

DISEASES AND DISORDERS

Molecular subtyping of Alzheimer's disease using RNA sequencing data reveals novel mechanisms and targets

Ryan A. Neff^{1,2,3,4}, Minghui Wang^{1,2,3}, Sezen Vatansever^{1,2,3}, Lei Guo^{1,2,3}, Chen Ming^{1,2,3}, Qian Wang^{1,2,3}, Erming Wang^{1,2,3}, Emrin Horgusluoglu-Moloch^{1,2,3}, Won-min Song^{1,2,3}, Aiqun Li^{1,3,5,6,7}, Emilie L. Castranio¹¹, Julia TCW^{5,8}, Lap Ho^{1,2,3}, Alison Goate^{1,3,5,6,8,10,11}, Valentina Fossati⁹, Scott Noggle⁹, Sam Gandy^{7,10,11}, Michelle E. Ehrlich^{10,12}, Pavel Katsel^{7,15}, Eric Schadt^{1,3}, Dongming Cai^{10,13}, Kristen J. Brennand^{1,3,5,6}, Vahram Haroutunian^{5,6,11,14,15}, Bin Zhang^{1,2,3,16*}

Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

Alzheimer's disease (AD), the most common form of dementia, is recognized as a heterogeneous disease with diverse pathophysiologic mechanisms. In this study, we interrogate the molecular heterogeneity of AD by analyzing 1543 transcriptomes across five brain regions in two AD cohorts using an integrative network approach. We identify three major molecular subtypes of AD corresponding to different combinations of multiple dysregulated pathways, such as susceptibility to tau-mediated neurodegeneration, amyloid- β neuroinflammation, synaptic signaling, immune activity, mitochondria organization, and myelination. Multiscale network analysis reveals subtype-specific drivers such as *GABRB2*, *LRP10*, *MSN*, *PLP1*, and *ATP6V1A*. We further demonstrate that variations between existing AD mouse models recapitulate a certain degree of subtype heterogeneity, which may partially explain why a vast majority of drugs that succeeded in specific mouse models do not align with generalized human trials across all AD subtypes. Therefore, subtyping patients with AD is a critical step toward precision medicine for this devastating disease.

INTRODUCTION

Alzheimer's disease (AD) is the most common form of dementia in the elderly, estimated to affect more than 5.8 million individuals in the United States and more than 50 million worldwide, with almost half of individuals aged over 75 years (1). AD is primarily characterized by progressive neurological decline, especially selectively targeted memory loss and cognitive dysfunction (2). The neuropathological manifestations of AD start long before apparent cognitive symptoms, however. It is traditionally understood that these include the accumulation of amyloid-beta ($A\beta$) peptide as extracellular plaques and hyperphosphorylated tau as intracellular neurofibrillary tangles (NFTs), typically identified on postmortem biopsy and used for definitive AD diagnosis. Both $A\beta$ and NFT accumulation typically

progress to targeted neuronal and synaptic loss, mainly in regions of the cerebral cortex and the hippocampus. Concurrent with the neuronal loss in AD, there is an additional coordinated breakdown across other brain cell types such as gliosis, demyelination, and inflammation that exacerbates cognitive dysfunction (3).

However, an increasingly growing body of evidence has demonstrated that AD is a heterogeneous disease caused by various pathophysiologic mechanisms beyond the typical dogma. For instance, up to one-third of patients with a clinical diagnosis of AD have no accumulation of $A\beta$ (4), while many of those diagnosed with AD at postmortem biopsy do not show cognitive impairment (5). Among those at risk of sporadic late-onset AD (LOAD), patients may carry a unique set of numerous genetic changes with greater risk for developing the disease, including clusterin (*CLU*), triggering receptor expressed on myeloid cells 2 (*TREM2*), and apolipoprotein E (*APOE*) variants (6). The interaction between specific LOAD risk alleles and changes in disease pathogenesis, however, remains elusive.

Furthermore, it is very challenging to predict the progression of AD, suggesting high heterogeneity in disease progression among patients with AD. There is growing evidence that disease progression and responses to interventions differ substantially within LOAD. For instance, patients with LOAD often branch into distinct groups including (i) slow versus rapid cognitive decliners (7), (ii) amnesic versus nonamnesic AD (8), (iii) executive versus cortical visual defect versus dysphasia-predominant AD (9), (iv) psychosis and depression-associated AD (10–12), and (v) metabolic dysfunction-associated AD modulated by abnormalities in insulin resistance, hormonal deficiencies, or homocysteinemia (13). Last, the relationship between the various forms of AD and other non-AD dementias such as primary age-related tauopathy (14), vascular contributions to cognitive impairment and dementia (15), and frontotemporal dementia (16) must be better understood. Therefore, identifying unique molecular subtypes of AD resistant to other comorbid

¹Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA. ²Mount Sinai Center for Transformative Disease Modeling, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA. ³Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA. ⁴Medical Scientist Training Program, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. ⁵Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, New York, NY 10029, USA. ⁶Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, New York, NY 10029, USA. ⁷Department of Psychiatry, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, New York, NY 10029, USA. ⁸Ronald M. Loeb Center for Alzheimer's Disease, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. ⁹New York Stem Cell Foundation Research Institute, New York, NY 10019, USA. ¹⁰Department of Neurology, Icahn School of Medicine at Mount Sinai, New York NY 10029, USA. ¹¹Alzheimer's Disease Research Center, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. ¹²Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. ¹³Neurology, JJ Peters VA Medical Center, 130 West Kingsbridge Road, Bronx, NY 10468, USA. ¹⁴Mental Illness Research, Education and Clinical Centers, JJ Peters VA Medical Center, 130 West Kingsbridge Road, Bronx, NY 10468, USA. ¹⁵Psychiatry, JJ Peters VA Medical Center, 130 West Kingsbridge Road, Bronx, NY 10468, USA. ¹⁶Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA.

*Corresponding author. Email: bin.zhang@mssm.edu

conditions may provide new insights into AD patient subpopulations and pave the way for precision medicines for AD.

Molecular biomarkers may hold the promise for improving methodologies for AD subtype identification and classification (17, 18). Some recent studies have highlighted the great advantages of using RNA sequencing (RNA-seq) to profile the transcriptome of the brain with neurodegenerative diseases. For instance, a multiomic molecular analysis of LOAD across four brain regions uncovered subnetworks and previously unidentified molecular drivers of the disease, including the vacuolar adenosine 5'-triphosphate (ATP)-dependent proton pump *ATP6V1A*, which are now shown to modulate cognitive function in *Drosophila* models of AD (19). In addition, molecular network analysis of LOAD brains has identified an excess of dysregulated genes that cannot be fully predicted by a single model of the disease (19). Nevertheless, only a limited number of published papers describe RNA-seq studies of the most relevant material, namely, human AD brains across multiple regions (20). Therefore, it is highly likely that specific subtype signatures of AD exist across large transcriptomic studies that have not yet been identified.

In this study, we investigate molecular subtypes of AD in the Mount Sinai/JJ Peters VA Medical Center Brain Bank (MSBB-AD) (21) study and replicate these findings in the Religious Orders Study–Memory and Aging Project (ROSMAP) (22, 23). The MSBB-AD cohort is a critical resource to understand postmortem molecular changes leading to regional vulnerability as the samples were extracted from four brain regions, including the frontal pole (FP), the superior temporal gyrus (STG), the parahippocampal gyrus (PHG), and the inferior frontal gyrus (IFG), of the same set of human. Understanding the genetic and molecular differences between molecular subtypes of AD within these data will provide novel insights into disease pathogenesis and offer new avenues for developing effective therapeutics.

RESULTS

The PHG carries the strongest molecular signal of LOAD

We obtained clinical and transcriptomic signatures from the MSBB-AD study of 364 human brains, including whole-transcriptome RNA-seq data from four brain regions (FP, STG, PHG, and IFG) from participants with AD that showed neurocognitive decline as measured by clinical dementia rating (CDR) score > 1 and nondemented controls (CDR = 0). An overview of the cohorts, methods, and analyses used in this study is provided in fig. S1. In addition, table S1 summarizes the clinical and pathologic phenotypes for the samples in the MSBB-AD cohort with the transcriptomic data from the PHG. We identified several confounding factors in the RNA-seq data, including age of death, race, and postmortem interval (PMI). To minimize reidentification of clinical and technical covariates (see Materials and Methods), we first correct the transcriptomic data for age of death, race, gender, PMI, batch number, and RNA integrity number (RIN) using a mixed-effects model.

To understand which brain regions and molecular processes are most vulnerable to dysregulation in AD, we perform differential gene expression analysis between AD and control, generating differentially expressed genes (DEGs) for each of the four brain regions in the MSBB-AD (see Materials and Methods). The PHG brain region has the largest number of DEGs [3571 genes; adjusted false discovery rate (FDR), 0.05] compared to the FP (3 genes), STG (1 gene), and IFG (181 genes). The results are consistent with previ-

ous DEG analyses (19) of the MSBB-AD transcriptomic data and suggest that the PHG is most vulnerable in AD as manifested by marked transcriptomic dysregulation. In addition, these findings are consistent with our previous pan-cortical atlas of AD (independent of the data described here), in which the PHG brain region showed the most substantial transcriptomic changes (24). Previous work has shown that the hippocampus is strongly associated with A β and tau accumulation and early memory loss in AD (24). We have also shown in previous studies of the MSBB-AD cohort that transcriptomic changes in the PHG region highly overlap the Kyoto Encyclopedia of Genes and Genomes (KEGG) (25) Alzheimer's and Parkinson's disease gene sets and are correlated with high A β plaque density (19), demonstrating that these changes are consistent with AD disease progression. Therefore, the PHG region carries the strongest molecular signature of AD.

Normalization of data by AD stages removes confounding signal of neuronal loss

Patients with more severe AD-associated dementia, such as those in a later stage of the disease, have been reported to have more neuronal loss at postmortem biopsy (3). Therefore, it is important to control for AD stage before transcriptomic analysis is performed between AD participants. Previous work by our laboratory has shown that brain cell-type surrogate proportion values (SPVs), including the proportion of neurons in a sample, can be inferred from bulk RNA-seq data when combined with measurements of brain cell type-specific gene expression patterns (26) and may serve as a marker of neuronal loss in AD.

To determine the extent of hippocampal neuronal loss in the MSBB-AD cohort, we perform cell-type proportion analysis (Materials and Methods) of bulk tissue transcriptomic data from the PHG region. As shown in fig. S2A, there is a strong relationship between PHG neuronal loss and CDR score in the MSBB-AD cohort, with astrogliosis and increased abundances of other cell types associated with disease progression. Currently, there is no universally accepted method to rectify for neuronal loss seen in AD, as this reduction in neurons is both the cause and effect of molecular changes leading to cognitive impairment. We examine both normalization by neuronal cell-type proportion SPVs and AD dementia (as measured by CDR score, range 0 to 6) to reduce the molecular signature of neuronal loss in AD, using a linear effect model (Materials and Methods). We show that after either normalization by neuronal cell-type proportion or dementia severity (fig. S2, A and B), there is no remaining significant correlation between cell type and AD stage. Furthermore, the remaining correlation after normalization by dementia severity is not reduced further with additional normalization for neuronal cell type, as shown in fig. S2C. Thus, both normalization by CDR and cell-type proportion SPVs effectively remove the confounding effects of neuronal loss along disease progression in the MSBB-AD cohort.

We further normalize the PHG transcriptomic data by CDR score, to remove the confounding effects of neuronal loss in later AD stages. Therefore, any identified differences between groups of AD participants should be distinct from previously identified clinical subtypes of AD that rely on these metrics.

Identification of AD putative subtypes from molecular data of the PHG

We seek to determine whether subgroups of AD participants demonstrate molecular and clinical differences. To robustly identify

subgroups of AD participants, we evaluate the performance of several sample clustering methods for AD participants using the normalized gene expression data within each brain region in the MSBB-AD cohort. We choose two classical clustering algorithms (hierarchical, *k*-means) as well as two novel network-based clustering algorithms [weighted sample gene network analysis (WSCNA) and multiscale embedded gene expression network analysis (MEGENA)] to group similar samples together into putative AD subtypes. WSCNA shows the best performance in terms of clustering quality and thus is adopted to identify AD subtypes for the subsequent analyses.

As shown in Fig. 1, WSCNA identifies five subtypes in the MSBB-AD (clusters A, B1, B2, C1, and C2) across all 151 participants with PHG transcriptomic data. On the basis of the dendrogram and the network similarity heatmap (Fig. 1, A and B), these five subtypes can be further grouped into three major classes of AD-labeled class A, class B (comprising subtypes B1 and B2), and class C (comprising subtypes C1 and C2). Each class of subtypes has a similar number of samples (47 in class A, 54 in class B, and 50 in class C) (Fig. 1C).

Cluster stability is defined here as the rate at which sample pairs group together into the same subtypes upon repeated re-clustering on a random subset of the input data (Materials and Methods). Subtypes from WSCNA clustering are generally stable (fig. S4), and sample pairs grouped together, on average, between 60 and 91% across all five detected AD subtypes. The class C subtypes have the strongest stability, followed by class A and class B; all subtypes demonstrate a cluster stability strongly above random clusters, which was empirically determined at a stability range between 20 and 30%. Therefore, these subtypes show specific robust molecular signals suitable for classification into stable subtypes.

Molecular signatures of putative AD subtypes

To characterize the molecular signatures of these AD subtypes, we identify DEGs for each of the five subtypes compared with nondemented controls (CDR = 0) from the RNA-seq data in the PHG region. The identified AD subtype molecular signatures are provided in data file S1. Each AD subtype has a specific transcriptomic signature of up- and down-regulated genes that distinguishes it from the others at a molecular level, suggesting a plurality of different mechanisms of AD. As shown by Fig. 1D, there is a clear separation of molecular signatures between the five AD subtypes that can be visually appreciated from the whole-transcriptome gene expression heatmap after identifying gene modules using weighted interaction network analysis (WINA) (24, 27).

Using mean gene expression levels grouped by Gene Ontology (GO) pathway as surrogate markers for the activity level of various molecular processes in the brain, we identify a number of differences in key AD-related pathways between the subtypes, providing insights into disease pathogenesis. As shown in Fig. 1E, we highlight significant deviations (Welsh's $P < 0.05$) of 74 AD-related signatures from previous studies and GO pathways (17, 28), including pathways related to A β , oxidative stress, tau NFT, and synaptic function, across the five AD subtypes and controls. First, we compare our subtype molecular signatures with the postmortem hippocampal transcriptional signatures of AD by Blalock *et al.* (28) (termed Blalock signatures). Overall, molecular signatures from all AD participants are consistent with Blalock; however, this consistency is not shared between each of the individual AD subtypes. We

observe that the direction of the gene expression changes in the class C subtypes is consistent with the Blalock signatures, while the changes in the class A subtype are opposite to Blalock. In contrast, the signature of the class B subtypes does not show significant enrichment of the Blalock signatures. Therefore, AD subtypes may be classified into three larger classes (classes A, B, and C), i.e., typical (class C), intermediate (class B), or atypical (class A), by molecular presentation when compared to the Blalock signatures of AD.

