Combinatorial analyses reveal that cellular composition changes have different impact on transcriptomic changes of cell type specific genes in Alzheimer's Disease brains

Shunian Xiang
Shenzhen University

Travis Johnson
Ohio State University

Tianhan Dong
Indianan University, School of Medicine

Zhi Huang
Purdue University

Michael Cheng
Indiana University Bloomington

Tianfu Wang
Shenzhen University

Kai Yang
Indiana University Department of Pediatrics

Dong Ni（✉ nidong@szu.edu.cn ）
Shenzhen University

Kun Huang（✉ kunhuang@iu.edu ）
Indiana University School of Medicine

Jie Zhang（✉ jizhan@iu.edu ）
Indiana University School of Medicine   https://orcid.org/0000-0001-6939-7905

Research article

Combinatorial analyses reveal that cellular composition changes have different impact on transcriptomic changes of cell type specific genes in Alzheimer's Disease brains

Shunian Xiang[1,2], Travis Johnson[3,4], Tianhan Dong[5], Zhi Huang[6], Michael Cheng[1], Tianfu Wang[2], Kai Yang[7], Dong Ni*[2], Kun Huang*[3] and Jie Zhang*[1,8]

[1]Department of Medical & Molecular Genetics, Indiana University, School of Medicine, Indianapolis, IN 46202, USA; email: xiangshunian@163.com (SX); miccheng@iu.edu (MC).

[2]Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China; email: tfwang@szu.edu.cn (TW)

[3]Department of Medicine, Indiana University, School of Medicine, Indianapolis, IN 46202, USA; email: kunhuang@iu.edu.

[4]Department of Biomedical Informatics, The Ohio State University, Columbus, OH 43210, USA; email: travis.johnson@osumc.edu.

[5]Department of Pharmacology, Indiana University, School of Medicine, Indianapolis, IN46202, USA; email: dongtian@iu.edu

[6]Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA; email: huang898@purdue.com

[7]Department of Pediatrics, Indiana University, School of Medicine, Indianapolis, IN 46202, USA; email: ky11@iu.edu

[8]Lead Contact

*Corresponding authors: Jie Zhang: jizhan@iu.edu; Kun Huang: kunhuang@iu.edu; Dong Ni: nidong@szu.edu.cn

Abstract

Background: Alzheimer's disease (AD) brains are characterized by progressive neuron loss and gliosis which involves mostly microglia and astrocytes. Comparative transcriptomic analysis on AD vs. normal brain tissues helps to identify key genes/pathways involved in AD initiation and progression. However, many such studies using bulk brain tissue samples have not considered cell composition changes in AD brains, which may lead to expression changes that are not due to transcriptional regulation.

Methods: Using five large transcriptomic datasets including 1,681 brain tissue samples (882 AD, 799 normal) in total, we first mined frequent co-expression network modules across them, then combined differential expression and differential co-expression analysis on the mined modules in AD versus normal brains. Integrated with cell type deconvolution analysis, we addressed the question of whether the module expression changes are due to altered cellular composition or transcriptional regulation. We then used four additional large AD/normal transcriptomic datasets to validate our findings.

Results: The integrative analysis revealed highly elevated expression level of microglia modules in AD without co-expression change. Decreased expression and elevated co-expression are observed for neuron modules in AD, while significant over-expression and co-expression perturbation are observed in astrocyte modules, all of which has not been previously reported. The expression levels of astrocyte modules also show the strongest correlation with the clinicopathological biomarkers among all cell type specific modules.

Conclusion: Further analysis indicated that the overall increased expression of the core microglia modules can be well explained by the increased microglia cell population in AD brains instead of *bona fide* microglia genes' upregulation. In contrast, the decreased expression and perturbed co-expression in AD neuron modules are due to both neuron cell loss and expression regulation of neuronal pathways including differentially expressed transcription factors such as BCL6 and STAT3, which previous study was not able to identify from the shadow of the cellular composition

change. Similarly, the strong changes in expression and co-expression in the astrocyte modules may be also due to a combinatory effect from astrogliosis and astrocyte gene activation in AD brains. In this work, we demonstrated that the combinatorial analyses not only provide a powerful approach to delineate the origin of transcriptomic changes in bulk tissue data, but also lead to a deeper understanding of genes in AD.

Key words: Alzheimer's Disease, transcriptomic analysis, differential co-expression, gene co-expression network, cell type deconvolution, microglia, astrocyte

Background

Alzheimer's disease (AD) is the most prevalent form of dementia, affecting about 40 million people with mostly elderlies worldwide today (Huang and Mucke, 2012). Despite decades of research, the disease etiology of AD remains unclear (Magalingam et al., 2018). While less than 10% cases of early on-set AD can be traced to genetic mutations in genes such as *PSEN1*, *PSEN2*, or *APP*, the cause of late-onset AD is still unknown (Robinson et al., 2017). The hallmarks of AD pathological changes include extracellular β-amyloid deposition, tau protein mediated intracellular neurofibrillary tangles, dystrophic neurites and neurons, as well as synapses death (Magalingam et al., 2018; Robinson et al., 2017). Besides neuron loss, the proliferation of the microglia and astrocytes is also observed in AD (Fakhoury, 2018). Microglia is the major immune cell type in the central nervous system. They usually conjugate and form clusters around amyloid plaques in the AD brains (Clayton et al., 2017; Solito and Sastre, 2012). The microglia-mediated neuroinflammation and its potential link to AD etiology has been a focus in this field recently (Lucin and Wyss-Coray 2009, Darweesh, Wolters et al. 2018, Tansey, Cameron et al. 2018, Patir, Shih et al. 2019). Previously, several studies on transcriptomic data from bulk brain tissues have identified differentially expressed genes or gene co-expression networks associated with microglia in AD samples (Celarain et al., 2016; Li et al., 2015; Sekar et al., 2015; Verheijen and Sleegers, 2018; Zhang et al., 2013). For example, Sekar *et al.* and Li *et al.* both identified differentially expressed immune response genes (Li et al., 2015; Sekar et al., 2015). Celarain *et al.* showed that up regulation of microglia gene *TREM2* correlates with 5-hydroxymethycytosine enrichment in Alzheimer's disease hippocampus (Celarain et al., 2016). Zhang *et al.* identified an up expressed immune and microglia-specific co-expression module in AD brains specifically (Zhang et al., 2013). Miller *et al.* carried out gene co-expression network analysis using WGCNA and uncovered up gene pathways consistent with a key role for early microglia activation in AD (Langfelder and Horvath, 2008)(Miller et al., 2008). Humphries *et al.* performed network analysis on AD transcriptome and identified a network involved in innate immune response specifically correlated to AD (Humphries et al., 2015). However, these studies have not accounted for the altered

population of different cell types in the affected brain regions when performing the gene differential expression or co-expression network analysis. Therefore, in bulk tissue transcriptomic analysis, we have to be very cautious and determine whether the widely observed down-regulation of neuron gene expression and up-regulated microglia expression patterns are due to changes in transcriptional regulation or simply due to cell population change of certain cell type (Srinivasan et al., 2016).

Although this has been noticed and concluded that in the bulk brain tissue samples from AD patients, mRNA level change reflects cellular composition change (Srinivasan et al., 2016), no further investigation has been carried out to find out that for a specific cell type, how the cellular composition change impacts their corresponding marker gene expression levels. Specifically, whether the cellular composition changes affect gene expression in a similar fashion, and whether *bona fide* transcriptional regulations can still be revealed on top of the cell population effect in bulk tissue samples. In our study, we re-examined the gene expression changes of AD compared to normal brain from bulk brain tissues in the five major large transcriptomic datasets (Table-1), accounting for the cell composition changes in the affected brain regions of AD. We aimed to separate the *bona fide* transcriptional regulation signals from the expression level changes caused by cell population changes in the observed differential gene expression and differential co-expression phenomena. This is achieved by combining the gene differential expression (DE) analysis with gene differential co-expression (DC) analysis on the identified gene co-expression network modules, along with the estimation of relative proportions of major brain cell types in AD and normal brain samples. Gene co-expression analysis has been applied in AD research to identify clusters of genes (modules) with specific functions that were dysregulated in the AD brains (Humphries et al., 2015; Miller et al., 2008; Zhang et al., 2013). It is known that co-expressed gene modules are often specific to certain cell type due to the genes involved in common pathways, functions, and/or regulated by common upstream regulators (Seyfried et al., 2017). A gene co-expression module specific to certain cell type with no or minimal correlation change between two conditions but high level of differential expression (DE) indicate that the DE is possibly due to the cell abundance change instead of expression dysregulation. In contrast, changes of gene co-

expression relationships (differential correlation aka DC) are more likely due to some of the genes in the module regulated by altered activities of transcription factors or other regulators. Therefore, combining DE and DC analyses on gene co-expression modules along with cell type proportion estimation can differentiate the effect of cellular composition changes from *bona fide* regulatory changes. It will also identify gene modules with high level of DE and DC change, which may be linked to key upstream regulation change. Therefore, this analysis result can help generate new hypotheses and open new direction for the pursuit of the elusive AD etiology.

We first performed frequent gene co-expression network (FGCN) mining on five large human AD transcriptomic datasets generated from five separate studies, which contain multiple brain regions mentioned above. We obtained FGCN modules which were conserved across multiple datasets, avoiding potential technical platform or systematic bias in any single study. Then, we computed the DE score and DC score for each module and partitioned the modules into different categories according to the high/low DC and DE scores. Next, to adjust for the cell type composition changes in AD and normal brain samples, the abundance for each of the five cell types (neuron, microglia, astrocytes, oligodendrocyte, and endothelial cells) were estimated in every sample using cell type deconvolution method BRETIGEA (Chikina et al., 2015; McKenzie et al., 2018). We observed a consistently decreased abundance of neuron cells and increased abundance of glia cells (microglia, astrocytes) for the AD samples across all five datasets. Two gene modules were identified as microglia core gene modules from AD and normal samples respectively. Both modules showed large increased gene expression levels with minimal changes in correlations between gene expression profiles in AD samples. Further analyses indicated that the increased expression of genes from the microglia core modules with high DE scores are largely due to microglia cell population increase observed in AD progression instead of *bona fide* upregulation of gene expression within microglia cells, which is a novel finding. We also identified neuron marker enriched modules with high DC and high DE scores, which we believe are due to the combinatorial effects of neuron cell proportion change and dysregulation of neuron associated pathways in AD. The two astrocyte modules identified with high expression and high co-expression change are likely due to the same combinatory effect, and their eigengene expressions are highly correlated with the

6

clinicopathological biomarkers of AD. Furthermore, we identified six transcription factors for the first time that are frequently differentially expressed among nine transcriptomic AD datasets and are predicted to target multiple genes in these neuron modules, which can shed light on the changed connectome in AD brains.

Methods

Data Sets and Sample Processing

Five transcriptomic datasets from one or multiple regions of brain tissue from both AD patients and normal persons were used in this study (Table 1). Two of them are microarray datasets GSE5281 (Liang et al., 2007), GSE48350 (Berchtold et al., 2008) from NCBI Gene Expression Omnibus (GEO) and three are RNA-seq datasets from the Accelerating Medicines Partnership-Alzheimer's Disease (AMP-AD) Knowledge Portal. The three AMP-AD RNA-seq datasets were designated as MSBB (Wang et al., 2018), ROSMAP (De Jager et al., 2018) and Mayo (Allen et al., 2016) by the project original name or research site, which contain transcriptome-wide FPKM values or raw read counts for AD and normal brain samples. Only samples from patients diagnosed as "AD" were considered as AD samples. Four additional transcriptomic datasets containing AD and normal human brain samples are used for validation of differential gene expression. The sample size and dataset details for both FGCN mining and validation sets were summarized in Table 1.

Raw microarray datasets were processed using R/Bioconductor package "affy" to generate normalized expression profiles of the brain samples by RMA normalization using default parameters (Gautier et al., 2004). For preprocessing, all datasets were pre-filtered to remove probes without gene annotation. For genes with multiple probes, we selected the probe with the highest expression value for that gene (Zhang et al., 2010, 2012). For the RNA-seq datasets, genes with more than 50% zero expression levels across samples from each condition (*i.e.,* AD or normal) were removed. Next, for both microarray and RNA-seq data, genes with variance in the bottom 20 percentile of the entire dataset were removed. Genes with mean expression value in the bottom 10

percentile were also removed. At the end of the processing, 10,931 genes that present in all five pre-filtered datasets were used for the following frequent co-expression network mining.

Frequent gene co-expression network (FGCN) construction and module detection

Frequent GCN modules were mined separately in AD and normal samples. For microarray datasets, Pearson correlation coefficient (PCC) was used as the correlation measure of every gene pair. For RNA-seq datasets, due to the large range of expression values and non-Gaussian distribution, spearman correlation coefficient (SCC), which is much less sensitive to outliers, was used as described in (Yu et al., 2004, 2007).

Because the range of correlation values varies substantially among different datasets, instead of applying a uniform threshold across datasets, the gene pairs with top 5 percentile correlation coefficient values (|PCC| or |SCC|, with p-values less than 0.05) within each dataset was selected for the frequent network mining as in previous similar studies (Zhang et al., 2010, 2012). For each selected gene pair, the frequency was computed across five datasets and used as the edge weight for FGCN mining.

Previously reported lmQCM (local maximized Quasi-Clique Merger) algorithm (Zhang and Huang, 2014) was used to mine the frequent gene co-expression modules in AD as well as in normal samples, which has been implemented by the online tool package TSUNAMI (Huang et al., 2019). lmQCM allows overlaps between modules and is capable of identifying relatively smaller co-expressed local modules than WGCNA. The lmQCM algorithm takes five parameters (t, λ, γ, β, and minimum module size) that were set as following: t = 1.0, λ = 1.0, γ = 0.81 (*i.e.,* the initial co-expressed gene pair of a FGCN module should be a gene pair showing up in at least four of the five datasets), β = 0.3, and minimum module size = 10. Prefixes "AD" and "N" were used to designate modules mined from AD samples and the ones from normal brain samples respectively.

