

## GENERAL ARTICLE

# Integrative genomics approach identifies conserved transcriptomic networks in Alzheimer's disease

Samuel Morabito<sup>1,3</sup>, Emily Miyoshi<sup>2,3</sup>, Neethu Michael<sup>2,3</sup> and Vivek Swarup<sup>2,3,\*</sup>,†

<sup>1</sup>Mathematical, Computational and Systems Biology (MCSB) Program, University of California, Irvine, CA 92697, USA, <sup>2</sup>Department of Neurobiology and Behavior, University of California, Irvine, CA 92697, USA and <sup>3</sup>Institute for Memory Impairments and Neurological Disorders (MIND), University of California, Irvine, CA 92697, USA

\*To whom correspondence should be addressed at: 3224 Biological Sciences III, Irvine, CA 92697, USA. Tel: (949)-824-3182; Email: vswarup@uci.edu

## Abstract

Alzheimer's disease (AD) is a devastating neurological disorder characterized by changes in cell-type proportions and consequently marked alterations of the transcriptome. Here we use a data-driven systems biology meta-analytical approach across three human AD cohorts, encompassing six cortical brain regions, and integrate with multi-scale datasets comprising of DNA methylation, histone acetylation, transcriptome- and genome-wide association studies and quantitative trait loci to further characterize the genetic architecture of AD. We perform co-expression network analysis across more than 1200 human brain samples, identifying robust AD-associated dysregulation of the transcriptome, unaltered in normal human aging. We assess the cell-type specificity of AD gene co-expression changes and estimate cell-type proportion changes in human AD by integrating co-expression modules with single-cell transcriptome data generated from 27 321 nuclei from human postmortem prefrontal cortical tissue. We also show that genetic variants of AD are enriched in a microglial AD-associated module and identify key transcription factors regulating co-expressed modules. Additionally, we validate our results in multiple published human AD gene expression datasets, which can be easily accessed using our online resource (<https://swaruplab.bio.uci.edu/consensusAD>).

## Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disease, which is symptomatically characterized by impairment in cognitive and executive functions, including memory loss (1,2). AD is pathologically characterized by the presence of neurofibrillary tangles and senile plaques, which are composed of hyperphosphorylated microtubule associated protein tau (MAPT; tau) inclusions and amyloid beta (A $\beta$ ) deposits, respectively (2,3). While aging is the major non-genetic risk factor for AD,

several genetic risk factors, including mutations in MAPT, have been found and associated with the disease (4–7). Despite the identification of causal genetic mutations and disease risk genes implicated by genome-wide association studies (GWAS) of AD (6,7), the molecular mechanisms of neurodegeneration are highly complex and are still poorly understood, impeding the design of effective therapeutic intervention.

Transcriptomics coupled with gene network analysis are powerful tools for investigating quantitative molecular

†Vivek Swarup, <http://orcid.org/0000-0003-3762-2746>

Received: May 6, 2020. Revised: July 10, 2020. Accepted: July 27, 2020

© The Author(s) 2020. Published by Oxford University Press. All rights reserved. For Permissions, please email: [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

phenotypes and pathways underlying disease progression in a genome-wide manner. These approaches have the potential to unravel previously unidentified disease-associated pathways, as well as highlighting disease regulators, which themselves serve as candidate targets for drug development. Within this context, the National Institute of Aging has developed a Target Identification and Preclinical Validation Project of the Accelerating Medicines Project–Alzheimer’s Disease (AMP-AD) consortium (8), whose goal is to integrate high-throughput genomic and molecular data from the diseased brain within a network driven structure (8,9). Under this program, several AD-specific large-scale RNA sequencing (RNA-seq) projects have been conducted. These projects have identified transcriptomic networks and splicing events that are altered in the AD cortex (9–15). However, these studies were limited to specific brain banks, and a holistic meta-analysis across hundreds of samples with integrative genomic approaches is pertinent to robustly identify disease-specific signatures in an unbiased manner.

Here we have taken a functional genomics and integrative systems biology approach to identify human AD-specific transcriptomic alterations conserved across multiple cohorts and multiple cortical regions. We applied consensus weighted gene co-expression network analysis (cWGCNA), a meta-analytical approach, to more than 1200 postmortem brain RNA-seq samples across brain banks and identified robust, disease relevant co-expression modules. With this approach, we were not only able to identify strong, disease-specific signatures across multiple cohorts but also were able to define brain-region specific alterations that are fundamental to disease pathophysiology. We found several neuronal and glial specific modules that are altered in AD but not in normal human aging, and one of the glial modules was enriched in AD candidate risk genes from AD GWAS datasets. We also extensively examined module enrichment of GWAS SNPs associated with other neurological disorders. We used generic and brain-specific ATAC-seq and ChIP-seq datasets to define transcription factors (TFs) regulating the neuronal and glial modules, identifying members of the NF- $\kappa$ B pathway as regulators of AD-associated co-expression changes. In addition, we have leveraged various orthogonal datasets, including DNA methylation, histone acetylation (H3K9ac) and expression and methylation quantitative loci (eQTLs and mQTLs), to provide a comprehensive picture of the genetic alterations associated with AD. While most co-expression modules were enriched in AD-associated epigenetic changes, one glial module was notably and consistently not enriched in these epigenetic marks.

Massive changes in cell composition are concurrent with the progression of AD, impairing accurate estimation of transcriptomic changes from bulk tissue data. To circumvent this problem, we generated single-nuclei transcriptomics data from aged, cognitively healthy postmortem human brain samples and used their transcriptomic signatures to accurately estimate cell-type proportion changes from bulk RNA-seq samples and to classify AD-specific co-expression modules by cell type. Finally, we confirm the robustness of our findings by computing the preservation of co-expression modules in multiple orthogonal published gene expression datasets of AD encompassing several brain regions. Such a comprehensive study across multi-scale datasets has not been performed in neurodegeneration and provides a framework for future meta-analytic integration studies. Together, our findings define a core set of transcriptomic changes altered during the progression of the disease and highlight several regulators of these changes, which may be used as therapeutic targets.

## Results

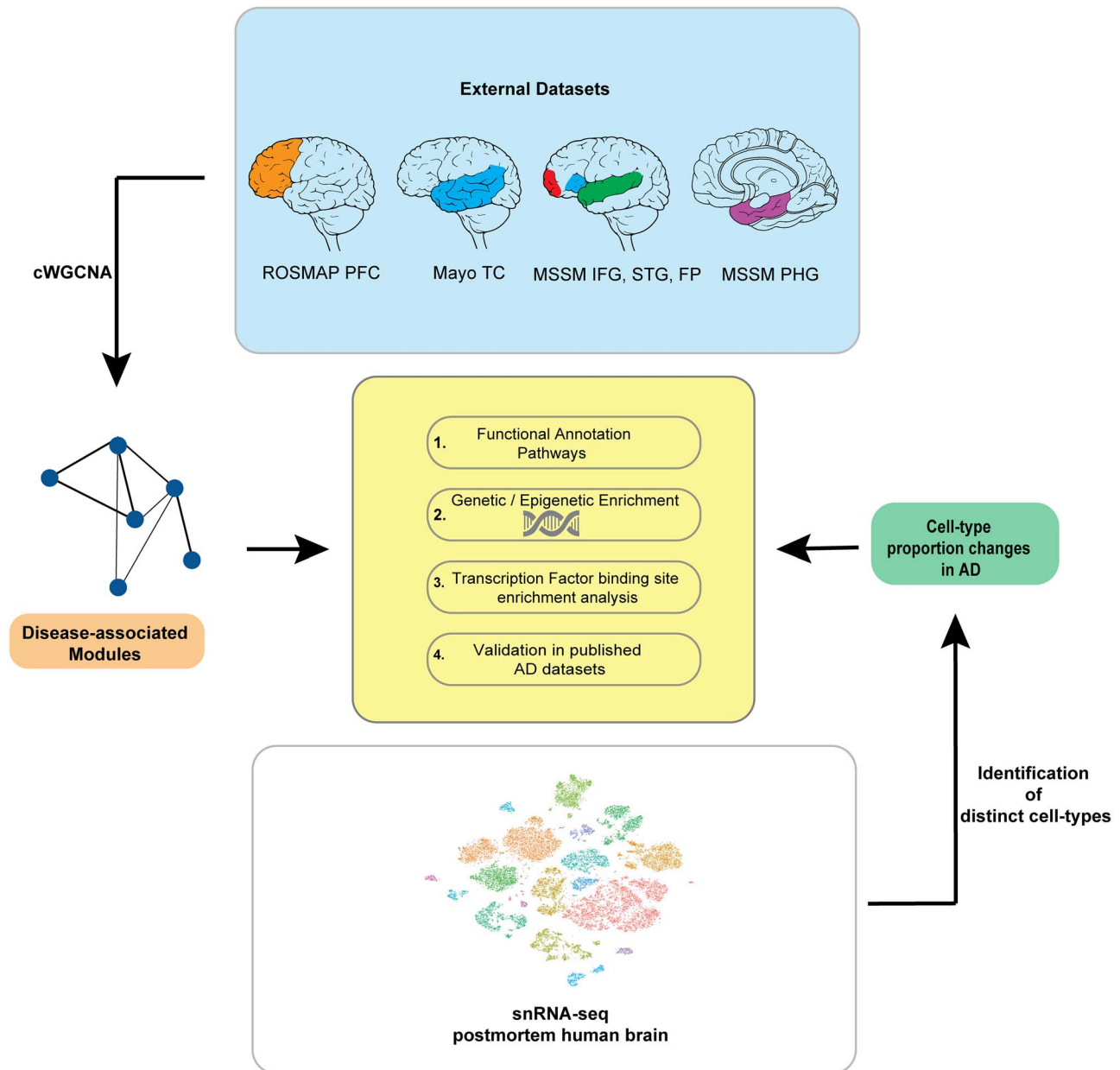
In order to identify robust, key transcriptional changes in AD, we analyzed RNA-seq data from three different studies—Mayo Clinic Brain Bank (Mayo) (10), the Religious Orders Study and Memory and Aging Project (ROSMAP) (14) and Mount Sinai School of Medicine (MSSM) (16). The Mayo dataset contained temporal cortex samples ( $n=160$ ) (10), while the ROSMAP dataset was composed of prefrontal cortex samples ( $n=632$ ) (14). The MSSM cohort included samples from four different regions—parahippocampal gyrus, inferior frontal gyrus, superior temporal gyrus and the frontal pole ( $n=476$ , Fig. 1) (16). Control samples were classified as Braak stage 0–I. Early-stage pathology samples were defined as Braak stage II–IV and CERAD score of possible AD, while late-stage pathology samples were Braak stage V–VI and CERAD score of probable and definite AD (see [Supplementary Material, Table S1](#) for all case information). All datasets were uniformly processed and regressed for cell-type proportion changes and other technical and biological covariates (see [Materials and Methods](#)).

cWGCNA (17,18) is a widely used analysis method to discover biologically relevant gene modules across multiple gene expression datasets. We performed cWGCNA on the three bulk tissue RNA-seq datasets—Mayo, ROSMAP and MSSM—to examine disease-related changes conserved across more than a thousand individuals from different cortical regions and different brain tissue repositories (Fig. 2A). We identified 10 consensus modules significantly correlated with AD diagnosis—four negatively (CM7, CM1, CM10 and CM12) and six positively correlated (CM16, CM5, CM4, CM8, CM9 and CM23;  $FDR < 0.05$ ; Fig. 2B, [Supplementary Material, Table S2](#)). To better understand these modules, we examined the relationships between co-expression modules using a bi-weighted mid-correlation (Fig. 2C). We found modules CM9 and CM23 were strongly positively correlated, but CM23 was strongly anti-correlated with CM1, providing us an idea of how these modules function together. In addition, cWGCNA modules were found to be concordant with gene modules constructed using several other co-expression and gene clustering algorithms (19–22) (see [Materials and Methods, Supplementary Material, Fig. S1](#)).

### Identification of cell-type clusters in the aged human brain

Due to the cell-type composition changes during the progression of the disease, profiling gene expression in only bulk tissue samples may obscure biologically relevant cell-type specific changes. While single-cell RNA-seq allows us to evaluate transcriptional changes within cell-types, it is prohibitively costly to execute on large cohorts (i.e. hundreds of individuals). To circumvent this issue, we developed a framework that leverages single-cell RNA-seq data from a smaller cohort in conjunction with co-expression network analysis in order to estimate cell-type specific transcriptomic changes in large, bulk tissue RNA-seq datasets.

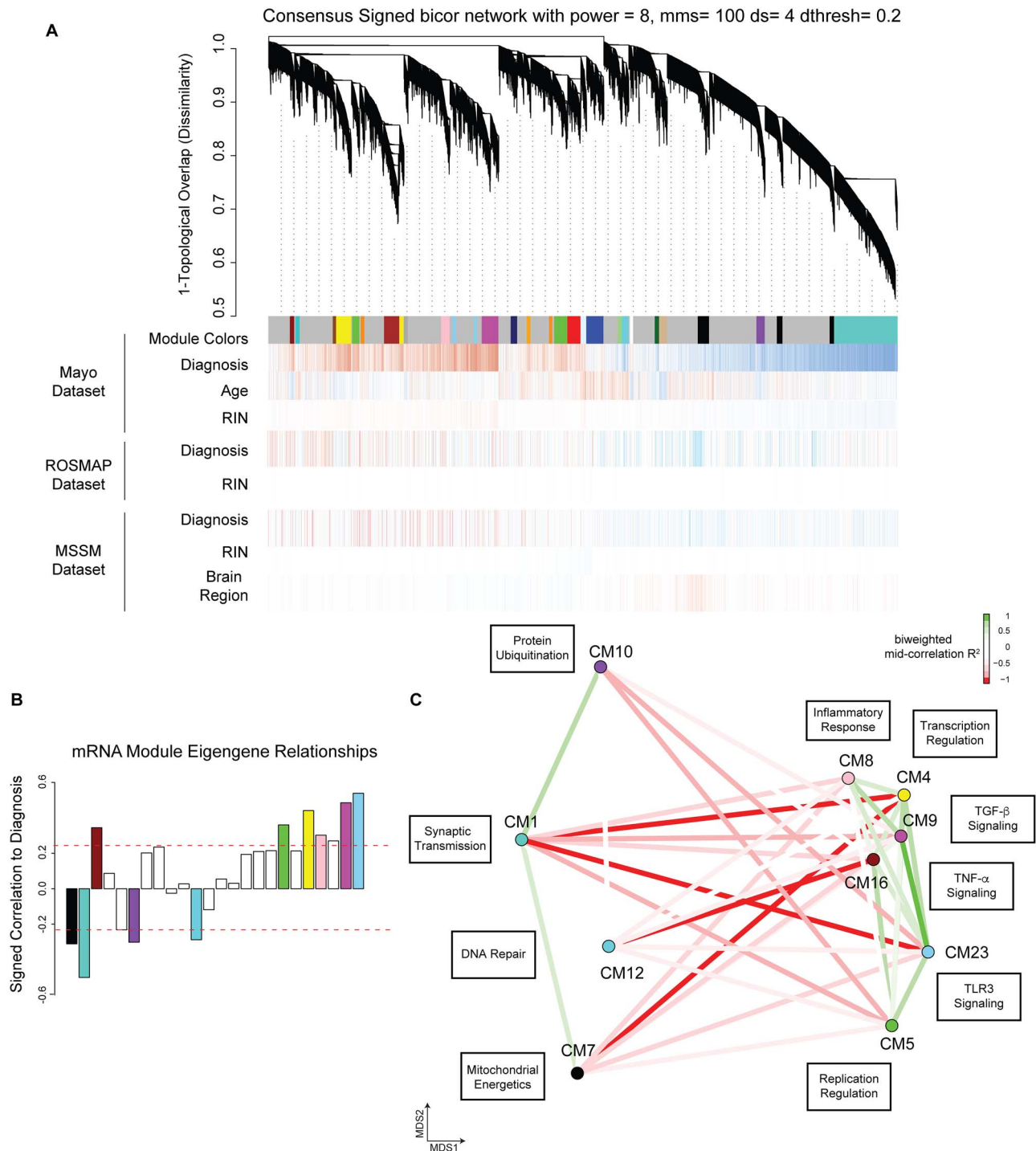
We isolated nuclei and performed single-nuclei RNA-seq (snRNA-seq,  $n=27\,321$  nuclei) on postmortem human brain tissue from aged, neurologically healthy controls ( $n=5, 67$  to 90+ years old, PFC, [Supplementary Material, Table S1](#)) to clarify cell-type proportions and the corresponding transcriptional profiles of the aged, adult human cortex. Through unsupervised Louvain community detection, we identified 18 transcriptionally distinct cell populations that we visualized using t-distributed stochastic neighbor embedding (t-SNE; Fig. 3A, [Supplementary Material,](#)