We observe only weak molecular enrichment of A β - and tau-related pathways across all AD participants as a group, but strong enrichment of these pathways in the subtypes. For instance, we see strong up-regulation of A β binding, clearance, and fiber formation pathways in subtype C1 and scavenger receptor activity in subtypes C1 and C2, while these same pathways are down-regulated in subtypes A and B1. On the other hand, tau-neighborhood genes ("GNF2_MAPT" pathway) are strongly up-regulated in subtypes A, B1, and B2 but down-regulated in C1 and C2. Tau protein binding and tau-related P35 pathway genes are up-regulated in the subtype A. Therefore, it is likely that AD subtypes may be characterized by either A β activity predominant (class C) or microtubule associated protein tau (MAPT)-activity predominant (class A + B) although they cannot fully explain all differences seen between the five subtypes.

The predicted subtypes also differ strongly in neuronal activity despite normalization for AD staging. We see broad down-regulation of glutamergic, γ -aminobutyric acid (GABA) related, glycinergic, and dendritic synaptic pathways in class C subtypes, with absent changes in cholinergic and dopaminergic synaptic pathways, suggesting that these synapse types are selectively resilient to AD subtype molecular changes. On the other hand, we observe strong up-regulation of these same synapse pathways in classes A and B, with the exception of up-regulation of glycinergic synapse in class A. This pattern is consistent with differences in synaptic excitation pathways between subtypes: Excitatory synapses are up-regulated in classes A and B but down-regulated in class C. These data suggest that AD subtypes may be split into those selectively vulnerable to synaptic depression (class C) versus synaptic excitation (classes A and B).

Dysregulated immune system activities, including reactive gliosis and the breakdown of the blood-brain barrier, have been repeatedly observed in AD brains (29). In subtypes B2, C2, and especially C1, immune-related pathways including the innate and adaptive immune response, immune system activation, inflammation, circulatory system development, and endothelial cell migration are up-regulated compared with normal controls. Such up-regulation coincides with increased expression of blood-brain barrier, basement membrane, and cell matrix adhesion genes. However, these immune response pathways are down-regulated in subtypes A and B1. These data and the findings relative to synaptic pathways suggest that disease progression across AD subtypes is characterized by either increased immune or synapse pathway activity.

Last, certain molecular pathways are subtype specific and thus provide greater insights into disease pathogenesis when considering other enriched AD pathways. For example, many protein degradation-related genes, including ubiquitination and polyubiquitination, protein catabolism, the proteasome, and proteins targeting for destruction, are up-regulated in subtype A while organic acid-related genes, including acid secretion and acidic amino acid transport, are specifically

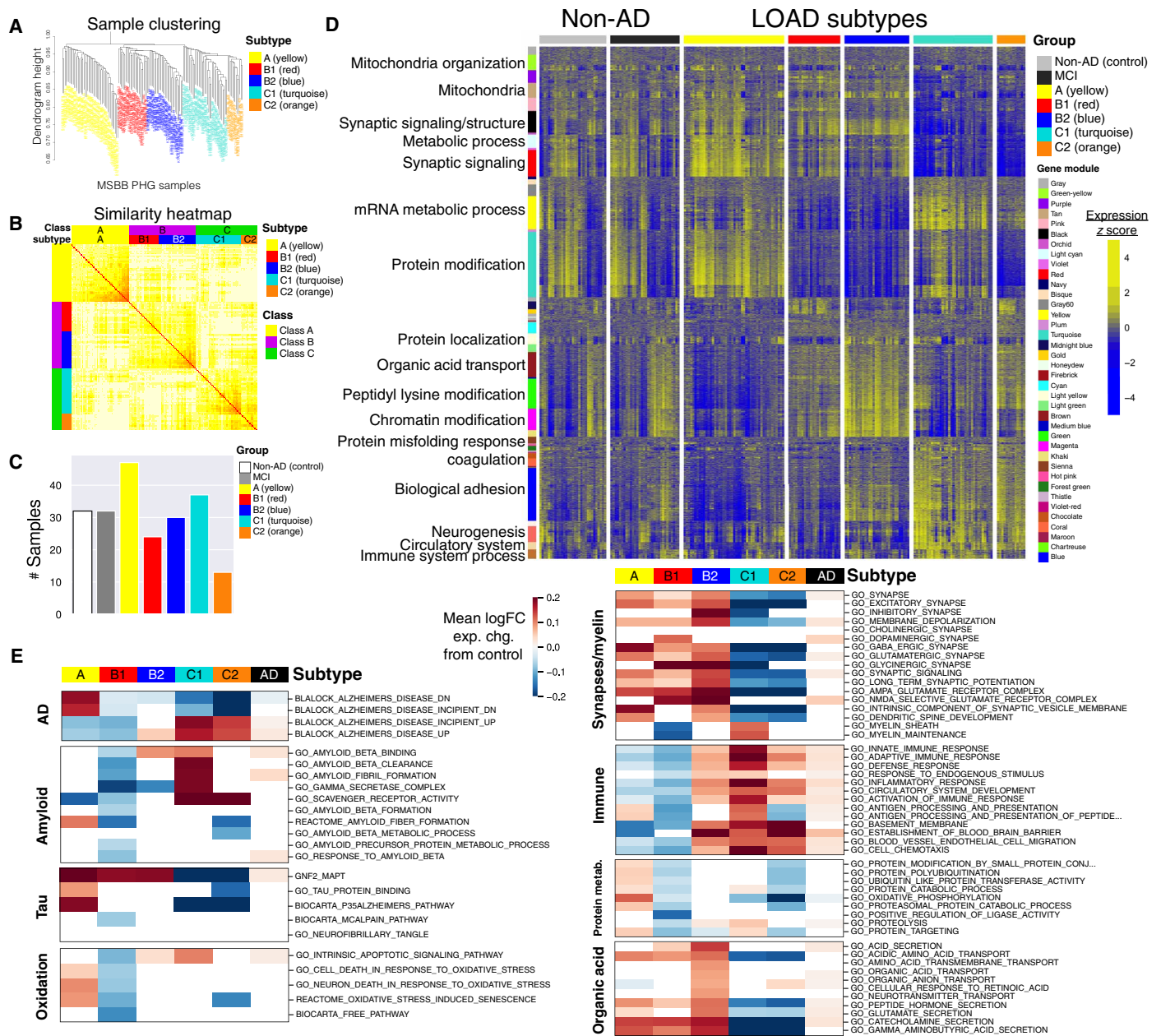


Fig. 1. Identification of five stable molecular subtypes of AD. (A and B) WSCNA clustering dendrogram and topological overlap matrix (TOM) heatmap, showing three major classes (A, B, and C) and five subtypes annotated as A, B1, B2, C1, and C2, corresponding to the yellow, red, blue, turquoise, and orange clusters, respectively. (C) Number of samples in each subtype, control (CDR = 0) and mild cognitive impairment (MCI) (CDR = 0.5). (D) Gene expression profiles of all the samples in the PHG from the MSBB-AD cohort. The samples on the columns are grouped by subtype, and the genes on the rows are grouped by WINA module. FC, fold change. (E) Change in mean expression level of various gene pathways for each AD subtype in comparison with the normal control samples. AD-related pathways, representing differential expression from previous AD studies, are derived from the MSigDB. Sets are grouped by major area of biological activity.

up-regulated in class B. Defects in the acidification of lysosomes in neurons have been previously associated with neuronal and synaptic loss in AD, as well as decreases in long-term potentiation (19).

Association of clinical and pathological phenotypes and APOE variants with putative AD subtypes

To understand the clinical characteristics of these molecularly defined AD subtypes, we examine the relationship between charac-

terized AD pathologic markers in the MSBB-AD study and each subtype. As shown in Fig. 2 (B and C), under the Kruskal-Wallis (KW) one-way analysis of variance (ANOVA) test, the AD subtypes are marginally associated with several clinical AD markers, including tau NFT levels in the medial frontal cortex (KW, $P = 0.041$), A β mean plaque density (KW, $P = 0.020$), and APOE e4 (KW, $P = 0.048$) and APOE e2 (KW, $P = 0.012$) allele counts. The “A β predominant” AD subtypes (class C), with a mean plaque density of 14.2/mm²,

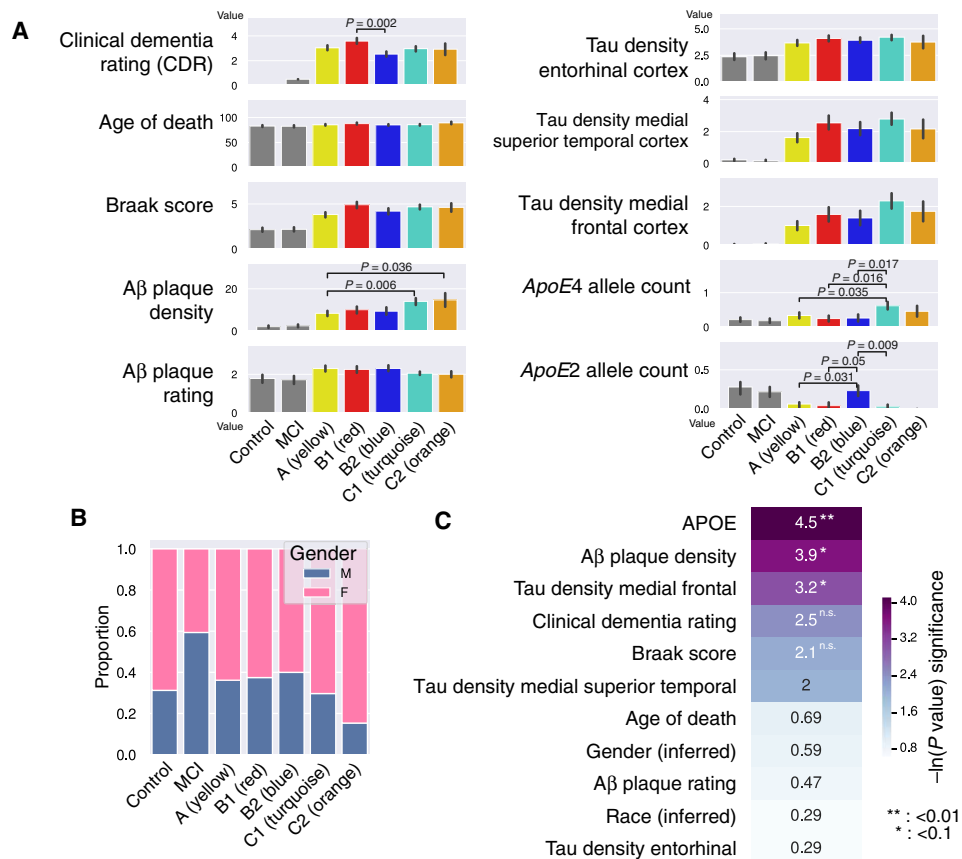


Fig. 2. Mean values of several clinical and pathologic traits across AD subtypes. (A) Bar plots of mean CDR, Braak score, Aβ plaque density, tau NFT densities (measured in the entorhinal cortex, the medial superior temporal cortex, and the medial frontal cortex), APOE4 allele count, and APOE2 allele count across five subtypes, control, and MCI. (B) Stacked bar chart of inferred biological sex from transcriptomic data for all the PHG samples, across five subtypes, control, and MCI. (C) Natural log-transformed P values from the KW ANOVA test of clinical, pathologic, and demographic variables. Significant tests are greater than ~3.0, which corresponds to an α of 0.05. n.s., not significant.

show a significantly larger Aβ plaque burden than both of the “tau-predominant” class A (mean, 8.4/mm²; Welsh’s *t* test, $P = 3.1 \times 10^{-3}$) and the class B (mean, 9.6/mm²; Welsh’s *t* test, $P = 0.018$) subtypes, despite no significant difference in cognitive decline as measured by CDR. Consistent with the preprocessing steps already performed on the PHG data, the subtypes do not show significant differences in previously corrected covariates. We do not see significant changes in CDR score (KW, $P = 0.082$; Fig. 2C), biological sex (KW, $P = 0.554$; Fig. 2, B and C), ethnicity (KW, $P = 0.748$), PMI (KW, $P = 0.502$), or age of death (KW, $P = 0.503$) across AD subtypes.

We also seek to better understand differences in APOE allele dosages between AD subtypes. We find that certain subtypes are preferentially enriched or depleted for the e4 and e2 alleles compared with others. For instance, the subtype C1 has a significantly increased APOE e4 allele dosage (median, 0.61 alleles per participant) compared with the subtypes A ($P = 0.035$ under Welsh’s *t* test), B1 ($P = 0.015$), and B2 ($P = 0.017$). This is consistent with the known influence of the e4 allele on AD pathogenesis, including the formation of Aβ plaques and NFTs, most similar to the molecular signature of the Aβ-predominant subtypes. On the other hand, subtype C2, which shares many molecular features with C1, does not have this association with APOE e4. Furthermore, subtype B2 has an increased APOE e2 allele dosage (median, 0.23 alleles per partici-

pant) compared with subtype A ($P = 0.031$) and C1 ($P = 0.0091$); however, similar to APOE e4 among the class C subtypes, the APOE e2 dosage is also much higher than subtype B1 ($P = 0.049$). Therefore, while APOE may modulate AD pathogenesis and contribute to some molecular signatures in a portion of subtypes, we show that APOE dosage is insufficient to fully explain all of the molecular similarities and differences between both related and distinct AD subtypes.

