

## Sex-Specific Cross Tissue Meta-Analysis Identifies Immune Dysregulation in Women with Alzheimer's Disease

Manish D Paranjpe<sup>1,2</sup>, Stella Belonwu<sup>1,3</sup>, Jason K Wang<sup>2</sup>, Tomiko Oskotsky<sup>1</sup>, Aarzu Gupta<sup>1</sup>, Alice Taubes<sup>1,5</sup>, Kelly Zalokusky<sup>5,6</sup>, Ishan Paranjpe<sup>1,4</sup>, Benjamin S Glicksberg<sup>4</sup>, Yadong Huang<sup>5,6</sup>, Marina Sirota<sup>1,7</sup>, for the Alzheimer's Disease Neuroimaging Initiative

1. UCSF Bakar Computational Health Sciences Institute, University of California San Francisco, San Francisco, CA, USA
2. Harvard-MIT Program in Health Sciences and Technology, Harvard Medical School, Boston, MA, USA
3. Pharmaceutical Sciences and Pharmacogenomics Graduate Program, University of California San Francisco, San Francisco CA, USA
4. Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY USA
5. The Gladstone Institute of Neurological Disease, San Francisco, CA, USA
6. Department of Neurology, University of California, San Francisco, USA
7. Department of Pediatrics, University of California San Francisco, San Francisco, CA, USA

**Corresponding author:** Marina Sirota, Assistant Professor, UCSF Bakar Institute for Computational Health Sciences, 550 16<sup>th</sup> Street, San Francisco, CA 94158 (marina.sirota@ucsf.edu)

**Word Count:** 4999

**Tables:** 3 Main Tables and 17 Supplementary Tables

**Figures:** 6 Figures

**References:** 67 References

## ABSTRACT

In spite of evidence of females having a greater lifetime risk of developing Alzheimer's Disease (AD) and greater apolipoprotein E4-related (apoE4) AD risk compared to males, molecular signatures underlying these findings remain elusive. We took a meta-analysis approach to study gene expression in the brains of 1,084 AD patients and age-matched controls and whole blood from 645 AD patients and age-matched controls. Gene-expression, network-based analysis and cell type deconvolution approaches revealed a consistent immune signature in the brain and blood of female AD patients that was absent in males. Machine learning-based classification of AD using gene expression from whole blood in addition to clinical features revealed an improvement in classification accuracy upon stratifying by sex, achieving an AUROC of 0.91 for females and 0.80 for males. These results help identify sex and apoE4 genotype-specific transcriptomic signatures of AD and underscore the importance of considering sex in the development of biomarkers and therapeutic strategies for AD.

## INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder and the most common cause of dementia<sup>1,2</sup>. It is pathologically characterized by the deposition of extracellular amyloid  $\beta$  (A $\beta$ ) and intracellular tau, otherwise referred to as plaques and neurofibrillary tangles, respectively<sup>3-5</sup>. AD is also marked by neuronal loss, impaired neurotransmitter signaling, neuroinflammation, and dysregulation of neuronal metabolism and immune response in the central nervous system<sup>6-8</sup>. AD prevalence increases dramatically with age, where the majority of cases are in individuals above the age of 65<sup>1,9</sup>. Although AD was identified more than a century ago<sup>10</sup>, its cause and pathophysiology are not fully understood, and there are no available treatments that aid in halting or reversing the disease<sup>11</sup>. Accordingly, it is of high priority to tackle AD, as it is projected to triple in incidence by 2050 as a consequence of population aging<sup>6,8,12</sup> and, to date, has no disease-modifying therapies.