Differential expression (DE) measurement of GCN modules between AD and normal samples

To determine if a gene module is differentially expressed between AD and normal samples within a dataset, we adopted the DE score measure for individual genes similar to Lui *et al.* (Lui et

al., 2015) and extended it to measure a group of genes in a FGCN module. For a module with m genes from AD and normal samples, D denotes AD group, while N denotes normal group. We adopted the widely used absolute t-value in t statistics to quantify the degree of differential expression of expression levels. The t-value for a given gene *i* is defined as:

$$|t_i| = \frac{|x_D - x_N|}{\sqrt{\frac{s_D^2}{n_D} + \frac{s_N^2}{n_N}}}$$

where x_D and x_N were mean expression levels in disease and normal states, n_D and n_N were samples sizes of disease and normal states, and s_D and s_N were standard deviations of expression levels in disease and normal states (Lui et al., 2015).

To adjust for the different module size, the module DE scores were normalized by dividing the sum of t-value of all genes in a specific module with the module size. The normalized DE score is considered the measure of the overall differential expression level for the module. Higher DE score indicates a larger overall difference expression of a module. For each frequent module, we calculated its DE scores in all five datasets and then get the median of them as the median DE score for that module.

Differential co-expression (DC) measurement of GCN modules between AD and normal samples

Similarly, to determine the differential co-expression level of a module in two different conditions (*i.e.,* AD vs. Normal) within a dataset, we first computed the differential co-expression measure of each pair of genes in the module (Lui et al., 2015). For a pair of genes, we used a differential co-expression measure, Z, which quantifies the correlation difference between two genes in AD and normal samples. Z_{ij}, between X_i and X_j is defined as:

$$Z_{ij} = \frac{|z_{ij}^N - z_{ij}^D|}{\sqrt{\frac{1}{n_N - 3} + \frac{1}{n_D - 3}}}$$

where n_D and n_N were samples sizes in the normal and AD disease conditions, z_{ij}^N and z_{ij}^D were the Fisher-transforms of the Pearson/Spearman correlation coefficients r_{ij}^N and r_{ij}^D, which were defined as:

$$z_{ij}^N = \frac{1}{2} ln \left| \frac{1 + r_{ij}^N}{1 - r_{ij}^N} \right|,$$

$$z_{ij}^D = \frac{1}{2} ln \left| \frac{1 + r_{ij}^D}{1 - r_{ij}^D} \right|.$$

We calculated Z values for every pair of genes in a given module, then normalize the score by dividing with the L_2 norm of Z values within a module. The resulted L_2 norm of Z values is considered the overall differential co-expression measure of the module, termed as *DC score*. Higher DC score indicates larger overall differential co-expression level of a module between AD and normal samples. For each frequent module, we calculated DC scores in all five datasets and then obtained the median of them as the median DE score of that module.

The original z-score as well as the modified DC score only take the absolute value of a correlation coefficient into account. In order to determine if the module gains or loses correlation in AD versus normal condition, we also measured the correlation change using our previously developed metric Centered Concordance Index (CCI, Han et al., 2016). CCI ranges from 0 to 1, and the larger the CCI value, the stronger the genes are correlated within a module. The CCI boxplot of a given module were plotted across five datasets to measure if a module gains or loses correlation in AD vs. normal condition, and the significance of CCI change was analyzed by t-test.

Categorizing GCN modules into high and low DE and DC groups

With DE and DC measures for all modules defined, we further classified the modules into high or low DE/DC category. The median of the DC/DE scores of all modules were used to define "low" and "high" in DC/DE (solid lines on Figure 2). Then we partitioned the modules into 4 categories for further analysis: high DC and high DE (HDC_HDE), high DC and low DE (HDC_LDE), low DC and high DE (LDC_HDE), low DC and low DE(LDC_LDE).

FGCN module functional enrichment analysis

ToppGene (Chen et al., 2009) was used to perform GO and pathway enrichment analysis of the FGCN modules. The functional enrichment terms were considered as significant enriched if they were with FDR adjusted p-value less than 0.05. Enriched transcription factors (TFs) targeting the module genes were obtained from Enrichr with database "TRANSFAC_and_JASPAR_PWMs" with cut-off for adjusted p-value being 0.05 (Kuleshov et al., 2016). To check if the module gene members are transcription factor or not, we compared the gene list against an online database TFcheckpoint (Tripathi et al., 2013) to uncover if any of the genes are TFs.

Differential gene expression analysis between AD and normal samples for individual datasets

We also performed differential gene expression (DGE) analysis for each of the five datasets with R package "limma" for two microarray datasets, as well as R package "limma-voom" (built for RNA-seq analysis) for three RNA-seq datasets respectively (Ritchie et al., 2015). Foldchange of 1.2 and FDR adjusted p-value < 0.05 were used as the threshold for differentially expressed genes selection. The differentially expressed genes were defined as differential expressed in at least two datasets. To confirm the DEGs from the above analysis, we further acquired four human brain transcriptomic datasets containing AD and normal samples from NCBI GEO database (GSE15222, GSE33000, GSE84422) and from Allen Brain Atlas Aging, dementia and TBI study (https://aging.brain-map.org/), and repeated the same DGE analysis on them. The results are updated in Table-3.

Brain cell type deconvolution on each individual dataset

The brain cell type-specific marker gene lists were obtained from McKenzie *et al.* (McKenzie et al., 2018) in which five human and mouse cell type-specific transcriptome-wide RNA expression datasets were compared and contrasted to identify consensus brain cell marker genes conserved across five datasets. We estimated brain cell type relative proportions using the R package "BRETIGEA" (McKenzie et al., 2018) with the marker gene lists from McKenzie *et al.* (McKenzie et al., 2018) to deconvolute five major cell types in the brain - namely neuron, microglia, astrocytes,

endothelia, and oligodendrocytes. The program outputs cell proportion indices, with a high index indicate a relatively high cell proportion. Cell proportion indices were obtained by BRETIGEA with default settings. Boxplots of specific cell type relative proportions in AD and normal samples (plot results from all regions if multiple regions exist in a dataset) were generated across five datasets. The change of a specific cell type relative proportions between AD and normal samples was analyzed by Wilcoxon rank sum test, with a significant difference as the FDR adjusted p-value less than 0.05 or 0.01 (* for p-value < 0.05, ** for p-value < 0.01 in the boxplot).

FGCN modules enrichment of specific cell type marker genes

The same sets of marker gene lists for neurons, astrocytes, oligodendrocytes, microglia, endothelial were used for the cell types enrichment analysis for AD as well as normal FGCN modules, which are listed in Supplementary S1. Enrichment of certain cell-type markers in a module was assessed by hypergeometric test, with a significant enrichment as FDR adjusted p-value less than 0.05.

FGCN structure change analysis by identifying gain or loss hub genes from normal to AD condition

In a weighted network, the degree of a node which labels the node strength is defined as the sum of weights. We examined the gene node degree distributions of the co-expression networks generated from the highly correlated gene pairs in the five datasets (see the FGCN mining section) and defined the nodes with the top 5 percentile degree as hubs genes in this network based on literature (Sun and Zhao, 2010). We compared the hub genes in AD and normal networks and defined the hub genes identified in normal network but not in AD as lost hub genes and the hub genes shown in AD network but not present in the normal one as gained hub genes.

FGCN module eigengene computing and the correlation with cell abundance/AD clinicopathological traits

Principle component analysis (PCA) was performed on the expression matrices of the identified FGCN modules from AD as well as normal samples. The first principle component (PC1) of the expression matrix of a specific module was computed as eigengene value of that module to represent the weighted average expression profile (Zhang and Huang, 2014; Zhang et al., 2010). We tested the association of the expression of each FGCN module with clinicopathological traits in MSBB dataset by calculating the Spearman correlation between each module's eigengene profile and three clinicopathological traits, namely Clinical Dementia Rating scale (CDR), Braak & Braak stage score (BB score), and neuropathology trait of amyloid plaque scale (Plaque_Mean). To check what extent the cell proportion change affects the module expression level, we also examined the correlation of module eigengene values with the corresponding estimated cell type relative proportions by Spearman correlation for all five datasets.

Examine the neuron proportion changes across different brain regions in AD and normal individuals respectively

In the analysis we observed some gene modules specifically enriched with neuron markers showed stronger correlation relationships in AD than in normal, which is intriguing given the perturbed neuron activities in AD. Since we performed the FGCN module mining with combined data for those datasets with multiple brain regions, we suspected that for neuron associated genes, their detected co-expression relationship may due to changes in neuron populations either across brain regions or across the cohort. Therefore, we first check if there are concordant cell population changes in different brain regions in a specific condition (AD or normal). Specifically, we examined if the estimated neuron or microglia population change correlates differently across brain regions in AD (comparing to normal brain) samples. The MSBB dataset was used as it contains matched samples of four brain regions from 56 AD patients and 40 normal persons. For each condition (AD or normal), we computed the CCI value that indicates the overall correlation of cell relative proportions across the four brain regions. We compared the CCI of neuron/microglia relative proportion across AD patients to that across normal subjects. Higher CCI for AD may lead to attribution of cell population changes in brain regions to the higher co-expression relationship for

13

neuron gene modules in AD while lower CCI for AD may indicate that the co-expression relationship may mainly be due to neuron population changes or regulatory changes among patients.

Identify genes regulated by enriched transcription factors in neuron modules

To identify genes that can be regulated by transcription factors listed in Table-3, we downloaded ChIP-seq data for BCL6 and STAT3 from (Yan et al., 2013 and Zhou et al. 2017), then overlapped the peak region from those data with our AD1 and N1 module gene list. The full list of genes with BCL6/STAT3 peaks in their promoter regions were listed in Supplementary Table-12.

Results

FGCN modules from AD and normal brains are separated into different groups by DE and DC levels

Gene co-expression network (GCN) modules are subsets of genes with tightly correlated expression profiles in a gene co-expression network. Genes in a GCN module are often co-regulated and participate in common biological processes or pathways in homogeneous tissues. Our frequent network mining approach identified frequent gene co-expression network (FGCN) modules detected across multiple datasets from independent AD studies, thus eliminating the potential systematic bias from cohort composition, experimental design, or technical platform in any single study. This approach allows us to identify *bona fide* co-expressed genes in a specific condition and eliminates spurious correlation from data/study bias (see Methods for detail). In this study, we obtained 15 FGCN modules in AD samples from the five large transcriptomic datasets mentioned previously (Table 1), ranging from AD1 (largest, 1,247 genes) to AD15 (smallest, 10 genes) and 9 FGCN modules in normal brains ranging from N1 (largest, 1,003 genes) to N9 (smallest, 10 genes)(Supplementary Table 1). The majority of these modules are specific to each condition (AD or normal) as determined by minimal gene overlaps (Supplementary Table 1), which is consistent with previous finding that AD is a complex disease with gene co-expression largely perturbed in AD brains compared with normal ones (Xiang et al., 2018; Zhang et al., 2013). Only three pairs of modules contain genes largely overlapped between AD and normal conditions

(Jaccard index > 0.3) and the highest overlapping is between AD1 and N1 (Jaccard index 0.47; Supplementary Table 1).

To characterize the module in overall differential expression and differential co-expression levels, we calculated DE score and DC score between AD and normal samples respectively for all modules using the formula modified from Lui et al. 2015 (Methods; for scores for all modules, please see Supplementary Table 2). In Fig. 1a, the DC and DE scores for all modules were plotted. Separated by the median levels of DE and DC, all 24 modules were partitioned into four quadrants. The modules with high DE and high DC scores are denoted as HDC_HDE, the ones with high DE and low DC scores are denoted as HDE_LDC, the ones with low DE and high DC scores are denoted as LDE_HDC, and the ones with low DE and low DC scores are denoted as LDE_LDC (Fig. 1a).

To further characterize the FGCN modules and explore their possible association with AD pathology in different partitions, we examined those modules for the enrichments of DE genes, including biological process/pathway and cell type specificity (Supplementary Table 3). The majority of the modules are enriched with specific cell types (Table-2) or specific biological processes (listed on the side of Fig. 1a). We also checked if the modules are enriched with differentially expressed genes between AD and normal samples (Supplementary Table 4-5). The four partitions of modules each shows distinct enrichment of differentially expressed genes and categories of biological processes (Table 2). HDE_HDC partition showed large perturbation in expression levels. Among them, nearly or more than half of the genes in AD1, N1, N2 and N8 were up expressed in AD, while more than 80% of the genes in AD5, AD11 and N5 were down expressed. These modules were significantly enriched in three categories of biological processes that all linked to AD pathology. More specifically, AD1, N1, N2 correspond to neuronal function (Uylings and de Brabander, 2002). N5 and AD11 are involved in angiogenesis and vasculature development, which is consistent with the previous finding that changes in vasculature leading to altered blood flow in the brain is a risk factor for AD (Dickstein et al., 2010; Kelleher and Soiza, 2013a; Koizumi et al., 2016). AD2 is linked to gliogenesis and cell motility. N8 is involved in gap junction, which supports the assumption that gap junction associated with blood brain barrier may be disrupted in AD to

allow the access of virus or bacteria to AD brain (Cai et al., 2018; Devanand, 2018; Montagne et al., 2017; Readhead et al., 2018).

The majority of genes from modules in the HDE_LDC partition were over-expressed and involved in immune response (AD3, N6), neurogenesis (AD14), or cell adhesion (AD8). In LDE_HDC partition, modules were involved in gliogenesis (AD7), ribosomal function (AD9), and myelination (AD12). LDE_LDC modules mainly contain genes with house-keeping functions such as translation (AD10, AD15, N7), interferon signaling (AD13), protein metabolism (AD4), lipid metabolism (N3), and gliogenesis (N4, AD6). No significant DE genes appear in these modules except for AD6, AD12, N3 and N4. All these indicate that the low DE modules were less associated with AD pathology.

Besides distinct DE patterns and biological processes of the four partitions, we observed that a large number of FGCN modules were specific to certain cell types (Fig. 1b, Table-2, Supplementary Table 6), especially for HDE_HDC modules. HDE_HDC modules were enriched with neuron (AD1, N1, N2) and astrocyte (AD2, N5) markers, while the HDE_LDC modules were enriched with microglia (AD3, N6) and endothelial markers (AD14, AD8). For LDE_LDC and LDE_HDC modules, the associated cell types were mainly oligodendrocytes (AD4-7).