A subset of postmortem Alzheimer’s brains with PHG transcriptomic data available ($n = 55$ of 151) has additional quantification of Aβ plaque and tau NFT amounts across multiple brain regions, as shown in fig. S6. We observed that tau NFT counts are significantly associated with the AD subtypes across the inferior parietal lobule (KW test, $P = 0.017$) and medial frontal gyrus (KW, $P = 0.034$). In these regions, both the class B and C subtypes have increased tau NFT burden. In contrast, Aβ plaque rating is significantly elevated in the inferior parietal lobule (KW, $P = 0.031$), medial frontal gyrus (KW, $P = 0.041$), and lateral frontal gyrus (KW, $P = 0.012$) in only the class C (Aβ-predominant) subtypes. These results are consistent with the previous signatures from the GO pathway analysis, indicating that class C subtypes are Aβ predominant, while class B subtypes are tau NFT predominant. It is likely that while class A shows increased MAPT pathway activity, it is resilient to the development

of tau NFTs, perhaps via increased protein degradation pathway activity. Therefore, these results suggest that class A subtypes are tau NFT resilient. As expected from the analysis on all samples, both CDR (KW, $P = 0.155$) and Braak score (KW, $P = 0.075$) are not associated with the AD subtypes. Therefore, AD staging is not associated with the changes in A β plaque and tau NFT levels in the subtypes.

Key network regulators of putative AD subtypes

The diverse molecular changes in the AD subtypes suggest distinct intrinsic molecular mechanisms underlying each subtype. To identify potential key regulators of the molecular changes in each AD subtype, we use a network biology approach that integrates multi-scale embedded gene coexpression network analysis (MEGENA) (30) and Bayesian causal network (BN) (31) inference (19, 24, 27, 32). Toward this end, we construct a coexpression network on the basis of all the AD samples in the PHG, which includes 22,291 genes and 61,152 edges, and a BN comprised of 21,577 genes and 23,554 edges. Key driver analysis (KDA; Materials and Methods) of each resulting network and the subtype DEG signatures identifies a ranked list of 955 up-regulated and 639 down-regulated key network regulator genes (KNRs) in the MEGENA network and a ranked list of 1226 up-regulated and 846 down-regulated KNRs in the BN. Last, the intersection of the BN and MEGENA network KNRs yields a subset of 233 up- and 164 down-regulated KNRs across the five subtypes (table S2). The full list of AD subtype molecular signatures, MEGENA and BN KNRs, as well as their intersection and combined P values under the aggregated Cauchy association test, is provided in data file S1.

Figure 3 (A and B) shows the up- and down-regulated key drivers of each subtype in the context of the MEGENA network, while Fig. 3 (C and D) shows the top 20 up- and down-regulated KNRs per subtype in the MEGENA network and the BN. Even the subtypes within each class (e.g., B and C) have subtype-specific drivers. As shown in Fig. 3 (A and B), each subtype's key drivers fall onto separate parts of the MEGENA network, suggesting that specific gene modules are subtype specific, and subtypes may be driven by a specific, yet diverse set of disease mechanisms that lead to AD. For instance, a large number of neuronal genes located at the center of the global network are down-regulated in the A β -predominant AD subtypes (C1 and C2) and are predicted to be pathogenic in AD. These down-regulated neuronal genes are predicted to be regulated by KNRs *GABRB2* (BN, $P = 7.2 \times 10^{-44}$), *SYT1* ($P = 3.6 \times 10^{-30}$), *ATP6V1A* ($P = 1.81 \times 10^{-27}$), and *SCN2A* ($P = 4.07 \times 10^{-116}$) in both models. On the other hand, the top right of the MEGENA network consists of many oligodendrocytic genes that are down-regulated in the class B subtypes and are predicted to be regulated by *PLP1* (BN, $P = 1.05 \times 10^{-14}$), *Ermin* (*ERMN*) ($P = 1.51 \times 10^{-32}$), *Quaking* (*QKI*) ($P = 6.95 \times 10^{-30}$), and *Stromal antigen 2* (*STAG2*) (8.89×10^{-27}) in both models. Last, the bottom right of the network is enriched for several down-regulated microglial, endothelial, and astrocytic genes that are driven by *LRP10* (BN $P = 2.15 \times 10^{-8}$), *TLN1* ($P = 3.71 \times 10^{-8}$), *LAMB2* ($P = 2.2 \times 10^{-8}$), *MYO1C* ($P = 5.2 \times 10^{-3}$), and *NOTCH1* ($P = 4.6 \times 10^{-12}$). Consistent with the observation that the AD subtypes in the classes A and C show opposite gene expression changes in known AD-associated gene signatures, many up-regulated neuronal KNRs (*GABRB2*, *LRP10*, *SYT1*, and *PREPL*) in the class A subtype are down-regulated in class C. Therefore, it is likely that both the class A and class C subtypes result from either inhibitory or excitatory dysregulation along a single axis in specific neuronal processes.

Last, we sought to investigate whether these subtype KNRs in the PHG are also key regulators in other brain regions in the MSBB-AD cohort. The same network analysis procedure is applied to the other regions, including the FP, IFG, and STG. As shown in Fig. 3E, many neuronal genes, including the KNRs *SCN2A*, *GABRB2*, *PLP1*, and *UGT8*, have a consistent direction (up- or down-regulation) in two or more brain regions in the MAPT pathway and A β -predominant AD subtypes. Furthermore, some key oligodendrocytic genes such as *PLP1* and *UGT8* remain as key network regulators for the class B subtypes. Therefore, although the PHG shows the greatest vulnerability, the subtype key regulators have consistent dysregulation in all the regions studied here.

Cell-type specificity of subtype molecular signatures

As previously discussed, to better understand brain cell-type specificity of transcriptomic changes in each AD subtype, we perform cell-type proportion analysis (Materials and Methods) on each sample using the brain cell-type marker signatures determined by cell type-specific sequencing previously conducted by our laboratory (26). Figure 4A shows significant and unique changes in the cell-type composition in the AD subtypes. For instance, class A has a small increase in neurons combined with a decrease of oligodendrocyte precursor cells (OPCs), astrocytes, and endothelial cells. On the other hand, class C has a substantial loss of neurons accompanied by an increase of oligodendrocytes, astrocytes, OPCs, and endothelial cells, opposite of the changes observed in class A. The class B subtypes have mixed changes in other cell types, but both have a small to moderate decrease in oligodendrocytes, consistent with the oligodendrocytic key regulators found through network analysis. These patterns are consistent with the reactive astrogliosis and microgliosis commonly seen in some patients with AD with immune system activation in response to misfolded or polymerized amyloid A β (3), which is indicated by the presence of inflammatory markers.

Next, using RNA-seq data derived from cultured brain cells (including neurons, astrocytes, microglia, endothelial cells, and oligodendrocytes), we examine the cell-type specificity of the key regulator genes of each AD subtype (Materials and Methods). As shown in Fig. 4 (B to F), we find that for the class C ("A β -predominant") subtypes, KNRs are up-regulated in microglia (*TLN1*, *MSN*, and *IL6R*), endothelial cells (*TAGLN2*), and astrocytes (*LRP10*, *GNA12*, and *LTBP3*) and down-regulated in neurons (*ATP6V1A*, *SCN2A*, *GABRB2*, and *NAPB*), consistent with neuroinflammatory destruction. On the other hand, the class A subtype has up-regulated neuronal regulators (*GABRB2*, *SYT1*, and *SCN2A*), indicating increased neuron remodeling and activity, and down-regulated KNRs in astrocytes, endothelial cells, and microglia (*LRP10*, *NOTCH1*, *MYO1C*, and *TLN1*). Only the class B shows marked down-regulation of genes in oligodendrocytes (*PLP1*, *UGT8*, *CLDND1*, *ERMN*, and *ENPP2*) with up-regulation in other cell types, suggesting that a demyelinating process may be relevant in this class.

Enrichment of known AD genetic risk markers with specific subtypes

To identify the influence of genetic determinants on AD subtype, we investigate the differences in polygenic risk score (PRS) between the predicted AD subtypes. We first compute each sample's PRS against the genome wide association study (GWAS) meta-analysis of Kunkle *et al.* (33) across AD using the PRSice R package (34). As non-European samples are excluded from the meta-analysis,

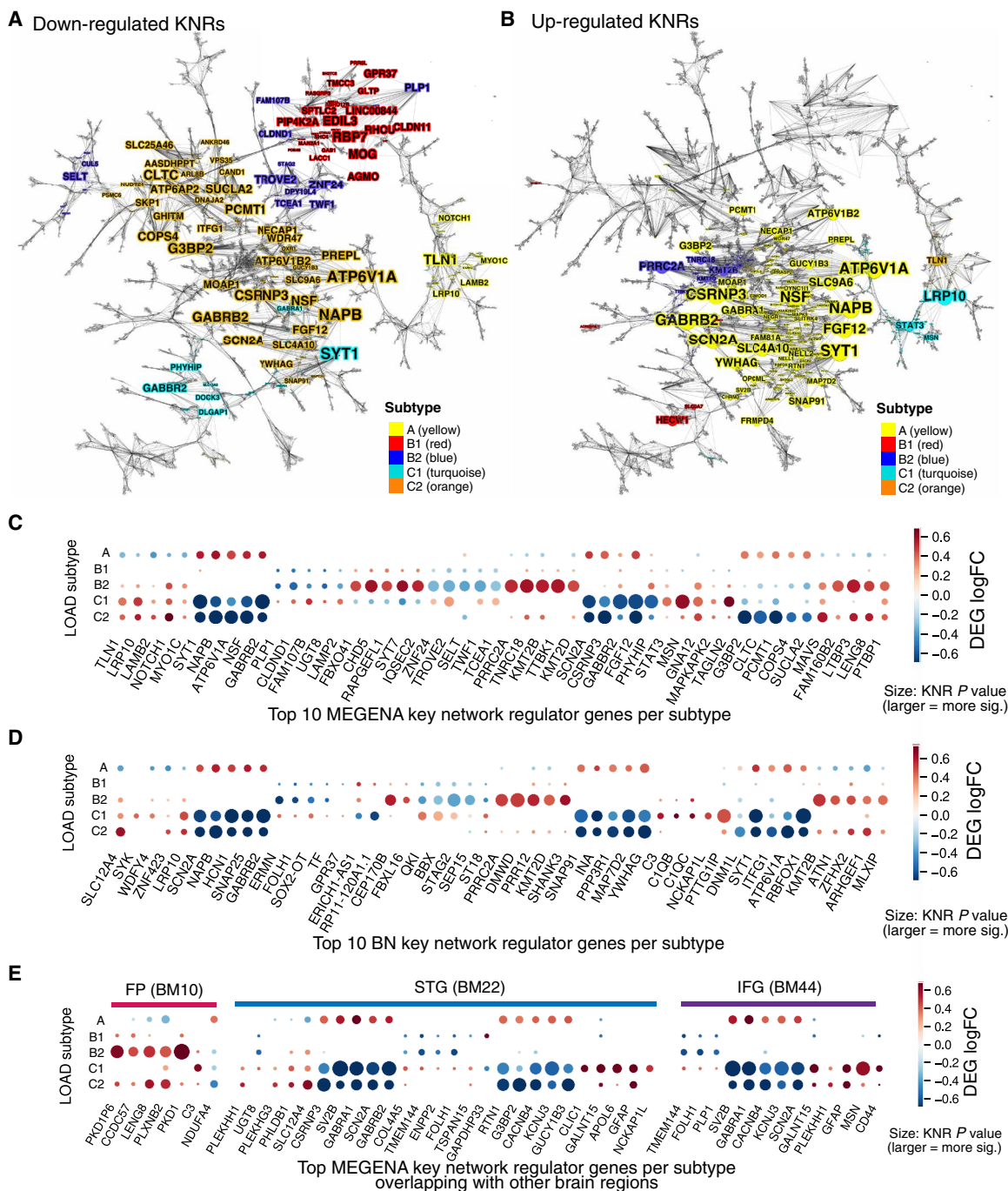


Fig. 3. MEGENA and BN-based key drivers of the AD subtypes. (A and B) Top down- and up-regulated MEGENA key drivers for each subtype plotted in its location in the network. Color of a node represents subtype (ties resolved; described below), while size of a node corresponds to the total number of genes in the two-hop network neighborhood around the gene, which are differentially expressed. Some genes are drivers for more than one subtype, and ties are resolved by coloring the node corresponding to the subtype with the smallest signature so as to preserve faint signals. (C) Heatmap of the top 20 down- and up-regulated MEGENA key drivers for each AD subtype, where size of the node represents KNR natural log P value (larger is more significant). (D) Heatmap of the top 20 down- and up-regulated BN key drivers for each AD subtype. (E) AD subtype key drivers in the MEGENA network supported by gene expression changes in other brain regions, which the region of overlap listed.

we exclude non-European individuals in the MSBB-AD during the PRS calculation. As shown in fig. S7, two of the three subtype classes show an increase in PRS burden compared with nondemented controls, with significant differences in class A (median PRS, 0.235; $P = 0.048$ under Welsh's t test) and C (PRS, 0.44; $P = 0.013$) sub-

types. While class B shows an increase in PRS, this difference is not significant from nondemented controls (PRS, 0.33; $P = 0.117$). In addition, we see an increased PRS burden across all AD samples (median PRS, 0.35) compared with nondemented controls (PRS, -0.43 ; $P = 0.016$). Despite these significant differences between

