Open access • Posted Content • DOI:10.1101/510420

Meta-analysis of the human brain transcriptome identifies heterogeneity across human AD coexpression modules robust to sample collection and methodological approach

Benjamin A. Logsdon, Thanneer M. Perumal, Vivek Swarup, Minghui Wang ...+40 more authors

Institutions: Sage Bionetworks, University of California, Los Angeles, Icahn School of Medicine at Mount Sinai, Institute for Systems Biology ...+13 more institutions

Published on: 03 Jan 2019 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Sample collection

Meta-analysis of the human brain transcriptome identifies heterogeneity across human AD coexpression modules robust to sample collection and methodological approach

Benjamin A. Logsdon[1,2*†], Thanneer M. Perumal[*1], Vivek Swarup[3], Minghui Wang[4], Cory Funk[5], Chris Gaiteri[6], Mariet Allen[7], Xue Wang[7,8], Eric Dammer[9], Gyan Srivastava[10], Sumit Mukherjee[1], Solveig K. Sieberts[1], Larsson Omberg[1], Kristen D. Dang[1], James A. Eddy[1], Phil Snyder[1], Yooree Chae[11], Sandeep Amberkar[12], Wenbin Wei[12], Winston Hide[12], Christoph Preuss[13], Ayla Ergun[14], Phillip J Ebert[15], David C. Airey[15], Gregory W. Carter[14], Sara Mostafavi[16], Lei Yu[6], Hans-Ulrich Klein[17], the AMP-AD Consortium[18], David A. Collier[15], Todd Golde[19], Allan Levey[9], David A. Bennett[6], Karol Estrada[20], Michael Decker[10], Zhandong Liu[21,22], Joshua M. Shulman[22,23], Bin Zhang[4], Eric Schadt[4], Phillip L. De Jager[17], Nathan D. Price[5], Nilüfer Ertekin-Taner[7,24], Lara M. Mangravite[1†]

*These authors contributed equally to this work.

[1]Sage Bionetworks, Seattle, WA, 98121, USA
[2]Lead contact
[3]Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA.
[4]Department of Genetics and Genomic Sciences, Mount Sinai Center for Transformative Disease Modeling, Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA
[5]Institute for Systems Biology, Seattle, WA, USA
[6]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA
[7]Department of Neuroscience, Mayo Clinic, Jacksonville, FL, 32224, USA.
[8]Department of Health Sciences Research, Mayo Clinic, Jacksonville, FL, 32224, USA.
[9]Emory University, Atlanta, GA, USA
[10]AbbVie, Boston, MA, USA
[11]Genentech, South San Francisco, CA, 94080, USA
[12]Sheffield Institute for Translational Neuroscience, University of Sheffield, Sheffield, UK
[13]The Jackson Laboratory, 600 Main Street, Bar Harbor, ME, USA 04609
[14]Biogen, Cambridge, MA, USA
[15]Eli Lilly, Indianapolis, IN, USA
[16]University of British Columbia, Vancouver, BC, Canada
[17]Center for Translational & Computational Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York, NY, USA
[18]Full list of consortia authors and affiliations (doi:10.7303/syn17114455).
[19]University of Florida, Gainesville, FL, USA
[20]Translational Genome Sciences, Biogen, Cambridge, MA 02142 USA
[21]Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA
[22]Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX, USA
[23]Departments of Neurology, Neuroscience, and Molecular & Human Genetics, Baylor College of Medicine, Houston, TX, USA
[24]Department of Neurology, Mayo Clinic, Jacksonville, FL, 32224, USA.

†Correspondence:

Lara M Mangravite, PhD
Sage Bionetworks
2901 3rd Avenue
Seattle, WA 98121
lara.mangravite@sagebionetworks.org

Benjamin A Logsdon, PhD
Sage Bionetworks
2901 3rd Avenue
Seattle, WA 98121
ben.logsdon@sagebionetworks.org

SUMMARY

Alzheimer's disease (AD) is a complex and heterogenous brain disease that affects multiple inter-related biological processes. This complexity contributes, in part, to existing difficulties in the identification of successful disease-modifying therapeutic strategies. To address this, systems approaches are being used to characterize AD-related disruption in molecular state. To evaluate the consistency across these molecular models, a consensus atlas of the human brain transcriptome was developed through coexpression meta-analysis across the AMP-AD consortium. Consensus analysis was performed across five coexpression methods used to analyze RNA-seq data collected from 2114 samples across 7 brain regions and 3 research studies. From this analysis, five consensus clusters were identified that described the major sources of AD-related alterations in transcriptional state that were consistent across studies, methods, and samples. AD genetic associations, previously studied AD-related biological processes, and AD targets under active investigation were enriched in only three of these five clusters. The remaining two clusters demonstrated strong heterogeneity between males and females in AD-related expression that was consistently observed across studies. AD transcriptional modules identified by systems analysis of individual AMP-AD teams were all represented in one of these five consensus clusters except ROS/MAP-identified Module 109, which was specific for genes that showed the strongest association with changes in AD-related gene expression across consensus clusters. The other two AMP-AD transcriptional analyses reported modules that were enriched in one of the two sex-specific Consensus Clusters. The fifth cluster has not been previously identified and was enriched for genes related to proteostasis. This study provides an atlas to map across biological inquiries of AD with the goal of supporting an expansion in AD target discovery efforts.

INTRODUCTION

Alzheimer's Disease (AD) is a debilitating neurodegenerative disease affecting more than 5 million Americans for which we lack effective long-term disease-modifying therapeutic strategies (Cummings et al., 2014). Several therapeutic mechanisms are under active evaluation in clinical trials (Kumar et al.,

2015) across the field - including the amyloid hypothesis. Because AD is likely to result from molecular dysregulation across a series of biological systems within the brain (De Strooper and Karran, 2016), there is some question as to whether therapeutic targeting of a single pathway will be sufficient to completely address the full burden of this disease. Furthermore, recent evidence suggests that AD may be a collection of conditions with multiple underlying causes that lead to similar symptomatic and pathological end points (Brenowitz et al., 2017; Winblad et al., 2016). For these reasons, there is need to pursue a diverse set of mechanistic hypotheses for therapeutic intervention.

Systems biology analysis can provide a rich mapping of the inter-related molecular dysregulations involved in AD that may be useful to guide drug target discovery towards a diverse set of complementary therapeutic mechanisms. Several systems-level analyses of human AD brain have been previously reported (Allen et al., 2018a, 2018b; Lu et al., 2014; Mostafavi et al., 2018; Seyfried et al., 2017; Zhang et al., 2013a). The importance of neuroinflammation in AD has been described across these (Patrick et al., 2017; Zhang et al., 2013a) and also from genetic studies (Carrasquillo et al., 2017; Efthymiou and Goate, 2017; Guerreiro et al., 2013; Jin et al., 2015; Jonsson et al., 2013; Raj et al., 2014; Sims et al., 2017), supporting a major focus on this pathway for therapeutic development as is currently underway (Ardura-Fabregat et al., 2017). In addition to neuroinflammation, other molecular pathways have been identified from systems biology studies of both RNA and protein abundance including multiple processes related to oligodendrocytic functions such as myelination (Allen et al., 2018a; McKenzie et al., 2017; Mostafavi et al., 2018; Seyfried et al., 2017). In several cases, these compelling observations arise from individual studies and, as of yet, their reproducibility is unknown.

This project aims to develop an atlas of AD-associated changes in molecular state that provides a mechanism to evaluate the consistency and robustness of systems analyses and the use of their findings to support AD target discovery. To this aim, we build on the resources and expertise gathered across the Accelerating Medicines Partnership in Alzheimer's Disease Target Discovery and Preclinical Validation project (AMP-AD – ampadportal.org). AMP-AD focuses on identification of AD disease drivers using

104 systems-level evaluation of disease state in human brain tissue. A major early outcome of this consortium
105 was the generation and public release of RNA-seq data generated across three sizable but distinct human
106 postmortem brain studies that are distributed through the AMP-AD Knowledge Portal –
107 https://ampadportal.org (Allen et al., 2016; Jager et al., 2018; Wang et al., 2018). Here, we use
108 coexpression meta-analysis across these studies to develop a robust systems-level molecular atlas of AD.
109 Coexpression network analysis is a commonly used data-driven approach to identify gene sets (or
110 modules) that are similarly co-expressed across samples in a data set (Langfelder and Horvath, 2008a).
111 These modules are often comprised of genes involved in biological processes that interact and/or exhibit
112 coordinated activity in response to molecular and cellular states, pathological processes, and other factors
113 (Gaiteri et al., 2015). Although distinct and important biology may be uniquely represented in any one of
114 these studies, meta-analysis provides a generalized illustration of the changes in transcriptional state
115 associated with AD in a manner that is robust to technical confound and study heterogeneity. An atlas
116 derived of cross-study AD-associated transcriptional modules can support target discovery by (1)
117 promoting target discovery across multiple distinct biological processes, (2) informing experimental
118 design for target validation studies, (3) creation of improved experimental models and assessment of
119 current experimental models (Wan et al., submitted), and (4) evaluating population heterogeneity in
120 disease pathophysiology that may impact therapeutic efficacy.
121
122 **RESULTS**
123 *AMP-AD collection of human RNA-seq data.* We analyzed existing transcriptional data generated from
124 post-mortem brain tissue homogenate from three separate sample sets including the Religious Order
125 Study and the Memory and Aging Project (ROSMAP) (Bennett et al., 2012b, 2012a; Jager et al., 2018),
126 the Mount Sinai Brain Bank (MSBB) RNA-seq study (Wang et al., 2018), and the Mayo RNA-seq study
127 (Allen et al., 2016) (Mayo). Samples were collected from seven distinct brain tissues - dorsolateral
128 prefrontal cortex (DLPFC) in ROSMAP; temporal cortex (TCX) and cerebellum (CBE) in Mayo, and
129 inferior frontal gyrus (IFG), superior temporal gyrus (STG), frontal pole (FP), and parahippocampal gyrus

(PHG) in MSBB. Several differences in data collection and processing protocols across studies were identified and accounted for during data processing and analysis (see **Table 1** and **Methods**).