While the exact cause and pathophysiology remain unknown, a number of mutations and genetic risk factors have been identified as associated with AD. Apolipoprotein E (apoE) is the most common genetic risk factor for late onset AD<sup>8,13-18</sup>. ApoE is a lipid binding protein, that plays a central role in lipid transport and metabolism. It is highly expressed in the brain, and is important for maintaining neuronal membranes during inflammation and damage. In humans, apoE has three isoforms, apoE2, apoE3, and apoE4, which are encoded by the three alleles,  $\epsilon$ 2,  $\epsilon$ 3, and  $\epsilon$ 4, of the apoE gene, respectively. The  $\epsilon$ 2 isoform has been shown to be protective against AD, while the  $\epsilon$ 4 isoform (apoE4) is associated with increasing the risk and lowering the age of onset for developing late onset AD in a gene dose-dependent manner<sup>19,20</sup>. Specifically, one copy

of the  $\epsilon 4$  isoform confers a 3 to 4-fold increased risk and 7 year decrease in age of onset, while two copies confers a 12 to 15-fold increased risk of AD, and a 14 year decrease in age of onset<sup>8,21</sup>.

Sex is another major risk factor in AD. Female sex is associated with increased AD incidence, exacerbated pathophysiology and increased rate of cognitive decline related to the disease progression<sup>8,22–25</sup>. It has been conjectured that the higher prevalence in females is a result of longer life span<sup>8,25</sup>. Alternatively, studies have alluded to sex-specific hormonal and metabolic changes that interplay with the onset and progression of AD dementia<sup>6,16,26</sup>. Sex also interacts with apoE isoform status, where females with the apoE4 isoform are at increased risk compared to males<sup>27–29</sup>. Despite the clear therapeutic potential to better understanding these pathophysiological patterns, there is still little understanding of the mechanisms underlying sex-specific differences in AD.

With the rising prevalence of AD, it is critical to facilitate the development of robust means to detect AD early and discover therapeutic interventions<sup>30–33</sup>. Technological innovations and the increasing availability of large transcriptomic datasets present worthwhile avenues to study and characterize the molecular underpinnings of AD stratified by sex. Here, we analyze publicly available gene expression datasets from over 1,500 brain and blood samples to characterize this highly complex disease. To derive sex-specific transcriptomic molecular signatures, we perform a meta-analysis, differential gene expression, weighted gene co-expression network analysis, pathway enrichment, and cell-type deconvolution in a large cohort of brain and blood samples from AD patients and healthy controls (Figure 1). We further characterize these signatures and apply machine learning to build a predictive model based on biomarkers identified in the blood of AD patients. Our findings reveal underlying mechanisms of sex differences, which provide clinical implications for identifying more accurate, and less

invasive biomarkers, as well as efficacious therapeutics tailored to better fit the complex molecular profiles in AD.