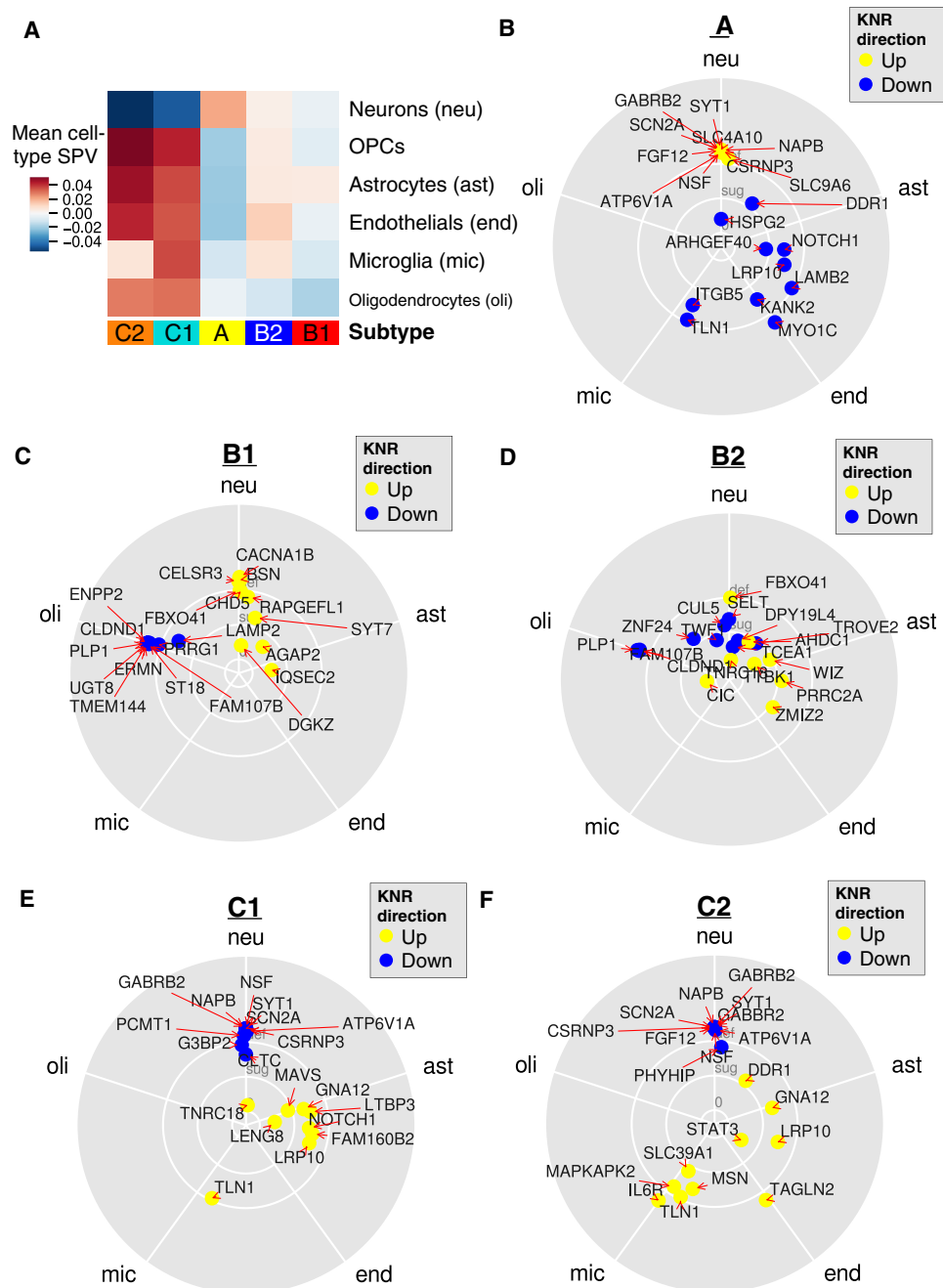


Fig. 4. Cell type-specific changes within each MSBB-AD subtype. (A) Mean change in cell-type proportion in each AD subtype, computed by averaging SPVs for the samples after cell-type deconvolution by BRETIGEA. (B to F) Vector addition of squared expression levels of AD subtype key regulators (up- and down-regulated) across five different cell types in a brain cell type-specific sequencing experiment (52).

individual AD subtypes and nondemented controls, there is no significant difference in PRS between the AD subtype classes. Therefore, these analyses demonstrate that genetic factors likely predispose individuals in the MSBB cohort to developing AD across each subtype, but these factors fail to adequately discriminate the molecular subtypes.

To better interrogate the intersection between known AD-associated genetic loci and the AD subtypes, we also intersect the AD risk genes compiled by the International Genomics of Alzheimer’s Project (IGAP) Consortium (6) and the predicted subtype-specific

key regulators. Forty-nine key regulators of the MEGENA network across all five MSBB subtypes have genetic loci associated with AD (IGAP gene-level significance, $P \leq 0.01$), including *AMPH*, *PICALM*, *MEF2C*, *EPDR1*, and *PSMC6* (table S2). *AMPH*, *MEF2C*, and *EPDR1* are down-regulated in the class C subtypes and up-regulated in class A. *PICALM* is down-regulated in subtype B2. *PSMC6* is down-regulated in both the class B and C subtypes.

AMPH, otherwise known as amphiphysin 1, is a vesicle cell surface protein important in clathrin-mediated endocytosis and is

primarily expressed in neurons. It is an important homolog of *BIN1*, which is highly expressed in oligodendrocytes as well as immune cell types and has been shown to be correlated with tau levels as well as glial fibrillary acidic protein (*GFAP*) and myelin basic protein (*MBP*) expression (35). In neurons, increased *AMPH* expression is associated with increased tau pathology, while increased *BIN1* expression is associated with decreased tau pathology (36). On the other hand, human autoantibodies to *AMPH*, which are observed in rare diseases (e.g., Stiff person syndrome), have been shown to induce defective presynaptic vesicle dynamics and composition, leading to decreased GABAergic transmission (37). Many GABA pathway genes are also predicted as key regulators in multiple subtypes (*GABRB2*, *GABRA1*, and *GABRA4*). Therefore, either up- or down-regulation of *AMPH* may lead to synaptic defects, either directly on the synaptic vesicles or through the secondary effects of tau accumulation. This is in line with our understanding of class C in which decreased *AMPH* may lead to decreased neuronal activity and up-regulated *AMPH* may increase tau pathway activity. Therefore, the effects of GABAergic signaling on cognitive dysfunction are likely critical to understanding and treating AD subtypes.

Subclassification of MSBB-AD samples with mild cognitive impairment (CDR = 0.5) recapitulates three major AD subtype classes

We have previously shown that AD subtypes are not correlated with dementia severity as measured by CDR score and that molecular signatures of these subtypes are independent from neuron degeneration and reactive gliosis seen in later stages of AD. We then seek to investigate whether participants with a mild cognitive impairment (MCI), without clinically defined AD, exhibit a subtype-specific signature consistent with the identified AD subtypes. On the basis of the clinical data available, we defined MCI for the MSBB-AD cohort as having a CDR score of 0.5, which corresponded to possible or mild dementia ($n = 32$).

To determine whether MCI samples group together with AD subtypes or whether MCI samples cluster separately on the basis of their molecular signatures, we repeat the WSCNA on the combined AD and MCI samples in the MSBB-AD cohort. As shown in the clustering dendrogram given in figs. S8 and S9, MCI samples are distributed in the branches corresponding with all five AD subtypes, including both A β and tau-predominant AD subtypes. We count the number of MCI samples distributed to each branch, labeled by the corresponding AD subtype (table S4). We observe that the distribution of the MCI samples across the subtypes is different than that of the AD samples (one-way ANOVA, $P = 4.7 \times 10^{-3}$, F statistic = 14.947, $df = 1$). The MCI samples are distributed more often to tau-predominant AD subtypes than A β -predominant AD subtypes. This could be due to a variety of reasons, including potential resilience to certain AD subtypes among the MCI group. Therefore, MCI samples may also be subclassified into different subtypes, providing additional insights into the molecular features of the disease.

Validation of MSBB-AD subtypes in ROSMAP shows conserved AD subtype molecular signatures

To verify the generalizability of AD subtypes identified in the MSBB-AD cohort, we perform the WSCNA-based subtyping analysis on the gene expression data from the dorsolateral prefrontal cortex (DLPFC) in the ROSMAP cohort ($n = 610$, with 391 AD cases). The ROSMAP cohort is an independent study of a differ-

ent brain region with more MCI ($n = 64$) and severe AD participants. The postmortem brains in the ROSMAP cohort were collected from individuals residing around multiple geographically distant sites across the United States, and participants enrolled in the study were evaluated multiple times over many years before death for cognitive impairment that was suggestive of AD (38). Cognitive impairment was measured by the mini-mental status exam (MMSE) at multiple time points, and pathologic factors such as A β and tau NFT burden were measured postmortem. In addition, the study includes three predominantly African-American and Hispanic communities in multiple locations (38). Therefore, reidentification of similar molecular subtypes of AD in the ROSMAP cohort should allow for greater generalization of the findings from the MSBB-AD alone.

Similar to the normalization process performed on the MSBB RNA-seq data, the ROSMAP gene expression data are corrected for batch effect, PMI, gender, RIN, and outliers, as well as dementia severity using MMSE scores to remove any potential effect of increasing neuronal loss with AD staging on the ROSMAP subtypes. Previous studies have shown a strong correspondence between CDR scores and MMSE (39), and therefore, MMSE would serve as a similar measure of dementia severity. As shown in fig. S10 (A and B), we measure the cell-type proportion of ROSMAP samples with increasing dementia severity before and after MMSE normalization. We demonstrate that—like in MSBB-AD—neuronal loss inferred from bulk transcriptomic data in ROSMAP is correlated with decreasing cognition (MMSE), and normalization using MMSE score can eliminate this observed bias and allows for stage-free subtyping. Furthermore, we show that cell-type proportion normalization is not sufficient to completely eliminate the bias from dementia severity, as a non-zero residual correlation still persists between MMSE and various cell types, and therefore, we choose MMSE normalization to correct for this.

As shown in Fig. 5A, we identify five subtypes from AD participants in this cohort, similar to those in the MSBB-AD cohort. Expression profiles from this cohort and from the subtypes in the PHG of the MSBB-AD cohort are significantly correlated (Fig. 5, B and C). First, by inspection of the up- and down-regulated gene modules alone, we observe two subtypes (blue and turquoise) with decreased synaptic signaling and increased immune response, one subtype (green) with increased synaptic signaling and protein modification activity, and two subtypes (yellow and red) that do not show changes in either pathway. Next, to quantify the similarity between the two sets of AD subtypes in the two independent cohorts, we perform correlation analysis of the mean gene expression profiles of these subtypes, followed by hierarchical clustering analysis. As shown in Fig. 5C, most subtypes in the ROSMAP cohort well match certain subtypes from the PHG in the MSBB-AD cohort, except the yellow subtype in the ROSMAP (Pearson correlation coefficients between 0.6 and 0.8). The clustering analysis further demonstrates that three major AD subtype classes (tau-predominant classes A and B, and A β -predominant class C) are highly conserved across both cohorts with different brain regions.

We then seek to determine whether the participants in the ROSMAP cohort are distributed in the AD subtypes at the same relative proportion as those in the MSBB-AD cohort. The distributions are roughly proportional to each other, with fewer samples in the C2 subtype than C1, with most samples either falling into class A or B (table S5). Excluding samples that do not match any subtype across both datasets, we perform a one-way ANOVA test across

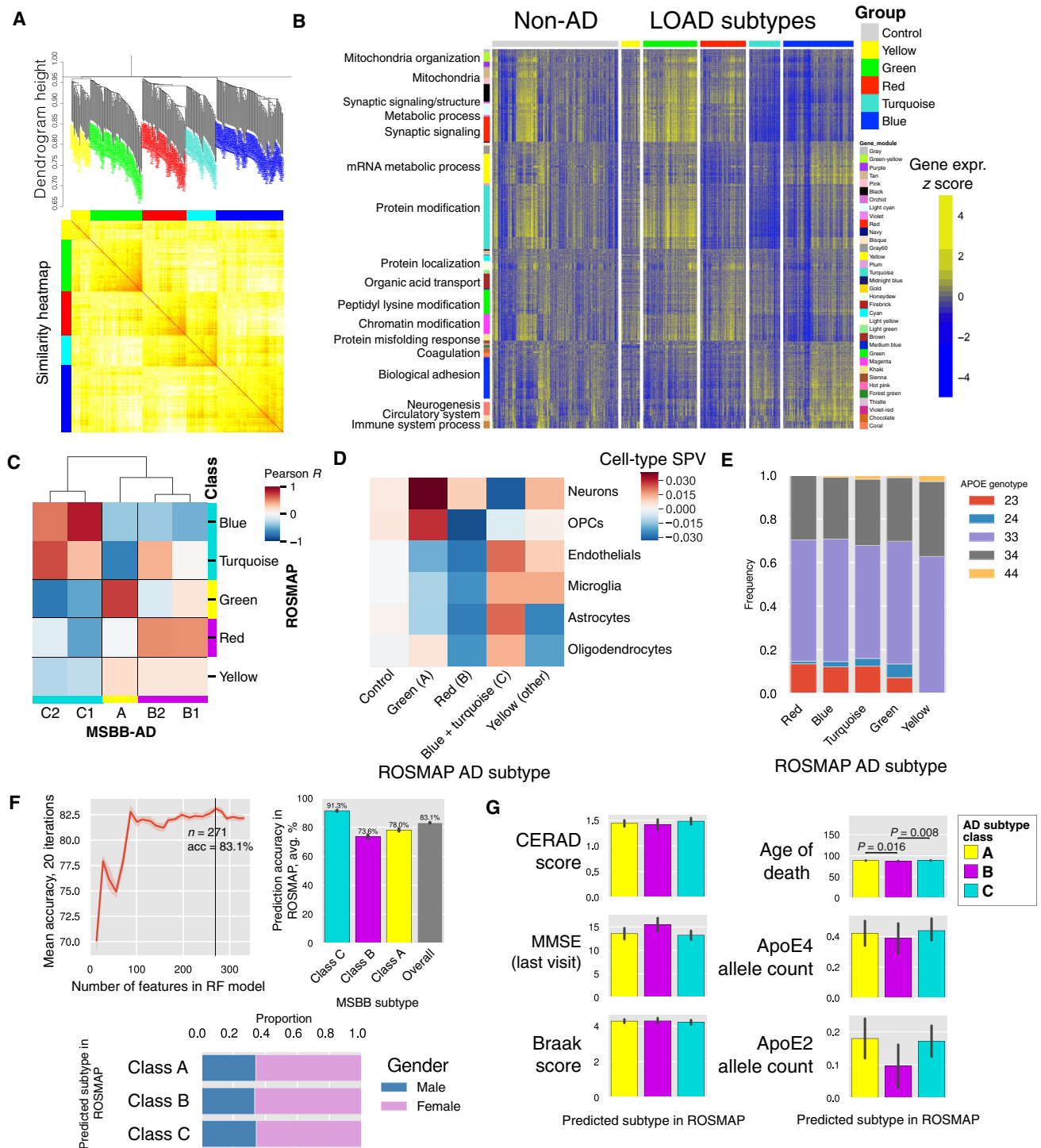


Fig. 5. Identification of the AD subtypes in the ROSMAP cohort. (A) WSCNA Clustering dendrogram and TOM similarity heatmap of the AD samples in the ROSMAP cohort. (B) Heatmap of gene expression across all the ROSMAP AD and control samples. Genes are organized by the WINA gene modules identified in the gene expression data of the PHG in the MSBB cohort. (C) Correlations between the MSBB and ROSMAP subtypes. First, mean gene expression across each AD subtype is computed for each gene, resulting in a 10 by 13,982 matrix of expression levels for all the 10 subtypes in the two cohorts (lowly expressed genes are excluded). Pearson correlation is then computed between each pair of subtypes, resulting in a 10 by 10 correlation matrix. (D) Cell-type SPV mean by ROSMAP AD subtype class, inferred as in (A) and (B), for MMSE-normalized data. (E) APOE genotype proportion by ROSMAP subtype. (F) Mean prediction accuracy of prediction of the ROSMAP subtypes by a classifier trained from the subtypes and the data from the RNA-seq data from the PHG in the MSBB cohort. RF classifiers are trained on the basis of different numbers of key regulators of the AD subtype classes identified in the MSBB cohort. The subtypes for each AD case in the ROSMAP cohort are independently determined as described in fig. S7 and are thus used as ground truth for evaluating the prediction by the predictors trained by the MSBB-AD data. (G) Bar plots of the association of clinical and pathologic phenotypes with the predicted subtypes of each ROSMAP sample.

both MSBB-AD and ROSMAP and find no significant difference in the distribution of samples. $P = 0.089$, F statistic = 5.01.