Development of AD-related transcriptional modules by consensus coexpression network analysis. To identify AD-related human transcriptional modules that were robustly observed in a generalized manner across methods and studies, we performed a consensus analysis for all seven tissue types using five coexpression analysis methodologies. These five distinct coexpression learning algorithms included: MEGENa (Song and Zhang, 2015), rWGCNA (Parikshak et al., 2016), metanetwork (**Methods**), WINA (Wang et al., 2016) and SpeakEasy (Gaiteri et al., 2015). Independent performance of each of the 5 methods across each of the 7 tissues identified 2,978 tissue-specific coexpression modules (CBE: 458, DLPFC: 450, FP: 393, IFG: 429, PHG: 370, STG: 336, TCX: 502, 10.7303/syn10309369.1). As expected, similar coexpression structure was observed in each data set across methods, as indicated by significant overlap in module memberships (**Figure S1**). Within each tissue, we next identified AD-related modules that were well-represented across analysis methodologies. This analysis was limited to those modules that were significantly enriched for differentially expressed genes related to AD based on a meta-analysis of differential expression across the seven brain regions (**Methods**). To do this, graph clustering (Pons and Latapy, 2005) was performed on all modules within a tissue with an edge betweenness community identification algorithm (Pons and Latapy, 2005) and weights from the Fisher's exact test estimate (**Methods**, **Figure S1**). This meta-analysis of coexpression modules and differential expression signatures identified 30 AD associated modules across the seven tissue types (CBE: 4, DLPFC: 4, FP: 4, IFG: 4, PHG: 5, STG: 4, TCX: 5, 10.7303/syn11932957.1).

To establish confidence that these consensus modules provided an improved and coherent representation of AD-altered biology, they were evaluated for enrichment of gene sets previously identified as relevant to AD (**Figure 1** and **Methods**). The consensus modules showed an improved percentage enrichment for previously published AD related gene sets or pathways relative to (a) the differential expression meta-analysis gene set (P-value = 0.036, Wilcoxon rank sum test), (b) the modules

defined by individual coexpression methods (P-value < $2x10^{-16}$ Wilcoxon rank sum test) and (c) those modules defined previously in the literature (Zhang et al., 2013b) (P-value $3.9x10^{-13}$ Wilcoxon rank sum test) (**Figure 1**). Evaluation across tissues demonstrated that these 30 consensus modules fell into five well-defined clusters that were highly preserved across study and tissue type and demonstrated a low degree of overlap between clusters (**Figure 2A**) based on a Fisher's exact test of gene membership overlap. These five 'consensus clusters' were used to represent distinct patterns of AD-related transcriptional state that were consistently observed.

Because previous studies have identified cell-type specific changes from transcriptional analysis, we next evaluated enrichment of cell-type specific gene signatures as identified from previously published cell type specific RNA-seq data from Zhang et al (Zhang et al., 2014). While most of the genes represented in these consensus clusters were not cell type specific, we did observe that genes represented in cell-type specific gene signatures were grouped by cluster. As shown in **Figure 2B**, cell-type specific gene sets clearly clustered into four of five consensus clusters: the astrocytic signature was enriched in Consensus Clusters A and B, the endothelia and microglial signatures were enriched in Consensus Cluster B, the neuronal signature was enriched in Consensus Cluster C, and the oligodendroglial signature was enriched in Consensus Cluster D. While we see significant enrichment for cell type specific gene-sets in Consensus Cluster A-D, these modules are large (2090 ± 1150 genes, **Table S3**), and show enrichment for a diverse set of biological processes beyond cell type specific processes (10.7303/syn11954640.1). Accordingly, Consensus Cluster E was not enriched for cell type specific signatures, but instead was consistently enriched for genes that were associated with proteostasis – including with the attenuation phase of the transcriptional response to heat shock (Abravaya et al., 1991; Fabregat et al., 2016), detection of unfolded protein (GO:0002235) (2015), response to unfolded protein (GO:0006986) (2015), and HSF1 activation (Cotto et al., 1996; Fabregat et al., 2016; Zuo et al., 1995) (**Table 2**).

Heterogeneity in expression of consensus clusters between females and males. To evaluate whether consensus clusters may prove useful in the identification of molecular heterogeneity in disease across

182 populations, we evaluated sex-specific differences across clusters in AD-related gene expression. Indeed,
183 sex-specific differential expression gene (DEG) sets – from a sex specific meta-analysis of differential
184 expression across the seven tissue types - were differentially enriched across the consensus clusters for
185 females vs. males (**Figure 2D**). Consensus Clusters A and B demonstrated similar direction of DEGs
186 across sex through enrichment was stronger in females vs. males. This suggests that transcriptional
187 changes related to neuroinflammatory processes were common across the sexes, albeit more pronounced
188 in females. In contrast, the Consensus Clusters C and D were strongly enriched for genes altered in
189 females but not in males, suggesting that the overall association of these clusters with AD-related
190 differential expression was predominantly driven by females. Consensus Cluster E was enriched for genes
191 that were down-regulated in both male and female AD cases. This last cluster was moderately enriched
192 for genes that were up-regulated in male AD cases, with no such enrichment observed in females. These
193 changes demonstrate that sex-specific changes in AD-related gene expression are heterogenous across
194 consensus cluster.
195
196 *Use of consensus modules as an atlas to evaluate diversity of AD target discovery efforts.* The consensus
197 clusters can be used in aggregate as an atlas to support the selection of a robust and diverse set of AD
198 therapeutic hypotheses for target discovery. To this aim, we next evaluated the enrichment across clusters
199 of AD biology under active investigation for target discovery (**Figure 2C**). Consensus Clusters A, B, and
200 C were enriched for AD pathways derived from the scientific literature (Amberger et al., 2015; Kanehisa
201 et al., 2017; Kutmon et al., 2016; Lambert et al., 2013; Mi et al., 2017; Nishimura, 2001; Safran et al.,
202 2010; Tryka et al., 2014), AD gene sets (Lambert et al., 2013; Tryka et al., 2014) derived from genetic
203 association analyses, and pathways related to therapeutic hypotheses currently undergoing active drug
204 development including the amyloid secretase pathway (neuronal cluster), the Presenilin amyloid
205 processing pathway (astrocytic cluster), and deregulation of *CDK5* pathway that is implicated in Tau
206 hyperphosphorylation (neuronal cluster).

We next evaluated the consensus clusters for enrichment of modules identified for AD target discovery by the individual teams within the AMP-AD consortium, as part of their systems biology-based target discovery programs. Each team had identified AD-related modules through study-specific analyses (Allen et al., 2018a; Johnson et al., 2018; McKenzie et al., 2017; Mostafavi et al., 2018). We evaluated how the primary findings from these individual analyses mapped onto the consensus clusters. First, we examined enrichment the results published from independent analysis of the ROS/MAP study (Mostafavi et al., 2018). This original analysis identified several modules that were associated with rate of cognitive decline including module 109. Unlike the modules previously identified within the literature or by the other teams (see below), module 109 membership did not group into a single consensus cluster. Instead, module 109 membership was enriched across 4 of the 5 consensus clusters. Strikingly, module 109 was the most strongly enriched for genes that are up in AD cases (FET OR, P-value: 9.8, $2x10^{-72}$), up in male AD cases (FET OR, P-value 7.4, $1.1x10^{-22}$), and up in female AD Cases (FET OR, P-value: 9.2,$2x10^{-78}$, **Table S5, Table S7, Table S8)** among all AD signatures we tested.