**Figure 1.** Schematic representation of the various analyses performed in the study. Consensus weighted gene co-expression network analysis (cWGCNA) with datasets from three different studies—Mayo Clinic Brain Bank (Mayo) temporal cortex (TC); Religious Orders Study and Memory and Aging Project (ROSMAP) prefrontal cortex (PFC); Mount Sinai School of Medicine (MSSM) para-hippocampal gyrus (PHG), inferior frontal gyrus (IFG), superior temporal gyrus (STG) and frontal pole (FP)—identified Alzheimer's disease (AD)-associated modules. Distinct cell-types were detected using single-nuclei RNA-sequencing (snRNA-seq) from postmortem human brain tissue revealing cell-type proportion changes in AD. Disease-associated modules and cell-type clusters were used for functional pathway analysis in addition to GWAS enrichment, transcription factor binding site enrichment and validation in published datasets.

Table S3). Canonical cell-type marker genes, such as *SLC17A7* for neurons, were examined in order to assign a cell type to each of the 18 clusters (Fig. 3B, Supplementary Material, Fig. S2A). The majority of the profiled nuclei was classified as excitatory neurons (Fig. 3C). This strategy identified both neuronal and non-neuronal subpopulations: seven excitatory neuron, four inhibitory neuron, two astrocyte and two oligodendrocyte subpopulations. We additionally found clusters consistent with endothelial cells, microglia and oligodendrocyte precursor cells (OPCs), and there was one cluster we were unable to confirm its cell-type origin.

Moreover, differential gene expression analysis identified gene markers specific to each cell-type cluster (Fig. 3D, Supplementary Material, Fig. S2B). These included previously established cell-type markers, such as *PVALB*, *NEUROD6* and *CX3CR1*, and layer-specific markers like *SV2C* and *HTR2C* (23,24). We also utilized the Allen Human Brain Atlas to verify the expression of these marker genes in the human cortex (Supplementary Material, Fig. S2C). Since we found multiple neuronal subpopulations, we aimed to characterize them further, distinguishing three selectively and highly co-expressed genes in each neuronal cluster (Fig. 3E, Supplementary Material, Fig. S2D). Based on

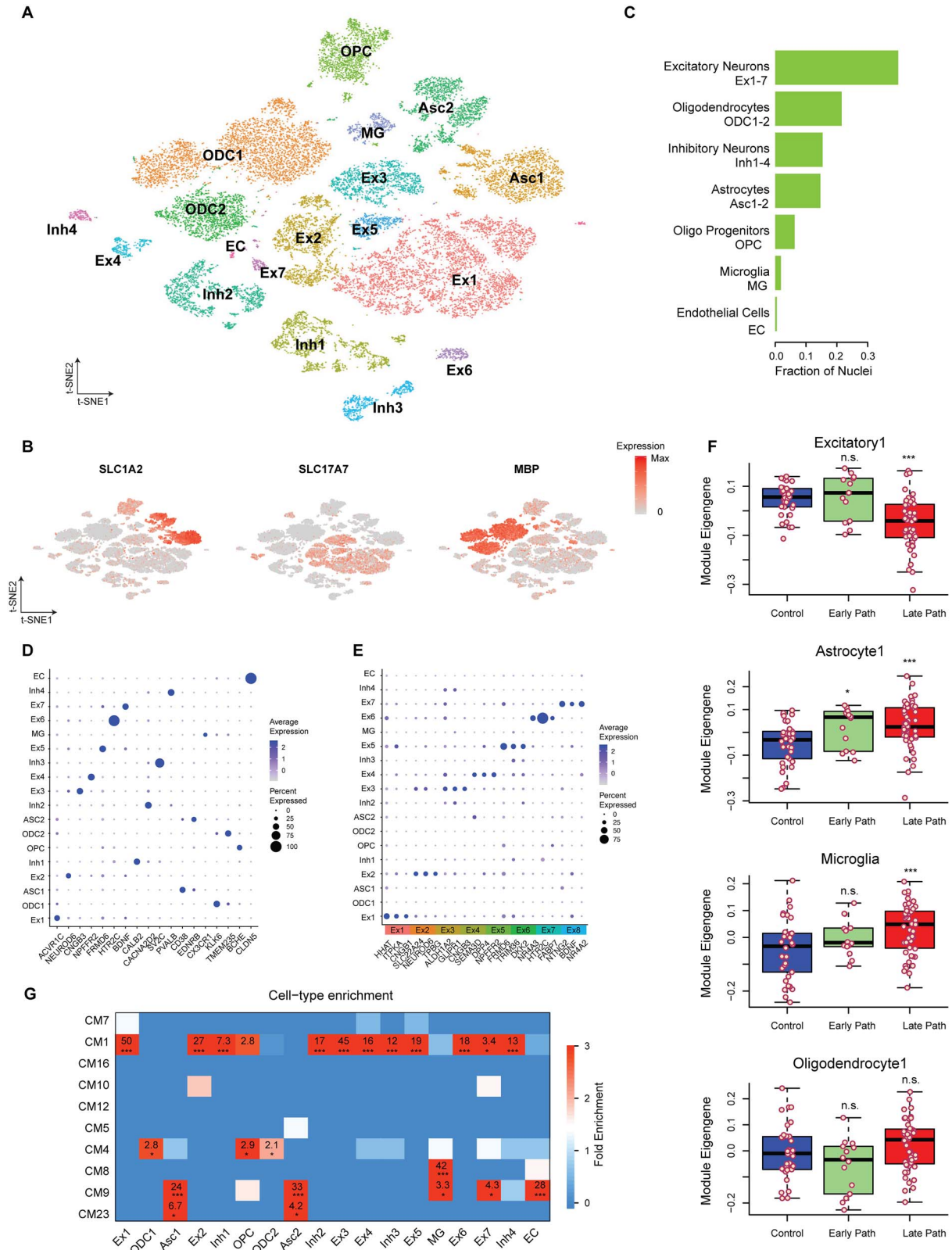


**Figure 2.** Consensus transcriptomic analysis identifies disease-relevant modules. (A) Consensus WGCNA dendrogram showing consensus modules from Mayo, ROSMAP and MSSM datasets. Red indicates positive correlation of covariate (diagnosis, age, RIN) with gene expression, and blue is negative correlation. (B) Signed correlation of mRNA module eigengenes with AD diagnosis. (C) Multidimensional scaling plot demonstrates relationships between consensus modules significantly correlated with diagnosis. Color signifies biweighted mid-correlation ( $R^2$ ) value. Modules are also annotated with summarizing terms.

these genes we were able to identify our excitatory neuron clusters 7 and 8 as deep layer neurons, while our inhibitory neuron cluster 3 likely originated from layer 3 neurons. In addition, we found inhibitory neuron subpopulations similar to Darmanis et al. (25)—our inhibitory neuron clusters 1 and 4,

CALB2-VIP and CRH-PVALB neurons, respectively. We also examined genes differentially expressed across the cell-type clusters in respect to sex, identifying X chromosome inactivation gene XIST upregulated in astrocytes from females (Supplementary Material, Fig. S3, Supplementary Material, Table S4). Additionally,





**Figure 3.** Single nuclei RNA-seq on the aged human brain elucidates bulk tissue RNA-seq analyses. (A) Cell-type clusters visualized with t-SNE plot. Cell types were labeled post hoc based on marker gene expression. (B) Marker gene expression in cell-type clusters visualized with same t-SNE plot as in (A). (C) Bar graph indicating fraction of nuclei identified as each cell type. (D–E) Dot plots showing differentially expressed cell-type markers (D) and excitatory neuronal markers (E) in cell-type clusters. Color denotes average gene expression, while size denotes percent of nuclei expressing the gene. (F) Boxplots depicting estimated cell-type abundance with pathological state for four different cell-type clusters—excitatory neuron cluster 1, astrocyte cluster 1, microglia cluster and oligodendrocyte cluster 1 (top to bottom)—in the Mayo TC dataset. (G) Enrichment of consensus modules in cell-type clusters identified by snRNA-seq.

NRXN1 was upregulated in several clusters from males and has been previously associated with male-specific behavioral changes (26,27).

### Clarifying AD bulk tissue RNA-seq changes with snRNA-seq

We reasoned that our cell-type clusters from snRNA-seq in the aged human brain could be leveraged to estimate AD-related changes in cell-type proportions across different brain regions and by disease stage. We devised a statistical framework to estimate the cell-type composition in bulk tissue RNA-seq data, not regressed for cell-type proportion changes, using our cell-type clusters (see [Materials and Methods](#)). Using the specificity score obtained from single-nuclei data, we independently estimated cell-type composition in each of the bulk RNA-seq datasets used. In the Mayo TC Set, we found that all excitatory and inhibitory neuronal subpopulations significantly decreased with late stage pathology ( $P < 0.001$ ), as expected; however, there was little change with early stage pathology ( $P > 0.05$ ; [Fig. 3F](#), [Supplementary Material, Fig. S4A](#)). On the other hand, microglia and astrocytes increased with late stage pathology ( $P < 0.001$ ). Microglia only tended to increase at early stage pathology ( $P > 0.05$ ), while both astrocyte subpopulations significantly increased ( $P < 0.05$ ). We also saw that at late stage pathology, OPCs greatly decreased ( $P < 0.001$ ), but both oligodendrocyte subpopulations tended to increase ( $P > 0.05$ ), suggesting the differentiation of OPCs into oligodendrocytes with increasing pathology. In addition, we used a published deconvolution algorithm, CIBERSORT (28), and found that our results were comparable to CIBERSORT (Storey's  $\pi_1 = 0.93$ ).

Notably, AD progresses in an anatomical manner (1,29), and gene expression changes may be causal or an effect of these region-specific cellular changes. While we found similar patterns in the other datasets, we noticed that RNA-seq datasets from temporal cortical regions demonstrated more pronounced changes in cell-type proportions than frontal cortical regions ([Supplementary Material, Fig. S4B](#), [Supplementary Material, Figs S5–S6](#)). This regional difference resembles the progression of pathology in which the temporal cortex is affected before the frontal cortex in AD (1,29).

Accordingly, we also utilized the cell-type clusters from snRNA-seq in order to further characterize the consensus modules we identified through cWGCNA. We classified the consensus modules as neuronal or non-neuronal based on enrichment of cell-type markers and additionally confirmed our classification by integrating our snRNA-seq results ([Fig. 3G](#)). For example, upregulated consensus module CM8 was selectively enriched in the microglia cell-type cluster, while CM23 was selectively enriched in both astrocyte clusters.

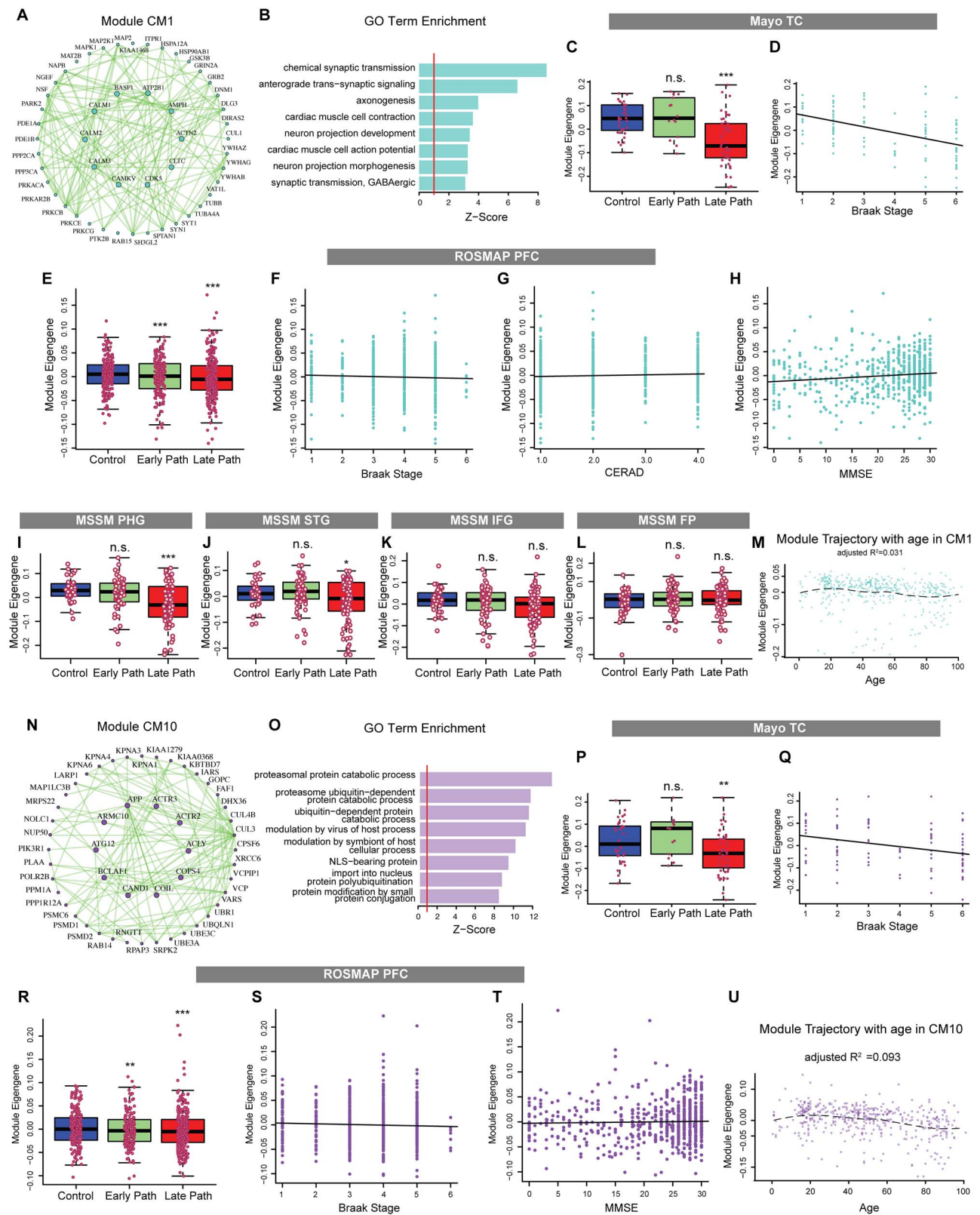
### Neuron-specific co-expression changes in AD

We next wanted to examine the AD-associated modules more closely and used experimentally derived databases of human protein–protein interactions (PPIs) from Inweb (30) and Biogrid (31) to guide our attention to important genes. We created integrated co-expression-PPI networks based on the expectation that a subset of the edges defined by co-expression are due to PPI (32–35); such interactions functionally annotate the networks' edges and provide orthogonal module validation with independent data (35). While gene expression and protein levels do not always correlate, we used a recently published

dataset of proteomics in human AD ( $n > 2000$  samples) to further investigate each cWGCNA module (36). We projected cWGCNA modules into a human AD proteomics dataset and observed agreement between module eigenprotein and module eigengene trajectories throughout AD progression (see [Materials and Methods](#), [Supplementary Material, Fig. S7](#)), demonstrating that AD-associated transcriptomic cWGCNA modules show similar expression patterns at the protein level. In these integrated networks, the edges between genes (nodes) represent both gene co-expression and PPI ([Fig. 4A, N](#)), permitting us to focus on hub genes observed at both the RNA and protein level.