As shown in Fig. 5D, the ROSMAP subtypes also exhibit predicted cell-type proportion changes in the frontal cortex. For example, the blue and turquoise subtypes (class C–like) show decreases in neuronal SPVs and increases in microglia and endothelial cell SPVs. On the other hand, the green subtype (class A–like) shows strong increases in neuronal and OPC SPVs, along with decreases in other cell types. Last, the red subtype (class B–like) shows strong decreases in OPC SPV only. Therefore, AD subtypes show strong, characteristic susceptibility to cell-type proportion changes that are consistent across subtypes identified in the two independent cohorts.

Similar to the MSBB subtypes, the ROSMAP subtypes are not associated with clinical and pathological traits, including biological sex, cognitive scores, age, Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) pathologic scores, MMSE, and Braak staging, suggesting that these subtypes are truly molecular subtypes with distinct expression patterns. As shown in Fig. 5E, *APOE* e2 allele dosage significantly decreases in the yellow (“other AD”) subtype compared to all other subtypes (Welsh’s t test, $P < 0.025$), which did not contain any *APOE* e2 individuals. On the other hand, *APOE* e4 allele dosage is not significantly associated with ROSMAP AD subtypes.

In addition, PRSs for ROSMAP subtypes and nondemented controls are also computed using the GWAS meta-analysis of Kunkle *et al.* (33) using the same methodology as for MSBB-AD, as shown in fig. S10C. Similarly to the MSBB-AD cohort, each of the ROSMAP classes A (median, 0.339, $P = 5.1 \times 10^{-4}$ under Welsh’s t test), B (median, 0.084, $P = 0.032$), and C (median, 0.191, $P = 1 \times 10^{-5}$) have an increased PRS compared to nondemented controls (median, -0.239); however, there are no significant PRS differences between subtypes.

Prediction of AD subtypes across different cohorts

We then develop a high-performance classifier to predict the MSBB-AD PHG-based subtype classes (A, B, and C) of the participants in the ROSMAP AD cohort. A random forest (RF) model is trained using all the AD samples in the MSBB PHG dataset and a range of 5 to 350 subtype key regulators as features until maximum accuracy is achieved. Additional details about the training and classification processes are given in Materials and Methods.

As shown in Fig. 5F, the performance of the RF model on the ROSMAP data peaks in accuracy with 271 features. This top-performing RF model was then selected for further validation and characterization. A list of molecular features used in the model, as well as top-performing features by class, is given in data file S2 and fig. S11. Using the ROSMAP-derived subtype categories as the ground truth, our classifier model reaches an accuracy of 83.1% (SD, 1.0%), with class C (A β predominant) AD subtype prediction peaking at 91.3% (SD, 1.0), suggesting that the identified subtypes are highly reproducible and predictable. Similar to the ROSMAP subtypes, the subtypes predicted by the classifier do not show a significant difference in cognition, CERAD score, MMSE, Braak score, *APOE*, or biological sex (Fig. 5G). There is a small but significant association with age of death under the KW test, but this was no longer significant after multiple testing correction ($P = 0.036$, adjusted $P = 0.252$), which is likely due to differences in age between the atypical and intermediate subtypes (Welsh’s t test, $P = 0.0048$). The difference in mean age between the groups is 1.6 years (87.6 years

versus 86.0 years for atypical and intermediate, respectively), which is relatively small and likely does not reflect marked differences in AD stages. In addition, there is no significant difference in age under the Welsh’s t test between the predicted A β - and tau-predominant subtypes even without multiple test correction.

Correspondence between human AD subtypes and AD mouse models

In the past two decades, several different mouse models of AD have been developed to characterize AD pathology, biology, and behavioral changes. Many of these mouse models perturb various AD-related proteins or regulatory genes such as A β , *APP*, *MAPT*, *PSEN1*, *TYROBP*, and *HDAC1*. Given the substantial differences in molecular signatures among the AD subtypes, we explore whether these AD subtypes’ transcriptomic signatures match the existing mouse model signatures.

We collect aligned RNA-seq data from 19 mouse model studies of AD that are publicly available at the Accelerating Medicines Partnership–Alzheimer’s Disease (AMP-AD) portal on Synapse.org and Gene Expression Omnibus, as listed in data file S3. Transcriptomic signatures for each mouse model are also identified. Many of these models may harbor multiple amyloid precursor protein (*APP*) variants and tau protein variants. The Swedish mice (*APPK670N/M671NL*) develop A β plaques near neurons. The Dutch (*APPE693Q*) mice accumulate soluble A β in perivascular cells at the blood-brain barrier. The 5XFAD mice recapitulate *APP* variants seen in familial forms of AD but do not have related tau NFT seen in AD. Tau protein variant *TauP301L* (“D35”) and *TauP301S* (“PS19”) mice develop hyperphosphorylated tau, as well as presenilin 1 (*PSEN1*) Δ exon9 and *M146V* variants. We also examine *APOE* variant mice and mouse models with mutant *HDAC1*, *TYROBP*, *TREM2*, *BIN1*, *CD2AP*, *CLU*, and *GFAP* alleles.

As shown in Fig. 6A, the class C subtypes (A β predominant) match to the 5XFAD (familial), *APP* Dutch (inflammatory), and *APP* Swedish (amyloid) mice, consistent with an amyloid-driven disease with increased immune and circulatory system activity, as well as previous findings that the class C subtypes may be driven by inflammatory processes (shown in Fig. 1E). Inflammation at the blood-brain barrier has been noted in AD and other age-related neurodegenerative diseases such as vascular dementia. On the other hand, the tau-predominant class A subtype has a gene signature opposite to those of 5XFAD and *APP* mouse models but consistent with that of the *TauP301L* model, in line with gene expression changes in known tau pathways. The remaining two class B subtypes (B1 and B2) show the strongest match with the *CLU* (*apoJ*-) mutant model and a good match with the *CD2AP* and *BIN1* mutant models. *Clusterin*, a secreted neuroprotective glycoprotein secreted primarily by astrocytes, has been shown to be increased in AD in response to tau-mediated neurodegeneration. *Clusterin* mutant mice have been shown to have less A β damage and neuritic dystrophy when bred with 5XFAD model mice versus controls (40). Therefore, the class B subtypes match the mouse models that carry tau-related neurodegenerative factors over amyloid-related factors.

We further examine the expression changes of the subtype-specific key regulators in the 5XFAD, *TauP301L*, and *CLU* mutant mouse models that match the three subtype classes. Figure 6B shows the gene expression levels of the top four key regulators from each of the five MSBB-AD subtypes in each mouse model. As expected, the gene expression differences between the AD subtypes across human participants are recapitulated in specific mouse models. Many

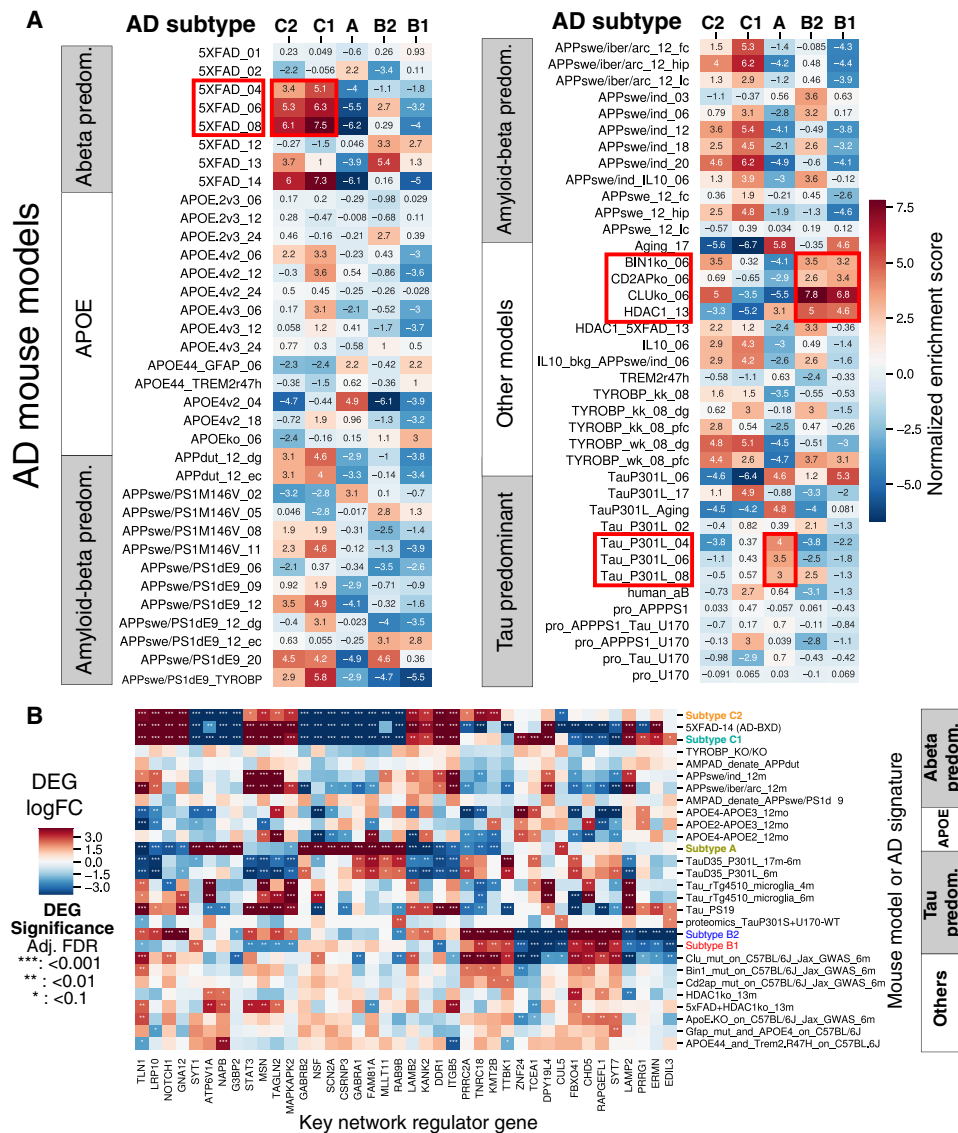


Fig. 6. Matching existing AD mouse models to the MSBB-AD subtypes. (A) GSEA enrichment of differential expression signatures of the identified AD subtypes (up- and down-regulated) for the gene signatures of the AD mouse models. Positive scores indicate strong consistency. (B) Gene expression of the top subtype key regulators across the mouse models, with significant DEGs shown.

KNRs of the Aβ- and tau-predominant subtypes have consistent expression changes in the respective human brain samples and the matched mouse models.

DISCUSSION

In this study, we systematically identify five molecular subtypes of AD in three major classes and subsequently characterize them with molecular signatures, network regulator genes, and matched mouse models. Through a novel clustering method, the AD subtypes are identified and cross-validated using independent postmortem brain RNA data. The hippocampal area demonstrates the greatest subtyping signal over the other regions, despite the fact that subtype effects are detectable across all brain regions studied here. These subtypes are independent of age and disease severity. These AD subtypes are well conserved across different independent co-

horts. A small number of predicted key regulators can be used to predict clinical features such as cognitive function or dementia. Each subtype has a unique set of key regulator genes, and many predicted subtype key regulators are also known as AD genetic risk genes. Last, we show that each existing mouse model of AD may match to a particular subset of human AD subtypes but not all subtypes simultaneously. This may partially explain how many existing clinical trials that showed promising efficacy in one particular mouse model later do not align with human trial results, assuming that study participants had consisted of a heterogeneous group of participants across many AD subtypes.

The molecular subtypes of AD cannot be fully explained by differences in postmortem pathologic variables such as Aβ and tau accumulation or by differences in APOE risk allele genotype between participants. These findings are consistent with previous studies that have shown that cognitive impairment is neither dependent on

nor fully assured by high levels of A β and tau accumulation (4, 5). It is more likely that A β and tau accumulation are often mediators or the end effects of neurodegeneration and inflammation, independent of hippocampal load. This is also consistent with recent meta-analyses of AD that have shown that the effects of APOE genotype on AD risk are significantly diminished in older individuals (>75 years, LOAD) (41).

The predicted subtypes and their molecular signatures as well as key regulators along with matched mouse models pave the way for developing novel therapeutics for AD toward precision medicine. The specific mouse models identified in this study that most closely match each of the AD subtypes may serve an important role in the validation of subtype key regulator genes and hit compounds. Additional work toward validating and better understanding the role of each subtype key regulator in its matching mouse model will provide great value and have a great impact on future studies of AD.

We observed many cell type-specific signatures and key regulator genes for each AD subtype, which are persistent even after the effects of dementia severity on neuronal loss and reactive gliosis are removed, suggesting that subpopulations of neuronal or glial cell types may contribute distinctively to different AD subtypes. Therefore, single-cell transcriptomic analysis of control and AD human brains will gain more insights into these cell types and the underlying molecular mechanisms, leading to the heterogeneity in AD. This will be a future direction of the research on AD subtype analysis.

Only about one-third of the AD cases carry consistent hallmarks (e.g., increased immune response and decreased synaptic signaling) of a “typical” AD presentation (class C), while the rest show opposite molecular gene regulation and other complex changes across multiple pathways and cell types (classes A and B). One potential avenue for research on the modulation of the hallmark signature across the subtypes is to look at the interaction between A β folding, tau, mitochondria, and lysosomal acidification, which are all dysregulated across the subtypes. Recent work has shown that certain A β oligomers may, depending on the presence of tau, block mitochondrial activity through lysosomal mTORC1 protein kinase activity but potentially stimulate mitochondrial activity in the absence of tau (42).

The bidirectional nature of certain key regulator genes (e.g., *GABRB2* and *ATP6V1A*) across the identified AD subtypes is notable but not unsupported by previous research. Recent work in AD mouse models has shown that the effects of α -synuclein on Alzheimer’s pathology and clinical phenotypes are important when the gene is up- and down-regulated. For instance, while reduced α -synuclein increases A β deposition in *APP* mutant mice, increased α -synuclein was shown in the same experiment to decrease spatial memory formation, increase neuronal cell cycle dysfunction, and increase specific tau conformational markers (43).

