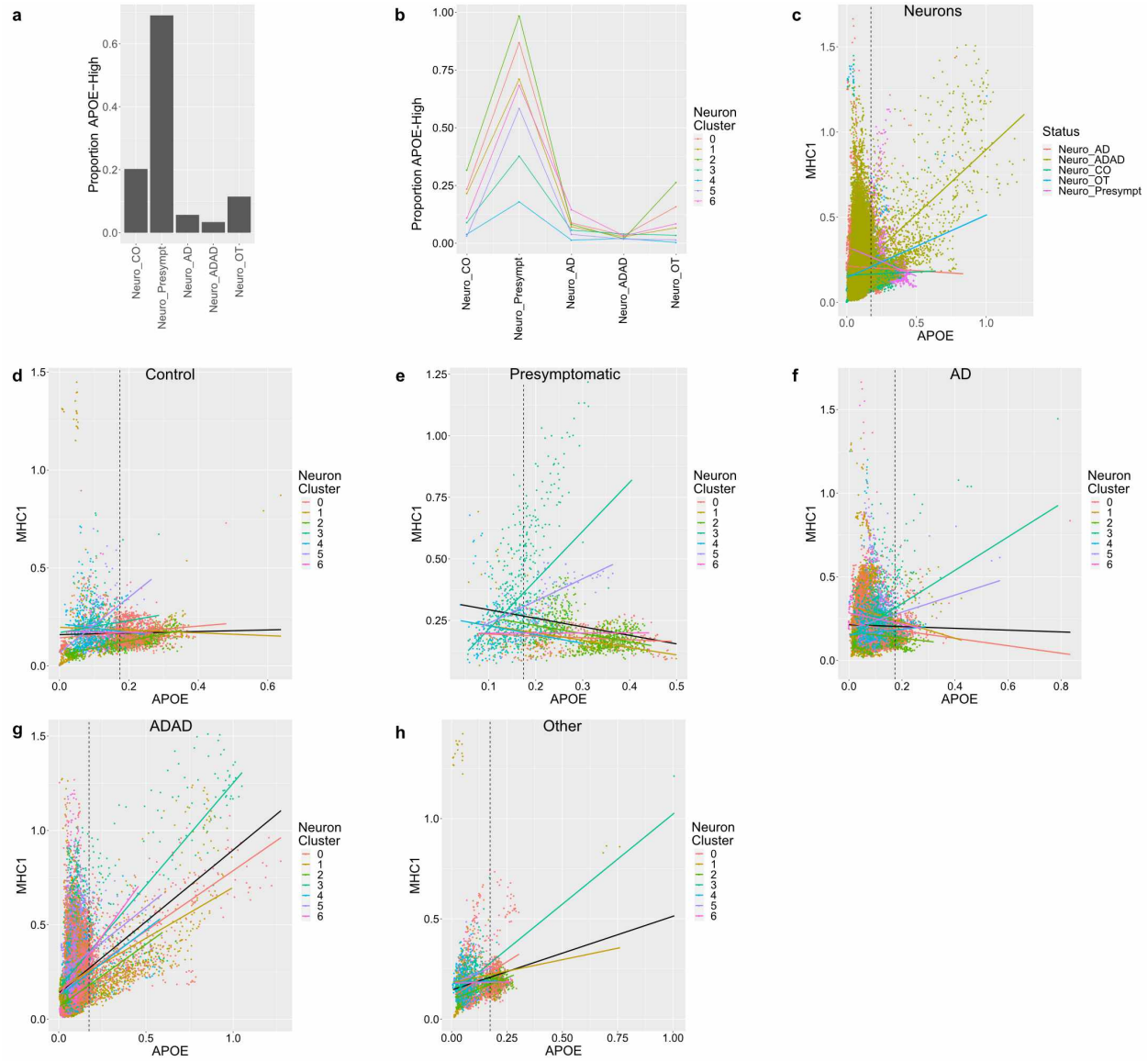


Figure 5. Neuronal APOE-high expression patterns.

a) Bar plot showing the average proportion of APOE-high neurons in each status. **b)** Average proportion of *APOE*-high neurons in each status split by subcluster. Clusters 0, 1, 2, 6 are excitatory and clusters 3, 4, 5 are inhibitory. **c-h)** Scatter plots and regression lines depicting the coexpression between *APOE* and *MHC-I* in neurons after MAGIC imputation. The vertical dotted line marks the *APOE*-high threshold. **d-h)** The black regression line depicts the overall correlation within the specified status.