Consistent with the neuronal dysfunction and neuronal loss seen in AD, we determined the downregulated consensus module CM1 was highly enriched in neuronal markers and our neuronal cell-type clusters ([Fig. 3G](#)). CM1's hub genes (CALM1, CALM2, CALM3, CAMKV, CLTC and ATP2B1) are indicative of calcium signaling dysregulation, providing evidence for the calcium hypothesis of AD (37) ([Fig. 4A](#), [Supplementary Material, Table S2](#)). GO analysis of CM1 demonstrated enrichment of terms like synaptic transmission and axonogenesis, further supporting its neuronal specificity ([Fig. 4B](#)). One of CM1's members was also PTK2B, which has been previously identified as an AD risk gene (7). In the Mayo TC dataset we found that CM1's module eigengene significantly decreased with late stage pathology ( $P < 0.001$ ) but was unchanged with early stage pathology ( $P > 0.05$ ; [Fig. 4C](#)), and the module's trajectory declined with increasing Braak stage ( $R = -0.47$ ,  $P = 1.4e-06$ ; [Fig. 4D](#)). Additionally, we examined CM1's trajectory in the ROSMAP PFC dataset and found a similar but weaker decline with increasing pathology ([Fig. 4E–G](#)). We also saw that CM1's module eigengene increased with MMSE score ( $R = 0.13$ ,  $P = 1.1e-03$ ; [Fig. 4H](#)). However, similar to the regional differences we found with cell-type proportions, we noticed that the decrease in CM1 module eigengene was greater in the MSSM datasets from temporal cortical regions (parahippocampal gyrus and superior temporal gyrus) compared to those from frontal cortical regions (inferior frontal gyrus and frontal plate; [Fig. 4I–L](#)). Further, we wanted to confirm that CM1 was associated with disease state rather than normal human aging, so we calculated a synthetic module eigengene in a North American Brain Expression Consortium (NABEC) dataset, comprising brain gene expression data from hundreds of normal human patients between the age of 16 and 100 years old (38). We found that CM1 was not correlated with age (adjusted  $R^2 = 0.031$ , [Fig. 4M](#)), supporting that the co-expression changes are specific to AD.

We then performed similar analyses on another downregulated consensus module, CM10, and we likewise identified this module as neuronal. One of CM10's hub genes is APP, classically studied in AD because it transcribes the precursor of A $\beta$  ([Fig. 4N](#)). Notably, CM10 is enriched in GO terms related to proteasomal degradation and ubiquitination ([Fig. 4O](#)), and CM10's hub gene CAND1 regulates neddylation (39). Neddylation has been found dysregulated in AD and may impact ubiquitination (39). Like CM1, we found the module eigengene decreased with late stage pathology ( $P < 0.01$ ), but we also surprisingly discovered CM10's module eigengene tended to increase with early stage pathology ( $P > 0.05$ ; [Fig. 4P](#)). This may suggest that in the early stages of disease, neurons attempt to degrade pathological protein aggregates. Again, we saw that the module's declining trajectory was more pronounced in the temporal cortex in comparison to the frontal cortex ([Fig. 4Q–T](#), [Supplementary Material, Fig. S8](#)) and CM10 was disease specific, rather than an effect of aging (adjusted  $R^2 = 0.093$ , [Fig. 4U](#)).



**Figure 4.** Consensus transcriptomics analysis identifies neuron-specific changes in disease. (A–B) Co-expression plot (A) and gene ontology term enrichment (B) for neuronal consensus module CM1. (C–D) Module eigengene trajectory with pathological state (C) and with Braak stage (D) for CM1 in Mayo TC dataset. (E–H) Module eigengene trajectory with pathological state (E), Braak stage (F), CERAD score (G) and MMSE score (H) for CM1 in ROSMAP PFC dataset. (I–L) Module eigengene trajectory with pathological state for CM1 in MSSM PHG (I), MSSM STG (J), MSSM IFG (K) and MSSM FP (L). (M) Module eigengene trajectory with age for CM1 in NABEC dataset. (N–O) Co-expression plot (N) and gene ontology term enrichment (O) for neuronal consensus module CM10. (P–Q) Module eigengene trajectory with pathological state (P) and with Braak stage (Q) for CM10 in Mayo TC dataset. (R–T) Module eigengene trajectory with pathological state (R), Braak stage (S) and MMSE score (T) for CM10 in ROSMAP PFC dataset. (U) Module eigengene trajectory with age for CM10 in NABEC dataset.



In addition, we examined consensus module CM7, which we also found neuronal and downregulated in AD. CM7's hub genes include ATP synthase subunits (ATP5G1, ATP5D, ATP5L, etc.), consistent with CM7's enrichment of GO terms related to mitochondrial function and ATP synthesis (Fig. 5A–B). While mitochondrial dysfunction and oxidative stress have been studied in neurodegeneration in the context of both neuronal and non-neuronal cell types (40–43), mitochondria in the synapse, specifically, have been found affected in AD (44), and the gene members of CM7 may contribute to this phenomenon. Similar to CM1 and CM10, we saw CM7's module eigengene declined with increasing pathology, and this decrease was greater in datasets from the temporal cortex compared to those from the frontal cortex (Fig. 5C–G, Supplementary Material, Fig. S8). Moreover, CM7 also showed no correlation with age, indicating the module as disease specific (adjusted  $R^2 = 0.0072$ , Fig. 5H).

### Glial-specific co-expression changes in AD

We next inspected modules that were classified as non-neuronal, specifically consensus modules, CM8, CM9, CM4 and CM5. These modules were positively correlated with AD diagnosis, and we identified CM8 as highly enriched in the microglia cell-type cluster from our snRNA-seq (Fig. 3G). Its hub genes (C1QA, C3, CASP1, CD14, etc.) and GO term enrichment of immune-related mechanisms support a critical role of neuroinflammation in AD (Fig. 5I–J). Microglia have been shown to secrete cytokines that induce a deleterious astrocytic state (A1)—one of these cytokines is C1Q (45). Similarly, members of consensus module CM9 included genes involved in some of the immune-related pathways recently implicated in AD, such as TGF $\beta$ /BMP (TGFBR2, BMPR1B), JAK/STAT (STAT3) and MAPK signaling (EGFR, FGFR1, FGF2; Fig. 5Q) (46,47). GO analysis indicated enrichment of terms related to positive regulation of signaling pathways (Fig. 5R). CM9 was highly enriched in astrocyte and endothelial cell-type clusters (Fig. 3G), and the JAK/STAT3 pathway previously has been shown to mediate astrocyte reactivity in AD (46).

In the Mayo TC dataset, both CM8 and CM9's module eigengenes significantly increased with late stage pathology ( $P < 0.001$ ), but while CM9 significantly increased in early stage pathology ( $P < 0.05$ ), there was little change for CM8 ( $P > 0.05$ , Fig. 5K, S). In addition, we again saw that the changes were more prominent in the temporal cortex than the frontal cortex (Fig. 5, Supplementary Material, Fig. S8), and both CM8 and CM9 demonstrated little correlation with age, supporting that these immune-related changes are specific to disease (adjusted  $R^2 = 0.0014$ , adjusted  $R^2 = 0.056$ ; Fig. 5P, X).

Further, consensus module CM4 contained gene members involved in various signaling pathways (NOTCH1, NFKB, ROCK1, RHOA, SMAD2, etc.; Fig. 6A), and GO analysis demonstrated enrichment of terms related to transcriptional regulation (Fig. 6B). With our snRNA-seq data, we determined this module to be enriched in OPC and both oligodendrocyte cell-type clusters (Fig. 3G), and this provided us a better functional understanding of CM4. It has been shown that the RhoA-ROCK pathway and Notch1 inhibit OPC differentiation (48–51); however, another member of CM4, ITGB1, has been demonstrated to promote myelination (52). Smad signaling also has been implicated in myelination (49,51). We additionally found that CM4's module eigengene increased with pathology, and this increase again was greater in datasets from the temporal cortex than those from the frontal cortex (Fig. 6C–G, Supplementary Material, Fig. S8). Moreover, we found CM4 was

not dependent on aging but was specific to disease (adjusted  $R^2 = 0.018$ ; Fig. 6h).

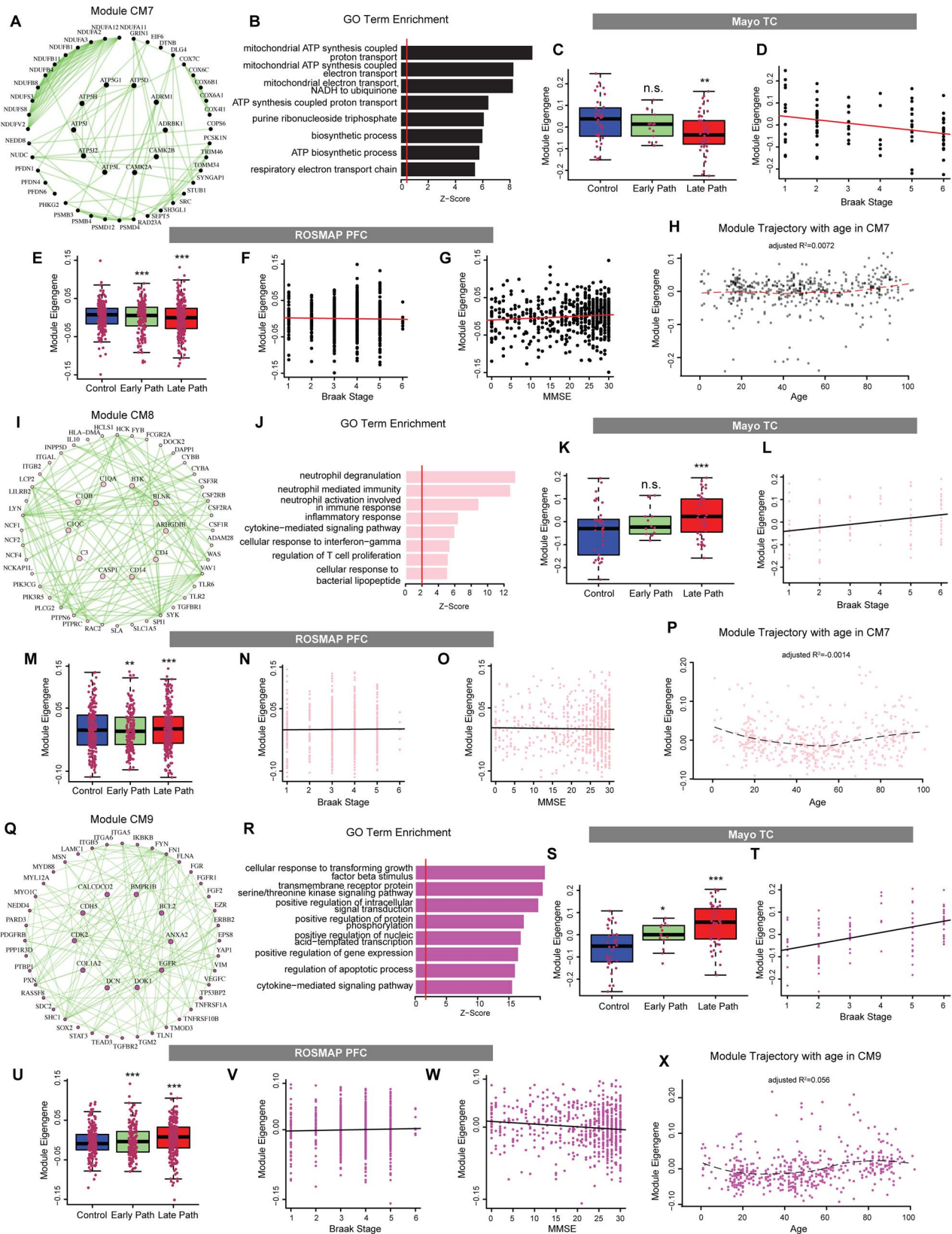
On the other hand, CM5's GO term enrichment and genes (HNRNPA2B1, PAN2, CLK2, etc.) were associated with microRNA (miRNA) processing (Fig. 6I–J). miRNAs have been largely studied in the context of development (53). For example, two of miRNA-125's predicted targets are SMAD4 and RHOA (members of CM4) (51), but miRNAs have been more recently demonstrated to play roles also in the immune system and neurodegeneration (34,54,55). Alterations in miRNA processing may mediate differential expression of miRNAs in AD. Similar to the previously mentioned non-neuronal modules, CM5's module eigengene increased with late stage pathology ( $P < 0.01$ ); however, the module eigengene also tended to decrease with early stage pathology ( $P > 0.05$ , Fig. 6K). We again saw that the increase in CM5's module eigengene with late stage pathology was more pronounced in the temporal cortex than frontal cortex and that CM5 showed little correlation with age (adjusted  $R^2 = 0.11$ ; Fig. 6L–P, Supplementary Material, Fig. S8). The remaining modules can be found in Supplementary Material, Figure S9.