Among all the clinical, pathological, and genetic factors tested, *APOE* genotype is highly associated with some of the AD subtypes. Besides *APOE*, additional genetic risk factors for AD, such as those identified by linkage analysis and GWAS in the IGAP study, hint at a plurality of mechanisms through which AD is developed (6, 44, 45). Furthermore, one study shows that some patients with MCI or depression before AD have a more gradual progression to the disease and that they have pathologic A β and tau cerebrospinal fluid (CSF) level changes that are not affected by *APOE* allele status, while the other forms of AD show a strong *APOE* association (46).

We demonstrate that participants with only MCI could be classified into all three subtype classes, suggesting that certain patients

may be selectively vulnerable for a certain disease-causing molecular process over others. One caveat of our analyses is that participants in both the MCI and AD groups are similarly aged; therefore, participants with MCI in the MSBB-AD likely demonstrate resilience to AD as a group and, as a result, may be in an earlier stage of the disease. While MCI participants are more often classified into tau-predominant subtypes than AD participants, it is unlikely that this is due to participant age in the MSBB-AD. Future work will investigate the progression of MCI to AD subtype via a prospective study to establish whether participants maintain a particular subtype designation from one stage of dementia to the next over a range of ages and clinical course of disease progression.

Drug repositioning, a process of matching known drug transcriptomic signatures to a disease signature in silico to identify hit candidates that produce the desired change in expression, may be useful in identifying compounds that change the level of top key regulator genes among certain AD subtypes that can be validated in their matched AD animal model system. For instance, a particular candidate compound identified using molecular signature data alone using an in silico drug repositioning algorithm (47) was able to increase the expression level of *ATP6V1A* and restore long-term potentiation and memory formation in both cultured neurons and AD mouse models (19). Therefore, AD subtype-specific signatures identified in the MSBB-AD cohort will play an important role not only in identifying new candidate drugs for treating AD but also in stratifying the patient population for suitable AD treatments.

To develop personalized treatments for AD, each patient should be classified by subtype. As we have shown, AD subtypes have very different transcriptomic signatures and therefore will likely require specialized treatments. Given that many subtype-specific key regulators have opposite directions in some AD subtypes, it is also possible that drugs that reduce AD symptoms in one subtype may exacerbate symptoms in another subtype. Furthermore, we construct a machine learning model for predicting AD subtypes that is effective and robust. However, since the prediction is based on brain transcriptomic data that are not available from vast majority of patients, unnecessary biomarkers for AD subtypes either from peripheral tissues (e.g., CSF and blood) or noninvasive brain imaging are highly desirable.

MATERIALS AND METHODS

AD cohorts

In this study, we used two AD cohorts of RNA-seq data: the MSBB-AD (21) study and the ROSMAP (22, 23). The MSBB-AD cohort includes RNA expression data in the following four different brain regions: FP (Brodmann area 10; $n = 265$ with 187 AD cases), STG (Brodmann area 22; $n = 240$ with 174 AD cases), PHG (Brodmann area 36; $n = 215$ with 151 AD cases), and the IFG (Brodmann area 44; $n = 222$ with 157 AD cases). Clinical phenotypes for each participant are also collected including age, race, sex, hypoxia-induced encephalopathy score, cognitive function scores, CDR, age of onset and death, and pathologic findings of tau and A β on biopsy. This cohort was specifically selected to include cases with either no neuropathology or only neuropathological lesions diagnostic of AD. Cases with mixed neuropathology, e.g., AD and cerebrovascular disease and AD with Lewy bodies, were specifically excluded from the study cohort. Controls were defined as those presenting with no cognitive impairment (i.e., CDR = 0) and no overt neuritic plaque or NFT involvement.

The ROSMAP (22, 23) includes whole-transcriptome RNA-seq data of the DLPFC from 615 participants including those with AD ($n = 391$), MCI ($n = 64$) and nondemented controls ($n = 160$) determined by a CERAD pathology score of definite AD or probable AD. Clinical and pathologic phenotypes, as well as demographic information, were also collected as well for each sample including MMSE scores (at time of diagnosis and last known), CERAD score, Braak score, cognitive score, APOE genotype, age of death, age at diagnosis, PMI, gender, race, education level, and if the participant was Spanish-speaking.

Clustering algorithm evaluation and cluster stability determination

We present a method to determine the empirical likelihood that a particular clustering is robust and apply it to each considered clustering algorithm and transcriptomic brain region. If true molecular subtypes exist and are reflected in the transcriptomic data from a certain brain region, then a robust clustering algorithm should produce the same set of clustered samples on repeated reclustering, even if fewer samples or molecular features are used to create the clustering. However, if no molecular signature is reflected in the transcriptomic data, or the clustering algorithm is not robust, then different sets of clustered samples may be produced upon repeated reclustering. On the MSBB-AD cohort data, we performed 50 rounds of bootstrapped reclustering using each of the four clustering algorithms, withholding 20% of the samples and genes per round. We calculate the rate at which pairs of samples shared the same cluster across all 50 bootstrapping rounds (e.g., a pair of samples clustered together in 35 of 50 bootstrapped clustering rounds would have a rate of 70%), defined here as the pairwise sample reclustering rate. We then calculate the average pairwise sample reclustering rate for all pairs of samples within the sample clusters identified by each algorithm, as well as the average rate of same-sized clusters drawn from a distribution of 100,000 random pairs of samples. We term these average rates the cluster stability rate and the null cluster stability rate, respectively. We then calculate the empirical likelihood that the cluster stability rate and the null cluster stability rate are the same, under the binomial distribution. Using this method, a specific subtype grouping is considered a putative subtype if its empirically adjusted P value is less than 0.05.

RNA-seq data normalization

MSBB-AD RNA-seq data were processed with the Spliced Transcripts Alignment to a Reference (STAR) aligner and normalized using mixed-model correction for batch effect, RIN, ribosomal RNA rate, exonic RNA rate, PMI, age of death, inferred race, and inferred sex. Label swaps were inferred and corrected or removed if resolution was not possible.

To remove the disease stage effect, CDR is corrected in the MSBB-AD gene expression data through linear model normalization. This was verified by performing a second round of linear model fitting between CDR and gene expression, which showed that no significant DEGs remained between participants with different CDR scores across all brain regions in the MSBB-AD.

ROSMAP DLPFC RNA-seq data were also normalized for age of death, gender, batch, RIN, and PMI using mixed-model correction. Data were then subsequently normalized for last known MMSE score using a linear model, and no genes are shown subsequently to have a correlation with MMSE ($R^2 = 0$).

Differential gene expression analysis of AD subtypes

As described here, we performed differential gene expression (DEG) analysis to determine the molecular signatures of each of the AD subtypes compared with nondemented (CDR = 0) controls, starting with the RNA-seq counts per million (CPM) data as input. The analysis was carried out separately for each comparison. Log-scaled (base 2) gene CPMs from samples in the comparison were first fit to a linear model using the `lmfit()` provided by the `limma` R package before contrasts were fit. Empirical Bayes statistics for differential expression were then calculated using the `eBayes()` R function, followed by the `topTable()` R function to output significant DEGs. P values were adjusted by q values provided by the `qvalue` Bioconductor package, using default parameters.

Clustering algorithms used in the establishment of putative AD subtypes

WSCNA identifies sample clusters by analyzing gene expression level correlations between pairs of samples to build a sample correlation network, which is then used to calculate topological overlap [topological overlap matrix (TOM)] score that can be used to cluster similar samples together via k -means clustering. WSCNA extends the WINA algorithm (48) to samples by transposing the input matrix so that sample-sample correlations are compared. Note that gene expression data are standardized to z scores so that expression differences do not inflate the correlation metric.

Network-based KNR analysis of molecular subtype signatures

We applied the key driver analysis (KDA) (49) to the multiscale embedded gene expression network analysis (MEGENA) network generated from PHG data in the MSBB. KDA first generates a sub-network NG , defined as the set of nodes in N that are no more than h layers away from the nodes in G , and then searches the h -layer neighborhood ($h = 1, \dots, H$) for each gene in NG ($HLN_{g,h}$) for the optimal h^* , such that

$$ES_h^* = \max(ES_{h,g}) \forall g \in N_{g,h} \in \{1, \dots, H\}$$

where $ES_{h,g}$ is the computed enrichment statistic for $HLN_{g,h}$. This results in a list of predicted key network regulatory hub genes that may alter the expression pattern of its surrounding nodes and result in the DEG pattern observed.

Machine learning subtype classifier across other cohorts

We developed an RF model to classify samples into each AD subtype using the MSBB-AD PHG brain region data for training and then validated this model on the ROSMAP data. All RF models were built using the `scipy` Python library, with initial parameters of 300 decision trees and a maximum tree depth of 8. Before model creation, both datasets are first corrected for cohort effect between MSBB-AD and ROSMAP, using the `ComBat` program, to reduce technical differences between studies. We divide classifier creation into three steps: feature selection, model training, and model validation. For the feature selection step, we first selected different numbers of top KNRs as features from each subtype ($n = 1$ to 80 features per subtype; total, 5 to 400 features). We then train multiple RF models to predict subtype classification within the MSBB-AD (PHG) cohort and evaluate the model accuracy using leave-one-out cross-validation between the predicted and observed subtypes. In the model training step, an RF model is created on all AD participants'

PHG samples in MSBB-AD using only the top-performing features identified in the previous step. Last, for the model validation step, the RF model created from the MSBB-AD data is applied to the ROSMAP data, and model accuracy is evaluated by comparing the predicted ROSMAP subtypes from the RF model and the observed ROSMAP subtypes from network-based clustering analysis. We increase the number of features used in the RF model until maximum validation accuracy is achieved, and the top-performing set of features from this model is retained.

BN construction

A BN was constructed by integrating genome-wide gene expression, SNP genotype, and known transcription factor (TF)–target relationships in the PHG in the MSBB-AD cohort, similar to our previous work (27). Briefly, we first computed expression quantitative trait loci (eQTLs) and then used a formal statistical causal inference test (50) to infer the causal probability between gene pairs associated with the same eQTL. The causal relationships inferred are used, together with TF-target relationships from the Encyclopedia of DNA Elements (ENCODE) project, as structural priors for building a causal gene regulatory network from the gene expression data through a Markov chain Monte Carlo (MCMC) simulation-based procedure (51). We followed a network averaging strategy in which 1000 networks are generated from the MCMC procedure starting with different random structure, and links that shared by more than 30% of the networks are used to define a final consensus network structure. To ensure the consensus network is a directed acyclic graph, an iterative deloop procedure was conducted, removing the most weakly supported link of all links involved in any loop. Following Zhang *et al.* (27), we performed KDA (49) on the consensus Bayesian network to identify key network regulatory genes that can potentially regulate a large number of downstream nodes.

Cell-type proportion analysis and cell-type normalization

To estimate the cell-type proportion of bulk tissue RNA-seq data used in this study, we perform a cell-type deconvolution on each sample using the brain cell-type marker signatures provided by the BRETIGEA R package (26). One thousand marker genes per cell type were used from the human brain cell marker gene set (neurons, endothelial, oligodendrocytes, microglia, astrocytes, and OPCs) to generate all surrogate cell-type proportion (SPV) estimates, except for OPCs, which only had 500 marker genes available. Normalization of the bulk RNA-seq by brain cell type was also performed by BRETIGEA, using the default parameters and the calculated SPV values from the previous step.

Cell-type specificity plots

To generate cell-type specificity plots, using the mean cell-type gene expression levels from Zhang *et al.* (52), we plotted each squared expression value as a vector from the center on a polar coordinate system. We then calculated the vector sum from each of the expression levels and multiply the final result by a scaling parameter to create a final point as the estimate of the cell-type specificity of any gene under consideration.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/7/2/eabb5398/DC1>

[View/request a protocol for this paper from Bio-protocol.](#)