In contrast, the modules identified by the other AMP-AD teams were all enriched within a single consensus cluster. The Mayo and Mt Sinai teams each identified a separate module that was enriched for oligodendrocyte signatures and that significantly overlapped with consensus cluster D (**Figure 2C, Table S4**). Notably, AD-related decreases in expression of genes within these two modules were reported by each team. Because we observed sex-specific differences in AD gene expression in consensus cluster D but results from sex-specific analyses were not reported by either team, we evaluated the Mayo and Mt. Sinai modules for sex-specificity. Indeed, the sex-specific pattern of AD-related expression were also consistently observed in both the Mayo and Mt. Sinai modules. For the Mayo module, the effect size for AD-related increased expression in females was much smaller than the effect size for AD-related decreased expression in males (**Table S8, S10**), supporting a modest decrease in expression based on combined analysis across all AD cases as was reported by the Mayo team (Allen et al., 2018a). The Mt. Sinai module demonstrated the same AD-related increased expression in females but no significant changes in males. The genes represented in the Mt Sinai module may be specific to the female signature

because Mt. Sinai module membership was derived from analysis of a sample set with a preponderance of female samples. Finally, we evaluated cluster enrichment for an RNA-binding module identified by the Emory AMP-AD team (Johnson et al., 2018) from systems analysis of proteomic data. This was enriched in the Consensus Cluster B suggesting that it is co-expressed with genes involved with synaptic function.

Finally, we evaluated enrichment for the one hundred genes nominated by the AMP-AD consortia as the first set of candidates for AD target evaluation (https://agora.ampadportal.org). Because these targets were selected in part based on analysis of these data, we expected to observe significant enrichments within the consensus clusters. Interestingly, significant enrichment was observed (adjusted P-value < 0.05), but this was specific to Consensus Clusters A, B, and C – those that were also enriched for previously known AD biology processes. This suggests that the initial round of AMP-AD target nominations was guided by data-driven analysis in combination with evaluation of prior biological knowledge. Subsequent nominations would benefit from an expansion into biology represented within the other consensus clusters that are equally robust but have been less studied in the context of AD – particularly biology within Consensus Cluster D that was observed across multiple independent analyses.

DISCUSSION

Identification of therapies for the treatment or prevention of AD has been hampered by many difficulties including a limited pipeline of well-validated targets (Kumar et al., 2015). In part, this is because target discovery and validation has been plagued by multiple issues including: (a) poor understanding of the complex inter-related biological processes that are dysregulated with AD on a systems level (De Strooper and Karran, 2016), (b) presence of other aging-related neuropathologies that confound interpretation of differential expression studies(De Jager et al., 2018), (c) evaluation of AD biology in experimental model systems that do not effectively recapitulate human disease (King, 2018), and (d) reliance on therapeutic hypotheses that show no efficacy in late stage clinical trials (Makin, 2018). Using molecular data collected from human brains, this project provides an overview of the systems-level models of AD state in human brain that can be used to inform identification and assessment of complementary target

hypotheses. Human brain transcriptional data collected from three cohorts were used to define a robust, reproducible set of human AD associated coexpression modules by consensus network analysis across five co-expression network methodologies. Consensus modeling provided a set of generalizable observations that robustly define AD-associated dysregulation in transcriptional state, and as such provide a resource to guide target selection and validation strategies.

The consensus analysis identified 5 consensus clusters that represented distinct patterns of AD-related changes in gene expression. The observed AD-associated changes in transcriptional state were consistently observed across all brain regions except cerebellum in terms of the differential expression patterns (the coexpression patterns are conserved). The reason why these signatures are not observed with AD in cerebellum is unknown but may be caused by region-specific differences in AD-associated transcriptional dysregulation, in AD pathology, and/or in cellular resilience to AD pathology (Stowell et al., 2018). The interpretation of these differences is also confounded by the basic differences between cerebellar cortex and cerebral cortex in terms of cell composition, cellular architecture, and function (von Bartheld et al., 2016).

These consensus clusters provide a general framework for evaluating heterogeneity in disease across populations. In this analysis, few observed a drastic difference in AD-related expression changes in females vs. males within each consensus cluster. For 4 of 5 Consensus Clusters, females exhibited significantly greater AD-associated expression changes as compared to males. This included greater increases in expression of the module clusters that were enriched for astrocytic, microglial, endothelial, and oligodendroglial signatures and greater reduction in expression of those enriched for neuronal signatures. In contrast, AD-associated alterations in expression of Consensus Cluster E was more prominent in males than in females. This cluster was enriched for response to unfolded protein and heat shock response.

Increasingly, the literature supports differences between males and females in AD progression, although it is unknown whether these are caused by differences in AD-mediated processes, in rate of progression within comparable processes, in pre-disease state or in some other cause (Mielke et al., 2014;

285 [CSL STYLE ERROR: reference with no printed form.]). Furthermore, we see evidence of sex specific
286 differences in genetic regulation of disease (Nazarian et al., 2018), including at the level of expression of
287 the oligodendrocyte myelinating cell module. A suggestive hypothesis is that AD genetic loci identified
288 to date are highly enriched in neuroinflammatory modules *precisely* because they show similar biology (at
289 least at the transcriptomic level) in both men and women (**Figure 2D**). This indicates that genetic
290 association analyses stratified by sex may further illuminate some of the missing genetic factors
291 underlying Alzheimer's disease. Preliminary evidence suggests that this may be the case: a genetic risk
292 score calculated in ROSMAP based on the 21 IGAP risk loci was associated with an eigengene in the
293 oligodendrocyte consensus cluster (specifically the DLPFC brown module) (adjusted p-value, Bonferroni:
294 10^{-2}), but only in females and was significantly different from males. Further disentangling the role of
295 sex and genetics in understanding disease heterogeneity will be key to development of efficacious
296 therapeutic interventions, especially if there are different underlying mechanisms driving disease etiology
297 between men and women.

298 Identification of conserved human AD-related consensus clusters provides several benefits in the
299 pursuit of AD target discovery. First, they highlight the distinct aspects of brain biology that are
300 dysregulated with AD– including several that are not currently under active investigation for drug
301 development. While not all dysregulated pathways are likely to be causative, these results suggest that a
302 broader range of therapeutic hypotheses exist and serves to guide researchers to areas of biology that may
303 merit further pursuit. In this manner, the consensus clusters were used to demonstrate three major areas of
304 AD-related biology that are under active pursuit for drug discovery and a fourth area of interest. This
305 fourth, represented by Consensus Cluster D, had a complex subcluster architecture that may contain
306 several biological processes of interest for further pursuit. Indeed, two of these subclusters were
307 independently identified and are under active pursuit by the AMP-AD Mayo and Mt. Sinai target
308 discovery teams respectively. We note that the use consensus methodologies is explicitly designed to
309 identify signatures of disease that are most robust to technical and study specific heterogeneity and, as
310 such, can provide evidence to support costly drug discovery programs. This does not preclude the

311 relevance of other interesting biology that was not recapitulated across studies due to small effect size or
312 uneven representation across studies based on differences in sample ascertainment.

313 In addition to evaluating diversity in AD therapeutic hypotheses, these human AD-related clusters
314 can be used to identify appropriate experimental model systems for further evaluation. Traditionally, AD
315 model systems have been developed through genetic perturbations of one or more AD-related pathways
316 (King, 2018). While none of these models provides a complete recapitulation of human disease, many
317 provide a useful framework to evaluate dysregulation within specific pathways. Since a subset of human
318 co-expression clusters display conserved co-expression and/or overlapping differential expression in
319 brains from AD mouse models, these human clusters may help to assess the appropriateness of AD
320 experimental models for pathway-specific evaluation – as well as to highlight other genetic perturbations
321 that may provide useful model systems to complement those that are commonly used(Mostafavi et al.,
322 2018; Neuner et al., 2018, Wan et al. submitted).

323 This analysis provides an important first step in developing a molecular framework to evaluate
324 and promote diversity in AD target discovery. There are a handful of caveats to the approach taken in this
325 study. First and foremost, this study focuses on transcriptional measures of disease response in post-
326 mortem brain and, as such, provides an initial but incomplete picture of the molecular response and
327 triggers of disease – including proteomic, epigenomic, and metabolomic signatures of disease. Previous
328 work indicates the correlation between transcriptomic and proteomic signatures of disease is relatively
329 modest (Pearson's r = 0.30) (Seyfried et al., 2017), and thus a more thorough integrative analysis is
330 warranted to determine the full space of molecular signature of disease progression. Additionally, all of
331 the samples were from post-mortem tissue which could potentially introduce non-AD specific effects due
332 to the state of the person at death (e.g. the effect of agonal state or preterminal decline in cognition
333 immediately prior to death). Because we adopted a case/control analytic strategy to enable a meta-
334 analysis across the three sources of data, each of which has a very different study design, we could not
335 consider individuals with intermediate phenotypes. As such, this analysis is limited to a syndromic
336 diagnosis of pathologic AD, further refined by including cognitive evaluations for the ROSMAP and

MSSM subjects. Given limitations of available neuropathologic phenotypes, we were not able to consider the possible impact of other aging-related pathologies on our results. Finally, because these studies focused on whole tissue analysis, we cannot resolve which observed changes are driven by differences in the cellular composition of the tissue samples between AD cases and controls (neuronal death and reactive gliosis), and which are due to actual differences in the cellular expression levels. In looking for consensus modules across multiple brain regions that are variably influenced by AD pathology and further characterizing these modules based on additional evidence for involvement in AD, we have identified robust changes that may not entirely be driven by the former. However further work is needed to refine the molecular changes and pathways associated with AD and the implications for specific central nervous system cell-types.