Cell type deconvolution analysis indicate significantly decreased neuron population and increased glia cells in AD brains

Mounting evidence showed that cell composition changes in AD brains compared to normal ones, with various extent of neuron loss and elevated gliosis (Andrade-Moraes et al., 2013; Gomez-Nicola and Perry, 2016; Jain et al., 2015). To quantify the cell type composition changes and their association with module differential expression or differential co-expression, we estimated the cell composition of five major brain cell types in AD and normal brain samples using BRETIGEA, a method that has been previously developed to successfully quantify the cellular composition (McKenzie et al., 2018). Our results clearly showed that neurons were consistently and significantly decreased in AD samples while microglia, astrocytes, and endothelia were all significantly increased in AD samples across all five datasets (Fig 2. a-d), which are consistent with previous

reports (Cotman and Su, 1996; Fakhoury, 2018; Gomez-Nicola and Perry, 2016). We also observed that oligodendrocytes were only significantly changed in two of the five datasets.

The highly differential expression in microglia modules in HDE_LDC section were largely due to cell population change

The overall differential expression of a GCN module may reflect the change of specific regulatory network/pathway, the representative cell population change, or the combinatory effect of both, any of which can reflect certain pathological changes in AD brains. Our combined DE-DC analysis provides an approach to differentiate the regulatory change and the cell population change. Specifically, the HDE_LDC category, where modules are highly differentially expressed but with minimal disruption of gene co-expression level, indicates that the overall gene expression changes of these modules may likely result from the change of representing cell population. This partition includes N6, AD3, and AD14 (Fig. 1b). Due to the relative high DC value of AD14 which is located near the DC median line, we only consider AD3 and N6 as HDE_LDC. AD3 and N6 are both microglia gene enriched modules. 17 of the 21 genes in N6 are overlapped with AD3, so we combined the two modules into one as the microglia module. This microglia module is enriched with innate immune system functions and several infectious disease pathways, but does not contain cytokine genes (Supplementary Fig. 1, Supplementary Table 1). Near 90% of genes were upregulated in AD samples in at least two of the five datasets (Table 2). Notably, genes in this microglia module are also among the core microglia genes in previous study (Patir et al., 2019), and one third of this module (14 out of 42) are also identified from a microglia subpopulation from single cell transcriptomic study that is highly correlated to AD pathology (Mathys et al., 2019). However, none of the 42 genes was among the recently identified disease-associated microglia genes (Keren-Shaul et al., 2017). The low DC scores indicate the correlation relationship within the module was not disrupted significantly in AD. Indeed, as shown in Fig. 3a, the networks of the microglia module in AD and normal samples showed comparably high density.

To further verify the co-expression relationships among genes within the microglia module, we used another correlation metric, centered concordance index (CCI, Han et al., 2016), to evaluate

the microglia module's gene inter-correlation in each of the five datasets separately. The CCI values of AD versus normal samples align very closely to the diagonal line in the upper right corner, which are high in both conditions, indicating these genes are highly correlated with each other in both AD and normal samples in all of the five studies (Fig. 3b). Consistent with our postulation that the over-expression of microglia core modules may be due to cell composition change in AD, a significant increase of estimated microglia relative proportions in AD compared to normal is observed across five datasets (Fig. 2b). Furthermore, the microglia module expression, as measured by the module's eigengene value, are directly in proportion to the estimated microglia cell relative proportions (PCC = 0.948, Fig. 3c for GSE48350; Supplementary Fig. 2 for other four datasets). As a comparison control, much lower correlation was seen between the eigengene values of the microglia module and neuron cell relative proportions (Fig. 3c for GSE48350; Supplementary Fig. 2 for other four datasets). To further test this hypothesis, we then calculated and plotted the DE score vs. cell proportion change for genes from the combined AD3 and N6 modules across all five datasets (Fig. 3d). The result shows a strong positive correlation between DE scores and microglial cell proportion changes (PCC = 0.885, p-value < 0.05). All of these results, consistently observed across five independent AD cohorts, confirm that the microglia core genes' co-expressions are not perturbed in the AD brains and the module over-expression are mostly due to the high microglia proportion in the AD bulk brain samples.

Since some of the previous differential gene expression or co-expression studies of AD (Miller et al., 2008; Zhang et al., 2013) performed on bulk transcriptomic data may not take cell abundance change into account, our findings indicated that some of the identified AD associated microglia gene upregulations may need re-evaluation.

Neuron modules showed high DE that may be due to decreased proportion of neurons and high DC that may indicate regulatory change and neuron hyperactivity.

The HDE_HDC partition in Fig. 1a contains gene modules that are highly disrupted in both gene expression and co-expression, which indicates that the gene expression changes of the modules are likely involving gene regulatory changes. We took two modules AD1 and N1 for further

investigation. AD1 and N1 are the largest modules in AD and normal samples respectively. Genes in these two modules largely overlap with each other (Supplementary Fig. 3). Gene ontology (GO) and pathway enrichment analysis on the modules shows that both of them are highly enriched with neuron markers for neuronal functions, as well as energy metabolism and mitochondrial functions (Fig. 4a). To further investigate the difference between the two neuron modules, we categorized the genes from the two modules into three groups: the overlapping genes, the genes unique to AD1, and the genes unique to N1 (Supplementary Fig. 3). The GO/pathway enrichment results for the three groups are shown in Supplementary Table 7-9. The genes shared between the two modules are enriched in AD pathway, synapse associated pathways, the citric acid cycle, and respiratory electron transport pathway. Interestingly, the genes unique to AD1 are enriched with more synapse and neuron pathways, such as protein-protein interactions at synapses, glutamate binding, activation of AMPA receptors and synaptic plasticity, and voltage gated channels activity, while ubiquitin-mediated proteolysis was found unique to N1. The findings from genes only present in A_M1 support the recent discovery that neuron hyperactivity, especially the glutamate signaling hyperactivity observed in AD brains (Stargardt et al., 2015),(Zott et al., 2019). The fact that ubiquitin-mediated proteolysis only preserved in N1 implies that the decoupling of such activity with neuron activity in AD brains. It may reflect a disrupted normal ubiquitin-mediated proteolysis pathway in AD, which has been observed before in neurodegenerative disorders, and has been attributed to the filamentous protein aggregation in the affected brains (Layfield et al., 2003).

As seen from Table 2, about half of the genes in the two neuron modules (47.5% of AD1 module, 44.8% in N1) are under-expressed in AD brains in at least two datasets. We further investigated if the large proportion of under-expressed genes could be due to an overall downregulation of the neuron associated genes, or due to a decreased proportion of neuronal cells. A consistently significant decrease of neuron cell population in AD compared to normal has been observed across all five datasets (Fig. 2a). To investigate whether the expression of neuron modules were significantly affected by the neuron cell population change, we check the correlation between the AD1/N1 eigengene values (module overall expression) and the neuron cell proportion across all samples. The result showed that expression of AD1/N1 modules were highly correlated

with neuron cell relative proportions in all five datasets (median PCC value 0.937 for AD1, and 0.928 for N1, Fig. 4c,e for GSE48350 dataset; Supplementary Fig. 4-5 for other four datasets), suggesting that neuron loss in AD accounts for the decrease of the neuronal gene expression significantly. As the control, much lower correlations were seen between the neuron module expression and the microglia cell relative proportions (Fig. 4c-d for GSE48350; Supplementary Fig. 4-5 for other four datasets).

The high DC scores of the neuron modules imply large changes in correlations of gene expression profiles in AD. To check which cohort (AD or normal) has a higher within-module gene correlation, we computed the CCIs for the neuron modules. To our surprise, the CCI values are consistently higher in AD brains for all five datasets (Fig. 4b), which means these genes are more correlated with each other in AD than in normal brains. Since three of the five datasets contain multiple brain regions, we first checked if the neuron proportion varies consistently across different brain regions in AD brains more than in normal ones, which could potentially lead to a high correlation for the gene expression profiles. As we know, AD is first observed in hippocampus region of the brain and gradually spread to the entire cortex (Anand and Dhikav, 2012). The extent of neuron loss differs in brain regions due to different extent of damage to each region in AD progression, but this difference due to neuron loss may not exist in normal aging brains, which could potentially result in high correlation of neuron module genes in AD. In the dataset of MSBB, where four brain regions (frontal pole, superior temporal gyrus, parahippocampal gyrus, inferior frontal gyrus) are present, we discovered that the inter-correlation of neuron relative proportion variations across the four brain regions (CCI 0.820) was actually lower among AD patients than the ones among normal people (CCI 0.944). As a contrast, the inter-correlation of microglia relative proportion variations across the four brain regions among AD patients is comparable to that across normal persons (CCI 0.856 for AD, 0.868 for normal). We therefore concluded that the variation of neuron relative proportions across regions in AD did not contribute significantly to the high DC value in AD1/N1 modules.

To further check if the correlational change among in AD neuron module is really increased, we examined the GCN structural change by checking the presence of loss of hub genes in AD vs

normal GCN. We obtained 547 hub genes each from AD as well as normal GCN networks respectively (designated by possessing top 5 percentile connections in each condition), then identified lost and gained hub genes in AD by comparing the two hub gene lists (Supplementary Fig. 6). The hub genes shown in normal GCN but not in AD GCN are denoted as *lost hub genes*, and the ones shown in AD GCN but not in normal GCN are denoted as *gained hub genes* in AD. We obtained 136 genes for each case (Supplementary Fig. 6). We then crosschecked these genes with AD1/N1 and found that most of the changed hub genes are in the neuron modules AD1 or N1. Specifically, AD1 contains 123 of the 136 gained hub genes in AD, and 84 of the 136 lost hub genes in AD; while N1 accounted for 68 of the 136 gained hub genes in normal and 96 of the 136 lost hub genes in normal. The result suggested that a large change of network connection occurs in neuron modules of AD compared to normal brains. In AD1, both gained and lost hub genes include multiple transcription factors (ie, *PEG3*, *ZNF365*, *HERC1*, *FBXW7*, *TERF2IP*, *STAT4*, *NLK*, *RBFOX2*, *CDK5*, *POLR2K*, *PSMD12* are gained hub transcription factors in AD; *BZW2*, *RNF6*, *ZBTB11*, *SORBS3*, *TRIM56*, *MORF4L1*, *CNOT7*, *RBCK1*, *SMYD3*, *RAN*, *KLHL12*, *ZC3H15*, and *EID1* are lost hub transcription factors in AD). These network structural changes and transcription factor changes not only indicate the DC change is *bona fide* correlational change among module genes, but also suggest that the DE changes observed in these two modules contain change from regulatory level.

To further confirm the expression regulation changes among the neuron modules, we checked if there are changed upstream regulators for the FGCN modules. We performed regulatory transcription factor enrichment analysis to identify regulatory cascades within each module using "TRANSFAC and JASPAR PWMs" database (Kuleshov et al., 2016). Many genes from AD1 and N1 were targeted by TFs that are frequently differentially expressed among the five datasets as well as the four additional validation datasets as shown in Table 3. Specifically, five TFs were predicted to regulate 116 to 204 genes in AD1 and six TFs were predicted to regulate 94 to 273 genes in N1 (for details see Supplementary Table 10). To verify if the TFs may regulate the predicted genes, we acquired the ChIP-seq peak data of the top differentially expressed TFs, Bcl6 and Stat3 (Yan et al, 2013 and Zhou et al, 2017). We then crosschecked AD/N1 genes promoter

region for binding peaks of these two TFs. We found 18 genes from AD1 and 56 genes from N1 have Bcl6 binding on their promoter regions; while 238 genes from AD1 and 685 genes from N1 have Stat3 binding sites on their promoters (Supplementary Table 12). Bcl6 is a transcription repressor that interact with Stat protein to modulate gene expression in B cells. It also plays a role in neurogenesis and neuron differentiation (ref). More experimental evidences from published animal model studies also indicated that constitutive expression of BCL6 in transgenic mice showed upregulated expressions of *ANXA6, MAPK6, HDAC9, PFKM, GABARAPL1* (Bedekovics et al., 2016), which are all from our AD1 module, and downregulates the expression of *CCND2* and *CDKN1B* in another BCL6-activating mice strain (Kusam et al., 2004), which are from the N1 module. Stat3 protein upregulates *ATR* expression to control DNA damage and affect apoptosis. It also binds to the promoter region of *OPA1* and upregulates *OPA1* expression [PMID: 28637784]. *OPA1* is essential for mitochondrial fusion and the mutation of this gene has been associate with neurodegenerative diseases [PMID: 26113818]. Both *OPA1* and *ATR* are members of AD1 and N1 modules.

The enriched differentially expressed TFs and their confirmed regulatory role in the M1 genes indicates that the high DE modules in this category are not only due to neuron loss, but also due to upstream regulatory changes. We therefore concluded that the overall decreased expression level of neuron associated genes in AD1/N1 are due to both decreased proportion of neuron cells and regulatory changes. However, without further data, it is difficult to quantify the degree of contribution for either cause.

Astrocyte modules (AD2 and N5) are also significantly differentially expressed and differentially co-expressed in this quadrant (Fig. 1a). In fact, the two astrocyte modules are the mostly elevated expressed modules. We have not further pinpointed the cause of these changes, but with the consistently increased proportions of astrocytes in AD brains among the five datasets (Fig. 2c), we speculated that the two causes (i.e., cell population and up-regulation) both account for the changes in astrocytes modules, similar as observed in neuron modules.