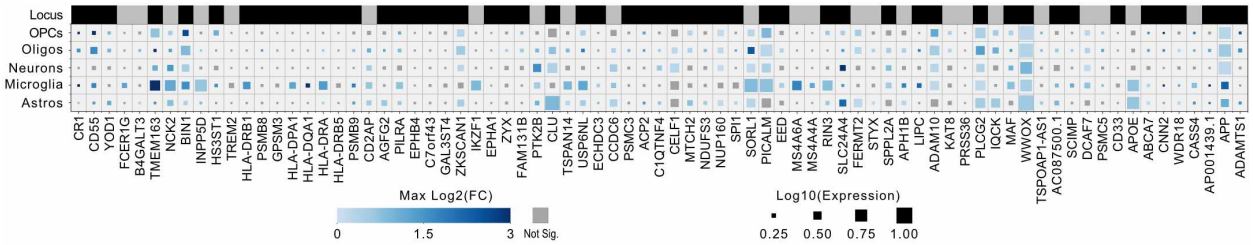
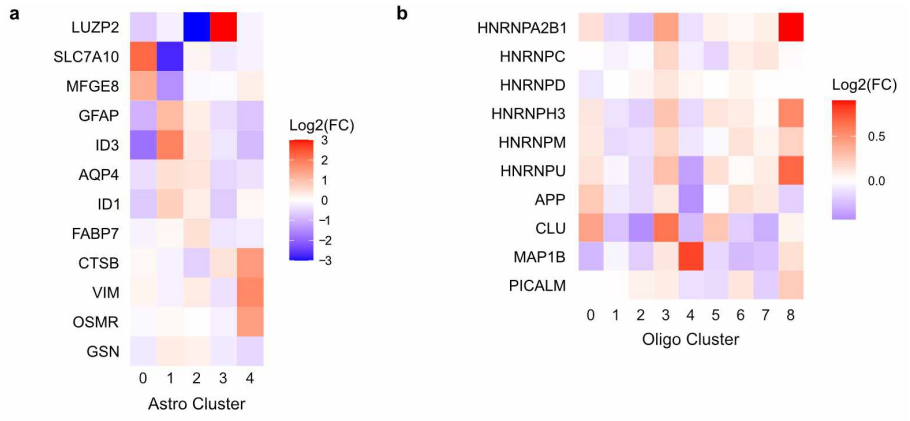