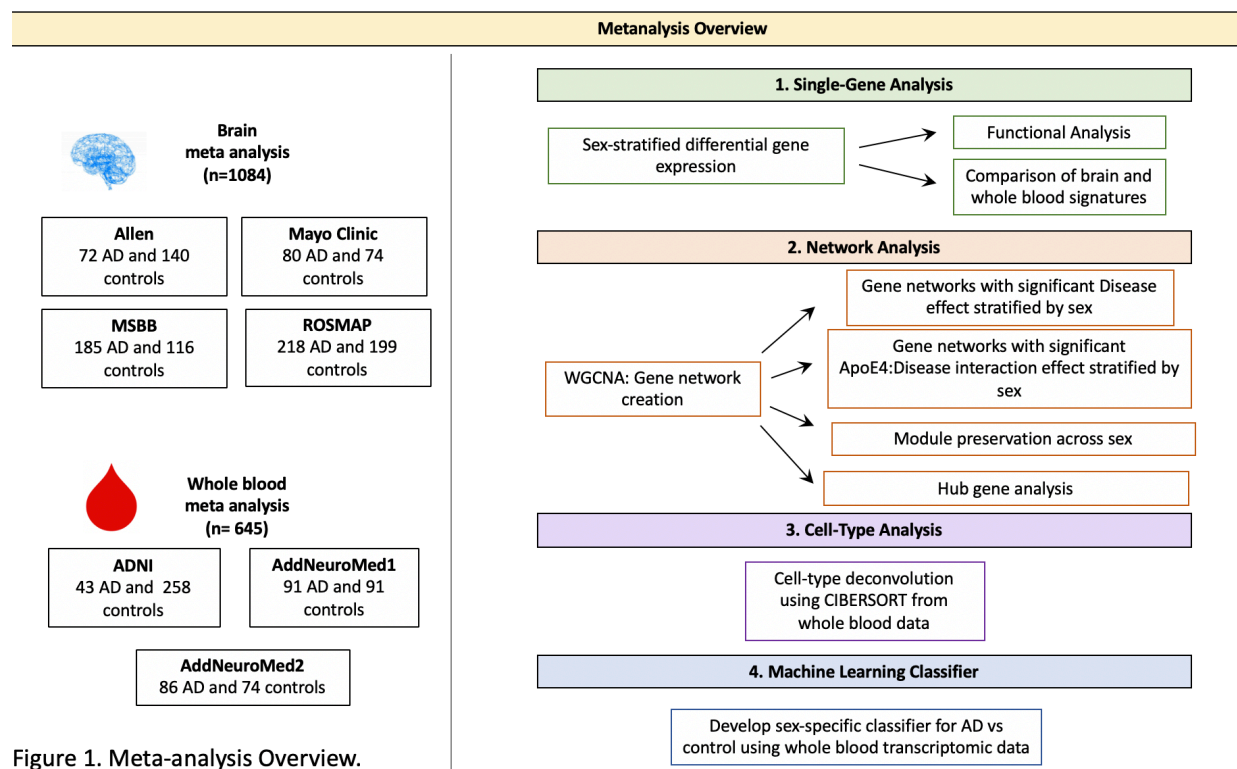


Figure 1. Meta-analysis Overview.

## METHODS

### Study Cohorts:

Publicly available RNA-sequencing (RNA-Seq) and microarray datasets from the Gene Expression Omnibus (GEO) and from consortium studies indexed on PubMed were searched for the key word “Alzheimer’s”. To minimize technical variability, brain samples were restricted to RNA-sequencing studies while blood analyses were restricted to microarray studies. Samples were curated to include bulk gene expression from subjects with Alzheimer’s or elderly healthy individuals with no history of neurodegenerative disease. Individuals with non-Alzheimer’s neurodegenerative diseases including Huntington’s and Parkinson’s were excluded. Brain

samples were restricted to the hippocampus, parietal cortex, temporal cortex and prefrontal cortex. Additional clinical covariates, including age, sex, apoE4 carrier status, education were recorded for the samples and used as covariates or stratification variables in subsequent analyses.

### **Gene Expression Meta-Analysis:**

Meta-analysis was conducted separately for brain and blood studies according to standard quality control, normalization, and batch correction procedures. All data processing was conducted using R (v3.6.1).

#### *Brain studies:*

Raw RNA-sequencing data were processed for the Mount Sinai Brain Bank (MSBB)<sup>34</sup>, Mayo Clinic RNAseq<sup>35</sup>, and Religious Orders Study and Memory and Aging Project (ROSMAP)<sup>36</sup> as previously described in the AMP-AD consortium project . Briefly, read alignment and counting was performed using STAR<sup>38</sup>. Alignment quality metrics were generated using PICARD<sup>39</sup>. For the Allen dataset, expected counts produced using RSEM were downloaded from the Allen Brain Atlas: Aging Dementia and TBI Study website<sup>40</sup>. Counts-per-million (CPM) were calculated for all studies. Genes with less than 1 CPM in at least 50% of samples across tissue diagnosis group were removed. Genes with missing gene length or GC content percentage metrics were removed. Library normalization was performed using conditional quantile normalization.

Following read alignment and normalization, studies were merged using common genes between the four studies. Mean value imputation was performed for missing gene expression values. Quantile normalization was performed across studies. The ComBat function from the *sva* package<sup>41</sup> was used to perform cross-study normalization, retaining variation in apoE4 carrier

status, sex, and diagnosis. Principal component analysis (PCA) plots were generated to evaluate successful batch correction and to detect outliers.