### Assessment of genetic risk within consensus modules

The previously described transcriptomic alterations can be a cause or consequence of the disease. To address this question, we used GWAS SNPs as causal anchors and integrated our module-level information to find which modules harbor these causal anchors. We examined the enrichment of GWAS genes in our modules using the Multi-marker Analysis of GenoMic Annotation (MAGMA) pipeline (56), which controls for confounds like gene length. Each gene was assigned a score based on the best P-value of a single nucleotide polymorphism from an AD GWAS study (5), and MAGMA P-value  $< 0.05$  was used as the cut-off to include a gene as an AD candidate. We found that AD candidate genes were highly enriched in the microglial module CM8, further asserting a key role of the immune system in AD (Fig. 7A, Supplementary Material, Table S5). Several candidate genes, including CSF2RB, MS4A6A, CD33, SPI1 and CD14, were mapped in CM8, and these have been previously implicated in AD (57–60). In addition, our findings were consistent with a previous study (61) in which they reported AD risk genes in a microglial module. We wanted to further confirm our results by examining the largest number of GWAS enrichment analysis to date, including other neurodegenerative and neurological disorders (7,62–99). We found that CM8 was enriched only in GWAS hits from an AD study and a multiple sclerosis study (Supplementary Material, Fig. S10). Interestingly, CM8 was not enriched in risk genes associated with other tauopathies (frontotemporal dementia and progressive supranuclear palsy).

In addition, we validated our results in data from genome-wide gene-based association analysis (GWGAS), employed by Jansen *et al.*, using MAGMA on a much larger cohort comprising 71 880 AD cases and 383 378 controls (6). Similar to our GWAS enrichment scores, we found that GWGAS genes were enriched in CM8 (Fig. 7B). Moreover, we inspected brain-specific eQTLs and mQTLs overlapping with AD GWAS SNPs (AD-associated brain eQTLs and mQTLs, see Materials and Methods) and found that CM8 was enriched in AD-associated brain eQTLs. On the other hand, astrocytic module, CM9, was enriched in AD-associated brain mQTLs; these results again highlight glial immune-related processes in AD.





**Figure 5.** Consensus transcriptomics analysis identifies neuron-specific mitochondrial changes and glial-specific immune-related changes in disease. (A–B) Co-expression plot (A) and gene ontology term enrichment (B) for neuronal consensus module CM7. (C–D) Module eigengene trajectory with pathological state (C) and with Braak stage (D) for CM7 in Mayo TC dataset. (E–G) Module eigengene trajectory with pathological state (E), Braak stage (F) and MMSE score (G) for CM7 in ROSMAP PFC dataset. (H) Module eigengene trajectory with age for CM7 in NABEC dataset. (I–J) Co-expression plot (I) and gene ontology term enrichment (J) for non-neuronal

## Integration with orthogonal datasets and drug databases

We sought to understand the regulatory landscape of these AD-associated modules and integrated multiple layers of data to gain novel insights. We performed a transcriptome-wide association study (TWAS) to identify genes with cis-regulated expression associated with pathological and clinical traits of AD (100). We used plaque and tangle density as quantitative measures of the pathological changes in AD along with clinical dementia rating (CDR) scores to find significant transcriptome-wide associations. Using hypergeometric overlap tests, we found that downregulated cWGCNA modules—CM1, CM7, CM10 and CM12—were enriched in TWAS genes anti-correlated with plaque and tangle density, and these modules also were enriched in TWAS genes positively correlated with cognitive decline (Fig. 7C). Conversely, upregulated modules—CM4, CM16 and CM23, for example—were enriched in TWAS genes positively correlated with plaque and tangle density and TWAS genes negatively correlated with cognitive decline. These findings confirm our module trajectories with pathology and cognition previously discussed, but interestingly, module enrichment of TWAS genes associated with tangle density was not only less but also sometimes incongruent to module enrichment of TWAS genes associated with plaque density. Interestingly, microglial module CM8 was enriched in TWAS genes positively correlated with plaque density but was not enriched in TWAS genes associated with tangle density. This suggests that the upregulation of cis-regulated microglial immune-related genes may be either driving or affected by A $\beta$  pathology but not by tau pathology. On the other hand, CM7, which we identified as neuron-specific and mitochondrial-related, was highly enriched in TWAS genes anti-correlated with plaque density but is also enriched in TWAS genes positively correlated with tau density, indicating contrasting relationships between tau and A $\beta$  pathology with neuronal mitochondrial function.

As a further step to characterize our modules and to identify co-expression changes associated with epigenetic changes, we leveraged histone H3 lysine 9 acetylation (H3K9ac) marks, which are found near active transcription start sites and enhancers, associated with plaque or tangle density (101). We found that all AD-associated modules except microglial module CM8 were enriched in H3K9ac marks significantly associated with tangles (Fig. 7D). In addition, only neuronal module CM1 and astrocytic module CM9 were enriched in plaque-associated H3K9ac marks at promoter regions. We also examined our consensus modules with a dataset of AD-associated DNA methylated regions (102,103) and found only upregulated modules CM5, CM16 and CM23 were enriched in AD-associated hypo-methylated DNA regions. However, downregulated modules CM10, CM12, CM1 and CM7 were enriched in hyper-methylated DNA regions. The notable lack of CM8 enrichment in both AD-associated H3K9ac marks and AD-associated DNA methylation hints at an intriguing possibility that microglial immune-related changes in disease may not be mediated by epigenetic changes.

Additionally, we queried the Library of Integrated Network-Based Cellular Signatures (LINCS) database to identify potential drugs to target each consensus module (see [Materials and Methods](#),

[Supplementary Material, Fig. S11](#)). For modules upregulated in AD, we queried for drugs that would inhibit module expression, whereas for modules downregulated in AD, we queried for drugs that activate module expression. Notably, we found several drugs previously reported to be beneficial in AD and other neurodegenerative diseases. For example, histone deacetylase inhibitor Vorinostat, found to modulate CM9, is currently in human clinical trials (NCT03056495). Both saracatinib, an inhibitor of the Src/abl family of kinases and a modulator of CM8, and resveratrol, a modulator of CM5, have been found to have beneficial effects in human clinical trials (104,105).

## Identification of TFs regulating consensus modules

Network analysis identifies co-expressed modules that are likely co-regulated (33,106,107). With this in mind, we performed transcription factor binding site (TFBS) enrichment analysis (see [Materials and Methods](#)) to identify TFs that may be regulating disease-associated modules. This method has been successfully used previously to accurately predict TFs (32). TFBS revealed that master TF—RELA, otherwise known as p65, appears to control both neuronal and non-neuronal modules (Fig. 7E–F). It becomes further evident that NF $\kappa$ B signaling may be mediating transcriptional changes in AD since we found TFs NKFB1 and RELB associated with non-neuronal modules. Previous studies have reported increased NF $\kappa$ B activation in AD and pointed to a relationship with A $\beta$  pathology (108–111). Moreover, Withaferin A has been identified as an inhibitor of NF $\kappa$ B signaling (112,113) and has shown therapeutic potential to protect against A $\beta$  toxicity (114), and we found Withaferin A is a modulator of CM9, CM4 and CM1 ([Supplementary Material, Fig. S11](#)).

## Validation of consensus modules in external datasets

Since modules can be affected by non-biological variables like differences in sample preparation, we needed to authenticate the existence of our modules of interest in other datasets.

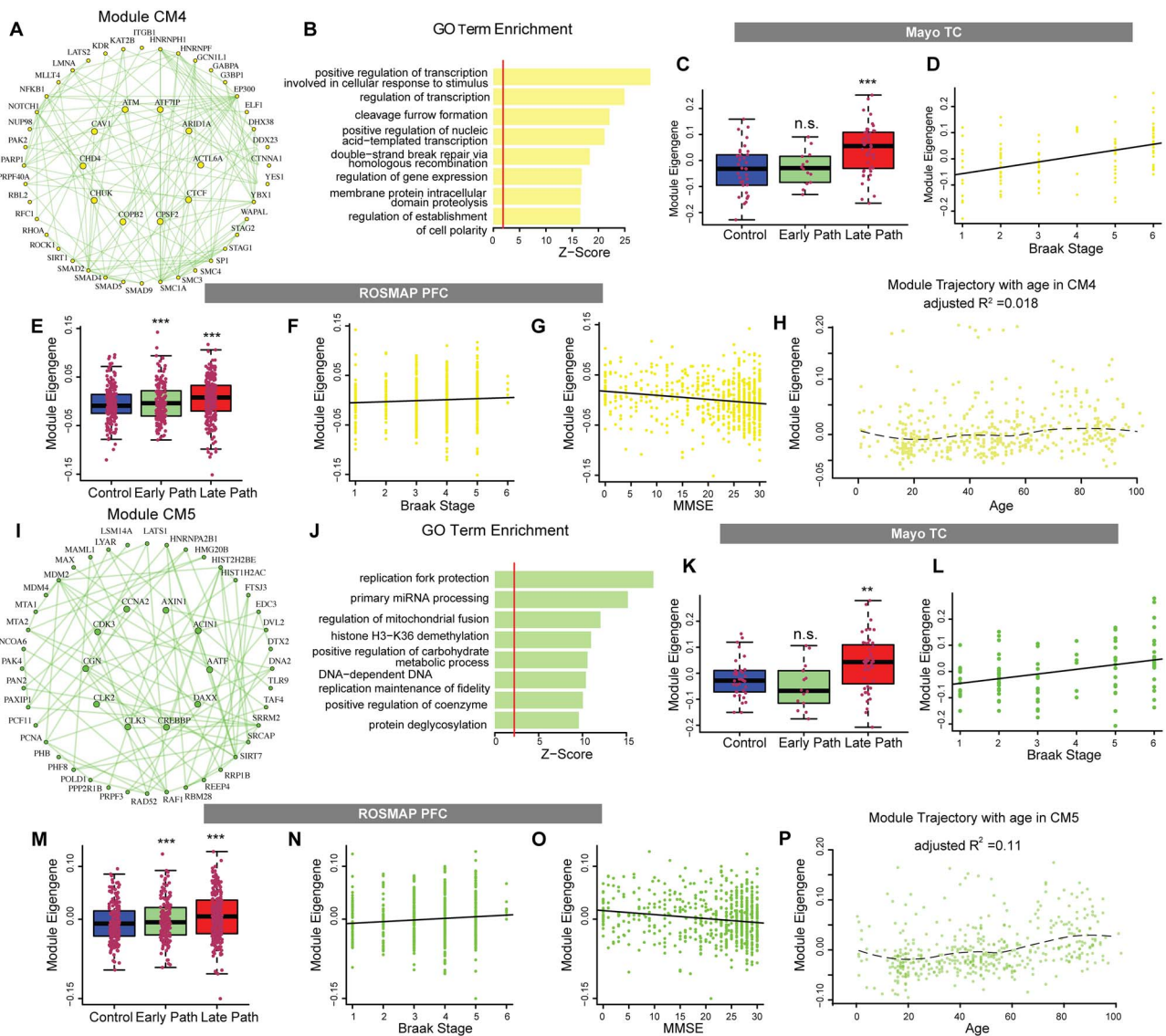
In order to do this, we utilized module preservation analysis, which helps us quantify the features of a particular module that are preserved across multiple networks (see [Materials and Methods](#)). We found that our consensus modules were highly preserved across published human RNA-seq and microarray datasets (61,115,116), indicated by Z-summary preservation (Fig. 8A, [Supplementary Material, Table S6](#)). Moreover, we examined the module eigengene trajectories of modules—CM10, CM1 and CM8—in these datasets and found comparable changes with diagnosis (Fig. 8B–D). As expected and also in line with our findings, we observed that changes were notably greater in the temporal cortex than the frontal cortex, again reflecting AD's progressive atrophy of different brain regions (1) ([Supplementary Material, Fig. S12](#)). These results altogether support the robustness of our findings.

## Discussion

Previous AD transcriptomic studies have been limited to tissue from a single brain repository or a single brain region (11,14,15);

---

consensus module CM8. (K–L) Module eigengene trajectory with pathological state (K) and with Braak stage (L) for CM8 in Mayo TC dataset. (M–O) Module eigengene trajectory with pathological state (M), Braak stage (N) and MMSE score (O) for CM8 in ROSMAP PFC dataset. (P) Module eigengene trajectory with age for CM8 in NABEC dataset. (Q–R) Co-expression plot (Q) and gene ontology term enrichment (R) for non-neuronal consensus module CM9. (S–T) Module eigengene trajectory with pathological state (S) and with Braak stage (T) for CM9 in Mayo TC dataset. (U–W) Module eigengene trajectory with pathological state (U), Braak stage (V) and MMSE score (W) for CM9 in ROSMAP PFC dataset. (X) Module eigengene trajectory with age for CM9 in NABEC dataset.



**Figure 6.** Consensus transcriptomics analysis identifies non-neuronal changes in disease. (A–B) Co-expression plot (A) and gene ontology term enrichment (B) for non-neuronal consensus module CM4. (C–D) Module eigengene trajectory with pathological state (C) and with Braak stage (D) for CM4 in Mayo TC dataset. (E–G) Module eigengene trajectory with pathological state (E), Braak stage (F) and MMSE score (G) for CM4 in ROSMAP PFC dataset. (H) Module eigengene trajectory with age for CM4 in NABEC dataset. (I–J) Co-expression plot (I) and gene ontology term enrichment (J) for non-neuronal consensus module CM5. (K–L) Module eigengene trajectory with pathological state (K) and with Braak stage (L) for CM5 in Mayo TC dataset. (M–O) Module eigengene trajectory with pathological state (M), Braak stage (N) and MMSE score (O) for CM5 in ROSMAP PFC dataset. (P) Module eigengene trajectory with age for CM5 in NABEC dataset.