These transcriptional AD-related module clusters represent an attractive mechanism to support translational research. Predictions of genes with an important role within an AD-related module cluster have been validated experimentally *in vitro* and *ex vivo* (Mostafavi et al., 2018; Yu et al., 2018; Zhang et al., 2013b). Within model systems, gene signatures for human AD clusters can serve as readouts to evaluate consequences of target engagement that are known to be relevant to human disease (Wan et al., submitted). Such experiments could also identify biochemical signatures – or consequences –associated with changes in human AD clusters that could be used to advance therapeutic hypotheses or identify endophenotypic biomarkers. While effectiveness of such approaches needs to be tested, such approaches are already underway in several programs including those using mouse, fly, and cell-based model systems to evaluate AD biology. An integrated, systems approach to AD target evaluation is a powerful opportunity to advance the field.

ACKNOWLEDGEMENTS

The results published here are in whole or in part based on data obtained from the AMP-AD Knowledge Portal (doi:10.7303/syn2580853). ROSMAP Study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by

NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984, U01AG46152, the Illinois Department of Public Health, and the Translational Genomics Research Institute. Mayo RNAseq Study data were provided by the following sources: The Mayo Clinic Alzheimer's Disease Genetic Studies, led by Dr. Nilufer Ertekin-Taner and Dr. Steven G. Younkin, Mayo Clinic, Jacksonville, FL using samples from the Mayo Clinic Study of Aging, the Mayo Clinic Alzheimer's Disease Research Center, and the Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01 AG003949, NINDS grant R01 NS080820, CurePSP Foundation, and support from Mayo Foundation. Study data includes samples collected through the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona. The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026 National Brain and Tissue Resource for Parkinson's Disease and Related Disorders), the National Institute on Aging (P30 AG19610 Arizona Alzheimer's Disease Core Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimer's Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson's Disease Consortium) and the Michael J. Fox Foundation for Parkinson's Research. MSBB data were generated from postmortem brain tissue collected through the Mount Sinai VA Medical Center Brain Bank and were provided by Dr. Eric Schadt from Mount Sinai School of Medicine. Furthermore, Emory study data were supported through funding by NIA grants P50 AG025688, U01 AG046161, and U01 AG061357.

AUTHOR CONTRIBUTIONS

BAL, TMP, VS, MW, CF, CG, MA, PS, YC, CF, XW performed differential and network expression analyses. BAL, JE, KDD, PJE, PS performed bioinformatic analyses. BAL, TMP, CG, MW, LMM, SKS, KDD, PE, LMM designed the analysis plan. BAL, TMP, MA, NET, LMM, AL, DAB, PLDJ, JMS, GWC wrote the manuscript. BAL, TMP, LMM, KD, CG, MA, ED, GS, SM, SA, WH, HUK, CP, MD,

389 KE, LY, AE, CP, GWC contributed to interpretation of analyses. DAC, TG, AL, DAB, KE, MD, ZL,

390 BZ, ES, PLDJ, NDP, NET conceived the human study design.

391

392 **DECLARATION OF INTERESTS**

393 All authors declare no competing conflicts of interests.

394

REFERENCES

Abravaya, K., Phillips, B., and Morimoto, R.I. (1991). Attenuation of the heat shock response in HeLa cells is mediated by the release of bound heat shock transcription factor and is modulated by changes in growth and in heat shock temperatures. Genes Dev. *5*, 2117–2127.

Adamcsek, B., Palla, G., Farkas, I.J., Derényi, I., and Vicsek, T. (2006). CFinder: locating cliques and overlapping modules in biological networks. Bioinformatics *22*, 1021–1023.

Ahn, Y.-Y., Bagrow, J.P., and Lehmann, S. (2010). Link communities reveal multiscale complexity in networks. Nature *466*, 761–764.

Allen, M., Carrasquillo, M.M., Funk, C., Heavner, B.D., Zou, F., Younkin, C.S., Burgess, J.D., Chai, H.-S., Crook, J., Eddy, J.A., et al. (2016). Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. Sci. Data *3*, 160089.

Allen, M., Wang, X., Burgess, J.D., Watzlawik, J., Serie, D.J., Younkin, C.S., Nguyen, T., Malphrus, K.G., Lincoln, S., Carrasquillo, M.M., et al. (2018a). Conserved brain myelination networks are altered in Alzheimer's and other neurodegenerative diseases. Alzheimers. Dement. *14*, 352–366.

Allen, M., Wang, X., Serie, D.J., Strickland, S.L., Burgess, J.D., Koga, S., Younkin, C.S., Nguyen, T.T., Malphrus, K.G., Lincoln, S.J., et al. (2018b). Divergent brain gene expression patterns associate with distinct cell-specific tau neuropathology traits in progressive supranuclear palsy. Acta Neuropathol.

Altay, G., and Emmert-Streib, F. (2010). Inferring the conservative causal core of gene regulatory networks. BMC Syst. Biol. *4*, 132.

Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F., and Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res. *43*, D789–D798.

Ardura-Fabregat, A., Boddeke, E.W.G.M., Boza-Serrano, A., Brioschi, S., Castro-Gomez, S., Ceyzériat, K., Dansokho, C., Dierkes, T., Gelders, G., Heneka, M.T., et al. (2017). Targeting Neuroinflammation to Treat Alzheimer's Disease. CNS Drugs *31*, 1057–1082.

von Bartheld, C.S., Bahney, J., and Herculano-Houzel, S. (2016). The search for true numbers of neurons

and glial cells in the human brain: A review of 150 years of cell counting. J. Comp. Neurol. *524*, 3865–3895.

Bennett, D.A., Schneider, J.A., Arvanitakis, Z., and Wilson, R.S. (2012b). Overview and findings from the religious orders study. Curr. Alzheimer Res. *9*, 628–645.

Bennett, D.A., Schneider, J.A., Buchman, A.S., Barnes, L.L., Boyle, P.A., and Wilson, R.S. (2012a). Overview and findings from the rush Memory and Aging Project. Curr. Alzheimer Res. *9*, 646–663.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. *2008*, P10008.

Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H., and Del Tredici, K. (2006). Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. Acta Neuropathol. *112*, 389–404.

Brenowitz, W.D., Keene, C.D., Hawes, S.E., Hubbard, R.A., Longstreth, W.T., Woltjer, R.L., Crane, P.K., Larson, E.B., and Kukull, W.A. (2017). Alzheimer's disease neuropathologic change, Lewy body disease, and vascular brain injury in clinic- and community-based samples. Neurobiol. Aging *53*, 83–92.

Carrasquillo, M.M., Allen, M., Burgess, J.D., Wang, X., Strickland, S.L., Aryal, S., Siuda, J., Kachadoorian, M.L., Medway, C., Younkin, C.S., et al. (2017). A candidate regulatory variant at the TREM gene cluster associates with decreased Alzheimer's disease risk and increased TREML1 and TREM2 brain gene expression. Alzheimers. Dement. *13*, 663–673.

Clauset, A., Newman, M.E.J., and Moore, C. (2004). Finding community structure in very large networks. Phys. Rev. E *70*, 066111.

Cotto, J.J., Kline, M., and Morimoto, R.I. (1996). Activation of heat shock factor 1 DNA binding precedes stress-induced serine phosphorylation. Evidence for a multistep pathway of regulation. J. Biol. Chem. *271*, 3355–3358.

Csardi, G., InterJournal, T.N.-, Systems, C., and 2006, undefined The igraph software package for complex network research. Necsi.Edu.

Cummings, J.L., Morstorf, T., and Zhong, K. (2014). Alzheimer's disease drug-development pipeline:

few candidates, frequent failures. Alzheimers. Res. Ther. *6*, 37.

De Strooper, B., and Karran, E. (2016). The Cellular Phase of Alzheimer's Disease. Cell *164*, 603–615.

Efthymiou, A.G., and Goate, A.M. (2017). Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. Mol. Neurodegener. *12*, 43.

Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., et al. (2016). The Reactome pathway Knowledgebase. Nucleic Acids Res. *44*, D481–D487.

Gaiteri, C., Chen, M., Szymanski, B., Kuzmin, K., Xie, J., Lee, C., Blanche, T., Chaibub Neto, E., Huang, S.-C., Grabowski, T., et al. (2015). Identifying robust communities and multi-community nodes by combining top-down and bottom-up approaches to clustering. Sci. Rep. *5*, 16361.

Guerreiro, R., Wojtas, A., Bras, J., Carrasquillo, M., Rogaeva, E., Majounie, E., Cruchaga, C., Sassi, C., Kauwe, J.S.K., Younkin, S., et al. (2013). *TREM2* Variants in Alzheimer's Disease. N. Engl. J. Med. *368*, 117–127.

Haury, A.-C., Mordelet, F., Vera-Licona, P., and Vert, J.-P. (2012). TIGRESS: Trustful Inference of Gene REgulation using Stability Selection. BMC Syst. Biol. *6*, 145.

Huynh-Thu, V.A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. PLoS One *5*, e12776.

Jack, C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., Holtzman, D.M., Jagust, W., Jessen, F., Karlawish, J., et al. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimer's Dement. *14*, 535–562.