HDE modules, especially astrocyte modules, are highly associated with AD neuropathological features and cognitive deterioration

Since HDE modules are highly enriched with neuronal and microglia functions, pathways, and genes that are previously linked to AD pathology, we examined the association of the expression levels of genes in these modules with neuropathological features and clinicopathological markers. We computed the correlation of the module eigengene values (module-level gene expression) to the clinicopathological markers including Clinical Dementia Rating scale (CDR), Braak & Braak stage score (BB score), and amyloid deposition plaque scale (Plaque_Mean) in MSBB dataset, in which all of these clinical attributes are available (Supplementary Table 11). Modules were ranked from high to low based on the average of correlation value to the three traits (Fig. 5a). We observed that high DE modules show significantly stronger association with the clinicopathological markers than the low DE modules (Fig. 5a). The expression of the top two modules, AD2 and N5, which are both highly enriched for astrocyte marker genes (Fig.1b), are strongly correlated with the three AD clinical biomarkers, consistent with the findings that astrocytes are involved in the sustained immune response and mitochondria-related function in AD brains (Fakhoury, 2018; Sekar et al., 2015). Modules enriched with neuron function and neuron cell markers (AD1 and N1) are significantly decreased in expression and negatively correlated with all three AD clinical traits, consistent with gradual loss of neuron and neuronal functions in AD progression. The expression of the aforementioned two microglia modules are positively correlated with all three clinical traits, consistent with the activated immune response observed in AD brains. The expression of endothelia-marker enriched modules (AD8,13,14) is weakly positively associated with the disease progression markers, and the oligodendrocytes modules (AD4-7, N3,4) have even weaker association with them. Among all FGCN modules which expression is highly correlated (correlation absolute values >0.3) with the clinical markers, only two modules N8 and AD11 are not specific in certain cell types (Fig. 5a). N8 is enriched with gap junction functions and is negatively correlated with all three traits, consistent with gap junction dysfunction observed in AD (Nakase and Naus, 2004). AD11 is enriched with angiogenesis functions, and genes are involved in vascular morphogenesis. It is positively correlated with CDR and Braak stage, supporting with the fact that neurovascular dysfunction plays important role in AD (Dickstein et al., 2010).

Furthermore, we also assessed the association of cell type relative proportions with clinicopathological markers, to check if the cell composition change patterns correlate with the disease progression, which shows the necessity of taking cell type deconvolution into consideration in the transcriptomic analysis with bulk tissue samples. This was achieved by computing the correlation of the cell proportion in the five major cell types (neuron, microglia, astrocytes, endothelial and oligodendrocytes) to the same three clinical biomarkers (CDR, BB score, Plaque_Mean) in MSBB dataset. We did observe high correlation of cell proportions with the three clinical biomarkers except for oligodendrocytes. More specifically, relative proportion of neuron was negatively correlated with clinical traits (Fig. 5b), consistent with the gradual neuron loss in AD progression (Uylings and de Brabander, 2002). Relative proportion of microglia and astrocytes were positively correlated with clinical traits (Fig. 5b), consistent with microgliosis and astrogliosis around amyloid plaques (Fakhoury, 2018). We also observed significant positive correlation of endothelial with AD progression, consistent with endothelial activation in AD (Grammas, 2011; Kelleher and Soiza, 2013b; Koizumi et al., 2016). No significant correlation of oligodendrocytes was observed, which indicated less association of oligodendrocytes with AD compared to the other four cell types (Fig. 5b).

Discussion

Due to the rapid development of next-generation sequencing, large cohorts of transcriptomic datasets from mixed cell types of human brains have accumulated quickly for AD research (Allen et al., 2016; De Jager et al., 2018; Wang et al., 2018). Although single cell sequencing technique start to emerge, issues such as the technical difficulty of cell dissociation and sample preparation, very high dropout rate for signal reading, and the scarcity of brain samples, made it unlikely to have large cohort of AD single cell studies available in the near future (Mathys et al., 2019). To fully take advantage of the existing bulk tissue transcriptomic datasets and identify AD associated pathways, comparative studies have been conducted between AD and normal brains (Miller et al., 2008; Zhang et al., 2013). Some studies also applied GCN network mining approach. For example, Zhang *et al.* applied weighted gene co-expression network analysis (WGCNA) and identified an AD

associated module enriched in microglial function (Zhang et al., 2013). Miller *et al.* adopted WGCNA to uncover disease-relevant expression patterns for major cell types (Miller et al., 2013). However, these studies all focused on just a single dataset. In this study, we mined frequent gene co-expression network (FGCN) on five independent transcriptomic gene expression datasets that encompass a total of 1,681 brain tissue samples from AD and normal human subjects to avoid potential system bias from a single dataset, and identified co-expressed modules that are consistently present in multiple brain regions in a large AD population. Member genes in every identified gene modules are consistently highly correlated with each other across all five datasets.

Many transcriptomic analyses have shown that there are significant upregulation of immune response genes and downregulation of neuronal function genes in AD (Miller et al., 2008; Zhang et al., 2013). It is widely accepted that microglia clusters near amyloid plaques and microglia-mediated inflammation are activated in in the progression of AD (Clayton et al., 2017; Hansen et al., 2018; Katsumoto et al., 2018). However, most of such studies used bulk brain samples containing mixed cell types without taking microglia cell abundance change into account. To fully and accurately infer AD associated transcriptomic changes, it is necessary to re-examine the effect of cell population changes on gene transcriptomic levels in AD research. In order to distinguish whether the transcriptomic changes are due to changes in gene regulation or cell composition, or both, we introduced a new approach by combining the differential expression and differential co-expression analysis, together with cell type deconvolution to characterize cell specific FGCN modules and differentiate the cause of transcriptomic changes. To the best of our knowledge, this is the first study to combine multiple transcriptomic datasets of AD for frequent network mining and the first one with combined DE and DC analyses along with cell abundance estimation in AD.

The highly enriched microglial marker modules show highly differential expression but minimal co-expression change in AD compared to normal samples. Together with the increased microglia cell population from normal to AD and other analysis, we demonstrated that the overall expression changes of these modules are largely due to cell population change instead of transcriptional regulation, which is consistent with the findings by Srinivasan et al. (Srinivasan, et al. 2016). However, this does not rule out the possibility of a subpopulation of disease-associated microglia

(DAM) cell that may exist in AD brains, as previously identified in (Keren-Shaul et al., 2017, Rangaraju et al., 2018). Actually, out of the 2,077 DAM genes identified in the gene co-expression study (Rangaraju et al. 2018), eight of them are co-expressed in the core microglia gene module we identified (AD3/N6, Supplementary Table 13). However, when only examining this subset of genes, the subset's expression correlation (measured by CCI) are not altered significantly as shown in the Supplementary Figure-7. Please note that the 2,077 DAM signature genes were from *APP* transgenic AD mouse model samples (Rangaraju et al., 2018), which may be more similar in terms of disease mechanism to early onset AD rather than late onset AD analyzed in this work. Regardless of the origin, their correlations remain unchanged, which indicates that, just like the rest of the microglia module genes, this subset of genes' relatively high DE can be explained by the increasing of microglia proportion.

In the recent published single cell transcriptomic analysis of 48 AD human samples (Mathys et al., 2019), a specific subtype of microglia is shown to have strong correlation between its cell abundance and the pathological phenotype. Interestingly, among the 77 signature genes for this cell subtype, 14 of them overlap with our combined microglia modules which contains 42 genes (Supplementary Table 13). This significant overlap (Fisher exact test p=4.38E-24) is consistent with our notion that these signature gene expression changes largely reflect cell population change instead of gene regulation change. The strong correlation between the 14 genes for the specific subtype with general microglia genes in our analysis suggests that the proportion of this subtype in general microglia population may remain consistent.

Please note that, the microglia core gene module does not include any cytokine gene, which were known to be upregulated in the neuroinflammation of AD brains (Morimoto et al., 2011). We do not rule out either the possibility that certain individual microglia genes, such as TREM2 or TYROBP, are up-regulated in AD brains in some samples as observed in (Celarain et al., 2016; Haure-Mirande et al., 2017; Ma et al., 2015). However, in the five datasets we have analyzed, the tightly co-expresed core microglia module as a whole, its expression change is largely due to the microglia proliferation in affected AD brains.

For the genes in AD1 and N1 modules, although they share many common genes, a lot of hub genes have changed. Several transcription factors that are differential expressed in AD may account for such change (Table-3), which differential expressions were further confirmed in additional four AD transcriptomic datasets from separate studies (Table-1 and 3). To check if they indeed regulate the genes in those two modules, we further analyzed BCL6 and STAT3 as the examples since they the most frequently differentially expressed TFs. We found these two TFs bind to a large number of genes from AD1 and N1 as indicated from ChIP-seq experiments, which provided strong evidences that they serve as the upstream regulators for those genes' expression.

Besides the experimental evidences from transgenic mice and that Bcl6 and Stat3 regulate M1 genes, as mentioned in the Results. We also found 21 of the AD1 genes are Stat3 upstream regulators, eight of which have been shown recently shown to involve in AD pathology in transgenic mouse model studies (*MAPK1* (Du et al., 2019), *RHEB* (Shahani et al., 2014), *SUMO1* (Knock et al., 2018), *VPS35* (Li et al., 2019), *GSK3B* (Llorens-Martin et al., 2017), *DYRK1A* (Branca et al., 2017), *EPHA4* Fu et al., 2014) or in Parkinson's disease (*PAK4* Won et al, 2019)). Combined with the fact that Bcl6 also regulates the expression of STAT3 (Peng et al., 2019), this further indicates a central role of Stat3's in the regulation of neuron (M1) modules from AD and normal samples.

Although we observed overall decreased expression of neuron modules which was consistent with the neuron loss in AD brains (Srinivasan et al, 2016), to our surprise, we also detected an increment in overall co-expression level of the neuron modules in AD brains across all five datasets, which is a new discovery. This increased correlation in the neuron module may be linked to the neuronal hyperactivity in AD previously observed in both human and mouse models (Stargardt et al., 2015). Evidence from both imaging and molecular studies revealed that neuronal hyperactivity, especially glutamate signaling, occurs in the cortex and hippocampus during or even before MCI, which modulates Aβ levels and triggers synaptic dysfunction in AD (Pomara and Bruno, 2018; Stargardt et al., 2015; Zott et al., 2019). Multiple experiments in mouse model of AD also showed clustering of hyperactive neurons near amyloid plaques and the reduction of neuronal hyperactivity can prevent further build-up of amyloid plaques and synapse loss (Busche et al., 2008, 2012; Lerdkrai et al., 2018; Nuriel et al., 2017). Since many genes that are uniquely presented in the AD

neuron module are also involved in glutamate signaling and synaptic activity (see Supplementary Table 2), we speculate that the strong co-expression among these genes as well as the upregulated TFs may imply some form of hyperactivity still remained for neurons in symptomatic AD brains. Such signals could be previously overlooked due to the overall decreased neuron population in late AD stage (previous studies indicated such hyperactivity only present in the asymptomatic phase of AD (Bossers, et al. 2010 ; Lau, et al. 2013)). In the future, we plan to further quantify the impact of neuron population change on expression signals and mitigate its effect to discover real expressional changes due to regulation.

Our results show that modules enriched with neuron gene markers are under-expressed while modules enriched with microglial, astrocytes, endothelial gene makers are over-expressed in AD brains. This is consistent with recent single-cell transcriptomic study on AD and normal brain samples that both up- and down-regulated DE genes were highly cell-type specific, with most genes perturbed in neuron or a specific glia cell type (Mathys et al., 2019). While previous analysis of AD bulk transcriptomic data may have overlooked the effect of cell type or cell population change, our study showed that cell type relative abundance changes actually play a major role in microglia and neuron gene differential expression in AD patient brains.

As for the other three cell types, our finding showed that oligodendrocyte cell is not ubiquitously changed across five datasets, which is a novel discovery. However, we detected significant cell abundance changes for astrocytes and endothelia cells in AD samples as well, which suggests that differential gene expression analysis on bulk tissue transcriptomic data also need to adjust for cell population change when genes are associated to those cell types. Such adjustment will help researchers pinpoint the *bona fide* mRNA regulatory changes associated with disease initiation and progression.

Due to the rarity of human brain samples and the difficulty in isolate specific cell types from the brain tissue, single cell transcriptomic analysis is still scarce in this field and most transcriptomic analysis are still using bulk tissue brain samples. With comparative analysis being the major type of transcriptomic analysis, the influence of signal changes resulted from cell composition change cannot be overestimated and our findings address the need for a clarification for this issue.

Conclusion

Through FGCN module mining, we identified cell type specific modules that are differentially expressed and/or differentially co-expressed in five large AD cohort studies. Applying further computational analyses on these gene modules, we demonstrated that the overall increased expression of the core microglia module can be well explained by the increased microglia cell population observed in AD brains instead of *bona fide* upregulation of microglia gene expression. We were also able to untangle the overall decreased expression in the neuron modules of AD to identify both the neuron cell loss and expression regulation of neuron associated pathways, both of which are a novel discoveries. The overall increased co-expression level of the neuron module may result from the neuron hyperactivity reported previously. Further network hub gene investigation confirmed the co-expression changes, and we identified several differentially expressed transcription factors (*BCL6* and *STAT3*) in multiple studies of AD brains that are potential key regulators for such changes. Furthermore, although not a focus in this study, we observed significant changes in the expression and co-expression levels for the two astrocyte modules, which may also result from a combinatory effect from cell population and the astrocyte gene activation in AD brains. More importantly, these two modules' expression highly correlate with the three universally used AD clinical markers, which indicates that astrocytes genes can serve as good biomarkers for disease progression. This work also demonstrated that the combination of DE, DC and cell type deconvolution analyses not only provides a powerful approach to delineate the origin of the transcriptomic change in bulk sample data, but also leads to a deeper understanding of the roles of the genes in the disease progress.

Abbreviations:

AD: Alzheimer's disease

$A\beta$: beta-amyloid protein

BB: Braak & Braak score

CCI: centered concordance index

CDR: clinical dementia rating

DC: differential co-expression

DE: differential expression

FGCN: Frequent gene co-expression network

GCN: gene co-expression network

lmQCM: local maximal quasi-clique merger

PCA: principal component analysis

PC1/2: principal component 1/2

PCC: Pearson correlation coefficient

SCC: Spearman correlation coefficient

TF: transcription factor

Declarations

Ethics approval and consent to participate

The work composed of only publicly available data, therefore no ethnic approval or consent is needed.

Consent for publication

All the authors consent for publication of this work.