REFERENCES AND NOTES

1. Alzheimer's Association, *2019 Alzheimer's Disease Facts and Figures* (2019); <https://alz.org/media/Documents/alzheimers-facts-and-figures-2019-r.pdf>.
2. M. S. Fiandaca, M. E. Mapstone, A. K. Cheema, H. J. Federoff, The critical need for defining preclinical biomarkers in Alzheimer's disease. *Alzheimers Dement.* **10**, S196–S212 (2014).
3. B. De Strooper, E. Karran, The cellular phase of Alzheimer's disease. *Cell* **164**, 603–615 (2016).
4. M. Sekiya, M. Wang, N. Fujisaki, Y. Sakakibara, X. Quan, M. E. Ehrlich, P. L. De Jager, D. A. Bennett, E. E. Schadt, S. Gandy, K. Ando, B. Zhang, K. M. Iijima, Integrated biology approach reveals molecular and pathological interactions among Alzheimer's A β 42, Tau, TREM2, and TYROBP in *Drosophila* models. *Genome Med.* **10**, 26 (2018).
5. D. Iacono, S. M. Resnick, R. O'Brien, A. B. Zonderman, Y. An, O. Pletnikova, G. Rudow, B. Crain, J. C. Troncoso, Mild cognitive impairment and asymptomatic alzheimer disease subjects: Equivalent β -amyloid and tau loads with divergent cognitive outcomes. *J. Neuropathol. Exp. Neurol.* **73**, 295–304 (2014).
6. J. C. Lambert, C. A. Ibrahim-Verbaas, D. Harold, A. C. Naj, R. Sims, C. Bellenguez, G. Jun, A. L. DeStefano, J. C. Bis, G. W. Beecham, B. Grenier-Boley, G. Russo, T. A. Thornton-Wells, N. Jones, A. V. Smith, V. Chouraki, C. Thomas, M. A. Ikram, D. Zelenika, B. N. Vardarajan, Y. Kamatani, C. F. Lin, A. Gerrish, H. Schmidt, B. Kunkle, N. F. F. F. Pasquier, V. Deramecourt, R. F. A. G. De Bruijn, N. Amin, A. Hofman, C. M. Van Duijn, M. L. Dunstan, P. Hollingworth, M. J. Owen, M. C. O'Donovan, L. Jones, P. A. Holmans, V. Moskvina, J. Williams, C. Baldwin, L. A. Farrer, S. H. Choi, K. L. Lunetta, A. L. Fitzpatrick, T. B. Harris, B. M. Psaty, J. R. Gilbert, K. L. Hamilton-Nelson, E. R. Martin, M. A. Pericak-Vance, J. L. Haines, V. Gudnason, P. V. Jonsson, G. Eiriksdottir, M. T. Bihoreau, M. Lathrop, O. Valladares, L. B. Cantwell, L. S. Wang, G. D. Schellenberg, A. Ruiz, M. Boada, C. Reitz, R. Mayeux, A. Ramirez, W. Maier, O. Hanon, W. A. Kukull, J. D. Buxbaum, D. Campion, D. Wallon, D. Hannequin, P. K. Crane, E. B. Larson, T. Becker, C. Cruchaga, A. M. Goate, D. Craig, J. A. Johnston, B. McGuinness, S. Todd, P. Passmore, C. Berr, K. Ritchie, O. L. Lopez, P. L. De Jager, D. Evans, S. Lovestone, P. Proitsi, J. F. Powell, L. Letenneur, P. Barberger-Gateau, C. Dufouil, J. F. Dartigues, F. J. Morón, D. C. Rubinsztein, P. St. George-Hyslop, K. Sleegers, K. Bettens, C. Van Broeckhoven, M. J. Huentelman, M. Gill, K. Brown, K. Morgan, M. I. Kambou, L. Keller, L. Fratiglioni, R. Green, A. J. Myers, S. Love, E. Rogava, J. Gallacher, A. Bayer, J. Clarimon, A. Lleo, D. W. Tsuang, L. Yu, D. A. Bennett, M. Tsolaki, P. Bossù, G. Spalletta, J. Collinge, S. Mead, S. Sorbi, B. Nacmias, F. Sanchez-Garcia, M. C. D. Naranjo, N. C. Fox, J. Hardy, P. Bosco, R. Clarke, C. Brayne, D. Galimberti, M. Mancuso, F. Matthews, S. Moebus, P. Mecocci, M. Del Zompo, H. Hampel, A. Pilotto, M. Bullido, F. Panza, P. Caffarra, M. Mayhaus, S. Pichler, W. Gu, M. Riemenschneider, L. Lannfelt, M. Ingelsson, H. Hakonarson, M. M. Carrasquillo, F. Zou, S. G. Younkin, D. Beekly, V. Alvarez, E. Coto, C. Razquin, P. Pastor, I. Mateo, O. Combarros, K. M. Faber, T. M. Foroud, H. Soininen, M. Hiltunen, D. Blacker, T. H. Mosley, C. Graff, C. Holmes, T. J. Montine, J. I. Rotter, A. Brice, M. A. Nalls, J. S. K. Kauwe, E. Boerwinkle, R. Schmidt, D. Rujescu, C. Tzourio, M. M. Nöthen, L. J. Launer, S. Seshadri, Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
7. S. L. Risacher, W. H. Anderson, A. Charil, P. F. Castelluccio, S. Shcherbinin, A. J. Saykin, A. J. Schwarz; Alzheimer's Disease Neuroimaging Initiative, Alzheimer disease brain atrophy subtypes are associated with cognition and rate of decline. *Neurology* **89**, 2176–2186 (2017).
8. D. E. Bredesen, Metabolic profiling distinguishes three subtypes of Alzheimer's disease. *Aging* **7**, 595–600 (2015).
9. J. S. Phillips, F. Da Re, L. Dratch, S. X. Xie, D. J. Irwin, C. T. McMillan, S. N. Vaishnavi, C. Ferrarese, E. B. Lee, L. M. Shaw, J. Q. Trojanowski, D. A. Wolk, M. Grossman, Neocortical origin and progression of gray matter atrophy in nonamnesic Alzheimer's disease. *Neurobiol. Aging* **63**, 75–87 (2018).
10. W. Qian, C. E. Fischer, T. A. Schweizer, D. G. Munoz, Association between psychosis phenotype and APOE genotype on the clinical profiles of Alzheimer's disease. *Curr. Alzheimer Res.* **15**, 187–194 (2018).
11. Y.-M. Lee, Y.-I. Chung, J.-M. Park, B.-D. Lee, E. Moon, H.-J. Jeong, J.-H. Kim, H.-J. Kim, C.-W. Mun, T.-H. Kim, Y.-H. Kim, E.-J. Kim, Decreased gray matter volume is associated with the subtypes of psychotic symptoms in patients with antipsychotic-naïve mild or moderate Alzheimer's disease: A voxel-based morphometry study. *Psychiatry Res. Neuroimaging* **249**, 45–51 (2016).
12. M. A. Rapp, M. Schnaider-Beeri, D. P. Purohit, D. P. Perl, V. Haroutunian, M. Sano, Increased neurofibrillary tangles in patients with Alzheimer disease with comorbid depression. *Am. J. Geriatr. Psychiatry* **16**, 168–174 (2008).
13. Y. Huang, L. Mucke, Alzheimer mechanisms and therapeutic strategies. *Cell* **148**, 1204–1222 (2012).
14. J. F. Crary, J. Q. Trojanowski, J. A. Schneider, J. F. Abisambra, E. L. Abner, I. Alafuzoff, S. E. Arnold, J. Attems, T. G. Beach, E. H. Bigio, N. J. Cairns, D. W. Dickson, M. Gearing, L. T. Grinberg, P. R. Hof, B. T. Hyman, K. Jellinger, G. A. Jicha, G. G. Kovacs, D. S. Knopman,

- J. Kofler, W. A. Kukull, I. R. Mackenzie, E. Masliah, A. McKee, T. J. Montine, M. E. Murray, J. H. Neltner, I. Santa-Maria, W. W. Seeley, A. Serrano-Pozo, M. L. Shelanski, T. Stein, M. Takao, D. R. Thal, J. B. Toledo, J. C. Troncoso, J. P. Vonsattel, C. L. White III, T. Wisniewski, R. L. Woljter, M. Yamada, P. T. Nelson, Primary age-related tauopathy (PART): A common pathology associated with human aging. *Acta Neuropathol.* **128**, 755–766 (2014).
15. Y. Chornenkyy, W.-X. Wang, A. Wei, P. T. Nelson, Alzheimer's disease and type 2 diabetes mellitus are distinct diseases with potential overlapping metabolic dysfunction upstream of observed cognitive decline. *Brain Pathol.* **29**, 3–17 (2019).
16. J. Bang, S. Spina, B. L. Miller, Frontotemporal dementia. *Lancet* **386**, 1672–1682 (2015).
17. E. M. Blalock, J. W. Geddes, K. C. Chen, N. M. Porter, W. R. Markesbery, P. W. Landfield, Incipient Alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2173–2178 (2004).
18. E. Courtney, S. Kornfeld, K. Janitz, M. Janitz, Transcriptome profiling in neurodegenerative disease. *J. Neurosci. Methods* **193**, 189–202 (2010).
19. M. Wang, A. Li, M. Sekiya, N. D. Beckmann, X. Qian, N. Schrode, M. B. Fernando, A. Yu, L. Zhu, J. Cao, L. Lyu, E. Horgusluoglu, Q. Wang, L. Guo, Y. Wang, R. Neff, W.-m. Song, E. Wang, Q. Shen, X. Zhou, C. Ming, S.-M. Ho, S. Vatansever, H. U. Kaniskan, J. Jin, M.-M. Zhou, K. Ando, L. Ho, P. A. Slesinger, Z. Yue, J. Zhu, S. Gandy, M. E. Ehrlich, D. Cai, V. Haroutunian, K. M. Iijima, E. Schadt, K. J. Brennan, B. Zhang, Molecular networks and key regulators of the dysregulated neuronal system in Alzheimer's disease. *bioRxiv* 788323 (2019).
20. N. A. Twine, K. Janitz, M. R. Wilkins, M. Janitz, Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLOS ONE* **6**, e16266 (2011).
21. M. Wang, N. D. Beckmann, P. Roussos, E. Wang, X. Zhou, Q. Wang, C. Ming, R. Neff, W. Ma, J. F. Fullard, M. E. Hauberg, J. Bendl, M. A. Peters, B. Logsdon, P. Wang, M. Mahajan, L. M. Mangravite, E. B. Dammer, D. M. Duong, J. J. Lah, N. T. Seyfried, A. I. Levey, J. D. Buxbaum, M. Ehrlich, S. Gandy, P. Katsel, V. Haroutunian, E. Schadt, B. Zhang, The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci. Data* **5**, 180185 (2018).
22. P. L. De Jager, Y. Ma, C. McCabe, J. Xu, B. N. Vardarajan, D. Felsky, H.-U. Klein, C. C. White, M. A. Peters, B. Lodgson, P. Nejad, A. Tang, L. M. Mangravite, L. Yu, C. Gaiteri, S. Mostafavi, J. A. Schneider, D. A. Bennett, A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci. Data* **5**, 180142 (2018).
23. D. A. Bennett, A. S. Buchman, P. A. Boyle, L. L. Barnes, R. S. Wilson, J. A. Schneider, Religious orders study and rush memory and aging project. *J. Alzheimers Dis.* **64**, S161–S189 (2018).
24. M. Wang, P. Roussos, A. McKenzie, X. Zhou, Y. Kajiwara, K. J. Brennan, G. C. De Luca, J. F. Cray, P. Casaccia, J. D. Buxbaum, M. Ehrlich, S. Gandy, A. Goate, P. Katsel, E. Schadt, V. Haroutunian, B. Zhang, Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease. *Genome Med.* **8**, 104 (2016).
25. H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, M. Kanehisa, KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
26. A. T. McKenzie, M. Wang, M. E. Hauberg, J. F. Fullard, A. Kozlenkov, A. Keenan, Y. L. Hurd, S. Dracheva, P. Casaccia, P. Roussos, B. Zhang, Brain cell type specific gene expression and co-expression network architectures. *Sci. Rep.* **8**, 8868 (2018).
27. B. Zhang, C. Gaiteri, L.-G. Bodea, Z. Wang, J. McElwee, A. A. Podtelezchnikov, C. Zhang, T. Xie, L. Tran, R. Dobrin, E. Fluder, B. Clurman, S. Melquist, M. Narayanan, C. Suver, H. Shah, M. Mahajan, T. Gillis, J. Mysore, M. E. MacDonald, J. R. Lamb, D. A. Bennett, C. Molony, D. J. Stone, V. Gudnason, A. J. Myers, E. E. Schadt, H. Neumann, J. Zhu, V. Emilsson, Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**, 707–720 (2013).
28. E. M. Blalock, H. M. Buechel, J. Popovic, J. W. Geddes, P. W. Landfield, Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease. *J. Chem. Neuroanat.* **42**, 118–126 (2011).
29. M. D. Sweeney, A. P. Sagare, B. V. Zlokovic, Blood–brain barrier breakdown in Alzheimer disease and other neurodegenerative disorders. *Nat. Rev. Neurol.* **14**, 133–150 (2018).
30. W. M. Song, B. Zhang, Multiscale embedded gene co-expression network analysis. *PLOS Comput. Biol.* **11**, e1004574 (2015).
31. M. Scutari, *Journal of Statistical Software Learning Bayesian Networks with the bnlearn R Package* (1996); www.jstatsoft.org/.
32. Q. Wang, Y. Zhang, M. Wang, W.-M. Song, Q. Shen, A. McKenzie, I. Choi, X. Zhou, P.-Y. Pan, Z. Yue, B. Zhang, The landscape of multiscale transcriptomic networks and key regulators in Parkinson's disease. *Nat. Commun.* **10**, 5234 (2019).
33. B. W. Kunkle, B. Grenier-Boley, R. Sims, J. C. Bis, V. Damotte, A. C. Naj, A. Boland, M. Vronskaya, S. J. van der Lee, A. Amalie-Wolf, C. Bellenguez, A. Frizzati, V. Chouraki, E. R. Martin, K. Sleegers, N. Badarinarayan, J. Jakobsdottir, K. L. Hamilton-Nelson, S. Moreno-Grau, R. Olaso, R. Raybould, Y. Chen, A. B. Kuzma, M. Hiltunen, T. Morgan, S. Ahmad, B. N. Vardarajan, J. Epelbaum, P. Hoffmann, M. Boada, G. W. Beecham, J. G. Garnier, D. Harold, A. L. Fitzpatrick, O. Valladares, M.-L. Moutet, A. Gerrish, A. V. Smith, L. Qu, D. Bacq, N. Denning, X. Jian, Y. Zhao, M. Del Zompo, N. C. Fox, S.-H. Choi, I. Mateo, J. T. Hughes, H. H. Adams, J. Malamon, F. Sanchez-Garcia, Y. Patel, J. A. Brody, B. A. Dombroski, M. C. D. Naranjo, M. Daniilidou, G. Eiriksdottir, S. Mukherjee, D. Wallon, J. Uphill, T. Aspelund, L. B. Cantwell, F. Garzia, D. Galimberti, E. Hofer, M. Butkiewicz, B. Fin, E. Scarpini, C. Sarnowski, W. S. Bush, S. Meslage, J. Kornhuber, C. C. White, Y. Song, R. C. Barber, S. Engelborghs, S. Sordon, D. Vojnovic, P. M. Adams, R. Vandenberghe, M. Mayhaus, L. A. Cupples, M. S. Albert, P. P. De Deyn, W. Gu, J. J. Himali, D. Beekly, A. Squassina, A. M. Hartmann, A. Orrelana, D. Blacker, E. Rodriguez-Rodriguez, S. Lovestone, M. E. Garcia, R. S. Doody, C. Munoz-Fernandez, R. Sussans, H. Lin, T. J. Fairchild, Y. A. Benito, C. Holmes, H. Karamujic-Comic, M. P. Frosch, H. Thonberg, W. Maier, G. Roschupkin, B. Ghetti, V. Giedraitis, A. Kawalia, S. Li, R. M. Huebinger, L. Kilander, S. Moebus, I. Hernandez, M. I. Kamboh, R. M. Brunini, J. Turton, Q. Yang, M. J. Katz, L. Concari, J. Lord, A. S. Beiser, C. D. Keene, S. Helisalmi, I. Kloszewska, W. A. Kukull, A. M. Koivisto, A. Lynch, L. Tarraga, E. B. Larson, A. Haapasalo, B. Lawlor, T. H. Mosley, R. B. Lipton, V. Solfrizzi, M. Gill, W. T. Longstreth, T. J. Montine, V. Frisardi, M. Diez-Fairen, F. Rivadeneira, R. C. Petersen, V. Deramecourt, I. Alvarez, F. Salani, A. Ciaramella, E. Boerwinkle, E. M. Reiman, N. Fievet, J. I. Rotter, J. S. Reisch, O. Hanon, C. Cupidi, A. G. A. Uitterlinden, D. R. Royall, C. Dufouil, R. G. Maletta, I. de Rojas, M. Sano, A. Brice, R. Cecchetti, P. S. George-Hyslop, K. Ritchie, M. Tsolaki, D. W. Tsuang, B. Dubois, D. Craig, C. K. Wu, H. Soininen, D. Avramidou, R. L. Albin, L. Fratiglioni, A. Germanou, L. G. Apostolova, L. Keller, M. Koutroumani, S. E. Arnold, F. Panza, O. Gkatzima, S. Asthana, D. Hannequin, P. Whitehead, C. S. Atwood, P. Caffarra, H. Hampel, I. Quintela, A. Carracedo, L. Lannfelt, D. C. Rubinsztein, L. L. Barnes, F. Pasquier, L. Frölich, S. Barral, B. McGinness, T. G. Beach, J. A. Johnston, T. L. Becker, P. Passmore, E. H. Bigio, J. M. Schott, T. D. Bird, J. D. Warren, B. F. Boeve, M. K. Lupton, J. D. Bowen, P. Proitsi, A. Boxer, J. F. Powell, J. R. Burke, J. S. K. Kauwe, J. M. Burns, M. Mancuso, J. D. Buxbaum, U. Bonuccelli, N. J. Cairns, A. McQuillin, C. Cao, G. Livingston, C. S. Carlson, N. J. Bass, C. M. Carlsson, J. Hardy, R. M. Carney, J. Bras, M. M. Carrasquillo, R. Guerreiro, M. Allen, H. C. Chui, E. Fisher, C. Masullo, E. A. Crocco, C. DeCarli, G. Bisceglia, M. Dick, L. Ma, R. Duara, N. R. Graff-Radford, D. A. Evans, A. Hodges, K. M. Faber, M. Scherer, K. B. Fallon, M. Riemschneider, D. W. Fardo, R. Heun, M. R. Farlow, H. Kölsch, S. Ferris, M. Leber, L.-W. Jin, G. Leonenko, L. M. Real, G. R. Jun, C. T. M. Baldwin, D. Grozeva, A. Karydas, G. Russo, J. A. Kaye, R. Kim, F. Jessen, N. W. Kowall, B. Vellas, J. H. Kramer, E. Vardy, F. M. LaFerla, K.-H. Jöckel, J. J. Lah, M. Dichgans, J. B. Leverenz, D. Mann, A. I. Levey, S. Pickering-Brown, A. P. Lieberman, N. Klopp, K. L. Lunetta, H. E. Wichmann, C. G. Lyketsos, K. Morgan, D. C. Marson, K. Brown, F. Martiniuk, C. Medway, D. C. Mash, M. M. Nöthen, E. Masliah, N. M. Hooper, W. C. McCormick, A. Daniele, S. M. McCurry, A. Bayer, A. N. McDavid, J. Gallacher, A. C. McKee, H. van den Bussche, M. Mesulam, C. Brayne, B. L. Miller, S. Riedel-Heller, C. A. Miller, J. W. Miller, A. Al-Chalabi, J. C. Morris, C. E. Shaw, A. J. Myers, J. Wiltfang, S. O'Bryant, J. M. Olichney, V. Alvarez, J. E. Parisi, A. B. Singleton, H. L. Paulson, J. Collinge, W. R. Perry, S. Mead, E. Peskind, D. H. Cribbs, M. Rossor, A. Pierce, N. S. Ryan, W. W. Poon, B. Nacmias, H. Potter, S. Sorbi, J. F. Quinn, E. Sacchinelli, A. Raj, G. Spalletta, M. Raskind, C. Caltagirone, P. Bossù, M. D. Orfei, B. Reisberg, R. Clarke, C. Reitz, A. D. Smith, J. M. Ringman, D. Warden, E. D. Roberson, G. Wilcock, E. Rogaevea, A. C. Bruni, H. J. Rosen, M. Gallo, R. N. Rosenberg, Y. Ben-Shlomo, M. A. Sager, P. Mecocci, A. J. Saykin, P. Pastor, M. L. Cuccaro, J. M. Vance, J. A. Schneider, L. S. Schneider, S. Slifer, W. W. Seeley, A. G. Smith, J. A. Sonnen, S. Spina, R. A. Stern, R. H. Swerdlow, M. Tang, R. E. Tanzi, J. C. Trojanski, J. C. Troncoso, V. M. Van Deerlin, L. J. Van Eldik, H. V. Vinters, J. P. Vonsattel, S. Weintraub, K. A. Welsh-Bohmer, K. C. Wilhelmsen, J. Williamson, T. S. Wingo, R. L. Woljter, C. B. Wright, C. E. Yu, L. Yu, Y. Saba, A. Pilotto, M. J. Bullido, O. Peters, P. K. Crane, D. Bennett, P. Bosco, E. Coto, V. Boccardi, P. L. De Jager, A. Lleo, N. Warner, O. L. Lopez, M. Ingelsson, P. Deloukas, C. Cruchaga, C. Graff, R. Gwilliam, M. Fornage, A. M. Goate, P. Sanchez-Juan, P. G. Kehoe, N. Amin, N. Ertekin-Taner, C. Berr, S. DeBette, S. Love, L. J. Launer, S. G. Younkin, J. F. Dartigues, C. Corcoran, M. A. Ikram, D. W. Dickson, G. Nicolas, D. Campion, J. A. Tschanz, H. Schmidt, H. Hakonarson, J. Clarimon, R. Munger, R. Schmidt, L. A. Farrer, C. Van Broeckhoven, M. C. O'Donovan, A. L. DeStefano, L. Jones, J. L. Haines, J. F. Deleuze, M. J. Owen, V. Gudnason, R. Mayeux, V. Escott-Price, B. M. Psaty, A. Ramirez, L. S. Wang, A. Ruiz, C. M. van Duijn, P. A. Holmans, S. Seshadri, J. Williams, P. Amouyel, G. D. Schellenberg, J.-C. Lambert, M. A. Pericak-Vance; Alzheimer Disease Genetics Consortium (ADGC); European Alzheimer's Disease Initiative (EADI); Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE); Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES), Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
34. J. Euesden, C. M. Lewis, P. F. O'Reilly, PRSice: Polygenic risk score software. *Bioinformatics* **31**, 1466–1468 (2015).
35. P. De Rossi, V. Buggia-Prévot, B. L. L. Clayton, J. B. Vasquez, C. van Sanford, R. J. Andrew, R. Lesnick, A. Botté, C. Deyts, S. Salem, E. Rao, R. C. Rice, A. Parent, S. Kar, B. Popko, P. Pytel,