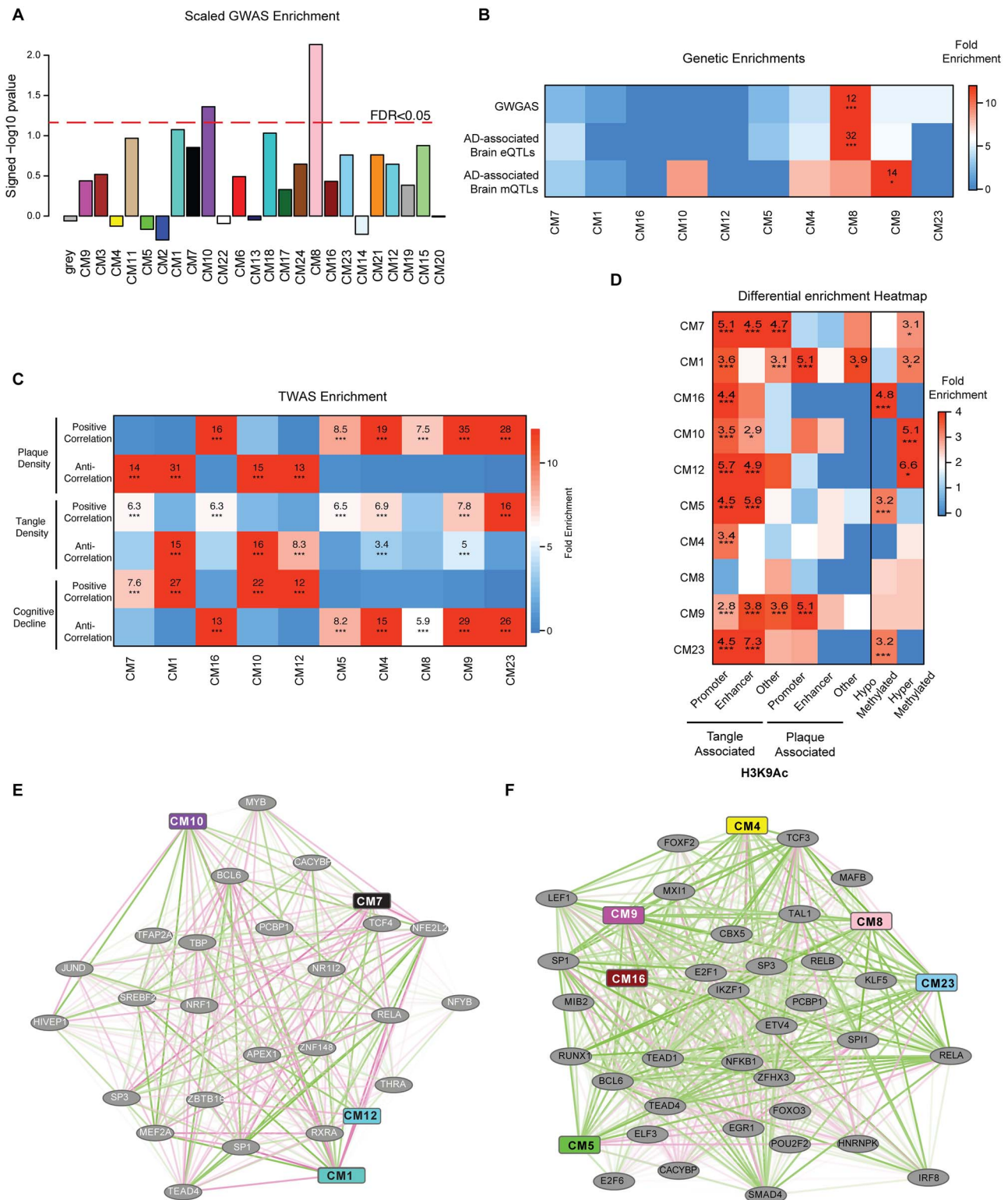
we performed cWGCNA (17,18) on three bulk tissue RNA-seq datasets (ROSMAP, Mayo and MSSM) in order to identify robust, conserved and biologically important transcriptomic alterations in AD across six cortical regions. We then integrated various genetic and epigenetic datasets, along with our own snRNA-seq data, in order to provide a more comprehensive picture of the genome-wide transcriptomic alterations occurring in AD and to begin to examine the associated gene regulatory programs.

A limitation of bulk tissue RNA-seq is that it does not address cell-type heterogeneity-associated changes that occur in disease, thus in order to probe the cell-type-specific transcriptional changes occurring in AD, we generated single-nuclei transcriptomic profiles on more than 27 000 nuclei from the aged human prefrontal cortex. We used a novel algorithm in order to estimate the cell-type abundance in bulk tissue RNA-seq data with our

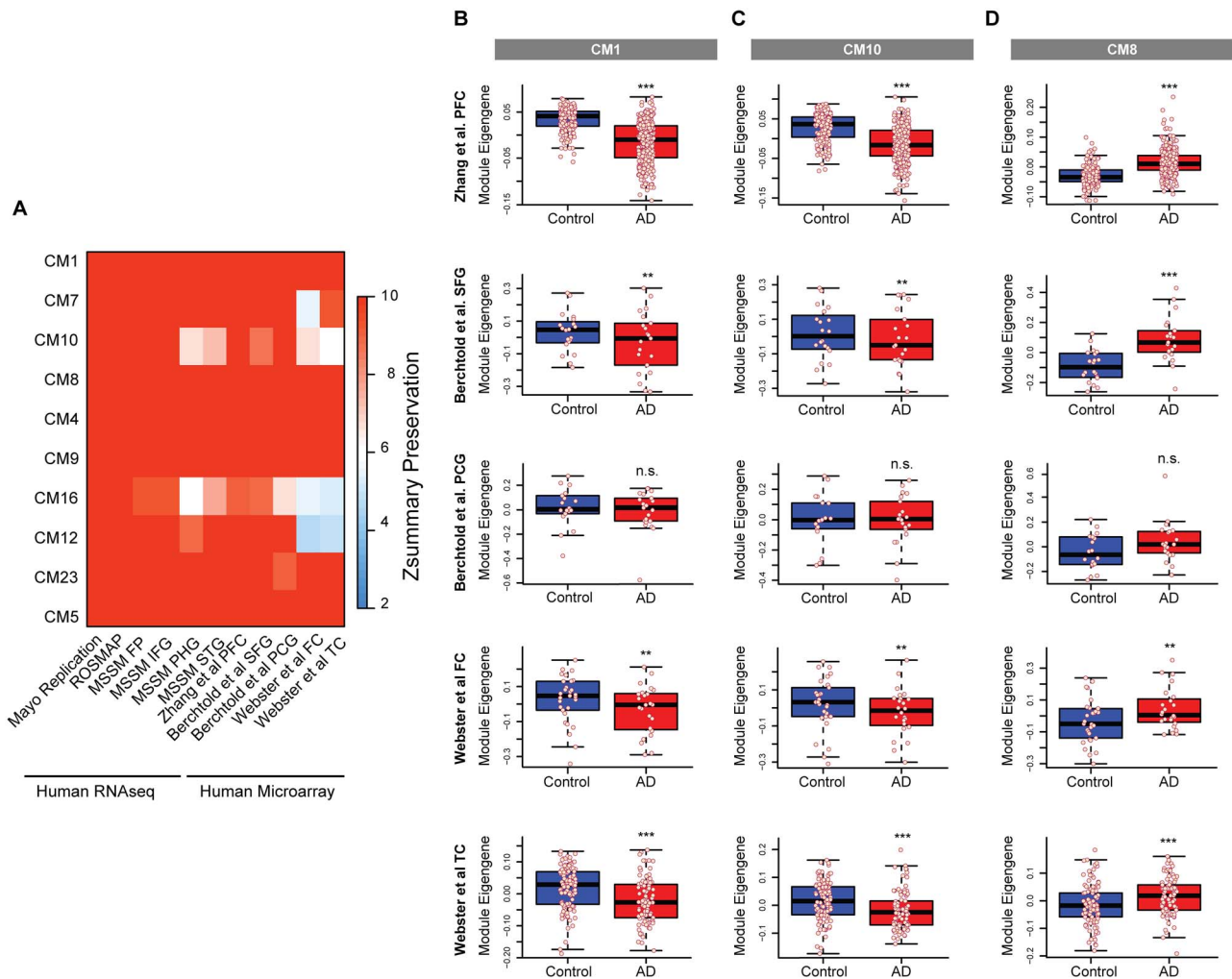
single-nuclei transcriptomic dataset. With this approach, we were able to show that the abundance of several excitatory and inhibitory neuronal clusters decreased with late stage pathology, whereas the abundance of microglial and astrocyte clusters increased. In addition, these changes in cell-type proportion were dependent on the brain region, modeling the progression of AD pathology in the brain (1,29).

Furthermore, we integrated our snRNA-seq data in order to help classify the cWGCNA modules based on cell type, allowing us to determine that AD co-expression changes related to calcium signaling (CM1), ubiquitination (CM10) and mitochondria function (CM7) are neuron specific. Gene set enrichment analysis showed that CM4 was associated with transcriptional regulation, but integration of our snRNA-seq data helped to clarify that this module was related to myelination and oligodendrocyte





**Figure 7.** Integration of orthogonal datasets to identify associated gene regulatory programs. (A) Mean scaled enrichment of AD genome-wide association study (GWAS) hits from Kunkle et al. (5) in consensus modules. (B) Heat map showing enrichment of consensus modules in genome-wide gene-based association analysis (GWGAS) (6) hits and AD-associated brain expression and methylation quantitative trait loci (eQTLs and mQTLs). (C) Heat map indicating enrichment of consensus modules with genes identified through transcriptome-wide association analysis (TWAS) (100). TWAS genes were measured in relation to plaque and tangle density, as well as clinical dementia rating (cognitive decline). (D) Heat map showing enrichment of consensus modules with H3K9 acetylation marks, correlated with tangle and plaque burden, in promoter, enhancer or other genomic regions and AD-associated hypermethylated and hypomethylated DNA regions. (E) Co-expression plot demonstrating relationships between neuronal consensus modules and transcription factors. Red edges reflect low correlation and green reflect high correlation. (F) Co-expression plot demonstrating relationships between non-neuronal consensus modules and transcription factors.



**Figure 8.** Validation of findings in external datasets. (A) Heat map indicating module preservation in other human RNA-seq and microarray datasets. (E–I) Box plot showing module eigengene trajectory with diagnosis of consensus modules CM10, CM8 and CM1 in datasets Zhang et al. PFC (61) (E), Berchtold et al. superior frontal gyrus (SFG, F) and post-central gyrus (115) (PCG, G), and Webster et al. FC (H) and TC (116) (I). The upper and lower lines represent the 75th and 25th percentile, respectively, and the central line represents the median.

maturation. While white matter changes have been previously implicated in AD, the associated molecular changes occurring have been largely unstudied (11,117–121), and a recent snRNA-seq study of AD also has highlighted myelination-related genes (122), supporting the importance of future work into understanding the gene members of CM4.

We also were able to confirm the role of astrocytes and microglia (CM9 and CM8) in immune-related processes in AD. Notably, we found that the microglial module CM8 was highly enriched in genetic risk signals associated with AD but not other tauopathies, suggesting that the concurrence of A $\beta$  and tau pathology may be important. However, recently, a transcriptomic study of microglia isolated from APP and tau transgenic mice investigated the differential effects of A $\beta$  and tau pathology, finding that A $\beta$  pathology results in an exacerbated inflammatory response from microglia, unseen in the tau transgenic-derived microglia (123). Similarly, we found that microglial module CM8 was enriched in TWAS genes associated with plaque density but not TWAS genes associated with tangle density, reinforcing the relationship between A $\beta$  pathology and microglial immune-related transcriptional changes.

Moreover, we wanted to clarify the epigenetic changes associated with AD co-expression changes, examining AD-associated DNA methylation regions and mQTLs, along with H3K9ac marks associated with tangle and plaque density. Interestingly, despite finding other modules significantly enriched in H3K9ac marks associated with pathology, we found that microglial module CM8 was not enriched in H3K9ac marks associated with either tangle or plaque density. Accordingly, CM8 was not enriched in AD-associated hypermethylated or hypomethylated DNA regions or AD-associated mQTLs, suggesting a lack of epigenetic regulation of CM8. Histone deacetylase inhibitors, however, have been shown to suppress microglial activation, which is associated with transcriptional changes (124,125), indicating that further work to elucidate the epigenetic mechanisms underlying AD is needed.

Since aging is the most important non-genetic risk factor for AD, we wanted to investigate if the AD-associated consensus modules are altered during normal human aging. Normal aging has been shown to alter glial and neuronal gene-expression changes in the brain (38); however, we hypothesized that these changes are distinct from pathological changes

occurring in neurodegenerative diseases. By calculating synthetic module eigengenes from AD-associated consensus modules in a NABEC dataset (38), we were able to confirm that the AD-associated modules were not significantly correlated with normal aging.

In addition, one of the biggest challenges in co-expression analysis is to find robust co-expressed modules that are altered in different cohorts and datasets. In order to tackle this problem, we used module preservation analysis, which is a powerful tool to find conserved modules in different datasets. Through module preservation analysis, we found that our AD-associated consensus modules are robustly preserved in multiple published AD datasets from different brain regions (61,115,116). This finding in combination with the modules' lack of correlation with normal human aging verified that we have identified disease-specific transcriptomic signatures.

Altogether, our study emphasizes the value of integrative analyses in order to fully understand the molecular changes occurring in disease. While single-cell RNA-seq provides the advantage of profiling gene expression changes at the cellular level, it is still not only expensive but also limited to mRNA. Thus, at this time, analyses incorporating both single-cell and bulk tissue transcriptomic data are important for thorough investigation of AD transcriptomic alterations. Additionally, we have examined epigenetic and TF regulation by integrating multiple orthogonal datasets, providing a framework for disease-associated gene regulatory networks in AD and further highlighting a need to employ integrative analyses of multiple types of data. We also hope that our findings will help pave way for accelerated drug development, a critical need in AD, since our study has identified robust co-expression modules and their hubs, in addition to TFs, which may be regulating these modules, all of which can be targets for novel therapeutics.

## Materials and Methods

### Data and code availability

RNA-seq data and details can be found on <https://www.synapse.org> using corresponding synapse (syn) IDs—Mayo RNA-seq (syn5550404), Mount Sinai Brain Bank (MSBB) (syn3159438) and ROSMAP (syn3219045). Single-cell data can be found on synapse website using synID# syn18915937. Additionally, we have created shiny apps to easily visualize gene, network-level and single-cell data, which can be accessed through our lab website at <https://swaruplab.bio.uci.edu/consensusAD>. Code for consensus WGCNA can be found here: <https://github.com/vswarup/ConsensusWGCNA>.

### Description of human samples and RNA-seq data

For single-cell analysis control human prefrontal cortex brain samples were obtained from UCI MIND's Alzheimer's Disease Research Center brain bank. These samples were obtained under UCI's Institutional Review Board. Post-mortem human brain tissues were obtained from three cohorts—Mayo study, MSBB study and ROSMAP study. All the human samples were obtained under their respective institutional review board. For details on human samples used in this study and their availability, please see [Supplementary Material, Table S1](#). RNA-seq data and details can be found on [synapse.org](https://www.synapse.org) website using corresponding synapse (syn) IDs—Mayo RNA-seq (syn5550404), MSBB (syn3159438) and ROSMAP (syn3219045).

### Other published datasets used in the study

For normal human aging, we used NABEC data ranging from 16 to 101 years old from frontal cortex (38,126,127). The gene-expression arrays are available from Gene Expression Omnibus (GEO)—GSE36192. Published AD microarray datasets were also downloaded from GEO including Zhang *et al.* prefrontal cortex data (61) (GSE44770), Berchtold *et al.* frontal and temporal cortex data (115) (GSE48350) and Webster *et al.* (116) (GSE15222).