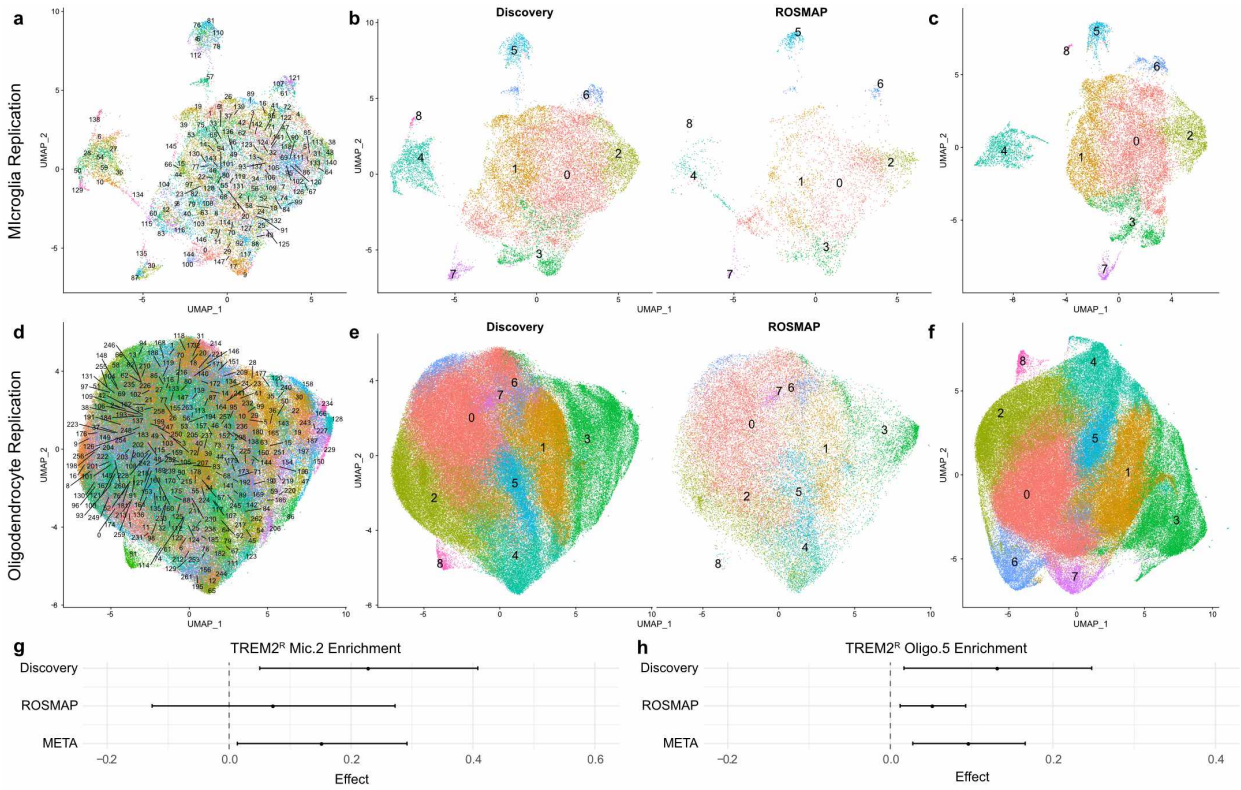
Figure 6. Cell type specific expression and cell state differential expression of prioritized genes help link GWAS hits to cell types.

A modified heatmap shows the average expression and max log2 fold change within cell types of prioritized genes in GWAS loci. At the top, the 'Locus' row indicates genes within the same locus using alternating black and gray rectangles. In the body of the plot, the color of the squares represents the max log2 fold change of the gene between cell states (subclusters) of that cell type. The log2 fold changes greater than three were set to three to preserve visual variability. Gray squares indicate no significant log2 fold change within that cell type. Square size represents log10 transformed gene expression. 77 of the 89 genes prioritized by Kunkle *et al*, Andrews *et al*. and Schwartzentruber *et al*. passed QC in our data.



Extended Figure 1. Gene expression patterns in astrocyte and oligodendrocytes clusters enriched in ADAD brains.

a) A heatmap shows the log2 fold change (Log2FC) of astrocyte marker genes taken from Habib *et al* within the astrocyte subclusters. *LUZP2, SLC7A10,* and *MFGE8* mark GFAP-low astrocytes (resting); *GFAP, ID3, AQP4, ID1,* and *FABP7* mark GFAP-high astrocytes (activated); *CTSB, VIM, OSMR,* and *GSN* mark disease associated astrocytes (DAA). **b)** A heatmap shows the Log2FC of HNRNP genes and AD risk genes controlled by the HNRNP gene family within the oligodendrocyte subclusters.



Extended Figure 2. *TREM2* associated clusters replicated in ROSMAP cohort.
a,d) UMAPs of discovery cohort (Knight ADRC and DIAN) and ROSMAP microglia (a) and oligodendrocytes (d) post integration. A resolution of 15 was used to identify a large number of clusters to facilitate the mapping of the discovery cohort cluster identities onto the ROSMAP nuclei. **b,e)** UMAPs of microglia (b) and oligodendrocytes (e) post cluster identity mapping, split by cohort. **c,f)** UMAPs of discovery cohort microglia (c) and oligodendrocytes (f) prior to ROSMAP integration. **g,h)** Tree plots of the Mic.2 (g) and Oligo.5 enrichment within TREM2 reduced activation variant carriers ($TREM2^R$). The linear regression effect values are shown for the independent discovery and ROSMAP cohorts as well as the merged meta-analysis results.