Jager, P. De, Ma, Y., McCabe, C., Xu, J., Vardarajan, B.N., Felsky, D., Klein, H.-U., White, C.C., Peters, M.A., Lodgson, B., et al. (2018). A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. BioRxiv 251967.

De Jager, P.L., Yang, H.-S., and Bennett, D.A. (2018). Deconstructing and targeting the genomic architecture of human neurodegeneration. Nat. Neurosci. *21*, 1310–1317.

Jin, S.C., Carrasquillo, M.M., Benitez, B.A., Skorupa, T., Carrell, D., Patel, D., Lincoln, S., Krishnan, S.,

473 Kachadoorian, M., Reitz, C., et al. (2015). TREM2 is associated with increased risk for Alzheimer's
474 disease in African Americans. Mol. Neurodegener. *10*, 19.
475 Johnson, E.C.B., Dammer, E.B., Duong, D.M., Yin, L., Thambisetty, M., Troncoso, J.C., Lah, J.J., Levey,
476 A.I., and Seyfried, N.T. (2018). Deep proteomic network analysis of Alzheimer's disease brain reveals
477 alterations in RNA binding proteins and RNA splicing associated with disease. Mol. Neurodegener. *13*,
478 52.
479 Jonsson, T., Stefansson, H., Steinberg, S., Jonsdottir, I., Jonsson, P. V., Snaedal, J., Bjornsson, S.,
480 Huttenlocher, J., Levey, A.I., Lah, J.J., et al. (2013). Variant of *TREM2* Associated with the Risk of
481 Alzheimer's Disease. N. Engl. J. Med. *368*, 107–116.
482 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives
483 on genomes, pathways, diseases and drugs. Nucleic Acids Res. *45*, D353–D361.
484 King, A. (2018). The search for better animal models of Alzheimer's disease. Nature *559*, S13–S15.
485 Krämer, N., Schäfer, J., and Boulesteix, A.-L. (2009). Regularized estimation of large-scale gene
486 association networks using graphical Gaussian models. BMC Bioinformatics *10*, 384.
487 Kumar, A., Singh, A., and Ekavali (2015). A review on Alzheimer's disease pathophysiology and its
488 management: an update. Pharmacol. Reports *67*, 195–203.
489 Kutmon, M., Riutta, A., Nunes, N., Hanspers, K., Willighagen, E.L., Bohler, A., Mélius, J., Waagmeester,
490 A., Sinha, S.R., Miller, R., et al. (2016). WikiPathways: capturing the full diversity of pathway
491 knowledge. Nucleic Acids Res. *44*, D488–D494.
492 Lambert, J.-C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., Jun, G.,
493 DeStefano, A.L., Bis, J.C., Beecham, G.W., et al. (2013). Meta-analysis of 74,046 individuals identifies
494 11 new susceptibility loci for Alzheimer's disease. Nat. Genet. *45*, 1452–1458.
495 Langfelder, P., and Horvath, S. (2008a). WGCNA: an R package for weighted gene co-expression
496 network analysis. BMC Bioinformatics *9*, 559.
497 Langfelder, P., and Horvath, S. (2008b). WGCNA: an R package for weighted correlation network
498 analysis. BMC Bioinformatics *9*, 559.

499 Logsdon, B.A., Gentles, A.J., Miller, C.P., Blau, C.A., Becker, P.S., and Lee, S.-I. (2015). Sparse
500 expression bases in cancer reveal tumor drivers. Nucleic Acids Res. *43*, 1332–1344.
501 Lu, T., Aron, L., Zullo, J., Pan, Y., Kim, H., Chen, Y., Yang, T.-H., Kim, H.-M., Drake, D., Liu, X.S., et
502 al. (2014). REST and stress resistance in ageing and Alzheimer's disease. Nature *507*, 448–454.
503 Makin, S. (2018). The amyloid hypothesis on trial. Nature *559*, S4–S7.
504 Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R.,
505 DREAM5 Consortium, M., Kellis, M., Collins, J.J., et al. (2012). Wisdom of crowds for robust gene
506 network inference. Nat. Methods *9*, 796–804.
507 Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano,
508 A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian
509 cellular context. BMC Bioinformatics *7 Suppl 1*, S7.
510 McKenzie, A.T., Moyon, S., Wang, M., Katsyv, I., Song, W.-M., Zhou, X., Dammer, E.B., Duong, D.M.,
511 Aaker, J., Zhao, Y., et al. (2017). Multiscale network modeling of oligodendrocytes reveals molecular
512 components of myelin dysregulation in Alzheimer's disease. Mol. Neurodegener. *12*, 82.
513 Meyer, P.E., Kontos, K., Lafitte, F., and Bontempi, G. (2007). Information-Theoretic Inference of Large
514 Transcriptional Regulatory Networks. EURASIP J. Bioinforma. Syst. Biol. *2007*, 1–9.
515 Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P.D. (2017).
516 PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data
517 analysis tool enhancements. Nucleic Acids Res. *45*, D183–D189.
518 Mielke, M.M., Vemuri, P., and Rocca, W.A. (2014). Clinical epidemiology of Alzheimer's disease:
519 assessing sex and gender differences. Clin. Epidemiol. *6*, 37–48.
520 Mostafavi, S., Gaiteri, C., Sullivan, S.E., White, C.C., Tasaki, S., Xu, J., Taga, M., Klein, H.-U., Patrick,
521 E., Komashko, V., et al. (2018). A molecular network of the aging human brain provides insights into the
522 pathology and cognitive decline of Alzheimer's disease. Nat. Neurosci. *21*, 811–819.
523 Nazarian, A., Yashin, A.I., and Kulminski, A.M. (2018). Genome-wide analysis of genetic predisposition
524 to Alzheimer's disease and related sex-disparities. BioRxiv 321992.

525 Neuner, S.M., Heuer, S.E., Huentelman, M.J., O'Connell, K.M.S., and Kaczorowski, C.C. (2018).
526 Harnessing Genetic Complexity to Enhance Translatability of Alzheimer's Disease Mouse Models: A
527 Path toward Precision Medicine. Neuron.
528 Nishimura, D. (2001). BioCarta. Biotech Softw. Internet Rep. *2*, 117–120.
529 Parikshak, N.N., Swarup, V., Belgard, T.G., Irimia, M., Ramaswami, G., Gandal, M.J., Hartl, C., Leppa,
530 V., Ubieta, L. de la T., Huang, J., et al. (2016). Genome-wide changes in lncRNA, splicing and regional
531 gene expression patterns in autism. Nature *540*, 423–427.
532 Patrick, E., Olah, M., Taga, M., Klein, H.-U., Xu, J., White, C.C., Felsky, D., Gaiteri, C., Chibnik, L.B.,
533 Mostafavi, S., et al. (2017). A cortical immune network map identifies a subset of human microglia
534 involved in Tau pathology. BioRxiv 234351.
535 Pons, P., and Latapy, M. (2005). Computing communities in large networks using random walks (long
536 version).
537 Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M.N., Replogle, J.M., Feng, T., Lee, M., Asinovski, N.,
538 Frohlich, I., et al. (2014). Polarization of the Effects of Autoimmune and Neurodegenerative Risk Alleles
539 in Leukocytes. Science (80-.). *344*, 519–523.
540 Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers
541 differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47–
542 e47.
543 Rosvall, M., and Bergstrom, C.T. (2008). Maps of random walks on complex networks reveal community
544 structure. Proc. Natl. Acad. Sci. U. S. A. *105*, 1118–1123.
545 Safran, M., Dalah, I., Alexander, J., Rosen, N., Iny Stein, T., Shmoish, M., Nativ, N., Bahir, I., Doniger,
546 T., Krug, H., et al. (2010). GeneCards Version 3: the human gene integrator. Database *2010*.
547 Seyfried, N.T., Dammer, E.B., Swarup, V., Nandakumar, D., Duong, D.M., Yin, L., Deng, Q., Nguyen,
548 T., Hales, C.M., Wingo, T., et al. (2017). A Multi-network Approach Identifies Protein-Specific Co-
549 expression in Asymptomatic and Symptomatic Alzheimer's Disease. Cell Syst. *4*, 60–72.e4.
550 Sims, R., van der Lee, S.J., Naj, A.C., Bellenguez, C., Badarinarayan, N., Jakobsdottir, J., Kunkle, B.W.,