### RNA-seq data processing

We performed standard alignment, QC and gene-expression analysis on each of the three datasets (ROSMAP, Mayo and MSSM) to ensure uniform processing of data. All samples were aligned to the human genome (GRCh38, GENCODE v25 gene model) using rnaSTAR aligner. Counts were obtained using rnaSTAR quantification and conditional quantile normalized FPKM data were used for downstream analysis. We used linear regression model to remove the effects of biological covariates like diagnosis, age, gender, post-mortem interval and technical variables like batch (depending on the cohort), RNA integrity number and sequencing biases and cell-type changes using gene-expression PCs from cell-type specific markers. The final model used was implemented in R as follows:

$$\begin{aligned} \text{lm}(\text{expression} \sim & \text{diagnosis} + \text{age} + \text{gender} + \text{PMI} + \text{batch} \\ & + \text{RIN} + \text{Seq.PC1} + \text{Seq.PC2} + \text{Neuron.PC1} \\ & + \text{Astrocytes.PC1} + \text{Microglia.PC1} + \text{Oligodendrocytes.PC1}) \end{aligned}$$

where Seq.PC1 and Seq.PC2 are sequencing PCs obtained from aggregating sequencing metrics obtained from Picard tools, and cell-type specific PCs are denoted as Neuron.PC1, Astrocytes.PC1, Microglia.PC1 and Oligodendrocytes.PC1. Cell-type specific genes were obtained from previously published work (34).

### Consensus WGCNA

cWGCNA was performed with ROSMAP, Mayo and MSSM datasets using WGCNA package in R (17). We employed a signed cWGCNA approach by calculating component-wise values for topological overlap for individual brain banks. First, bi-weighted mid-correlations were calculated for all pairs of genes, and then a signed similarity matrix was created. In the signed network, the similarity between genes reflects the sign of the correlation of their expression profiles. The signed similarity matrix was then raised to power  $\beta$  ( $\beta = 12$ ) to emphasize strong correlations and reduce the emphasis of weak correlations on an exponential scale. The resulting adjacency matrix was then transformed into a topological overlap matrix as described elsewhere (33). Modules were defined using specific module cutting parameters, which included minimum module size of 100 genes, deepSplit=4 and threshold of correlation=0.2. Modules with correlation greater than 0.8 were merged together. We used first principal component of the module, called module eigengene, to correlate with diagnosis and other variables. Hub genes were defined using intra-modular connectivity (kME) parameter of the WGCNA package. Gene-set enrichment analysis was done using enrichR package (128). Potential drugs targeting consensus modules were identified using enrichR to query the LINCS ([www.clue.io](http://www.clue.io)) chemical perturbations database.



## Comparison to other gene-clustering algorithms

In addition to cWGCNA, we constructed gene modules using several alternate methods to demonstrate that these modules are detected using different techniques. For this analysis, we used samples from the Mayo cohort to construct gene modules using DICER, DiffCoEx, MEGENA and self-organizing maps (SOMs). We compared the overlap between the modules resulting from these algorithms to the top 200 genes in each cWGCNA module by intra-modular connectivity (kME) using a hypergeometric test (phyper function in R). Default parameters were used for DICER (v0.6.0) and MEGENA (v1.3.7) in their respective R packages. To perform DiffCoEx, we used the `diffcoex` function from the R package `MODifieRDev` (<https://github.com/ddeweerd/MODifieRDev>) with default parameters. We constructed a SOM of the scaled gene expression data using the `som` function from the R package `kohonen` (v 3.0.10) using a grid of 40 by 40 nodes, whereby each gene in the expression dataset is assigned to one of these nodes throughout the SOM training routine. Gene modules are then constructed by hierarchically clustering the resulting SOM nodes using the `hclust` function in R. We then visualized bi-weighted mid correlations between genes and relevant biological variables such as AD diagnosis, age at death and sex on the SOM to ensure that the SOM gene modules were capturing relevant biological features.

**Proteomic Module Eigengenes**—Processed proteomic AD dataset (36) was downloaded from synapse website (syn20933797). Consensus AD transcriptomic module definitions were used to calculate synthetic module eigengenes using the `moduleEigengenes` function in WGCNA package and plotted as boxplots stratified by control, asymptomatic AD and AD samples.

**Co-expression PPI networks**—Co-expression PPI networks were created as described previously (34). Briefly, we used two PPI resources, InWeb (30,129) and BioGRID (31) and using the union of the two resources, the subset of compiled PPIs between genes in a module was extracted and all edges were counted. The PPI dataset was matrix-multiplied with co-expression data from the RNA-seq data, the edges that were present both in PPI and co-expression datasets were eventually kept. This approach allowed us to infer tissue and species specificity in the PPI network. The PPI network was then visualized using the `igraph` package in R.

## Module Preservation Analysis

To understand the functional relationship of modules and its network properties in orthogonal datasets, we performed module preservation analysis. Module definitions from the cWGCNA analysis was used and Z-statistics were calculated using the module preservation function in WGCNA package in R (130).

## Genetic enrichment with eQTLs, mQTLs and GWAS

Brain specific eQTLs were generated from databases including the CommonMind Consortium Portal, AMP-AD, BRAINEAC and xQTL Serve. A false discovery rate of 0.05 was used to define significant eQTL associations. eQTLs that overlapped with AD associated SNPs from either Lambert *et al.* or Jansen *et al.* GWAS datasets (6,7) were annotated as AD-associated eQTLs. Similarly, methylation QTLs (mQTLs) defined in AMP-AD, and CommonMind Consortium Portal were annotated as AD-associated mQTLs if they overlapped with AD GWAS data. We also used GWAS data using MAGMA (56). GWAS enrichment analysis in modules was done using MAGMA pipeline. Briefly, GWAS summary statistics was obtained for AD from IGAP website

(7). Using MAGMA approach, we assigned each gene a score based on the best P-value of an SNP in a given GWAS study within 20 kb of the gene, and then set a P-value cut-off at 0.05 to define the gene as included in the common variant set related to AD. We performed enrichment analysis with logistic regression controlling for gene length, LD blocks and other biases (33).

## TWAS Enrichment Analysis

We performed AD TWAS using the FUSION package (<http://gusevlab.org/projects/fusion/>) (100) with gene-expression measured in ROSMAP PFC data and GWAS summary statistics from AD GWAS (7). We chose the best model of accuracy in the FUSION package (best cis-eQTL, best linear unbiased predictor, Bayesian sparse linear mixed model (131), Elastic-net regression, LASSO regression) and used for downstream association analyses. TWAS association statistics were Bonferroni corrected per GWAS, gene and intron separately. Association for TWAS was assessed for plaque and tangle densities and genome-wide significant genes were used for enrichment analysis. Similarly, CDR reported in ROSMAP dataset was used for association with cognitive decline.

## Single-cell Library Preparation

Five control human brain samples were processed for nuclei isolation following modified version of the published protocol (132). All procedures were carried on ice. Approximately 50 mg of the brain sample was homogenized in 500  $\mu$ l chilled Nuclei EZ lysis buffer (Sigma-Aldrich), using a motorized tissue grinder and disposable pellet pestles (Fisher Scientific) by stroking ~10–20 times. The homogenate was then incubated for 5 min after adding 1 ml of the Nuclei EZ lysis buffer. Homogenate was then filtered using a 70  $\mu$ m MACS strainer (Miltenyi Biotec). The flow through was centrifuged at 500 g for 5 min at 4°C and the supernatant was carefully removed. The nuclei pellet was then re-suspended in another 1.5 ml of Nuclei EZ lysis buffer and incubated for 5 min. The cell suspension was again centrifuged (500 g, 5 min, 4°C) and the supernatant was removed. A total of 500  $\mu$ l of Nuclei wash and re-suspension buffer (NWR buffer: 1 $\times$  PBS (Fisher Scientific), 1% BSA (Sigma-Aldrich) and 0.2 U/ $\mu$ l RNase inhibitor (New England Bio Labs)) was then added and incubated without re-suspending, to allow buffer interchange. After the incubation, 1 ml of NWR buffer was added and the nuclei were re-suspended. The resuspension was centrifuged (500 g, 5 min, 4°C) and the supernatant was removed. The nuclei pellet was then gently resuspended in 1.4 ml of NWR buffer, centrifuged (500 g, 5 min, 4°C), removed supernatant and was gently re-suspended in another 500  $\mu$ l of NWR buffer. All nuclei were collected by washing the walls of the centrifuge tube. Nuclei re-suspension was then filtered using 40- $\mu$ m Flowmi cell strainer (Bel-Art). The number of nuclei was then counted using an automated cell counter (Bio-Rad). Samples were directly sorted by FACS using DAPI positive gates at the UCI Immunology FACS core in the 10X buffer and libraries were prepared using Single-cell RNA-seq library kit (v3, 10X Genomics). Samples were pooled and sequenced at an average read-depth of 100 000 reads/cell.

## Single-cell Clustering and analysis

Raw sequencing Illumina bcl2 files were demultiplexed using 10X Genomics Cell Ranger software. Further analysis was performed in Seurat V3 package in R (133). We used `sctransform`

function in Seurat to normalize the data and top 75 PCs were used for clustering and visualization using Fourier transformation t-SNE (Fitsne (134)). Differential expression between clusters were performed using MAST package in Seurat (135). Additionally, differential expression was performed to compare cells originating from male and female samples within each cluster using MAST.

### Cell-type composition analysis

For cell-type composition analysis, unregressed FPKM data which did not account for cell-type principal components, were used. Using the most variable genes in a cluster, we performed WGCNA analysis and determined the intramodular connectivity (kME) of each gene. kME values are correlation of first principal components of a module/cluster to the expression of that gene. Thus, higher kME values can often explain most of the variance in the data because of their high correlation to PC1. Using Fisher's transformation, we converted kME values to specificity scores (Z-scores). We then used the top one percentile of genes by their kME values and calculated their module eigengene in respective bulk RNA-seq dataset. These module eigengene values were then plotted to understand the cell-composition changes in control and AD samples.

### Transcription factor binding site enrichment

TFBS enrichment was done using pipeline published elsewhere (106). Briefly, the computational pipeline identifies the TFs (in the promoter region) in the given gene list using experimentally validated TFs. The pipeline used TF position-weighted matrix from JASPAR and TRANSFAC databases as well as included experimental associated datasets obtained from ENCODE (136) and ChEA (137). The pipeline uses clover algorithm and corrects for background using human chromosome 22.

### Supplementary Material

Supplementary Material is available at HMG online.

### Acknowledgements

This work was supported by University of California Irvine startup funds and American Federation of Aging Research (AFAR) young investigator award to V.S. For ROSMAP dataset, the study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported by National Institute on Aging [P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984, U01AG46152]; the Illinois Department of Public Health; and the Translational Genomics Research Institute. The results published here are in whole or in part based on data obtained from the AMP-AD Knowledge Portal (doi:10.7303/syn2580853). For Mayo dataset, the study data were provided by the following sources: The Mayo Clinic Alzheimer's Disease Genetic Studies, led by Dr Nilufer Taner and Dr Steven G. Younkin, Mayo Clinic, Jacksonville, FL, using samples from the Mayo Clinic Study of Aging, the Mayo Clinic Alzheimer's Disease Research Center and the Mayo Clinic Brain Bank. Data collection was supported by National Institute on Aging [AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01 AG003949]; National Institute of Neurological Disorders and Stroke [R01 NS080820]; CurePSP

Foundation; and support from Mayo Foundation. Study data include samples collected through the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona. The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke [U24 NS072026 to National Brain and Tissue Resource for Parkinson's Disease and Related Disorders]; the National Institute on Aging [P30 AG19610 to Arizona Alzheimer's Disease Core Center]; the Arizona Department of Health Services (contract 211002, Arizona Alzheimer's Research Center); the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson's Disease Consortium); and the Michael J. Fox Foundation for Parkinson's Research. The results published here are in whole or in part based on data obtained from the AMP-AD Knowledge Portal (doi:10.7303/syn2580853). For MSSM dataset, the data were generated from postmortem brain tissue collected through the Mount Sinai VA Medical Center Brain Bank and were provided by Dr Eric Schadt from Mount Sinai School of Medicine. The results published here are in whole or in part based on data obtained from the AMP-AD Knowledge Portal (doi:10.7303/syn2580853). In addition, we are grateful to Thanneer Perumal and Ben Logsdon at Sage bionetworks for processing the fastq and bam files used in the analysis.

*Conflict of Interest statement.* The authors declare no conflict of interests.

### References

1. Braak, H. and Braak, E. (1991) Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.*, **82**, 239–259.
2. Vinters, H.V. (2015) Emerging concepts in Alzheimer's disease. *Annu. Rev. Pathol. Mech. Dis.*, **10**, 291–319.
3. Hardy, J. and Selkoe, D.J. (2002) The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science*, **297**, 353–356.
4. Desikan, R.S., Schork, A.J., Wang, Y., Witoelar, A., Sharma, M., McEvoy, L.K., Holland, D., Brewer, J.B., Chen, C.H., Thompson, W.K. et al. (2015) Genetic overlap between Alzheimer's disease and Parkinson's disease at the MAPT locus. *Mol. Psychiatry*, **20**, 1588–1595.
5. Kunkle, B.W., Grenier-Boley, B., Sims, R., Bis, J.C., Damotte, V., Naj, A.C., Boland, A., Vronskaya, M., van der Lee, S.J., Amlie-Wolf, A. et al. (2019) Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A $\beta$ , tau, immunity and lipid processing. *Nat. Genet.*, **51**, 414–430.
6. Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., Sealock, J., Karlsson, I.K., Hägg, S., Athanasiu, L. et al. (2019) Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.*, **51**, 404–413.
7. Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B. et al. (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.*, **45**, 1452–1458.
8. Hodes, R.J. and Buckholtz, N. (2016) Accelerating medicines partnership: Alzheimer's disease (AMP-AD) knowledge portal aids Alzheimer's drug discovery through open data sharing. *Expert Opin. Ther. Targets*, **20**, 389–391.