- S. Estus, G. Thinakaran, Predominant expression of Alzheimer's disease-associated BIN1 in mature oligodendrocytes and localization to white matter tracts. *Mol. Neurodegener.* **11**, 59 (2016).
36. S. Calafate, W. Flavin, P. Verstreken, D. Moechars, Loss of Bin1 promotes the propagation of tau pathology. *Cell Rep.* **17**, 931–940 (2016).
37. C. Werner, M. Pauli, S. Doose, A. Weishaupt, H. Haselmann, B. Grünwald, M. Sauer, M. Heckmann, K. V. Toyka, E. Asan, C. Sommer, C. Geis, Human autoantibodies to amphiphysin induce defective presynaptic vesicle dynamics and composition. *Brain* **139**, 365–379 (2016).
38. D. A. Bennett, J. A. Schneider, Z. Arvanitakis, R. S. Wilson, Overview and findings from the religious orders study. *Curr. Alzheimer Res.* **9**, 628–645 (2012).
39. R. Perneckzy, S. Wagenpfeil, K. Komossa, T. Grimmer, J. Diehl, A. Kurz, Mapping scores onto stages: Mini-mental state examination and clinical dementia rating. *Am. J. Geriatr. Psychiatry* **14**, 139–144 (2006).
40. Jax Mouse Model 005642 - B6.Cg-*Clu^{tm1.Jokh}/J* (2020); www.jax.org/strain/005642.
41. L. Liu, R. J. Caselli, Age stratification corrects bias in estimated hazard of *APOE* genotype for Alzheimer's disease. *Alzheimer's Dementia* **4**, 602–608 (2018).
42. A. Norambuena, H. Wallrabe, R. Cao, D. B. Wang, A. Silva, Z. Svindrych, A. Periasamy, S. Hu, R. E. Tanzi, D. Y. Kim, G. S. Bloom, A novel lysosome-to-mitochondria signaling pathway disrupted by amyloid- β oligomers. *EMBO J.* **37**, e100241 (2018).
43. S. S. Khan, M. LaCroix, G. Boyle, M. A. Sherman, J. L. Brown, F. Amar, J. Aldaco, M. K. Lee, G. S. Bloom, S. E. Lesné, Bidirectional modulation of Alzheimer phenotype by alpha-synuclein in mice and primary neurons. *Acta Neuropathol.* **136**, 589–605 (2018).
44. A. Ruiz, S. Heilmann, T. Becker, I. Hernández, H. Wagner, M. Thelen, A. Mauleón, M. Rosende-Roca, C. Bellenguez, J. C. Bis, D. Harold, A. Gerrish, R. Sims, O. Sotolongo-Grau, A. Espinosa, M. Alegret, J. L. Arrieta, A. Lacour, M. Leber, J. Becker, A. Lafuente, S. Ruiz, L. Vargus, O. Rodríguez, G. Ortega, M. A. Dominguez, R. Mayeux, J. L. Haines, M. A. Pericak-Vance, L. A. Farrer, G. D. Schellenberg, V. Chouraki, L. J. Launer, C. Van Duijn, S. Seshadri, C. Antúnez, M. M. Breteler, M. Serrano-Rios, F. Jessen, L. Tárraga, M. M. Nöthen, W. Maier, M. Boada, A. L. Ramírez, Follow-up of loci from the International Genomics of Alzheimer's Disease Project identifies *TRIP4* as a novel susceptibility gene. *Transl. Psychiatry* **4**, e358 (2014).
45. C. M. Karch, C. Cruchaga, A. M. Goate, Alzheimer's disease genetics: From the bench to the clinic. *Neuron* **83**, 11–26 (2014).
46. O. Hansson, H. Zetterberg, P. Buchhave, E. Londos, K. Blennow, L. Minthon, Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: A follow-up study. *Lancet Neurol.* **5**, 228–234 (2006).
47. X. Zhou, M. Wang, I. Katsyv, H. Irie, B. Zhang, EMUDRA: Ensemble of multiple drug repositioning approaches to improve prediction accuracy. *Bioinformatics* **34**, 3151–3159 (2018).
48. A. T. McKenzie, S. Moyon, M. Wang, I. Katsyv, W.-M. Song, X. Zhou, E. B. Dammer, D. M. Duong, J. Aaker, Y. Zhao, N. Beckmann, P. Wang, J. Zhu, J. J. Lah, N. T. Seyfried, A. I. Levey, P. Katsel, V. Haroutunian, E. E. Schadt, B. Popko, P. Casaccia, B. Zhang, Multiscale network modeling of oligodendrocytes reveals molecular components of myelin dysregulation in Alzheimer's disease. *Mol. Neurodegener.* **12**, 82 (2017).
49. B. Zhang, J. Zhu, in *Proceedings of the World Congress on Engineering 2013* (International Association of Engineers), vol. 2, pp. 1309–1312.
50. J. Millstein, B. Zhang, J. Zhu, E. E. Schadt, Disentangling molecular relationships with a causal inference test. *BMC Genet.* **10**, 23 (2009).
51. J. Zhu, M. C. Wiener, C. Zhang, A. Fridman, E. Minch, P. Y. Lum, J. R. Sachs, E. E. Schadt, Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. *PLoS Comput. Biol.* **3**, 692–703 (2007).
52. Y. Zhang, S. A. Sloan, L. E. Clarke, C. A. Caneda, C. A. Plaza, P. D. Blumenthal, H. Vogel, G. K. Steinberg, M. S. B. Edwards, G. Li, J. A. Duncan, S. H. Cheshier, L. M. Shuer, E. F. Chang, G. A. Grant, M. G. H. Gephart, B. A. Barres, Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron* **89**, 37–53 (2016).

Acknowledgments: We thank M. Peters for assisting with the preparation of documents and files for upload and sharing via the AD Knowledge Portal. **Funding:** Data collection was supported through funding by National Institute on Aging (NIA) (grants U01AG046170, RF1AG054014, RF1AG057440, R01AG057907, U01AG052411, R01AG062355, U01AG058635, R01AG068030, P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984, U01AG46152, U01AG52411, K01AG062683, and U01AG058635), the Illinois Department of Public Health, and the Translational Genomics Research Institute. **Ethics statement:** All IRB guidelines were followed with human or animal subjects' data when applicable. **Author contributions:** B.Z. perceived the concept and designed the study. R.A.N. performed the research. M.W., S.V., L.G., C.M., Q.W., E.W., E.H.-M., and W.-m.S. curated data and provided software tools. A.L., E.L.C., J.TCW., L.H., A.G., V.F., S.N., S.G., M.E.E., P.K., E.S., D.C., K.J.B., and V.H. participated in the discussion of the results. R.A.N. and B.Z. wrote the paper. All the authors reviewed and revised the paper. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The human postmortem sequencing data are available via the AD Knowledge Portal (<https://adknowledgeportal.synapse.org>). The AD Knowledge Portal is a platform for accessing data, analyses, and tools generated by the Accelerating Medicines Partnership Alzheimer's Disease (AMP-AD) Target Discovery Program and other NIA-supported programs to enable open-science practices and accelerate translational learning. The data, analyses, and tools are shared early in the research cycle without a publication embargo on secondary use. Data are available for general research use according to the following requirements for data access and data attribution (<https://adknowledgeportal.synapse.org/DataAccess/Instructions>). For access to content described in this manuscript see: <https://doi.org/10.7303/syn23521805>. In addition, all data needed to evaluate the conclusions presented in the paper and the Supplementary Materials are available from the authors upon request.

Submitted 2 March 2020

Accepted 12 November 2020

Published 6 January 2021

10.1126/sciadv.abb5398

Citation: R. A. Neff, M. Wang, S. Vatansever, L. Guo, C. Ming, Q. Wang, E. Wang, E. Horgusluoglu-Moloch, W.-m. Song, A. Li, E. L. Castranio, J. TCW, L. Ho, A. Goate, V. Fossati, S. Noggle, S. Gandy, M. E. Ehrlich, P. Katsel, E. Schadt, D. Cai, K. J. Brennand, V. Haroutunian, B. Zhang, Molecular subtyping of Alzheimer's disease using RNA sequencing data reveals novel mechanisms and targets. *Sci. Adv.* **7**, eabb5398 (2021).