551 Boland, A., Raybould, R., Bis, J.C., et al. (2017). Rare coding variants in PLCG2, ABI3, and TREM2
552 implicate microglial-mediated innate immunity in Alzheimer's disease. Nat. Genet. *49*, 1373–1384.
553 Song, W.-M., and Zhang, B. (2015). Multiscale Embedded Gene Co-expression Network Analysis. PLOS
554 Comput. Biol. *11*, e1004574.
555 Stowell, R.D., Wong, E.L., Batchelor, H.N., Mendes, M.S., Lamantia, C.E., Whitelaw, B.S., and
556 Majewska, A.K. (2018). Cerebellar microglia are dynamically unique and survey Purkinje neurons *in*
557 *vivo*. Dev. Neurobiol. *78*, 627–644.
558 Traag, V.A., and Bruggeman, J. (2009). Community detection in networks with positive and negative
559 links. Phys. Rev. E *80*, 036115.
560 Tryka, K.A., Hao, L., Sturcke, A., Jin, Y., Wang, Z.Y., Ziyabari, L., Lee, M., Popova, N., Sharopova, N.,
561 Kimura, M., et al. (2014). NCBI's Database of Genotypes and Phenotypes: dbGaP. Nucleic Acids Res.
562 *42*, D975–D979.
563 Wang, M., Roussos, P., McKenzie, A., Zhou, X., Kajiwara, Y., Brennand, K.J., De Luca, G.C., Crary,
564 J.F., Casaccia, P., Buxbaum, J.D., et al. (2016). Integrative network analysis of nineteen brain regions
565 identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's
566 disease. Genome Med. *8*, 104.
567 Wang, M., Beckmann, N.D., Roussos, P., Wang, E., Zhou, X., Wang, Q., Ming, C., Neff, R., Ma, W.,
568 Fullard, J.F., et al. (2018). The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic
569 data in Alzheimer's disease. Sci. Data *5*, 180185.
570 Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence
571 assessments and item tracking. Bioinformatics *26*, 1572–1573.
572 Winblad, B., Amouyel, P., Andrieu, S., Ballard, C., Brayne, C., Brodaty, H., Cedazo-Minguez, A.,
573 Dubois, B., Edvardsson, D., Feldman, H., et al. (2016). Defeating Alzheimer's disease and other
574 dementias: a priority for European science and society. Lancet. Neurol. *15*, 455–532.
575 Yu, L., Petyuk, V.A., Gaiteri, C., Mostafavi, S., Young-Pearse, T., Shah, R.C., Buchman, A.S.,
576 Schneider, J.A., Piehowski, P.D., Sontag, R.L., et al. (2018). Targeted brain proteomics uncover multiple

pathways to Alzheimer's dementia. Ann. Neurol. *84*, 78–88.

Zhang, B., Gaiteri, C., Bodea, L.-G., Wang, Z., McElwee, J., Podtelezhnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., et al. (2013a). Integrated Systems Approach Identifies Genetic Nodes and Networks in Late-Onset Alzheimer's Disease. Cell *153*, 707–720.

Zhang, B., Gaiteri, C., Bodea, L.-G., Wang, Z., McElwee, J., Podtelezhnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., et al. (2013b). Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell *153*, 707–720.

Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keeffe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N., et al. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J. Neurosci. *34*, 11929–11947.

Zuo, J., Rungger, D., and Voellmy, R. (1995). Multiple layers of regulation of human heat shock transcription factor 1. Mol. Cell. Biol. *15*, 4319–4330.

Mitochondrial bioenergetic deficit precedes Alzheimer's pathology in female mouse model of Alzheimer's disease. Natl. Acad Sci.

(2015). Gene Ontology Consortium: going forward. Nucleic Acids Res. *43*, D1049–D1056.

FIGURE LEGENDS

Figure 1 - Percentage of total pairwise module by literature curated AD gene set associations significantly enriched (FDR <=0.05) for 12 known AD gene sets with standard errors shown.

Figure 2 - A) Overlap between identified 30 AD associated coexpression modules - there are 5 predominant clusters identified across brain regions – Consensus Cluster A, B, C, D, and E. B) Cell type enrichments of the 30 AD associated coexpression modules. C) Enrichment for curated AD gene sets within the 30 AD associated coexpression modules. D) Enrichments for differentially expressed genes based on the meta-analysis in the 30 AD associated coexpression modules.

TABLE LEGENDS

Table 1 - Data characteristics of the AMP-AD human RNA-seq datasets.

Table 2 - Enrichment for heat shock response and unfolded protein response pathways for non-cell type specific modules. Fisher's Exact Test odds ratio and adjusted p-values of gene set enrichment are shown.

METHODS

Study design and data collection. Details of sample collection, postmortem sample descriptions, tissue and RNA preparation, library preparation and sequencing, and sample QC are provided in previously published work (Allen et al., 2016; Jager et al., 2018; Wang et al., 2018).

AD definition and cross-study harmonization. Sub-samples were selected to harmonize the LOAD case - control definition across the three studies for all differential expression analyses. To compare analysis results across studies and to get an understanding of LOAD biology across different tissues, we harmonized the LOAD definition across three studies. The motivation was to define LOAD cases as those with both clinical and neuropathological evidence for definitive late onset Alzheimer's disease - i.e. a high burden of neurofibrillary tangles, neuritic amyloid plaques, and cognitive impairment with little evidence of other pathology (Jack et al., 2018). Controls were concordantly defined as patients with a low burden of plaques and tangles, as well as no evidence of cognitive impairment if available. As such, for the ROSMAP study, we had individuals with a Braak neurofibrillary tangle (NFT) score (Braak et al., 2006) greater than or equal to 4, CERAD score less than or equal to 2, and a cognitive diagnosis of probable AD with no other causes as LOAD cases, Braak less than or equal to 3, CERAD score greater than or equal to 3, and cognitive diagnosis of 'no cognitive impairment' as LOAD controls. For MSBB, we analogously defined LOAD cases as those with CDR score greater than or equal to 1, Braak score greater than or equal to 4, and CERAD neuritic and cortical plaque score greater than or equal to 2 as LOAD cases, and CDR scores less than or equal to 0.5, Braak less than or equal to 3, and CERAD less than or equal to 1 as LOAD controls. It is to note here that the definitions of CERAD differs between ROSMAP and MSBB studies. For the Mayo Clinic RNASeq study, cases were defined based on neuropathology, with LOAD cases being based on Braak score greater than or equal to 4 and CERAD neuritic and cortical plaque score greater than 1 whereas LOAD controls being those defined as Braak less than or equal to 3, and CERAD less than 2. Further details concerning the diagnosis in the Mayo RNASeq study have been previously published (Allen et al., 2018a).

RNA-Seq Reprocessing, library normalization and covariates adjustment. To avoid some of the technical variabilities arising due to RNA-seq alignment and quantification, and also to account for some of the technical variabilities we reprocessed and realigned all the RNA-Seq reads from the source studies (Allen et al., 2016; Jager et al., 2018; Wang et al., 2018). The reprocessing was done using a consensus set of tools with only library type-specific parameters varying between pipelines. Picard (https://broadinstitute.github.io/picard/) was used to generate FASTQs from source BAMs. Generated FASTQ reads were aligned to the GENCODE24 (GRCh38) reference genome using STAR and gene counts were computed for each sample. To evaluate the quality of individual samples and to identify potentially important covariates for expression modeling, we computed two sets of metrics using the CollectAlignmentSummaryMetrics and CollectRnaSeqMetrics functions in Picard.

To account for differences between samples, studies, experimental batch effects and unwanted RNA-Seq specific technical variations we performed library normalization and covariate adjustments for each study separately using fixed/mixed effects modeling. The workflow consist of following steps: (i) gene filtereing: Genes that are expressed more than 1 CPM (read Counts Per Million total reads) in at least 50% of samples in each tissue and diagnosis category was used for further analysis, (ii) conditional quantile normalisation, was applied to account for variations in gene length and GC content, (iii) sample outlier detection using principal component analysis and clustering, (iv) Covariates identification and adjustment, where confidence of sampling abundance were estimated using a weighted linear model using voom-limma package in bioconductor (Ritchie et al., 2015). For most analyses, we perform a variant of fixed/mixed effect linear regression as shown here: gene expression ~ Diagnosis + Sex + covariates + (1| Donor) or gene expression ~ Diagnosis x Sex + covariates + (1|Donor), where each gene in linearly regressed independently with Diagnosis, variable explaining the AD status of an individual, identified covariates and donor information as random effect. Observation weights (if any) were calculated using the voom-limma (Ritchie et al., 2015) pipeline. So that observations with higher presumed precision will be up-weighted in the linear model fitting process. All these workflows were applied separately for each of the three studies.

Meta-Differential Expression Analysis. All the differential and meta-differential expression analysis were performed as weighted fixed/mixed effect linear models using the voom-limma (Ritchie et al., 2015) package in R. For each gene, linear regression was fit with biological and technical covariates that were associated with the top principal components of the expression data, as identified above. Two of the three studies - MSBB and Mayo RNAseq - obtained more than one tissue from the same donors. Therefore, except ROSMAP study, donor-specific effects were explicitly modeled as random effects. Different models were built for understanding the effects of diagnosis and sex-specific diagnosis effects. Depending on the model, coefficients related to either diagnosis or diagnosis time sex was statistically tested for being non-zero, implying an estimated effect for the primary variable of interest is above and beyond any other effect from the covariates. This test produces t-statistic (then moderated in a Bayesian fashion) and corresponding p-value. P-values were then adjusted for multiple hypothesis testing using false discovery rate (FDR) estimation, and the differentially expressed genes were determined as those with an estimated FDR below, or at, 5% with a corresponding absolute expression and fold-change cutoffs. To identify genes with evidence for change in expression across studies, we next performed a meta-analysis using a random effect and fixed effect models using rmeta r package (https://cran.r-project.org/web/packages/rmeta/index.html). The random effect model was selected as a conservative approach to correct for variation across studies.