Availability of data and materials

All major analysis results are attached in submission and available for public. Further request can be fulfilled by contacting corresponding authors.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JZ, DN and KH designed the study. SX collected, prepared the data and did analysis, drafted the manuscript. TJ contributed to the estimation of brain cell type relative proportions and differential gene expression analysis. MC contributed to the gene functional enrichment analysis. ZH contributed to figure generation, TD contributed to the ChIP-seq and analysis and literature search for module regulators, JZ, KH, DN, TW, ZH, TJ, TD revised the manuscript for all important intellectual content. All authors discussed the results and contributed to the final manuscript. All authors read and approved the final manuscript.

Reference

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.

Anand, K.S., and Dhikav, V. (2012). Hippocampus in health and disease: An overview. Ann Indian Acad Neurol *15*, 239–246.

Andrade-Moraes, C.H., Oliveira-Pinto, A.V., Castro-Fonseca, E., da Silva, C.G., Guimarães, D.M., Szczupak, D., Parente-Bruno, D.R., Carvalho, L.R.B., Polichiso, L., Gomes, B.V., et al. (2013). Cell number changes in Alzheimer's disease relate to dementia, not to plaques and tangles. Brain *136*, 3738–3752.

Berchtold, N.C., Cribbs, D.H., Coleman, P.D., Rogers, J., Head, E., Kim, R., Beach, T., Miller, C., Troncoso, J., Trojanowski, J.Q., et al. (2008). Gene expression changes in the course of normal brain aging are sexually dimorphic. Proc. Natl. Acad. Sci. U.S.A. *105*, 15605–15610.

Bossers, K., Wirz, K.T.S., Meerhoff, G.F., Essing, A.H.W., van Dongen, J.W., Houba, P., Kruse, C.G., Verhaagen, J., and Swaab, D.F. (2010). Concerted changes in transcripts in the prefrontal cortex precede neuropathology in Alzheimer's disease. Brain *133*, 3699–3723.

Branca C., Shaw D.M., Belfiore R., Gokhale V., Shaw A.Y., Foley C., Smith B., Hulme C., Dunckley T., Meechoovet B., Caccamo A., and Oddo S. (2017). Aging Cell. *16*(5):1146-1154.

Dyrk1 inhibition improves Alzheimer's disease-like pathology.

Busche, M.A., Eichhoff, G., Adelsberger, H., Abramowski, D., Wiederhold, K.-H., Haass, C., Staufenbiel, M., Konnerth, A., and Garaschuk, O. (2008). Clusters of hyperactive neurons near amyloid plaques in a mouse model of Alzheimer's disease. Science *321*, 1686–1689.

Busche, M.A., Chen, X., Henning, H.A., Reichwald, J., Staufenbiel, M., Sakmann, B., and Konnerth, A. (2012). Critical role of soluble amyloid-β for early hippocampal hyperactivity in a mouse model of Alzheimer's disease. Proc. Natl. Acad. Sci. U.S.A. *109*, 8740–8745.

Cai, Z., Qiao, P.-F., Wan, C.-Q., Cai, M., Zhou, N.-K., and Li, Q. (2018). Role of Blood-Brain Barrier in Alzheimer's Disease. J. Alzheimers Dis. *63*, 1223–1234.

Celarain, N., Sánchez-Ruiz de Gordoa, J., Zelaya, M.V., Roldán, M., Larumbe, R., Pulido, L., Echavarri, C., and Mendioroz, M. (2016). TREM2 upregulation correlates with 5-hydroxymethycytosine enrichment in Alzheimer's disease hippocampus. Clin Epigenetics *8*, 37.

Chen, J., Bardes, E.E., Aronow, B.J., and Jegga, A.G. (2009). ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. *37*, W305-311.

Chikina, M., Zaslavsky, E., and Sealfon, S.C. (2015). CellCODE: a robust latent variable approach to differential expression analysis for heterogeneous cell populations. Bioinformatics *31*, 1584–1591.

Clayton, K.A., Van Enoo, A.A., and Ikezu, T. (2017). Alzheimer's Disease: The Role of Microglia in Brain Homeostasis and Proteopathy. Front Neurosci *11*.

Cotman, C.W., and Su, J.H. (1996). Mechanisms of neuronal death in Alzheimer's disease. Brain Pathol. *6*, 493–506.

Darweesh, S. K. L., F. J. Wolters, M. A. Ikram, F. de Wolf, D. Bos and A. Hofman (2018). Inflammatory markers and the risk of dementia and Alzheimer's disease: A meta-analysis. Alzheimers Dement *14*(11): 1450-1459.

De Jager, P.L., Ma, Y., McCabe, C., Xu, J., Vardarajan, B.N., Felsky, D., Klein, H.-U., White, C.C., Peters, M.A., Lodgson, B., et al. (2018). A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. Sci Data *5*, 180142.

Devanand, D.P. (2018). Viral Hypothesis and Antiviral Treatment in Alzheimer's Disease. Curr Neurol Neurosci Rep *18*, 55.

Dickstein, D.L., Walsh, J., Brautigam, H., Stockton, S.D., Gandy, S., and Hof, P.R. (2010). Role of vascular risk factors and vascular dysfunction in Alzheimer's disease. Mt. Sinai J. Med. *77*, 82–102.

Du Y., Du Y., Zhang Y., Huang Z., Fu M., Li J., Pang Y., Lei P., Wang Y.T., Song W., He G., and Dong Z. (2019). MKP-1 reduces Aβ generation and alleviates cognitive impairments in Alzheimer's disease models. Signal Transduct Target Ther. *4*:58.

Fakhoury, M. (2018). Microglia and Astrocytes in Alzheimer's Disease: Implications for Therapy. Curr Neuropharmacol *16*, 508–518.

Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics *20*, 307–315.

Gomez-Nicola, D., and Perry, V.H. (2016). Analysis of Microglial Proliferation in Alzheimer's Disease. Methods Mol. Biol. *1303*, 185–193.

Grammas, P. (2011). Neurovascular dysfunction, inflammation and endothelial activation: Implications for the pathogenesis of Alzheimer's disease. J Neuroinflammation *8*, 26.

Fu A.K., Hung K.W., Huang H., Gu S., Shen Y., Cheng E.Y., Ip F.C., Huang X., Fu W.Y., Ip N.Y. (2014). Blockade of EphA4 signaling ameliorates hippocampal synaptic dysfunctions in mouse models of Alzheimer's disease. Proc Natl Acad Sci U S A. *111*(27):9959-64.

Han, Z., Zhang, J., Sun, G., Liu, G., and Huang, K. (2016). A matrix rank based concordance index for evaluating and detecting conditional specific co-expressed gene modules. BMC Genomics *17*, 519.

Hansen, D.V., Hanson, J.E., and Sheng, M. (2018). Microglia in Alzheimer's disease. J. Cell Biol. *217*, 459–472.

Haure-Mirande, J.-V., Audrain, M., Fanutza, T., Kim, S.H., Klein, W.L., Glabe, C., Readhead, B., Dudley, J.T., Blitzer, R.D., Wang, M., et al. (2017). Deficiency of TYROBP, an adapter protein for TREM2 and CR3 receptors, is neuroprotective in a mouse model of early Alzheimer's pathology. Acta Neuropathol. *134*, 769–788.

Huang, Y., and Mucke, L. (2012). Alzheimer mechanisms and therapeutic strategies. Cell *148*, 1204–1222.

Huang, Z., Han, Z., Wang, T., Shao, W., Xiang, S., Salama, P., Rizkalla, M., Huang, K., and Zhang, J. (2019). TSUNAMI: Translational Bioinformatics Tool Suite For Network Analysis And Mining. BioRxiv 787507.

Humphries, C.E., Kohli, M.A., Nathanson, L., Whitehead, P., Beecham, G., Martin, E., Mash, D.C., Pericak-Vance, M.A., and Gilbert, J. (2015). Integrated whole transcriptome and DNA methylation analysis identifies gene networks specific to late-onset Alzheimer's disease. J. Alzheimers Dis. *44*, 977–987.

Jain, P., Wadhwa, P.K., and Jadhav, H.R. (2015). Reactive Astrogliosis: Role in Alzheimer's Disease. CNS Neurol Disord Drug Targets *14*, 872–879.

Katsumoto, A., Takeuchi, H., Takahashi, K., and Tanaka, F. (2018). Microglia in Alzheimer's Disease: Risk Factors and Inflammation. Front Neurol *9*, 978.

Kelleher, R.J., and Soiza, R.L. (2013a). Evidence of endothelial dysfunction in the development of Alzheimer's disease: Is Alzheimer's a vascular disorder? Am J Cardiovasc Dis *3*, 197–226.

Kelleher, R.J., and Soiza, R.L. (2013b). Evidence of endothelial dysfunction in the development of Alzheimer's disease: Is Alzheimer's a vascular disorder? Am J Cardiovasc Dis *3*, 197–226.

Keren-Shaul, H., Spinrad, A., Weiner, A., Matcovitch-Natan, O., Dvir-Szternfeld, R., Ulland, T.K., David, E., Baruch, K., Lara-Astaiso, D., Toth, B., et al. (2017). A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. Cell *169*, 1276-1290.e17.

Knock E., Matsuzaki S., Takamura H., Satoh K., Rooke G., Han K., Zhang H., Staniszewski A., Katayama T., Arancio O., and Fraser P.E. (2018). SUMO1 impact on Alzheimer disease pathology in an amyloid-depositing mouse model. Neurobiol Dis. *110*:154-165.

Koizumi, K., Wang, G., and Park, L. (2016). Endothelial Dysfunction and Amyloid-β-Induced Neurovascular Alterations. Cell. Mol. Neurobiol. *36*, 155–165.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res *44*, W90–W97.

Kusam S., Vasanwala F.H., and Dent A.L. (2004). Transcriptional repressor BCL-6 immortalizes germinal center-like B cells in the absence of p53 function. Oncogene. *23*(3):839-44.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Lau, P., Bossers, K., Janky, R., Salta, E., Frigerio, C.S., Barbash, S., Rothman, R., Sierksma, A.S.R., Thathiah, A., Greenberg, D., et al. (2013). Alteration of the microRNA network during the progression of Alzheimer's disease. EMBO Mol Med *5*, 1613–1634.

Layfield, R., Cavey, J.R., and Lowe, J. (2003). Role of ubiquitin-mediated proteolysis in the pathogenesis of neurodegenerative disorders. Ageing Res. Rev. *2*, 343–356.

Lerdkrai, C., Asavapanumas, N., Brawek, B., Kovalchuk, Y., Mojtahedi, N., Olmedillas Del Moral, M., and Garaschuk, O. (2018). Intracellular Ca2+ stores control in vivo neuronal hyperactivity in a mouse model of Alzheimer's disease. Proc. Natl. Acad. Sci. U.S.A. *115*, E1279–E1288.

Li J.G., Chiu J., Praticò D. (2019). Full recovery of the Alzheimer's disease phenotype by gain of function of vacuolar protein sorting 35. Mol Psychiatry. doi: 10.1038/s41380-019-0364-x. [Epub ahead of print]

Li, X., Long, J., He, T., Belshaw, R., and Scott, J. (2015). Integrated genomic approaches identify major pathways and upstream regulators in late onset Alzheimer's disease. Sci Rep *5*, 12393.

Liang, W.S., Dunckley, T., Beach, T.G., Grover, A., Mastroeni, D., Walker, D.G., Caselli, R.J., Kukull, W.A., McKeel, D., Morris, J.C., et al. (2007). Gene expression profiles in anatomically and functionally distinct regions of the normal aged human brain. Physiol. Genomics *28*, 311–322.

Llorens-Martín M., Jurado J., Hernández F., and Avila J. (2014). GSK-3β, a pivotal kinase in Alzheimer disease. Front Mol Neurosci. 7:46.

Lucin, K. M. and T. Wyss-Coray (2009). Immune activation in brain aging and neurodegeneration: too much or too little? Neuron *64*(1): 110-122.

Lui, T.W.H., Tsui, N.B.Y., Chan, L.W.C., Wong, C.S.C., Siu, P.M.F., and Yung, B.Y.M. (2015). DECODE: an integrated differential co-expression and differential expression analysis of gene expression data. BMC Bioinformatics *16*, 182.

Ma, J., Jiang, T., Tan, L., and Yu, J.-T. (2015). TYROBP in Alzheimer's disease. Mol. Neurobiol. *51*, 820–826.

Magalingam, K.B., Radhakrishnan, A., Ping, N.S., and Haleagrahara, N. (2018). Current Concepts of Neurodegenerative Mechanisms in Alzheimer's Disease. Biomed Res Int *2018*, 3740461.

Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J.Z., Menon, M., He, L., Abdurrob, F., Jiang, X., et al. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. Nature *570*, 332–337.

McKenzie, A.T., Wang, M., Hauberg, M.E., Fullard, J.F., Kozlenkov, A., Keenan, A., Hurd, Y.L., Dracheva, S., Casaccia, P., Roussos, P., et al. (2018). Brain Cell Type Specific Gene Expression and Co-expression Network Architectures. Sci Rep *8*, 8868.

Miller, J.A., Oldham, M.C., and Geschwind, D.H. (2008). A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. J. Neurosci. *28*, 1410–1420.

Miller, J.A., Woltjer, R.L., Goodenbour, J.M., Horvath, S., and Geschwind, D.H. (2013). Genes and pathways underlying regional and cell type changes in Alzheimer's disease. Genome Med *5*, 48.

Montagne, A., Zhao, Z., and Zlokovic, B.V. (2017). Alzheimer's disease: A matter of blood-brain barrier dysfunction? J. Exp. Med. *214*, 3151–3169.

K. Morimoto, J. Horio, H. Satoh, L. Sue, T. Beach, S.Arita, I. Tooyama, and Y. Konishi. (2011). Expression profiles of cytokines in the brains of Alzheimer's disease (AD) patients, compared to the brains of non-demented patients with and without increasing AD pathology. J Alzheimers Dis. *25*(1): 59–76.

Nakase, T., and Naus, C.C.G. (2004). Gap junctions and neurological disorders of the central nervous system. Biochim. Biophys. Acta *1662*, 149–158.

Narayanan M., Huynh J.L., Wang K., Yang X., et al. (2014). Common dysregulation network in the human prefrontal cortex underlies two neurodegenerative diseases. Mol Syst Biol. *10*:743.