#### *Blood studies:*

Study data were downloaded from GEO for the AddNeuroMed datasets<sup>42</sup> or the Alzheimer's Disease Neuroimaging Initiative Consortium<sup>43</sup> (ADNI) for the ADNI dataset and processed. Raw data were not available for the ADNI dataset and therefore normalized expression data were used for all studies. Outlier removal was performed on individual studies by removing probes whose mean expression was outside 1.5 times the interquartile range. Probe IDs were mapped to gene symbols. Expression value of probes mapping to the same gene were reported as the median of all probes mapping to that gene<sup>44</sup>. Quantile normalization was performed across studies. Similar to the brain data analysis, the ComBat function from the *sva* package was used to perform cross-study normalization, retaining variation in apoE4 carrier status, sex and diagnosis. Principal component analysis (PCA) plots were generated to evaluate successful batch correction.

#### **Differential Gene Expression Analysis:**

All differential gene expression analyses were performed separately for brain and blood samples. The Limma package<sup>45</sup> was used to determine differentially expressed genes between cases and controls all together and stratified by sex. In each model, age and apoE4 carrier status were included as covariates to minimize confounding. An additional covariate of education was used in the blood analyses. Education was not available for all brain samples and therefore was not included as a covariate. A cutoff false discovery rate (FDR) of 0.05 and fold change (FC) of greater than or equal to 1.2 was used for brain analyses. Fold changes were calculated using the

individual study data before merging and weighted by sample size. For blood analyses, a FC cutoff was not used to maximize gene discovery. Significant overlap between up- and down-regulated genes between males and females was assessed using a hypergeometric test. Functional enrichment analysis of gene lists was carried out by overrepresentation analysis using the KEGG<sup>46</sup> database of biological pathways.

## **Network Analysis:**

### *Weighted Gene Co-Expression Network Analysis:*

In order to detect gene network level differences, network analysis was performed using Weighted Gene Co-Expression Network Analysis (WGCNA)<sup>47</sup>. All analyses were performed separately for brain and blood samples. In signed WGCNA, a module is defined as a set of genes whose expression is highly correlated in the same direction. Signed gene co-expression networks were created separately for male and female samples to identify sex-specific gene modules. Module Z-summary scores were computed to assess module preservation between male and female networks, as described previously<sup>48</sup>. A Z-summary score greater than ten was considered to be strong evidence of preservation between the two networks. A score between two and ten was considered to represent weak to moderate evidence of preservation, as previously described<sup>48</sup>.

Association between module gene expression and case/control status was assessed by relating the module eigengenes, defined as the first principal component of the genes in a given module, to case/control status using linear regression. Age, apoE4 carrier status, and education (for blood samples) were used as covariates to minimize confounding. An additional analysis identifying apoE-by-disease interaction effects was performed by adding the interaction term: apoE4 carrier status:case/control status to the previous model. Significant modules were



characterized by performing functional gene enrichment using the KEGG database of biological pathways<sup>31</sup>.

#### *Hub Gene Analysis:*

To identify central regulators of gene expression, we identified hub genes within significant modules, as described previously<sup>47</sup>. Hub genes were defined as genes with gene significance (the correlation between the gene expression and case/control status) greater than 0.2 and module membership (the correlation between gene expression and module eigengene) greater than 0.8, as previously described<sup>47</sup>. We also restricted hub genes to those that were differentially expressed in AD vs control. Network visualization using the STRING v11<sup>51</sup> database was used to assess evidence for protein-protein interactions between hub genes.

#### **Cell-type Deconvolution**

CIBERSORT<sup>52</sup> was applied to the transcriptomic signatures generated in the blood meta-analysis to deconvolve gene expression data into cell type composition and identify sex-specific dysregulation of immune cell types between cases and controls. CIBERSORT applies a linear support vector regression method to solve the problem:  $m = f \times B$  where  $m$  is an input mixture of gene expression data for a given sample,  $f$  is a vector consisting of fractions of each cell type in the mixture and  $B$  is a matrix of reference gene expression profiles. A gene expression profile of 22 reference cell populations was built using differential gene expression of purified or enriched cell populations from the authors of CIBERSORT.