Network Inference and Module Identification. We apply five distinct network module identification methodologies to each of the seven tissue specific expression data sets. This includes MEGENa (Song and Zhang, 2015), WINA (Wang et al., 2016), metanetwork, rWGCNA (Parikshak et al., 2016), and speakEasy (Gaiteri et al., 2015) to characterize a comprehensive landscape of transcriptomic variation across the seven brain regions and three studies. Briefly, MEGENa (Song and Zhang, 2015) is a method that infers a sparse graph based on a distance to define multiscale module definitions from coexpression data. Speakeasy is a label propagation method to identify robust coexpression modules that are identified both top up and bottom down (Gaiteri et al., 2015), rWGCNA is a version of WGCNA (Langfelder and Horvath, 2008b) that includes bootstrapping to identify robust modules, WINA is also a

684 variation on WGCNA that includes a modified tree cutting method to identify modules (Wang et al.,
685 2016). The metanetwork inference methodology is inspired by the DREAM5 method (Marbach et al.,
686 2012), where ensemble inference methodologies were identified as more robust for identification of gene-
687 gene interactions from coexpression data (Marbach et al., 2012).

688 *Metanetwork coexpression graph learning algorithm.* We construct a statistical network of gene
689 co-expression using an ensemble network inference algorithm. Briefly, we apply nine distinct gene co-
690 expression network inference methodologies ARACNe (Margolin et al., 2006), Genie3 (Huynh-Thu et al.,
691 2010), Tigress (Haury et al., 2012), Sparrow (Logsdon et al., 2015), Lasso (Krämer et al., 2009), Ridge
692 (Krämer et al., 2009), mrnet (Meyer et al., 2007), c3net (Altay and Emmert-Streib, 2010) and WGCNA
693 (Langfelder and Horvath, 2008b) and rank the edge lists from each method based on the method specific
694 edge weights, identify a mean rank for each edge across methods, then identify the total number of edges
695 supported by the data with Bayesian Information Criterion for local neighborhood selection with linear
696 regression. The ensemble approach is inspired by work DREAM consortia (Marbach et al., 2012)
697 showing that ensemble methods are better at generating robust gene expression networks across
698 heterogeneous data-sets.

699 *Metanetwork module identification methodology*. We identify metanetwork modules in each
700 tissue type based on the inferred network topology with a consensus clustering algorithm (Wilkerson and
701 Hayes, 2010) applied to multiple individual module identification methods. We ran individual network
702 clustering methods applied to each of the seven network topologies. These methods included CFinder
703 (Adamcsek et al., 2006), GANXiS (Gaiteri et al., 2015), a fast greedy algorithm (Clauset et al., 2004),
704 InfoMap (Rosvall and Bergstrom, 2008), LinkCommunities (Ahn et al., 2010), Louvain (Blondel et al.,
705 2008), Spinglass (Traag and Bruggeman, 2009), and Walktrap (Pons and Latapy, 2005), methods. All
706 implementations are from the igraph package (Csardi et al.) in R.

707 *Aggregate module identification.* For all 2978 modules identified across tissues (**Supplementary**
708 **Table S1**, 10.7303/syn10309369.1), we first identify which modules are enriched for >=1 AD specific
709 differential expressed gene set from the DEG meta-analysis (10.7303/syn11914606). This restricts the

total number of individual modules to 660 that show evidence of differential expression as a function of disease status. Next, we construct a within tissue module graph using a Fisher's exact test for pairwise overlap of gene sets between each pair of these 660 individual modules. An example of this graph is shown in **Figure S1**. We then apply the edge betweenness graph clustering method (Pons and Latapy, 2005) to identify aggregate modules from these module graphs that represent meta modules that are both differentially expressed and identified by multiple independent module identification algorithms. With this approach we identify 30 aggregate module definitions (10.7303/syn11932957.1) across the seven tissue types and three studies.

Enrichment analyses. Aggregate modules were interpreted using functional and cell type enrichment analysis. We performed a battery of enrichment tests to understand biological functionality, including evaluating primary hypotheses previously implicated by genetic findings in AD research, performing exploratory analyses of a large number of gene sets (such as those obtained from Gene Ontology), and performing enrichment for brain tissue specific cell types. We started by curating three categories of gene sets for analyzing the differential expression data and network modules: 1) a small group of pathways and gene sets previously implicated in genome-wide genetic studies of AD ("hypothesis-driven"), 2) a collection of thousands of "hypothesis-free" gene sets from large databases like GO, Wikipathways and Reactome, that would allow us to potentially characterize novel biology arising in brain expression related to AD, and 3) Brain specific cell type markers to potentially understand the changes in various cell type fractions (Zhang et al., 2014). AMP-AD specific gene sets were constructed by taking the union of gene set definitions reported in each of the following reports: RNA-binding protein modules (Johnson et al., 2018), oligodendroglial modules from MSSM (McKenzie et al., 2017), AD vs Control oligodendroglial modules in the Mayo RNAseq study (Allen et al., 2018a), and Module 109 from the ROSMAP study (Mostafavi et al., 2018).

Genes not measured in our data are filtered from the annotated gene sets. Annotated gene sets with less than 10% of genes expressed in our data sets were removed. Fisher's exact test was used to test enrichment of each gene set with the annotated set. Resulting p-values were corrected independently for

each set using Benjamini-Hochberg method for significance testing, owing to the differences in their hypothesis. Gene sets that had a minimum overlap of at least 3 genes were considered for further interpretation.

Statistics, code and data availability. All computation and calculations were carried out in the R language for statistical computing (version 3.3.0 - 3.5.1). Significance levels for p-values were set at 0.05 (unless otherwise specified), and analyses were two-tailed. An R package with all code for the metanetwork algorithm is available at https://github.com/Sage-Bionetworks/metanetwork, and a toolkit for integrating metanetwork with AWS high performance compute cluster cfncluster, and Synapse is available here https://github.com/Sage-Bionetworks/metanetworkSynapse. Furthermore, all code used to generate aggregate modules and figures are available in this R package: https://github.com/Sage-Bionetworks/AMPAD, with the following notebook collating the primary results: https://github.com/Sage-Bionetworks/AMPAD/blob/master/manuscript_analyses.Rmd.

SUPPLEMENTAL INFORMATION LEGENDS

Figure S1 - Clustering of individual AD coexpression modules. Similar coexpression structure was observed within each data set across methods, as indicated by significant overlap in module memberships. This module graph shows individual modules that are significantly enriched for at least one DEG meta-analysis signature in DLPFC (ROSMAP). Each node is a module, and an edge is drawn between modules if there is a statistically significant overlap of genes between the two modules. The edge betweenness clustering algorithm identifies four robust meta modules, which are colored green, purple, red and blue respectively.

Table S1 - Counts of number of individual coexpression modules identified by method and brain region (10.7303/syn10309369.1).

Table S2 – Study demographics for each of the AMP-AD studies for samples with available bulk homogenate RNA-seq data.

Table S3 – Module assignment to consensus clusters and module size.

Table S4 – Gene set enrichment results for aggregate modules compared to AMP-AD derived gene sets.

Table S5 – Gene set enrichment results for aggregate modules and AD gene sets against genes up-regulated in AD from the differential expressed gene sets from the random effect meta-analysis of differential expression.

772

773 **Table S6 -** Gene set enrichment results for aggregate modules and AD gene sets against genes
774 down-regulated in AD from the differential expressed gene sets from the random effect meta-
775 analysis of differential expression.

776

777 **Table S7** - Gene set enrichment results for aggregate modules and AD gene sets against genes
778 up-regulated in male AD from the differential expressed gene sets from the random effect meta-
779 analysis of differential expression.

780

781 **Table S8** - Gene set enrichment results for aggregate modules and AD gene sets against genes
782 up-regulated in female AD from the differential expressed gene sets from the random effect
783 meta-analysis of differential expression.