Nuriel, T., Angulo, S.L., Khan, U., Ashok, A., Chen, Q., Figueroa, H.Y., Emrani, S., Liu, L., Herman, M., Barrett, G., et al. (2017). Neuronal hyperactivity due to loss of inhibitory tone in APOE4 mice lacking Alzheimer's disease-like pathology. Nat Commun *8*, 1464.

Patir, A., Shih, B., McColl, B.W., and Freeman, T.C. (2019). A core transcriptional signature of human microglia: Derivation and utility in describing region-dependent alterations associated with Alzheimer's disease. Glia *67*, 1240–1253.

Peng C., Hu Q., Yang F., Zhang H., Li F., and Huang C. (2019). BCL6-Mediated Silencing of PD-1 Ligands in Germinal Center B Cells Maintains Follicular T Cell Population. J Immunol. *202*(3):704-713.

Pomara, N., and Bruno, D. (2018). Pathological Increases in Neuronal Hyperactivity in Selective Cholinergic and Noradrenergic Pathways May Limit the Efficacy of Amyloid-β-Based Interventions in Mild Cognitive Impairment and Alzheimer's Disease. J Alzheimers Dis Rep *2*, 165–167.

Rangaraju S., Dammer E.B., Raza S.A., Rathakrishnan P., Xiao H., Gao T., Duong D.M., Pennington M.W., Lah J.J., Seyfried N.T., and Levey A.I. (2018). Identification and therapeutic modulation of a pro-inflammatory subset of disease-associated-microglia in Alzheimer's disease. Mol Neurodegener. *13*(1):24.

Readhead, B., Haure-Mirande, J.-V., Funk, C.C., Richards, M.A., Shannon, P., Haroutunian, V., Sano, M., Liang, W.S., Beckmann, N.D., Price, N.D., et al. (2018). Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus. Neuron *99*, 64-82.e7.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Robinson, M., Lee, B.Y., and Hane, F.T. (2017). Recent Progress in Alzheimer's Disease Research, Part 2: Genetics and Epidemiology. J. Alzheimers Dis. *57*, 317–330.

Sekar, S., McDonald, J., Cuyugan, L., Aldrich, J., Kurdoglu, A., Adkins, J., Serrano, G., Beach, T.G., Craig, D.W., Valla, J., et al. (2015). Alzheimer's disease is associated with altered expression of genes involved in immune response and mitochondrial processes in astrocytes. Neurobiol. Aging *36*, 583–591.

Seyfried, N.T., Dammer, E.B., Swarup, V., Nandakumar, D., Duong, D.M., Yin, L., Deng, Q., Nguyen, T., Hales, C.M., Wingo, T., et al. (2017). A Multi-network Approach Identifies Protein-Specific Co-expression in Asymptomatic and Symptomatic Alzheimer's Disease. Cell Syst *4*, 60-72.e4.

Shahani N., Pryor W., Swarnkar S., Kholodilov N., Thinakaran G., Burke R.E., and Subramaniam S. (2014). Rheb GTPase regulates β-secretase levels and amyloid β generation. J Biol Chem. *289*(9):5799-808.

Solito, E., and Sastre, M. (2012). Microglia Function in Alzheimer's Disease. Front Pharmacol *3*.

Srinivasan, K., Friedman, B.A., Larson, J.L., Lauffer, B.E., Goldstein, L.D., Appling, L.L., Borneo, J., Poon, C., Ho, T., Cai, F., et al. (2016). Untangling the brain's neuroinflammatory and neurodegenerative transcriptional responses. Nat Commun 7, 11295.

Stargardt, A., Swaab, D.F., and Bossers, K. (2015). Storm before the quiet: neuronal hyperactivity and Aβ in the presymptomatic stages of Alzheimer's disease. Neurobiol. Aging *36*, 1–11.

Sun, J., and Zhao, Z. (2010). A comparative study of cancer proteins in the human protein-protein interaction network. BMC Genomics *11*, S5.

Tansey, K. E., D. Cameron and M. J. Hill (2018). Genetic risk for Alzheimer's disease is concentrated in specific macrophage and microglial transcriptional networks. Genome Medicine 10(1).

Tripathi, S., Christie, K.R., Balakrishnan, R., Huntley, R., Hill, D.P., Thommesen, L., Blake, J.A., Kuiper, M., and Lægreid, A. (2013). Gene Ontology annotation of sequence-specific DNA binding transcription factors: setting the stage for a large-scale curation effort. Database (Oxford) *2013*, bat062.

Uylings, H.B.M., and de Brabander, J.M. (2002). Neuronal changes in normal human aging and Alzheimer's disease. Brain Cogn *49*, 268–276.

Verheijen, J., and Sleegers, K. (2018). Understanding Alzheimer Disease at the Interface between Genetics and Transcriptomics. Trends Genet. *34*, 434–447.

Wang M., Roussos P., McKenzie A., Zhou X., Kajiwara Y., Brennand K.J., De Luca G.C., Crary J.F., Casaccia P., Buxbaum J.D., Ehrlich M., Gandy S., Goate A., Katsel P., Schadt E., Haroutunian V., and Zhang B. (2016). Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease. Genome Med. *8*(1):104.

Wang, M., Beckmann, N.D., Roussos, P., Wang, E., Zhou, X., Wang, Q., Ming, C., Neff, R., Ma, W., Fullard, J.F., et al. (2018). The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. Scientific Data *5*, 180185.

Webster J.A., Gibbs J.R., Clarke J., Ray M., Zhang W., Holmans P., Rohrer K., Zhao A., Marlowe L., Kaleem M., McCorquodale D.S., Cuello C., Leung D., Bryden L., et al. (2009) Genetic control of human brain transcript expression in Alzheimer disease., Am J Hum Genet. *84*(4):445-58.

Won S.Y., Park J.J., Shin E.Y., Kim G. (2019). PAK4 signaling in health and disease: defining the PAK4-CREB axis. Exp Mol Med. *51*(2):1-9.

Xiang, S., Huang, Z., Wang, T., Han, Z., Yu, C.Y., Ni, D., Huang, K., and Zhang, J. (2018). Condition-specific gene co-expression network mining identifies key pathways and regulators in the brain tissue of Alzheimer's disease patients. BMC Med Genomics *11*, 115.

Yan J., Enge M., Whitington T., Dave K., Liu J., Sur I., Schmierer B., Jolma A., Kivioja T., Taipale M., and Taipale J. (2013). Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. Cell. *154*(4):801-13.

Yu, H., Greenbaum, D., Xin Lu, H., Zhu, X., and Gerstein, M. (2004). Genomic analysis of essentiality within protein networks. Trends Genet. *20*, 227–231.

Yu, H., Kim, P.M., Sprecher, E., Trifonov, V., and Gerstein, M. (2007). The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comput. Biol. *3*, e59.

Zhang, J., and Huang, K. (2014). Normalized lmQCM: An Algorithm for Detecting Weak Quasi-Cliques in Weighted Graph with Applications in Gene Co-Expression Module Discovery in Cancers. Cancer Inform *13*, 137–146.

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

Zhang, J., Xiang, Y., Ding, L., Keen-Circle, K., Borlawsky, T.B., Ozer, H.G., Jin, R., Payne, P., and Huang, K. (2010). Using gene co-expression network analysis to predict biomarkers for chronic lymphocytic leukemia. BMC Bioinformatics *11 Suppl 9*, S5.

Zhang, J., Lu, K., Xiang, Y., Islam, M., Kotian, S., Kais, Z., Lee, C., Arora, M., Liu, H.-W., Parvin, J.D., et al. (2012). Weighted frequent gene co-expression network mining to identify genes involved in genome stability. PLoS Comput. Biol. *8*, e1002656.

Zhou K.R., Liu S., Sun W.J., Zheng L.L., Zhou H., Yang J.H., and Qu L.H. (2017). ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. Nucleic Acids Res. *45*(D1):D43-D50.

Zott, B., Simon, M.M., Hong, W., Unger, F., Chen-Engerer, H.-J., Frosch, M.P., Sakmann, B., Walsh, D.M., and Konnerth, A. (2019). A vicious cycle of β amyloid-dependent neuronal hyperactivation. Science *365*, 559–565.

Figure legends

Figure 1. **FGCN modules from AD and normal brains are classified according to DE and DC levels**. a. Differential expression score (DE) vs. differential co-expression score (DC) of 24 frequent gene co-expression modules (module names are in the parenthesis), majority of which are enriched with specific biological functions as indicated by matching colors of dots and the text on the side of the plot. Multiple colors of a module name indicate multiple functions. Most of the modules are also enriched for specific cell type markers: Neu: Neuron module; Ast: astrocyte module; Mic: microglia module; Endo: endothelia module; Oligo: oligodendrocyte module; unknown: modules not enriched with cell type-specific markers. **b**. Cell-type enrichment for each module was assessed by cross-referencing module genes against lists of cell type signatures of neurons, microglia, astrocytes, oligodendrocytes and endothelial. X-axis is labeled with module numbers. 24 FGCN modules were numbered as: 1-15 correspond to AD1 to AD15; 16-24 correspond to N1 to N9. The significance of the cell type signature enrichment in the modules is measured by Fisher's exact test and corrected for multiple comparisons by B&H FDR; horizonal blue lines indicate the threshold of FDR 0.05. There is no color match between panels a and b.

Figure 2. Cell type deconvolution analysis indicate significantly decreased neurons and increased glia cells in AD brains. The significance of cell type population changes between AD and normal is measured by Wilcoxon rank sum test and indicated by * if p value less than 0.05, and ** if p value less than 0.01. a. Estimated neuron relative proportions in AD and normal samples. b. Estimated microglia relative proportions in AD and normal samples. c. Estimated astrocytes relative proportions in AD and normal samples. d. Estimated endothelial relative proportions in AD and normal samples. e. Estimated oligodendrocytes relative proportions in AD and normal samples.

Figure 3. The highly differential expression in HDE_LDC modules were largely due to cell population change. a. Network visualization of the microglia module from the FGCN networks in AD and normal conditions (for legibility, only gene links that appears at least in four out of the five studies are shown). The genes with blue color in normal and red color in AD networks were up

expressed in AD in at least two of the five datasets. b. CCI of the microglia module in AD vs normal samples in each of the five datasets. c. The Pearson correlation of microglia module eigengenes with the estimated microglia cell relative proportion in GSE48350 dataset, d. The Pearson correlation of microglia module eigengenes with the estimated neuron cell relative proportions. e. The DE scores of the microglia module with mean microglia relative proportion changes between AD and normal samples across five datasets. The gray area indicates the region defined by 95% confidence intervals.

Figure 4. Neuron modules showed high DE that may be due to decreased proportion of neurons and high DC that may indicate regulatory change. a. Function enrichment of AD1/N1. b. CCI of neuron modules (AD1/N1) in AD vs. Normal samples in each of the five datasets. c. In GSE48350 dataset, the Pearson correlation of neuron module AD1 expression (eigengene values) with estimated neuron relative proportions. d. The Pearson correlation of AD1 eigengene values with estimated microglia relative proportions in GSE48350. e. In GSE48350 dataset, the Pearson correlation of neuron module N1 eigengene values with neuron relative proportions. f, The Pearson correlation of N1 eigengene values with microglia relative proportions in GSE48350.

Figure 5. HDE modules are associated with AD neuropathology and cognitive status. a. Correlation of FGCN module eigengene values with three clinical traits. Spearman correlation coefficient (SCC) and p value between module eigengene expression and clinical dementia rating (CDR), neurofibrillary tangle burden (Braak & Braak score), beta-amyloid burden (PlaqueMean) are displayed respectively. The modules are ranked by the mean scc value and marked by corresponding cell types. ast: astrocytes, mic: microglia, neu: neuron. b. Correlation of brain cell type relative abundance with clinical traits. scc and p value between estimated brain cell type relative proportions and each of CDR, Braak & Braak score, PlaqueMean are displayed respectively. The brain cell types are ranked by the mean scc value. ast: astrocytes, mic: microglia, neu: neuron, end: endothelia, oli: oligodendrocytes.

Tables

Table 1 Summary of AD and normal brain transcriptomic datasets used in this study.

Dataset	Platform	AD subject	Normal subject	Brain region
GSE5281 (Liang et al., 2007)	Microarray, Affymetrix HU133	87	74	Middle Temporal Gyrus (MTG), Posterior Cingulate Cortex (PCC), Entorhinal Cortex (EC), Hippocampus (HC), Superior Frontal Cortex (SFG), Primary Visual Cortex (VCX)
GSE48350 (Berchtold et al., 2008)	Microarray, Affymetrix HU133	80	173	Entorhinal Cortex (EC), Post-central Gyrus (PCG2), Superior Frontal Gyrus (SFG), Hippocampus (HC)
ROSMAP (De Jager et al., 2018)	RNA-seq	224	201	Dorsolateral prefrontal cortex (DLPFC)
MSBB (Wang et al., 2018)	RNA-seq	409	273	Frontal pole (BM10), Superior temporal gyrus (BM22), Parahippocampal gyrus (BM36), Inferior frontal gyrus (BM44)
Mayo (Allen et al., 2016)	RNA-seq	82	78	Temporal cortex (TCX)
Validation Datasets				
GSE33000 (Narayanan et al., 2014)	Rosetta/Merck Human 44k 1.1 microarray	310	157	Dorsolateral prefrontal cortex (DLPFC)
GSE15222 (Webster et al., 2009)	Sentrix HumanRef-8 Expression BeadChip	176	187	Whole brain with majority of cortex
GSE84422 (Wang et al., 2016)	Microarray, Affymetrix HU133 & HU133_Plus_2	60	65	19 regions
Allen Brain Dementia data (Allen Brain Atlas)	RNA-seq	43	60	HC, temporal cortex (TCX), parietal cortex, forebrain, white matter

Table 2 Summary of the characteristics of all FGCN modules from AD and normal samples.