CIBERSORT was used to deconvolve gene expression data from pooled male and female data, male only samples, and female only samples. In each condition, differences in cell type proportions between cases and controls were compared using a linear regression model adjusting

for age, sex (in the pooled male female analysis), and apoE4 carrier status. An additional analysis identifying apoE4-by-disease interaction effects was performed by adding the interaction term: apoE4 carrier status:case/control status to the previous model. A cutoff FDR of 0.05 was deemed significant.

## **Classification of Healthy and Alzheimer's Disease Patients**

A linear support vector machine (SVM) model with  $l_1$  regularization to enforce feature sparsity was used to classify Alzheimer's patients and healthy controls based on blood gene expression data. To assess the relative value of stratifying by sex in increasing model performance, we compared the performance of three models built using pooled male and female samples, male samples only, and female samples only. We also compared the performance of a 'clinical model' with age, sex (for male and female pooled samples), and apoE4 carrier status information to a 'clinical + molecular model' which included age, sex (for male and female pooled samples), apoE4 carrier status, and transcriptomic data from the blood meta-analysis.

For each model, data were split into 75% training/validation and 25% test sets using a class balancing procedure to maintain a constant case/control ratio across training/validation and test sets. A random search over the space  $10^{-4}$  to  $10^4$  with five-fold cross validation was used to optimize the C hyper-parameter, or the degree of regularization penalty applied for misclassified points. Receiver operating characteristic (ROC) curves were generated from the test set. Model performance was assessed using the area under the ROC curves. Feature importance was determined using the absolute value of the model coefficients corresponding to the vector coordinates orthogonal to the model hyperplane.

## **RESULTS**

### **Study Cohort Characteristics**

We obtained four publicly available RNA-seq data sets (Allen Brain Institute Aging Dementia and TBI study, Mayo Clinic RNA-seq, MSBB, and ROSMAP) from the brain (temporal cortex, parietal cortex, prefrontal cortex, and hippocampus) and three microarray datasets from whole blood (AddNeuroMed cohort 1, AddNeuroMed cohort 2 and ADNI). After outlier removal, we included a total of 1,084 brain samples (58% female; 26% apoE4 carriers) and 645 blood samples (58% female; 38% apoE4 carriers) in our analysis. Table 1 shows a summary of sample annotations including number of cases and controls, apoE carrier status, and number of males and females for brain datasets and blood datasets.

In the brain datasets, compared to controls, AD patients were significantly older (mean  $\pm$  SD for AD:  $86.5 \pm 6.0$  years and controls:  $84.8 \pm 7.4$  years; two sample t-test,  $P < 0.001$ ), more likely to be apoE4 carriers (AD: 38% carriers vs controls: 15% carriers; Chi-squared test,  $P < 0.001$ ), and more likely to be females (AD: 65% female vs controls: 51% female; Chi-squared test,  $P < 0.001$ ).

In the blood datasets, compared to controls, AD patients were significantly older (mean  $\pm$  SD for AD:  $77.0 \pm 7.1$  years and controls:  $74.7 \pm 5.7$  years; two sample t-test,  $P < 0.001$ ), more likely to be apoE4 carriers (AD: 60% carriers vs controls: 27% carriers; Chi-squared test,  $P < 0.001$ ), more likely to be females (AD: 64% female vs controls: 55% female; Chi-squared test,  $P < 0.001$ ), and had more years of education (mean  $\pm$  SD for AD:  $9.4 \pm 4.8$  years and controls:  $13.9 \pm 4.7$  years; two sample t-test,  $P < 0.001$ ).

Studies were merged and batch corrected using ComBat resulting in 13,500 common genes across 1,084 samples for brain studies and 3,371 common genes across 645 samples for blood studies. Supplementary Figure S1 and S2 show PCA plots before and after batch correction, demonstrating successful data merging and batch effect removal.