9. Mukherjee, S., Perumal, T., Daily, K., Sieberts, S., Omberg, L., Preuss, C., Carter, G., Mangravite, L. and Logsdon, B. (2019) Identifying and ranking potential driver genes of Alzheimer's disease using multi-view evidence aggregation. *Bioinformatics*, **35**, i568–i576.
10. Allen, M., Carrasquillo, M.M., Funk, C., Heavner, B.D., Zou, F., Younkin, C.S., Burgess, J.D., Chai, H.-S., Crook, J., Eddy, J.A. et al. (2016) Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci. Data*, **3**, 160089.
11. Allen, M., Wang, X., Burgess, J.D., Watzlawik, J., Serie, D.J., Younkin, C.S., Nguyen, T., Malphrus, K.G., Lincoln, S., Carrasquillo, M.M. et al. (2018) Conserved brain myelination networks are altered in Alzheimer's and other neurodegenerative diseases. *Alzheimers Dement.*, **14**, 352–366.
12. De Jager, P.L., Ma, Y., McCabe, C., Xu, J., Vardarajan, B.N., Felsky, D., Klein, H.-U., White, C.C., Peters, M.A., Lodgson, B. et al. (2018) A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci. Data*, **5**, 180142.
13. Gaiteri, C., Mostafavi, S., Honey, C.J., De Jager, P.L. and Bennett, D.A. (2016) Genetic variants in Alzheimer disease—molecular and brain network approaches. *Nat. Rev. Neurol.*, **12**, 413–427.
14. Mostafavi, S., Gaiteri, C., Sullivan, S.E., White, C.C., Tasaki, S., Xu, J., Taga, M., Klein, H.-U., Patrick, E., Komashko, V. et al. (2018) A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat. Neurosci.*, **21**, 811–819.
15. Raj, T., Li, Y.I., Wong, G., Humphrey, J., Wang, M., Ramdhani, S., Wang, Y.-C., Ng, B., Gupta, I., Haroutunian, V. et al. (2018) Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nat. Genet.*, **50**, 1584–1592.
16. Wang, M., Beckmann, N.D., Roussos, P., Wang, E., Zhou, X., Wang, Q., Ming, C., Neff, R., Ma, W., Fullard, J.F. et al. (2018) The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci. Data*, **5**, 180185.
17. Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.*, **9**, 559.
18. Zhang, B. and Horvath, S. (2005) A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.*, **4**, Article 17.
19. Chiu, D.S. and Talhouk, A. (2018) diceR: an R package for class discovery using an ensemble driven approach. *BMC Bioinform.*, **19**, 11.
20. Song, W.-M. and Zhang, B. (2015) Multiscale embedded gene co-expression network analysis. *PLoS Comput. Biol.*, **11**, e1004574.
21. Kohonen, T. (1990) The self-organizing map. *Proc. IEEE*, **78**, 1464–1480.
22. Tesson, B.M., Breitling, R. and Jansen, R.C. (2010) DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. *BMC Bioinform.*, **11**, 497.
23. Lake, B.B., Ai, R., Kaeser, G.E., Salathia, N.S., Yung, Y.C., Liu, R., Wildberg, A., Gao, D., Fung, H.L., Chen, S. et al. (2016) Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science*, **352**, 1586–1590.
24. Zeng, H., Shen, E.H., Hohmann, J.G., Oh, S.W., Bernard, A., Royall, J.J., Glatfelter, K.J., Sunkin, S.M., Morris, J.A., Guillozet-Bongaarts, A.L. et al. (2012) Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell*, **149**, 483–496.
25. Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Hayden Gephart, M.G., Barres, B.A. and Quake, S.R. (2015) A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci.*, **112**, 7285–7290.
26. Laarakker, M.C., Reinders, N.R., Bruining, H., Ophoff, R.A. and Kas, M.J.H. (2012) Sex-dependent novelty response in neurexin-1 $\alpha$  mutant mice. *PLoS One*, **7**, e31503.
27. Grayton, H.M., Missler, M., Collier, D.A. and Fernandes, C. (2013) Altered social behaviours in neurexin 1 $\alpha$  knockout mice resemble core symptoms in neurodevelopmental disorders. *PLoS One*, **8**, e67114.
28. Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M. and Alizadeh, A.A. (2015) Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods*, **12**, 453–457.
29. Scahill, R.I., Schott, J.M., Stevens, J.M., Rossor, M.N. and Fox, N.C. (2002) Mapping the evolution of regional atrophy in Alzheimer's disease: unbiased analysis of fluid-registered serial MRI. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 4703–4707.
30. Lage, K., Karlberg, E.O., Størling, Z.M., Olason, P.I., Pedersen, A.G., Rigina, O., Hinsby, A.M., Tümer, Z., Pociot, F., Tommerup, N. et al. (2007) A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.*, **25**, 309–316.
31. Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A. and Tyers, M. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.*, **34**, D535–D539.
32. Oldham, M.C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S. and Geschwind, D.H. (2008) Functional organization of the transcriptome in human brain. *Nat. Neurosci.*, **11**, 1271–1282.
33. Parikshak, N.N., Swarup, V., Belgard, T.G., Irimia, M., Ramaswami, G., Gandal, M.J., Hartl, C., Leppa, V., Ubieta, L.d.l.T., Huang, J. et al. (2016) Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature*, **540**, 423–427.
34. Swarup, V., Hinz, F.I., Rexach, J.E., Noguchi, K.-I., Toyoshima, H., Oda, A., Hirai, K., Sarkar, A., Seyfried, N.T., Cheng, C. et al. (2019) Identification of evolutionarily conserved gene networks mediating neurodegenerative dementia. *Nat. Med.*, **25**, 152–164.
35. Parikshak, N.N., Gandal, M.J. and Geschwind, D.H. (2015) Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.*, **16**, 441–458.
36. Johnson, E.C.B., Dammer, E.B., Duong, D.M., Ping, L., Zhou, M., Yin, L., Higginbotham, L.A., Guajardo, A., White, B., Troncoso, J.C. et al. (2020) Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nat. Med.*, **26**, 769–780.
37. Workgroup, A.S.A.C.H. and Khachaturian, Z.S. (2017) Calcium hypothesis of Alzheimer's disease and brain aging: a framework for integrating new evidence into a comprehensive theory of pathogenesis. *Alzheimers Dement.*, **13**, 178–182.e117.
38. Soreq, L., Rose, J., Soreq, E., Hardy, J., Trabzuni, D., Cookson, M.R., Smith, C., Ryten, M., Patani, R. and Ule, J. (2017) Major shifts in glial regional identity are a transcriptional hallmark of human brain aging. *Cell Rep.*, **18**, 557–570.



39. Chen, Y., Neve, R.L. and Liu, H. (2012) Neddylation dysfunction in Alzheimer's disease. *J. Cell. Mol. Med.*, **16**, 2583–2591.
40. DuBoff, B., Feany, M. and Götz, J. (2013) Why size matters—balancing mitochondrial dynamics in Alzheimer's disease. *Trends Neurosci.*, **36**, 325–335.
41. Lezi, E. and Swerdlow, R.H. (2012) Mitochondria in neurodegeneration. *Adv. Exp. Med. Biol.*, **942**, 269–286.
42. Joshi, A.U., Minhas, P.S., Liddelow, S.A., Hailleslassie, B., Andreasson, K.I., Dorn, G.W. and Mochly-Rosen, D. (2019) Fragmented mitochondria released from microglia trigger A1 astrocytic response and propagate inflammatory neurodegeneration. *Nat. Neurosci.*, **22**, 1635–1648.
43. Phatnani, H. and Maniatis, T. (2015) Astrocytes in neurodegenerative disease. *Cold Spring Harb. Perspect. Biol.*, **7**.
44. Du, H., Guo, L., Yan, S., Sosunov, A.A., McKhann, G.M. and Yan, S.S. (2010) Early deficits in synaptic mitochondria in an Alzheimer's disease mouse model. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 18670–18675.
45. Liddelow, S.A., Guttenplan, K.A., Clarke, L.E., Bennett, F.C., Bohlen, C.J., Schirmer, L., Bennett, M.L., Münch, A.E., Chung, W.-S., Peterson, T.C. et al. (2017) Neurotoxic reactive astrocytes are induced by activated microglia. *Nature*, **541**, 481–487.
46. Ben Haim, L., Ceyzeriat, K., Carrillo-de Sauvage, M.A., Aubry, F., Auregan, G., Guillemier, M., Ruiz, M., Petit, F., Houitte, D., Faivre, E. et al. (2015) The JAK/STAT3 pathway is a common inducer of astrocyte reactivity in Alzheimer's and Huntington's diseases. *J. Neurosci.*, **35**, 2817–2829.
47. von Bernhardt, R., Cornejo, F., Parada, G.E. and Eugénin, J. (2015) Role of TGF $\beta$  signaling in the pathogenesis of Alzheimer's disease. *Front. Cell. Neurosci.*, **9**.
48. Baer, A.S., Syed, Y.A., Kang, S.U., Mitteregger, D., Vig, R., French-Constant, C., Franklin, R.J.M., Altmann, F., Lubec, G. and Kotter, M.R. (2009) Myelin-mediated inhibition of oligodendrocyte precursor differentiation can be overcome by pharmacological modulation of Fyn-RhoA and protein kinase C signalling. *Brain*, **132**, 465–481.
49. Emery, B. and Lu, Q.R. (2015) Transcriptional and epigenetic regulation of oligodendrocyte development and myelination in the central nervous system. *Cold Spring Harb. Perspect. Biol.*, **7**, a020461.
50. van Tilborg, E., de Theije, C.G.M., van Hal, M., Wagenaar, N., de Vries, L.S., Benders, M.J., Rowitch, D.H. and Nijboer, C.H. (2018) Origin and dynamics of oligodendrocytes in the developing brain: implications for perinatal white matter injury. *Glia*, **66**, 221–238.
51. Elbaz, B. and Popko, B. (2019) Molecular control of oligodendrocyte development. *Trends Neurosci.*, **42**, 263–277.
52. Quintela-López, T., Ortiz-Sanz, C., Serrano-Regal, M.P., Gaminde-Blasco, A., Valero, J., Baleriola, J., Sánchez-Gómez, M.V., Matute, C. and Alberdi, E. (2019) A $\beta$  oligomers promote oligodendrocyte differentiation and maturation via integrin  $\beta$ 1 and Fyn kinase signaling. *Cell Death Dis.*, **10**, 445.
53. Ivey, K.N. and Srivastava, D. (2015) microRNAs as developmental regulators. *Cold Spring Harb. Perspect. Biol.*, **7**, a008144.
54. Lu, L.-F. and Liston, A. (2009) MicroRNA in the immune system, microRNA as an immune system. *Immunology*, **127**, 291–298.
55. Qiu, L., Tan, E.K. and Zeng, L. (2015) microRNAs and neurodegenerative diseases. *Adv. Exp. Med. Biol.*, **888**, 85–105.
56. de Leeuw, C.A., Mooij, J.M., Heskes, T. and Posthuma, D. (2015) MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.*, **11**, e1004219.
57. Shang, D.S., Yang, Y.M., Zhang, H., Tian, L., Jiang, J.S., Dong, Y.B., Zhang, K., Li, B., Zhao, W.D., Fang, W.G. and Chen, Y.H. (2016) Intracerebral GM-CSF contributes to transendothelial monocyte migration in APP/PS1 Alzheimer's disease mice. *J. Cereb. Blood Flow Metab.*, **36**, 1978–1991.
58. Proitsi, P., Lee, S.H., Lunnon, K., Keohane, A., Powell, J., Troakes, C., Al-Sarraj, S., Furney, S., Soininen, H., Kloszewska, I. et al. (2014) Alzheimer's disease susceptibility variants in the MS4A6A gene are associated with altered levels of MS4A6A expression in blood. *Neurobiol. Aging*, **35**, 279–290.
59. Griciuc, A., Serrano-Pozo, A., Parrado, A.R., Lesinski, A.N., Asselin, C.N., Mullin, K., Hooli, B., Choi, S.H., Hyman, B.T. and Tanzi, R.E. (2013) Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. *Neuron*, **78**, 631–643.
60. Huang, K.-L., Marcora, E., Pimenova, A.A., Di Narzo, A.F., Kapoor, M., Jin, S.C., Harari, O., Bertelsen, S., Fairfax, B.P., Czajkowski, J. et al. (2017) A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease. *Nat. Neurosci.*, **20**, 1052–1061.
61. Zhang, B., Gaiteri, C., Bodea, L.-G., Wang, Z., McElwee, J., Podtelezhnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R. et al. (2013) Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*, **153**, 707–720.
62. Ferrari, R., Hernandez, D.G., Nalls, M.A., Rohrer, J.D., Ramasamy, A., Kwok, J.B.J., Dobson-Stone, C., Brooks, W.S., Schofield, P.R., Halliday, G.M. et al. (2014) Frontotemporal dementia and its subtypes: a genome-wide association study. *Lancet Neurol.*, **13**, 686–699.
63. Chen, J., Yu, J.-T., Wojta, K., Wang, H.-F., Zetterberg, H., Blennow, K., Yokoyama, J.S., Weiner, M.W., Kramer, J.H., Rosen, H. et al. (2017) Genome-wide association study identifies MAPT locus influencing human plasma tau levels. *Neurology*, **88**, 669–676.
64. Group, P.S.P.G.S., Höglinger, G.U., Melhem, N.M., Dickson, D.W., Sleiman, P.M.A., Wang, L.-S., Klei, L., Rademakers, R., de Silva, R., Litvan, I. et al. (2011) Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat. Genet.*, **43**, 699–705.
65. The Alzheimer's Disease Neuroimaging, I., The, C.C., Epigen, Imagen, Sys, Hibar, D.P., Stein, J.L., Renteria, M.E., Arias-Vasquez, A., Desrivieres, S. et al. (2015) Common genetic variants influence human subcortical brain structures. *Nature*, **520**, 224–229.
66. Middeldorp, C.M., Hammerschlag, A.R., Ouwens, K.G., Groen-Blokhuis, M.M., Pourcain, B.S., Greven, C.U., Pappa, I., Tiesler, C.M.T., Ang, W., Nolte, I.M. et al. (2016) A genome-wide association meta-analysis of attention-deficit/hyperactivity disorder symptoms in population-based pediatric cohorts. *J. Am. Acad. Child Adolesc. Psychiatry*, **55**, 896–905.e896.
67. Zayats, T., Athanasiu, L., Sonderby, I., Djurovic, S., Westlye, L.T., Tamnes, C.K., Fladby, T., Aase, H., Zeiner, P., Reichborn-Kjennerud, T. et al. (2015) Genome-wide analysis of attention deficit hyperactivity disorder in Norway. *PLoS One*, **10**, e0122501.
68. Pappa, I., St Pourcain, B., Benke, K., Cavadino, A., Hakulinen, C., Nivard, M.G., Nolte, I.M., Tiesler, C.M.T., Bakermans-Kranenburg, M.J., Davies, G.E. et al. (2016) A genome-wide approach to children's aggressive behavior: the EAGLE consortium. *Am. J. Med. Genet.*, **171**, 562–572.