Module category & name	Module size	Up-reg. genes in AD vs. Normal	Down-reg. genes in AD vs. Normal	Enriched cell type
HDE_HDC				
AD1	1,247	29	593	neuron
AD2	197	159	0	astrocyte
AD7	21	17	0	oligodendrocyte
AD11	14	12	0	-
N1	1,003	10	449	neuron
N2	66	1	31	neuron
N5	25	23	0	astrocyte
N8	13	0	9	-
HDE_LDC				
N6	21	19	0	microglia
AD3	38	33	0	microglia
AD14	10	9	0	endothelia
N9	12	2	0	-
LDE_HDC				
AD5	29	8	0	oligodendrocyte
AD8	21	21	0	endothelia
AD9	21	0	0	-
AD12	11	9	0	oligodendrocyte
LDE_LDC				
AD4	32	0	0	oligodendrocyte
AD6	29	20	0	oligodendrocyte
AD10	17	0	0	-
AD13	10	9	0	endothelia
AD15	10	0	0	-
N3	47	26	0	oligodendrocyte
N4	35	25	0	oligodendrocyte
N7	19	0	0	-

Table 3 Neuron modules AD1/N1 are predicted to be targeted by differentially expressed TFs.

Module name	Enriched TF	Differential expression in AD vs. Normal	Number of predicted target genes in the module	Fisher's exact test adjusted p-value
AD1	BCL6	Up in 7/9 datasets	119/1247	0.0095
	JUND	Up in 4/9 datasets	116/1247	0.0148
	ZBTB16	Up in 5/9 datasets	128/1247	0.0200
	STAT3	Up in 7/9 datasets	204/1247	0.0449
	MYB	Down in 4/9 datasets	135/1247	8.87E-05
N1	ZBTB16	Up in 5/9 datasets	112/1003	0.0058
	BCL6	Up in 7/9 datasets	97/1003	0.022
	LEF1	Up in 6/9 datasets	273/1003	0.0337
	STAT3	Up in 7/9 datasets	160/1003	0.0432
	JUND	Up in 4/9 datasets	94/1003	0.0359
	MYB	Down in 4/9 datasets	113/1003	6.271e-05

Supplementary Information Excel Table Titles

Table 1. FGCN module genes from AD/normal brain samples and the overlap (Jaccard index) between each pair of AD and normal FGCN modules. (sup-Table1.xlsx)

Table 2. DE and DC score summary for all AD/Normal FGCN modules. (sup-Table2.xlsx)

Table 3. Top 10 enriched GO/pathway terms of each AD/normal FGCN module. (sup-Table3.xlsx)

Table 4. Frequently down expressed genes (in at least two datasets) between AD and normal samples, merged for all five datasets. (sup-Table4-5.xlsx)

Table 5. Frequently up expressed genes (in at least two datasets) between AD and normal samples, merged for all five datasets. (sup-Table4-5.xlsx)

Table 6. Enrichment of gene markers in five major cell types in all AD/normal FGCN modules. (sup-Table6.xlsx)

Table 7. Enriched pathway/GO terms for overlapping genes of AD1 and N1. (sup-Table7-9.xlsx)

Table 8. Enriched pathway/GO terms in genes uniquely to AD1. (sup-Table7-9.xlsx)

Table 9. Enriched pathway/GO terms in genes uniquely to N1. (sup-Table7-9.xlsx)

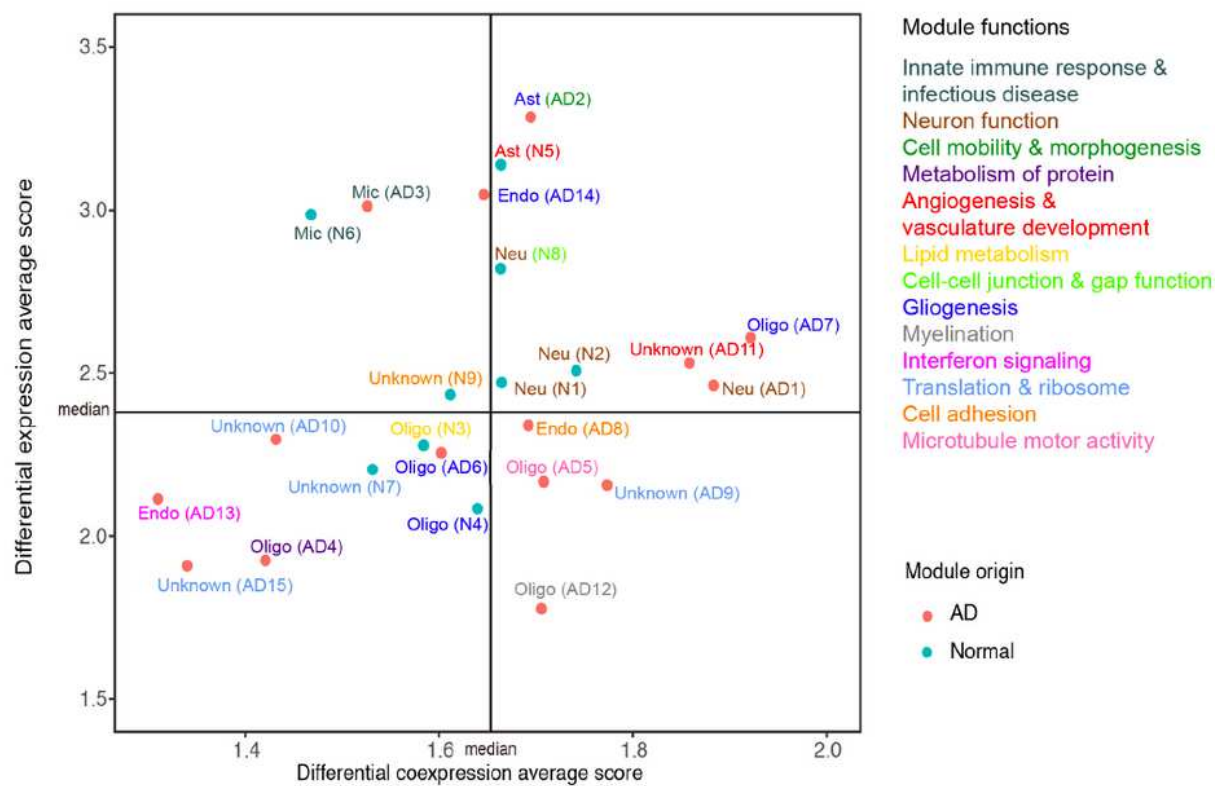Table 10. Enriched transcription factors and their targets for each FGCN modules. (sup-Table10.xlsx)

Table 11. Correlation of FGCN module eigengene with clinicopathological measurements in MSBB dataset. (sup-Table11.xlsx)

Table 12. Genes with BCL6/STAT3 ChIP-seq peaks in their promoter regions. (sup-Table12.xlsx)

Table 13. The overlap of AD3/N6 microglia module with two previously identified disease associated microglia modules (sup-Table13.xlsx).
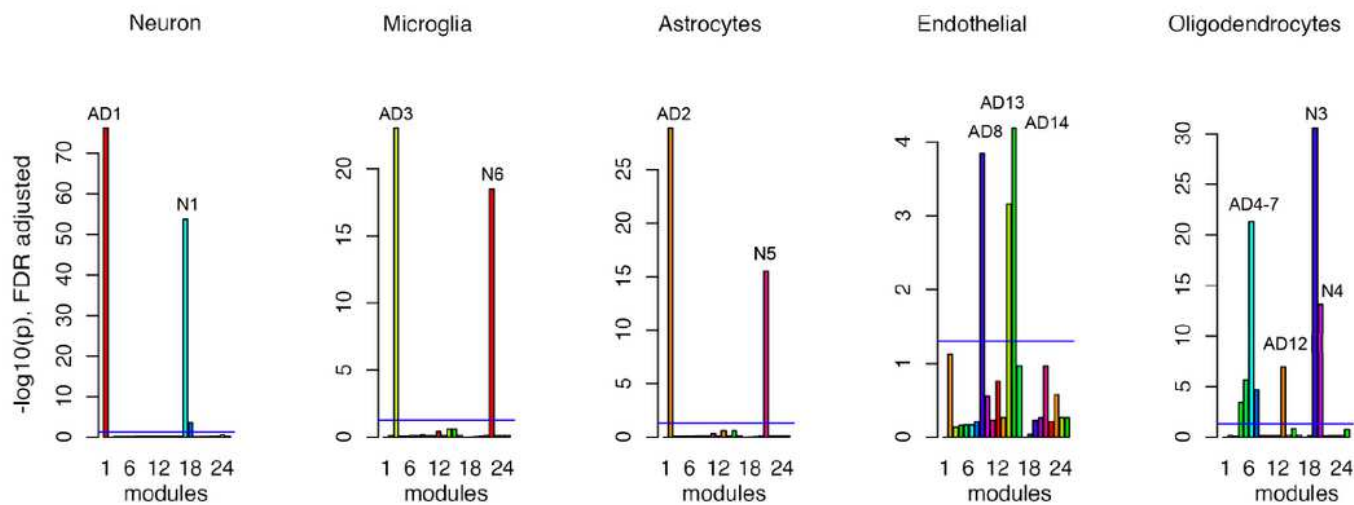
Figures

Figure 1

FGCN modules from AD and normal brains are classified according to DE and DC levels. a. Differential expression score (DE) vs. differential co-expression score (DC) of 24 frequent gene co-expression modules (module names are in the parenthesis), majority of which are enriched with specific biological

functions as indicated by matching colors of dots and the text on the side of the plot. Multiple colors of a module name indicate multiple functions. Most of the modules are also enriched for specific cell type markers: Neu: Neuron module; Ast: astrocyte module; Mic: microglia module; Endo: endothelia module; Oligo: oligodendrocyte module; unknown: modules not enriched with cell type-specific markers. b. Cell-type enrichment for each module was assessed by cross-referencing module genes against lists of cell type signatures of neurons, microglia, astrocytes, oligodendrocytes and endothelial. X-axis is labeled with module numbers. 24 FGCN modules were numbered as: 1-15 correspond to AD1 to AD15; 16-24 correspond to N1 to N9. The significance of the cell type signature enrichment in the modules is measured by Fisher's exact test and corrected for multiple comparisons by B&H FDR; horizonal blue lines indicate the threshold of FDR 0.05. There is no color match between panels a and b.
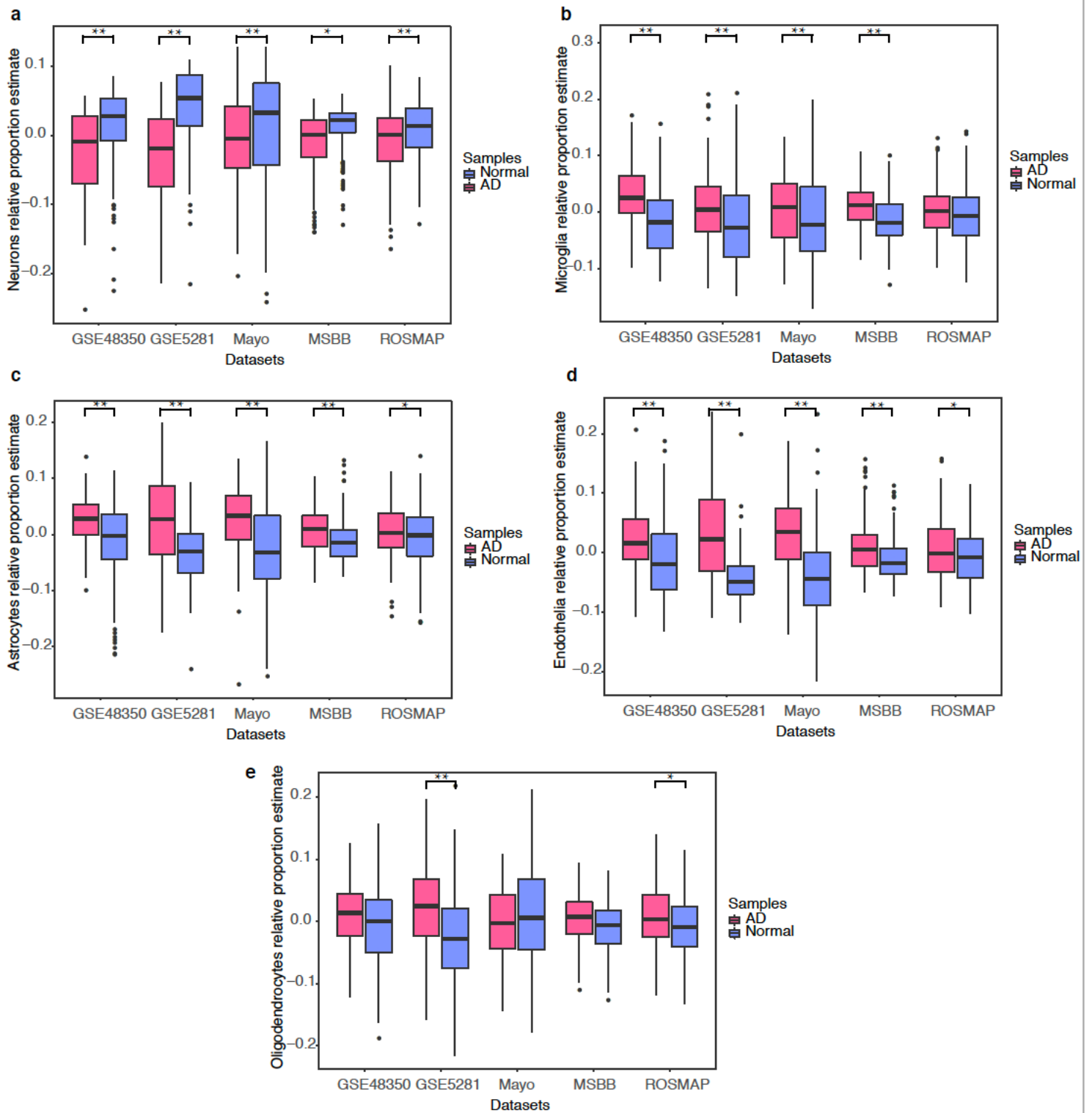
Figure 2