**Table 1: Meta-analysis Study Characteristics**

Study	Accession	Total participants	AD, no. (%)	AD		CN	
				Female /Male (% Female)	apoE4 /No (% Yes)	YesFemale /Male (% Female)	apoE4 Yes /No (% Yes)
<b>Brain Transcriptomic Studies</b>							
Allen	<a href="https://aging.brain-map.org/">https://aging.brain-map.org/</a>	212	72 (34)	29/43 (40)	22/50 (31)	54/86 (39)	19/121 (14)
Mayo Clinic RNA-Seq	syn5550404	154	80 (52)	49/31 (61)	42/38 (53)	36/38 (49)	9/65 (12)
MSBB	GSE52564	301	185 (62)	131/54 (71)	63/122 (34)	57/59 (49)	16/100 (13)
ROSMAP	syn3219045	417	218 (52)	151/67 (70)	83/135 (38)	122/77 (61)	33/166 (17)
<b>Sum</b>		<b>1084</b>	<b>555 (52)</b>	<b>360/195 (65)</b>	<b>210/345 (38)</b>	<b>269/260 (51)</b>	<b>77/452 (15)</b>
<b>Whole Blood Transcriptomic Studies</b>							
ADNI	<a href="http://adni.loni.usc.edu/">http://adni.loni.usc.edu/</a>	301	43 (14)	17/26 (40)	32/11 (74)	135/125 (52)	71/189 (27)
AddNeuroMed1	GSE63060	182	91 (50)	65/26 (71)	52/39 (57)	55/36 (60)	30/61 (33)
AddNeuroMed2	GSE63061	160	86 (43)	59/27 (69)	47/39 (55)	45/29 (61)	15/59 (20)
<b>Sum</b>		<b>645</b>	<b>220 (34)</b>	<b>141/79 (64)</b>	<b>131/89 (60)</b>	<b>235/190 (55)</b>	<b>116/309 (27)</b>

### Differential Gene Expression in the Brain Identifies a Distinct Sex-Specific Signature of AD

We observed distinct AD-associated transcriptomic signatures in the brain in males and females. A total of 981 genes were differentially expressed in females, including 583 upregulated genes and 398 downregulated genes ( $FC > 1.2$ ,  $q < 0.05$ ; Figures 2A-B; Supplementary Table 1). In males, 513 genes were differentially expressed, including 415 upregulated genes and 98 downregulated genes ( $FC > 1.2$ ,  $q < 0.05$ ; Figures 2A-B; Supplementary Table 1). Altogether,

631 genes were uniquely dysregulated in females, including 309 upregulated genes and 327 downregulated genes. In males, 166 genes were uniquely dysregulated, including 141 upregulated genes and 27 downregulated genes. There was a significant overlap of dysregulated genes across males and females ( $P < 0.05$ ; hypergeometric test).

Next, we characterized the transcriptomic signatures observed in the brains of male and female AD patients. In females, among upregulated AD genes, we found 69 enriched pathways, many of them relating to components of the innate and adaptive immune system (Table 2; Supplementary Table 2). Several upregulated HLA system genes including HLA-DPB1, HPA-DRA, HLA-DOA, HLA-DRB5, HLA-DMA, HLA-DPA1 contributed to enrichment of a number of pathways relating to response to infection (Table 3). Components of the complement system including C1QA, C4B, and C4A were also uniquely dysregulated in females (Table 4; Supplementary Table 2). We also observed an enrichment of genes in the MAPK signaling pathway including MRAS, MK2, and MK3. Downregulated AD genes in females were enriched for a number of neurological signaling pathways including synaptic vesicle exocytosis, neuroactive ligand-receptor activation, and GnRH signaling (Table 2; Supplementary Table 3).

Strikingly, we observed an enrichment of fewer immune-related pathways in males with AD. Among upregulated genes in male AD patients, we found 12 enriched pathways, including amoebiasis and cytokine-cytokine receptor interaction, suggestive of adaptive and innate immune activation (Table 2; Supplementary Table 4). Similar to females, we also observed an enrichment of the MAPK signaling pathway, including MAP4K4 and MK2, in males. Among downregulated genes in male AD patients, we did not identify significantly enriched pathways. For a full list of enriched pathways, refer to Supplementary Tables S2-S4.