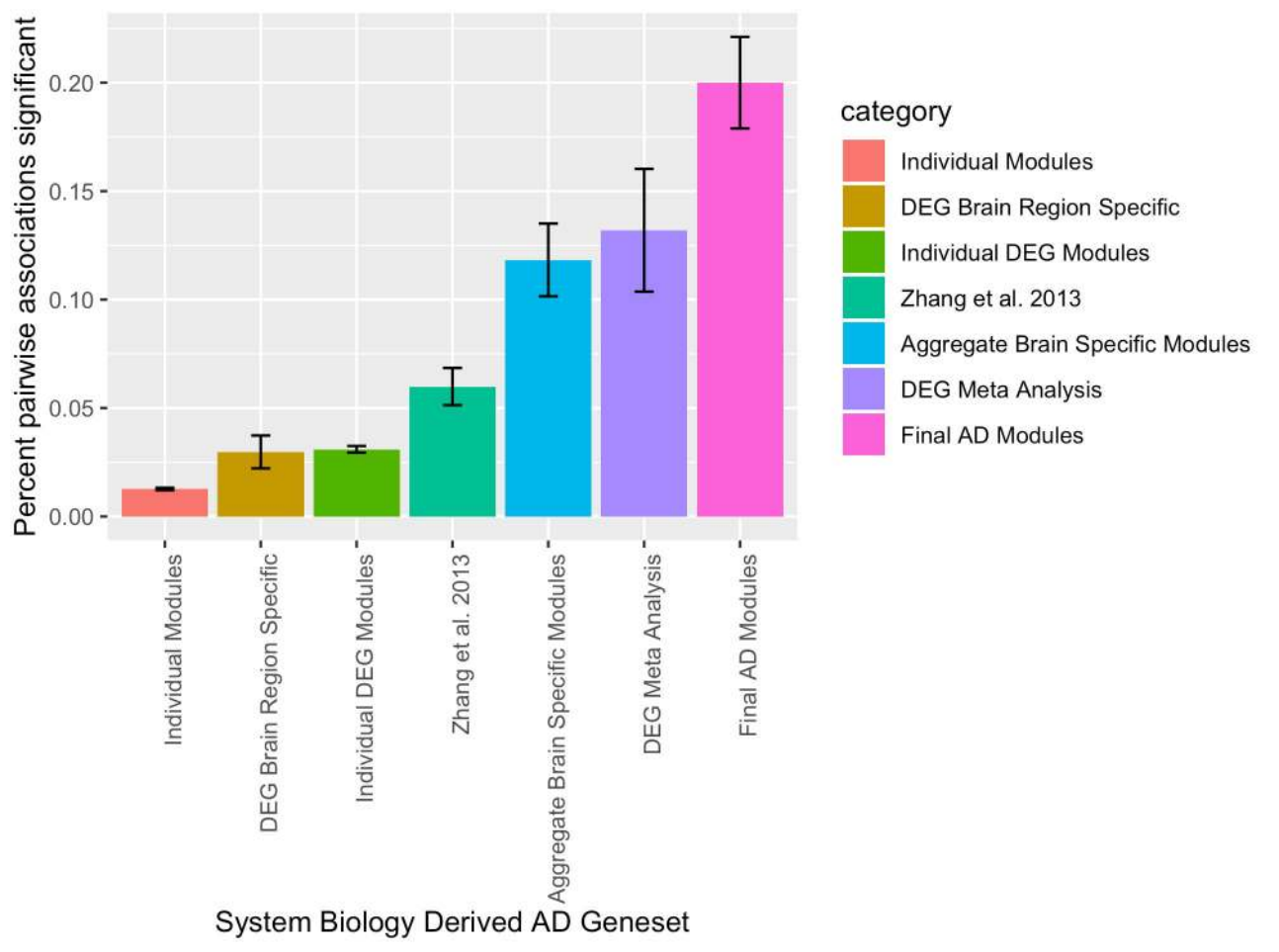
784

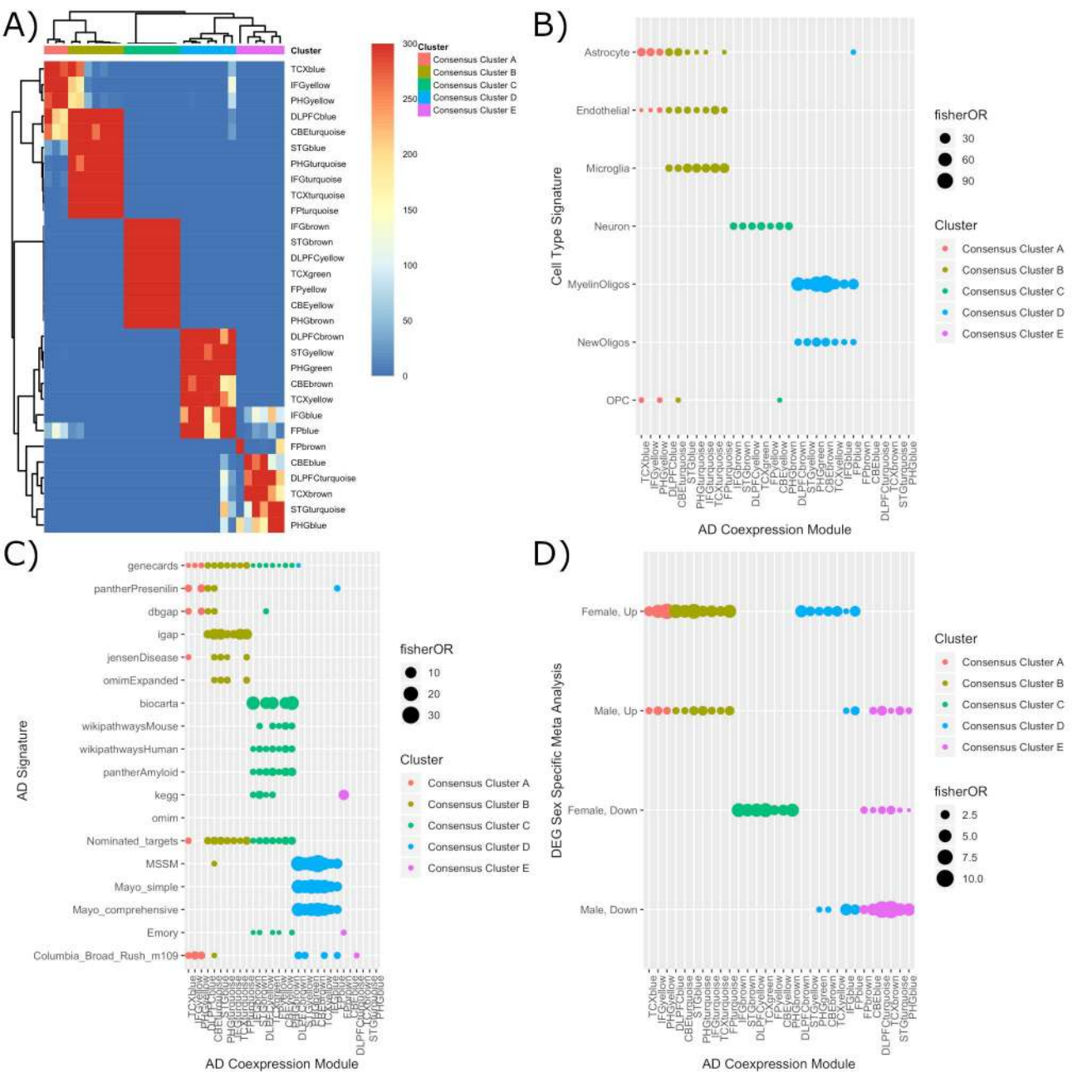
785 **Table S9 -** Gene set enrichment results for aggregate modules and AD gene sets against genes
786 down-regulated in female AD from the differential expressed gene sets from the random effect
787 meta-analysis of differential expression.

788

789 **Table S10 -** Gene set enrichment results for aggregate modules and AD gene sets against genes
790 down-regulated in male AD from the differential expressed gene sets from the random effect
791 meta-analysis of differential expression.

792


Percent pairwise associations significant
0.00
0.05
0.10
0.15
0.20
Individual Modules
DEG Brain Region Specific
Individual DEG Modules
Zhang et al. 2013
Aggregate Brain Specific Modules
DEG Meta Analysis
Final AD Modules
System Biology Derived AD Geneset
category
Individual Modules
DEG Brain Region Specific
Individual DEG Modules
Zhang et al. 2013
Aggregate Brain Specific Modules
DEG Meta Analysis
Final AD Modules


A)
Cluster
Consensus Cluster A
Consensus Cluster B
Consensus Cluster C
Consensus Cluster D
Consensus Cluster E
300
250
200
150
100
50
0
TCXblue
IFGyellow
PHGyellow
DLPFCblue
CBEturquoise
STGblue
PHGturquoise
IFGturquoise
TCXturquoise
FPturquoise
IFGbrown
STGbrown
DLPFCyellow
TCXgreen
FPyellow
CBEyellow
PHGbrown
DLPFCbrown
STGyellow
PHGgreen
CBEbrown
TCXyellow
IFGblue
FPblue
FPbrown
CBEblue
DLPFCturquoise
TCXbrown
STGturquoise
PHGblue
B)
Astrocyte
Endothelial
Microglia
Neuron
MyelinOligos
NewOligos
OPC
Cell Type Signature
AD Coexpression Module
fisherOR
30
60
90
Cluster
Consensus Cluster A
Consensus Cluster B
Consensus Cluster C
Consensus Cluster D
Consensus Cluster E
C)
genecards
pantherPresenilin
dbgap
igap
jensenDisease
omimExpanded
biocarta
wikipathwaysMouse
wikipathwaysHuman
pantherAmyloid
kegg
omim
Nominated_targets
MSSM
Mayo_simple
Mayo_comprehensive
Emory
Columbia_Broad_Rush_m109
AD Signature
AD Coexpression Module
fisherOR
10
20
30
Cluster
Consensus Cluster A
Consensus Cluster B
Consensus Cluster C
Consensus Cluster D
Consensus Cluster E
D)
Female, Up
Male, Up
Female, Down
Male, Down
DEG Sex Specific Meta Analysis
AD Coexpression Module
Cluster
Consensus Cluster A
Consensus Cluster B
Consensus Cluster C
Consensus Cluster D
Consensus Cluster E
fisherOR
2.5
5.0
7.5
10.0