Cell type deconvolution analysis indicate significantly decreased neurons and increased glia cells in AD brains. The significance of cell type population changes between AD and normal is measured by Wilcoxon rank sum test and indicated by * if p value less than 0.05, and ** if p value less than 0.01. a. Estimated neuron relative proportions in AD and normal samples. b. Estimated microglia relative proportions in AD and normal samples. c. Estimated astrocytes relative proportions in AD and normal

samples. d. Estimated endothelial relative proportions in AD and normal samples. e. Estimated oligodendrocytes relative proportions in AD and normal samples.

frequent GCN network structure of the microglia module in normal samples

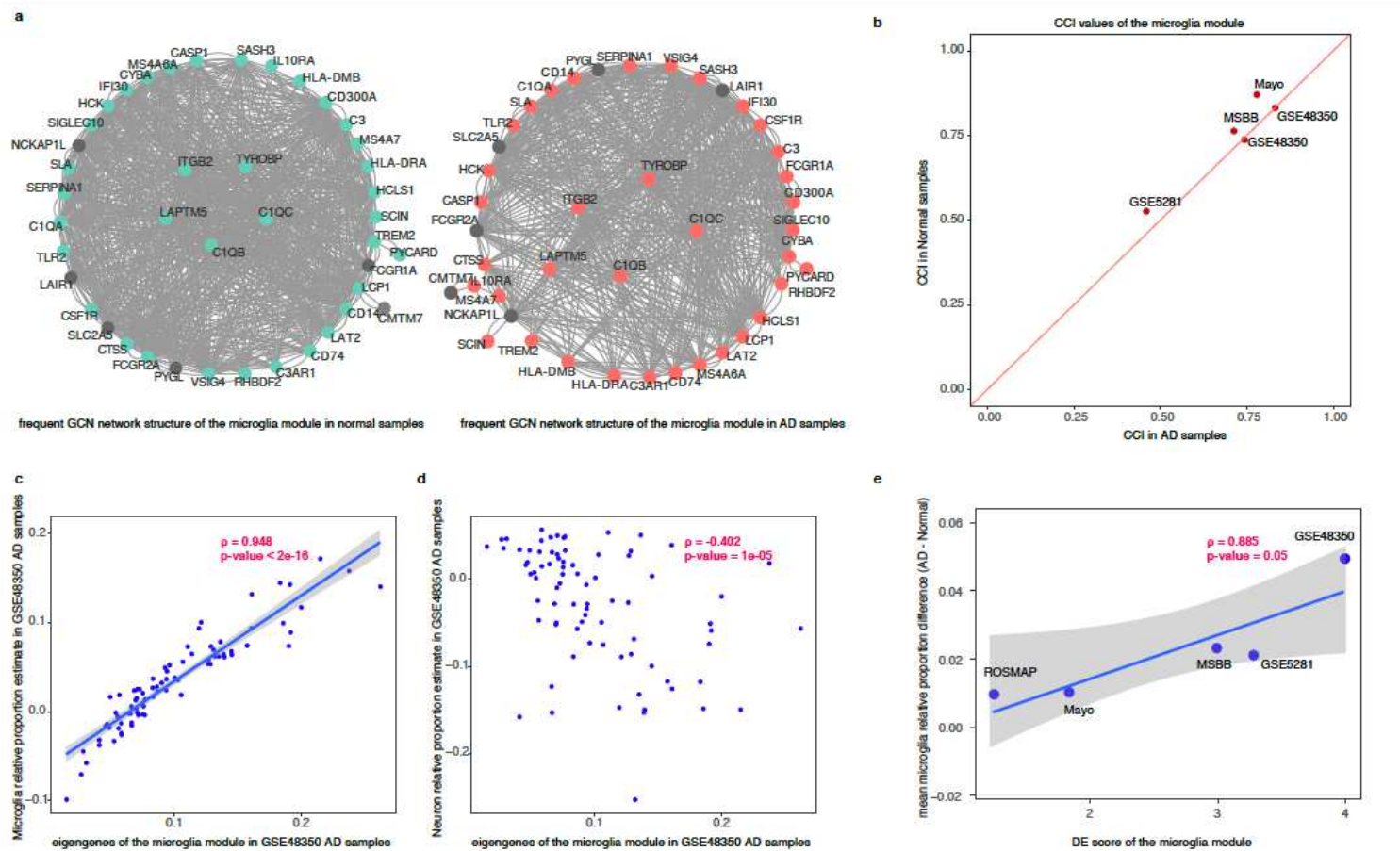frequent GCN network structure of the microglia module in AD samples

Figure 3

The highly differential expression in HDE_LDC modules were largely due to cell population change. a. Network visualization of the microglia module from the FGCN networks in AD and normal conditions (for legibility, only gene links that appears at least in four out of the five studies are shown). The genes with blue color in normal and red color in AD networks were up expressed in AD in at least two of the five datasets. b. CCI of the microglia module in AD vs normal samples in each of the five datasets. c. The Pearson correlation of microglia module eigengenes with the estimated microglia cell relative proportion in GSE48350 dataset, d. The Pearson correlation of microglia module eigengenes with the estimated neuron cell relative proportions. e. The DE scores of the microglia module with mean microglia relative proportion changes between AD and normal samples across five datasets. The gray area indicates the region defined by 95% confidence intervals.
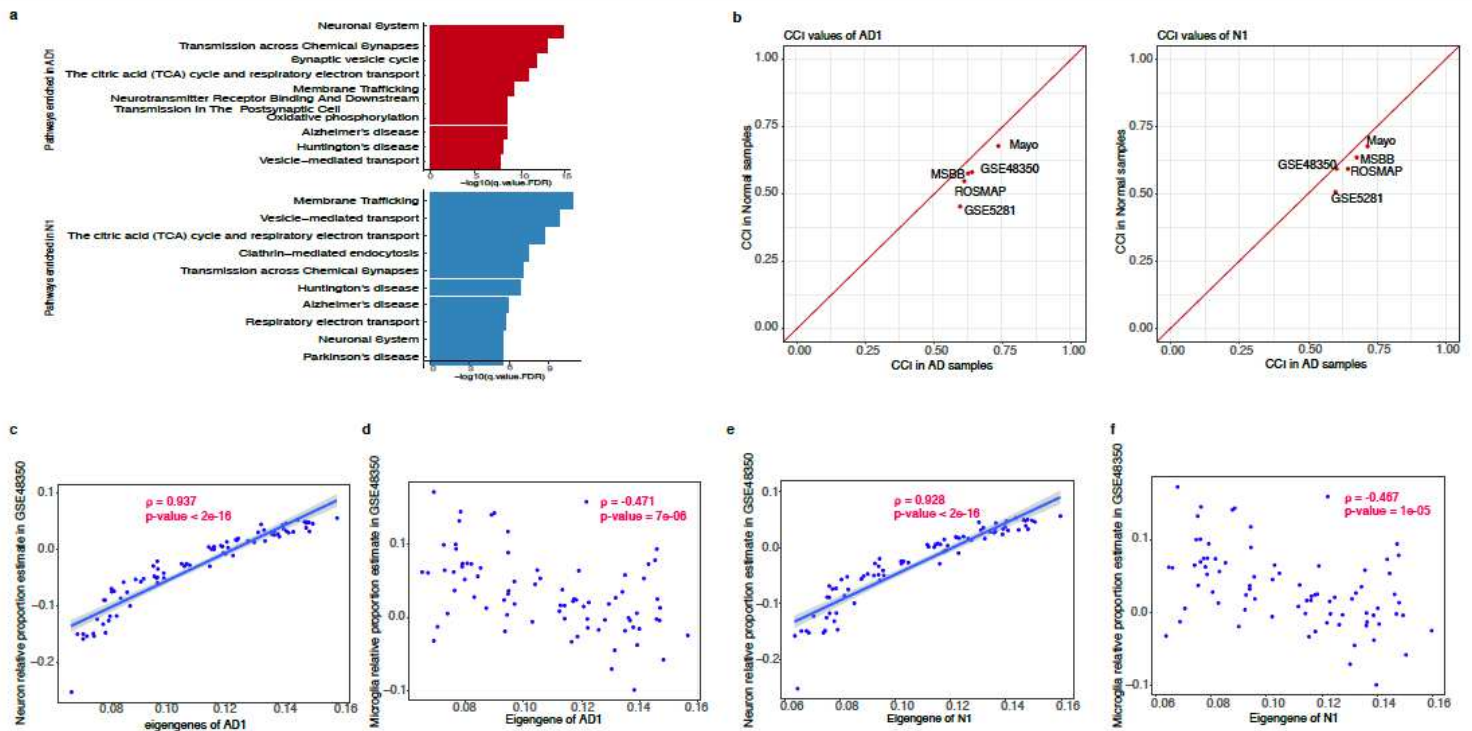
Figure 4

Neuron modules showed high DE that may be due to decreased proportion of neurons and high DC that may indicate regulatory change. a. Function enrichment of AD1/N1. b. CCI of neuron modules (AD1/N1) in AD vs. Normal samples in each of the five datasets. c. In GSE48350 dataset, the Pearson correlation of neuron module AD1 expression (eigengene values) with estimated neuron relative proportions. d. The Pearson correlation of AD1 eigengene values with estimated microglia relative proportions in GSE48350. e. In GSE48350 dataset, the Pearson correlation of neuron module N1 eigengene values with neuron relative proportions. f, The Pearson correlation of N1 eigengene values with microglia relative proportions in GSE48350.
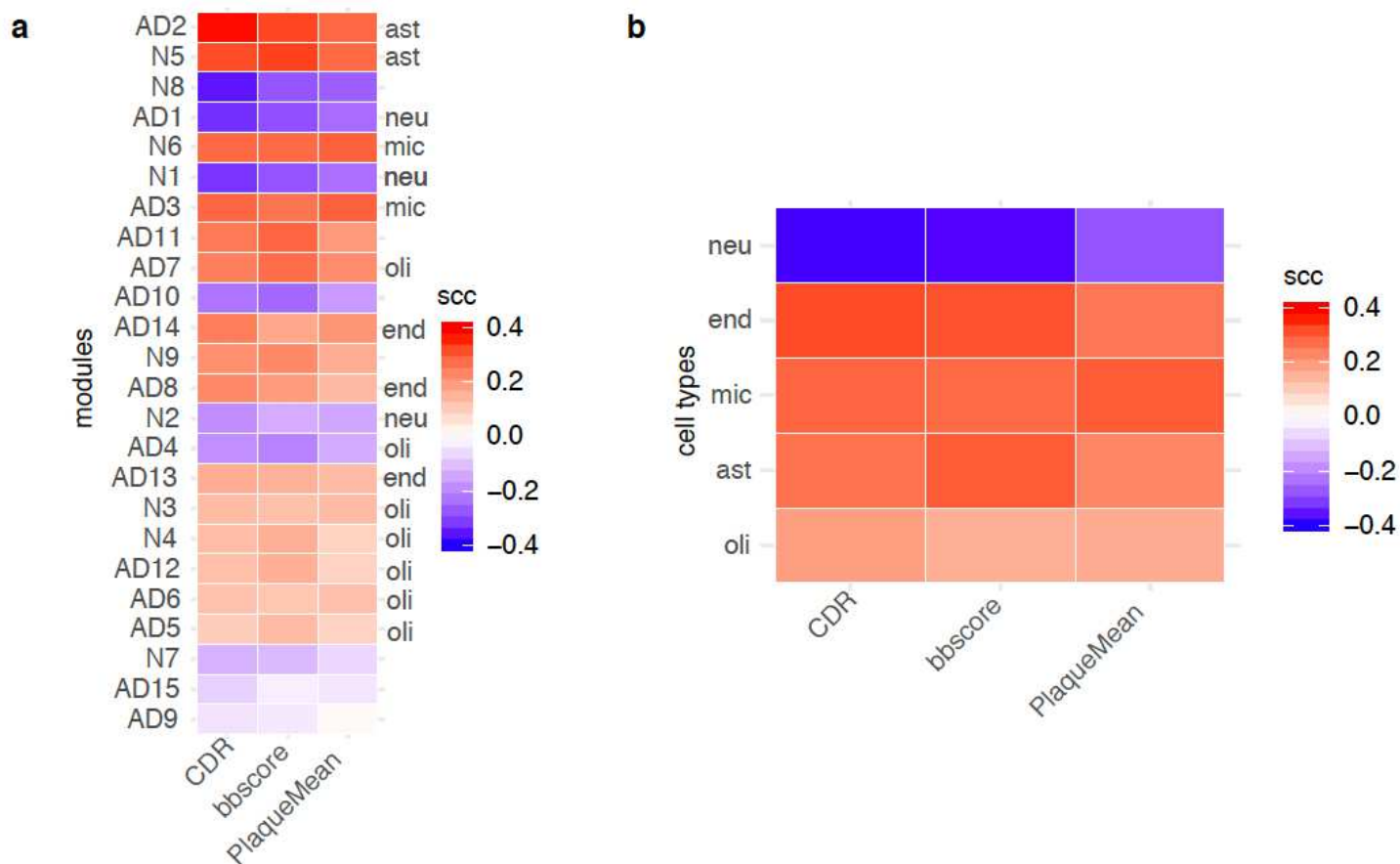
Figure 5

HDE modules are associated with AD neuropathology and cognitive status. a. Correlation of FGCN module eigengene values with three clinical traits. Spearman correlation coefficient (SCC) and p value between module eigengene expression and clinical dementia rating (CDR), neurofibrillary tangle burden (Braak & Braak score), beta-amyloid burden (PlaqueMean) are displayed respectively. The modules are ranked by the mean scc value and marked by corresponding cell types. ast: astrocytes, mic: microglia, neu: neuron. b. Correlation of brain cell type relative abundance with clinical traits. scc and p value between estimated brain cell type relative proportions and each of CDR, Braak & Braak score, PlaqueMean are displayed respectively. The brain cell types are ranked by the mean scc value. ast: astrocytes, mic: microglia, neu: neuron, end: endothelia, oli: oligodendrocytes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supTable79.xlsx
- supTable3.xlsx
- supTable2.xlsx
- supTable1.xlsx

- SupplementaryFiguresTables.docx
- supTable45.xlsx
- supTable13.xlsx
- SupTable12.xlsx
- supTable11.xlsx
- supTable10.xlsx
- supTable6.xlsx