69. Agrawal, A., Freedman, N.D. and Bierut, L.J. (2011) Genome-wide association studies of alcohol intake—a promising cocktail? *Am. J. Clin. Nutr.*, **93**, 681–683.
70. Registry, P., Group, S., Registry, S., Consortium, F.S., Consortium, S., Group, N.S., van Rheenen, W., Shatunov, A., Dekker, A.M., McLaughlin, R.L. et al. (2016) Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat. Genet.*, **48**, 1043–1048.
71. Hinney, A., Kesselmeier, M., Jall, S., Volckmar, A.L., Föcker, M., Antel, J., Gcan, Wtccc, Heid, I.M., Winkler, T.W. et al. (2017) Evidence for three genetic loci involved in both anorexia nervosa risk and variation of body mass index. *Mol. Psychiatry*, **22**, 192–201.
72. Otowa, T., Hek, K., Lee, M., Byrne, E.M., Mirza, S.S., Nivard, M.G., Bigdeli, T., Aggen, S.H., Adkins, D., Wolen, A. et al. (2016) Meta-analysis of genome-wide association studies of anxiety disorders. *Mol. Psychiatry*, **21**, 1391–1399.
73. Weiner, D.J., Wigdor, E.M., Ripke, S., Walters, R.K., Kosmicki, J.A., Grove, J., Samocha, K.E., Goldstein, J.I., Okbay, A., Bybjerg-Grauholm, J. et al. (2017) Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat. Genet.*, **49**, 978–985.
74. Autism Spectrum Disorders Working Group of The Psychiatric Genomics, C (2017) Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol. Autism*, **8**, 21.
75. The International Consortium for Blood Pressure Genome-Wide Association, S (2011) Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*, **478**, 103–109.
76. Mühleisen, T.W., Leber, M., Schulze, T.G., Strohmaier, J., Degenhardt, F., Treutlein, J., Mattheisen, M., Forstner, A.J., Schumacher, J., Breuer, R. et al. (2014) Genome-wide association study reveals two new risk loci for bipolar disorder. *Nat. Commun.*, **5**, 3339.
77. Kirkpatrick, R.M., McGue, M., Iacono, W.G., Miller, M.B. and Basu, S. (2014) Results of a “GWAS plus”: general cognitive ability is substantially heritable and massively polygenic. *PLoS One*, **9**, e112390.
78. Global Lipids Genetics, C (2013) Discovery and refinement of loci associated with lipid levels. *Nat. Genet.*, **45**, 1274–1283.
79. Cross-Disorder Group of the Psychiatric Genomics, C (2013) Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.*, **45**, 984–994.
80. Network and Pathway Analysis Subgroup of Psychiatric Genomics, C (2015) Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.*, **18**, 199–209.
81. Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R. et al. (2019) Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.*, **51**, 431–444.
82. LifeLines Cohort, S., Okbay, A., Baselmans, B.M.L., De Neve, J.-E., Turley, P., Nivard, M.G., Fontana, M.A., Meddens, S.F.W., Linnér, R.K., Rietveld, C.A. et al. (2016) Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.*, **48**, 624–633.
83. Okbay, A., Beauchamp, J.P., Fontana, M.A., Lee, J.J., Pers, T.H., Rietveld, C.A., Turley, P., Chen, G.-B., Emilsson, V., Meddens, S.F.W. et al. (2016) Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, **533**, 539–542.
84. International League Against Epilepsy Consortium on Complex, E (2014) Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. *Lancet Neurol.*, **13**, 893–903.
85. The Electronic Medical, R., Genomics, C., The, M.C., The, P.C., The LifeLines Cohort, S., Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H. et al. (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.*, **46**, 1173–1186.
86. Hibar, D.P., Adams, H.H.H., Jahanshad, N., Chauhan, G., Stein, J.L., Hofer, E., Renteria, M.E., Bis, J.C., Arias-Vasquez, A., Ikram, M.K. et al. (2017) Novel genetic loci associated with hippocampal volume. *Nat. Commun.*, **8**, 13624.
87. International Multiple Sclerosis Genetics, C., International, I.B.D.G.C., Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C. et al. (2015) Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.*, **47**, 979–986.
88. Hammerschlag, A.R., Stringer, S., de Leeuw, C.A., Sniekers, S., Taskesen, E., Watanabe, K., Blanken, T.F., Dekker, K., te Lindert, B.H.W., Wassing, R. et al. (2017) Genome-wide association analysis of insomnia complaints identifies risk genes and genetic overlap with psychiatric and metabolic traits. *Nat. Genet.*, **49**, 1584–1592.
89. Benke, K.S., Nivard, M.G., Velders, F.P., Walters, R.K., Pappa, I., Scheet, P.A., Xiao, X., Ehli, E.A., Palmer, L.J., Whitehouse, A.J.O. et al. (2014) A genome-wide association meta-analysis of preschool internalizing problems. *J. Am. Acad. Child Adolesc. Psychiatry*, **53**, 667–676 e667.
90. Consortium, C. (2015) Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*, **523**, 588–591.
91. Major Depressive Disorder Working Group of the Psychiatric, G.C (2013) A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry*, **18**, 497–511.
92. Andlauer, T.F.M., Buck, D., Antony, G., Bayas, A., Bechmann, L., Berthele, A., Chan, A., Gasperi, C., Gold, R., Graetz, C. et al. (2016) Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. *Sci. Adv.*, **2**, e1501678.
93. Tobacco and Genetics, C (2010) Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.*, **42**, 441–447.
94. Consortium, A., Consortium, U.K., Zheng, H.F., Forgetta, V., Hsu, Y.H., Estrada, K., Rosello-Diez, A., Leo, P.J., Dahia, C.L., Park-Min, K.H. et al. (2015) Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature*, **526**, 112–117.
95. International Parkinson's Disease Genomics, C., and Me Research, T, Chang, D., Nalls, M.A., Hallgrímsson, I.B., Hunkapiller, J., van der Brug, M., Cai, F., Kerchner, G.A., Ayalon, G. et al. (2017) A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat. Genet.*, **49**, 1511–1516.
96. Zhang, X.-J., Huang, W., Yang, S., Sun, L.-D., Zhang, F.-Y., Zhu, Q.-X., Zhang, F.-R., Zhang, C., Du, W.-H., Pu, X.-M. et al. (2009) Psoriasis genome-wide association study identifies susceptibility variants within LCE gene cluster at 1q21. *Nat. Genet.*, **41**, 205–210.
97. Consortium, G., Consortium, C., Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N.,

- Legge, S.E., Bishop, S. et al. (2018) Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.*, **50**, 381–389.
98. Schizophrenia Working Group of the Psychiatric Genomics, C (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, **511**, 421–427.
99. Go, M.J., Lee, Y., Park, S., Kwak, S.H., Kim, B.-J. and Lee, J. (2016) Genetic-risk assessment of GWAS-derived susceptibility loci for type 2 diabetes in a 10 year follow-up of a population-based cohort study. *J. Hum. Genet.*, **61**, 1009–1012.
100. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Peninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A. et al. (2016) Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.*, **48**, 245–252.
101. Klein, H.-U., McCabe, C., Gjonneska, E., Sullivan, S.E., Kaskow, B.J., Tang, A., Smith, R.V., Xu, J., Pfenning, A.R., Bernstein, B.E. et al. (2019) Epigenome-wide study uncovers large-scale changes in histone acetylation driven by tau pathology in aging and Alzheimer's human brains. *Nat. Neurosci.*, **22**, 37–46.
102. De Jager, P.L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L.C., Yu, L., Eaton, M.L., Keenan, B.T., Ernst, J., McCabe, C. et al. (2014) Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.*, **17**, 1156–1163.
103. Lunnon, K., Smith, R., Hannon, E., De Jager, P.L., Srivastava, G., Volta, M., Troakes, C., Al-Sarraj, S., Burrage, J., Macdonald, R. et al. (2014) Methyloic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat. Neurosci.*, **17**, 1164–1170.
104. van Dyck, C.H., Nygaard, H.B., Chen, K., Donohue, M.C., Raman, R., Rissman, R.A., Brewer, J.B., Koeppe, R.A., Chow, T.W., Rafii, M.S. et al. (2019) Effect of AZD0530 on cerebral metabolic decline in Alzheimer disease: a randomized clinical trial. *JAMA Neurol.*, **76**, 1219.
105. Moussa, C., Hebron, M., Huang, X., Ahn, J., Rissman, R.A., Aisen, P.S. and Turner, R.S. (2017) Resveratrol regulates neuro-inflammation and induces adaptive immunity in Alzheimer's disease. *J. Neuroinflammation*, **14**, 1.
106. Parikshak, N.N., Luo, R., Zhang, A., Won, H., Lowe, J.K., Chandran, V., Horvath, S. and Geschwind, D.H. (2013) Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*, **155**, 1008–1021.
107. Chandran, V., Coppola, G., Nawabi, H., Omura, T., Versano, R., Huebner, E.A., Zhang, A., Costigan, M., Yekkirala, A., Barrett, L. et al. (2016) A systems-level analysis of the peripheral nerve intrinsic axonal growth program. *Neuron*, **89**, 956–970.
108. Rao, J.S., Keleshian, V.L., Klein, S. and Rapoport, S.I. (2012) Epigenetic modifications in frontal cortex from Alzheimer's disease and bipolar disorder patients. *Transl. Psychiatry*, **2**, e132–e132.
109. Terai, K., Matsuo, A. and McGeer, P.L. (1996) Enhancement of immunoreactivity for NF- $\kappa$ B in the hippocampal formation and cerebral cortex of Alzheimer's disease. *Brain Res.*, **735**, 159–168.
110. Bales, K.R., Du, Y., Dodel, R.C., Yan, G.M., Hamilton-Byrd, E. and Paul, S.M. (1998) The NF-kappaB/Rel family of proteins mediates Abeta-induced neurotoxicity and glial activation. *Brain Res. Mol. Brain Res.*, **57**, 63–72.
111. Kaltschmidt, B., Uherek, M., Volk, B., Baeuerle, P.A. and Kaltschmidt, C. (1997) Transcription factor NF- $\kappa$ B is activated in primary neurons by amyloid peptides and in neurons surrounding early plaques from patients with Alzheimer disease. *Proc. Natl. Acad. Sci.*, **94**, 2642–2647.
112. Heyninck, K., Lahtela-Kakkonen, M., Van der Veken, P., Haegeman, G. and Vanden Berghe, W. (2014) Withaferin A inhibits NF-kappaB activation by targeting cysteine 179 in IKK $\beta$ . *Biochem. Pharmacol.*, **91**, 501–509.
113. Oh, J.H. and Kwon, T.K. (2009) Withaferin A inhibits tumor necrosis factor- $\alpha$ -induced expression of cell adhesion molecules by inactivation of Akt and NF- $\kappa$ B in human pulmonary epithelial cells. *Int. Immunopharmacol.*, **9**, 614–619.
114. Tiwari, S., Atluri, V.S.R., Yndart Arias, A., Jayant, R.D., Kaushik, A., Geiger, J. and Nair, M.N. (2018) Withaferin A suppresses beta amyloid in APP expressing cells: studies for tat and cocaine associated neurological dysfunctions. *Front. Aging Neurosci.*, **10**, 291.
115. Berchtold, N.C., Coleman, P.D., Cribbs, D.H., Rogers, J., Gillen, D.L. and Cotman, C.W. (2013) Synaptic genes are extensively downregulated across multiple brain regions in normal human aging and Alzheimer's disease. *Neurobiol. Aging*, **34**, 1653–1661.
116. Webster, J.A., Gibbs, J.R., Clarke, J., Ray, M., Zhang, W., Holmans, P., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M. et al. (2009) Genetic control of human brain transcript expression in Alzheimer disease. *Am. J. Hum. Genet.*, **84**, 445–458.
117. Araque Caballero, M.Á., Suárez-Calvet, M., Duering, M., Franzmeier, N., Benzinger, T., Fagan, A.M., Bateman, R.J., Jack, C.R., Levin, J., Dichgans, M. et al. (2018) White matter diffusion alterations precede symptom onset in autosomal dominant Alzheimer's disease. *Brain*, **141**, 3065–3080.
118. Nasrabady, S.E., Rizvi, B., Goldman, J.E. and Brickman, A.M. (2018) White matter changes in Alzheimer's disease: a focus on myelin and oligodendrocytes. *Acta Neuropathol. Commun.*, **6**, 22.
119. Cai, Z. and Xiao, M. (2016) Oligodendrocytes and Alzheimer's disease. *Int. J. Neurosci.*, **126**, 97–104.
120. Desai, M.K., Mastrangelo, M.A., Ryan, D.A., Sudol, K.L., Narrow, W.C. and Bowers, W.J. (2010) Early oligodendrocyte/myelin pathology in Alzheimer's disease mice constitutes a novel therapeutic target. *Am. J. Pathol.*, **177**, 1422–1435.
121. Dong, Y.-X., Zhang, H.-Y., Li, H.-Y., Liu, P.-H., Sui, Y. and Sun, X.-H. (2018) Association between Alzheimer's disease pathogenesis and early demyelination and oligodendrocyte dysfunction. *Neural Regen. Res.*, **13**, 908.
122. Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J.Z., Menon, M., He, L., Abdurrob, F., Jiang, X. et al. (2019) Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*, **570**, 332–37.
123. Sierksma, A., Lu, A., Mancuso, R., Fattorelli, N., Thrupp, N., Salta, E., Zoco, J., Blum, D., Buée, L., De Strooper, B. and Fiers, M. (2020) Novel Alzheimer risk genes determine the microglia response to amyloid- $\beta$  but not to TAU pathology. *EMBO Mol. Med.*, **12**, e10606.
124. Kannan, V., Brouwer, N., Hanisch, U.-K., Regen, T., Eggen, B.J.L. and Boddeke, H.W.G.M. (2013) Histone deacetylase inhibitors suppress immune activation in primary mouse microglia: HDACi inhibit microglia immune activation. *J. Neurosci. Res.*, **91**, 1133–1142.



125. Datta, M., Staszewski, O., Raschi, E., Frosch, M., Hagemeyer, N., Tay, T.L., Blank, T., Kreutzfeldt, M., Merkler, D., Ziegler-Waldkirch, S. et al. (2018) Histone deacetylases 1 and 2 regulate microglia function during development, homeostasis, and neurodegeneration in a context-dependent manner. *Immunity*, **48**, 514–529 e516.
126. Gibbs, J.R., van der Brug, M.P., Hernandez, D.G., Traynor, B.J., Nalls, M.A., Lai, S.-L., Arepalli, S., Dillman, A., Rafferty, I.P., Troncoso, J. et al. (2010) Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet.*, **6**, e1000952.
127. Kumar, A., Gibbs, J.R., Beilina, A., Dillman, A., Kumaran, R., Trabzuni, D., Rytten, M., Walker, R., Smith, C., Traynor, B.J. et al. (2013) Age-associated changes in gene expression in human brain and isolated neurons. *Neurobiol. Aging*, **34**, 1199–1209.
128. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A. et al. (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.*, **44**, W90–W97.
129. Rossin, E.J., Lage, K., Raychaudhuri, S., Xavier, R.J., Tatar, D., Benita, Y., International Inflammatory Bowel Disease Genetics, C., Cotsapas, C. and Daly, M.J. (2011) Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.*, **7**, e1001273.
130. Langfelder, P., Luo, R., Oldham, M.C. and Horvath, S. (2011) Is my network module preserved and reproducible? *PLoS Comput. Biol.*, **7**, e1001057.
131. Zhou, X., Carbonetto, P. and Stephens, M. (2013) Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet.*, **9**, e1003264.
132. Habib, N., Avraham-Davidi, I., Basu, A., Burks, T., Shekhar, K., Hofree, M., Choudhury, S.R., Aguet, F., Gelfand, E., Ardlie, K. et al. (2017) Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods*, **14**, 955–958.
133. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R. (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.*, **36**, 411–420.
134. Linderman, G.C., Rachh, M., Hoskins, J.G., Steinerberger, S. and Kluger, Y. (2019) Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data. *Nat. Methods*, **16**, 243–245.
135. Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M. et al. (2015) MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.*, **16**, 278.
136. Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T.W., Greven, M.C., Pierce, B.G., Dong, X., Kundaje, A., Cheng, Y. et al. (2012) Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.*, **22**, 1798–1812.
137. Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R. and Ma'ayan, A. (2010) ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics*, **26**, 2438–2444.