Lastly, we performed a non-stratified analysis comparing gene expression between AD and control samples irrespective of sex. Statistical models were adjusted for sex, apoE4 carrier status, and age. A total of 662 genes were upregulated and 430 genes were downregulated in patients with AD compared to controls (Figure S3, Table 2; Supplementary Table 1. Upregulated genes were enriched for several pathways previously implicated in AD including PI3K-Akt signaling and MAPK signaling as well as a number of immune related pathways including Staphylococcus aureus infection, human papillomavirus infection, and malaria (Supplementary Table S5). Several components of the complement system, including C4B, C4A, C1R, C3AR1, and C5AR1 also contributed to this enrichment (Supplementary Table S6). In our analysis of downregulated genes, we found several pathways related to neuroreceptor signaling and GABAergic transmission were enriched including the genes GABRA1, GNG3, GNG2, SLC32A1, GABRD, and GABRG2 (Supplementary Table S6).

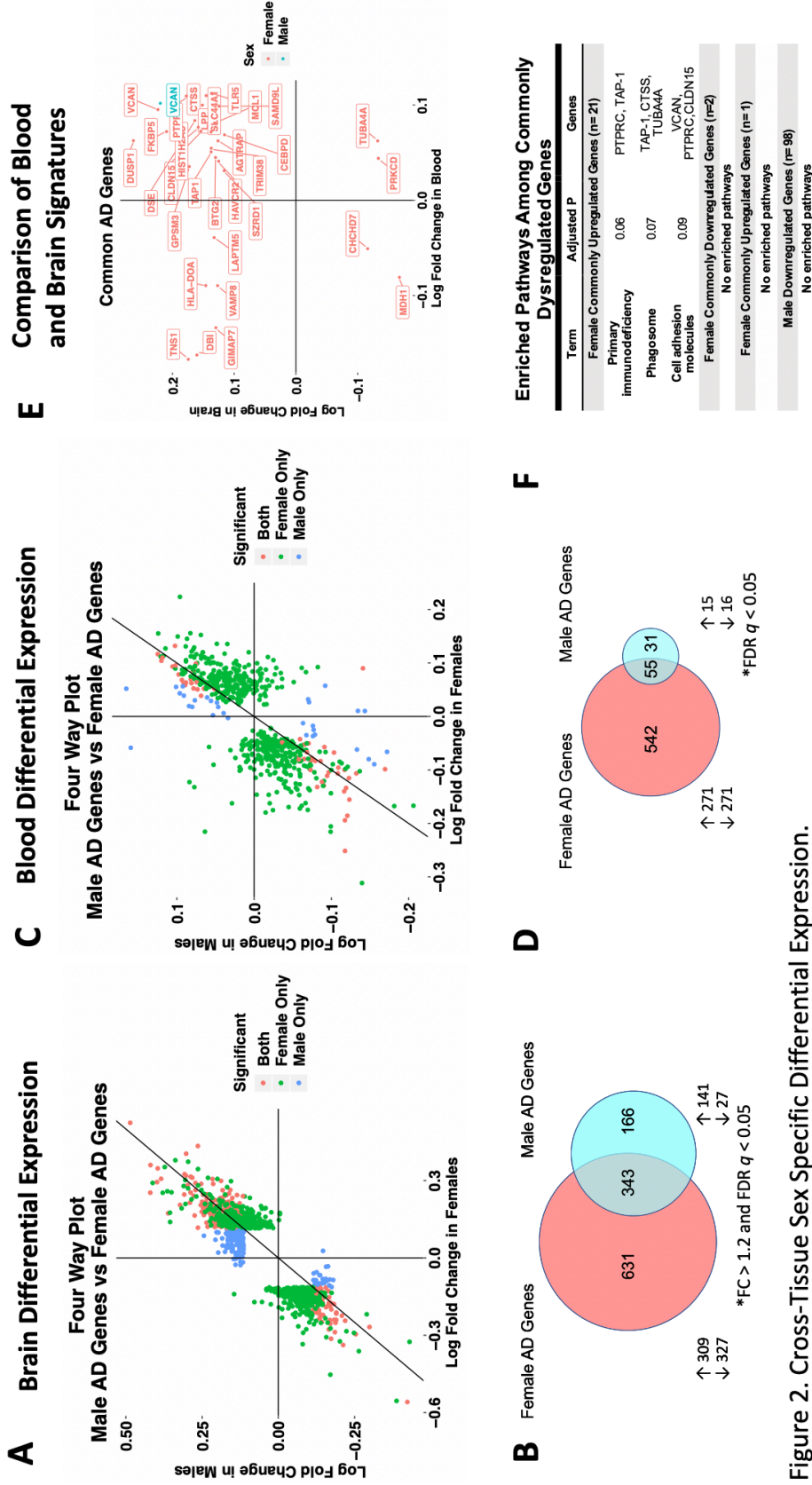


Figure 2. Cross-Tissue Sex Specific Differential Expression.



**Table 2: Enriched Pathways in Brain**

Term	Adjusted P	Genes
<b>Female Upregulated Genes (n= 583)</b>		
<b>Staphylococcus aureus infection</b>	<0.001	C1QB;C1QA;HLA-DRB5;CFH;CFI;PTAFR;C4B;C4A;HLA-DMA;HLA-DMB;FCGR2A;HLA-DPB1;HLA-DRA;MASP1;HLA-DOA;HLA-DPA1;C1QC
<b>MAPK signaling pathway</b>	<0.001	CSF1;FLT4;HSPB1;FGF1;FGF2;EGFR;RASGRP3;MECOM;RPS6KA1;GNA12;MAP3K20;CD14;MAP3K6;PDGFRB;TGFB2;ANGPT2;IL1R1;DUSP1;HGF;GNG12;NFKB1;TGFB1;GADD45G;TGFB2;EFNA1;MRAS;MAPKAPK3;MAPKAPK2;TAB2;MYD88;EPHA2;HSPA1A
<b>PI3K-Akt signaling pathway</b>	<0.001	CDKN1A;CSF1;ITGB5;FLT4;TNC;LPAR3;FGF1;THBS2;FGF2;EGFR;SPP1;ITGB8;IL6R;MCL1;PDGFRB;ANGPT2;SYK;LAMB2;NOS3;HGF;FN1;GNG12;OSMR;GNG11;NFKB1;EFNA1;COL1A2;ITGA10;ITGA6;SGK1;TLR4;EPHA2;TLR2
<b>Leishmaniasis</b>	<0.001	TGFB2;HLA-DRB5;NFKB1;HLA-DMA;HLA-DMB;FCGR2A;HLA-DPB1;HLA-DRA;TAB2;HLA-DOA;TLR4;MYD88;HLA-DPA1;TLR2
<b>Inflammatory bowel disease (IBD)</b>	<0.001	TGFB2;HLA-DRB5;NFKB1;HLA-DMA;HLA-DMB;HLA-DPB1;HLA-DRA;HLA-DOA;TLR5;TLR4;IL18R1;HLA-DPA1;TLR2
<b>Toxoplasmosis</b>	<0.001	TGFB2;HLA-DRB5;LAMB2;NFKB1;HLA-DMA;HLA-DMB;HLA-DPB1;HLA-DRA;TAB2;ITGA6;HLA-DOA;TLR4;MYD88;HLA-DPA1;TLR2;HSPA1A
<b>Cell adhesion molecules (CAMs)</b>	<0.001	HLA-DRB5;SDC2;HLA-E;CLDN11;OCLN;VCAN;HLA-DMA;PTPRC;HLA-DMB;CLDN15;HLA-DPB1;HLA-DRA;ITGB8;ITGA6;CD58;HLA-DOA;CD34;HLA-DPA1
<b>Epstein-Barr virus infection</b>	<0.001	LYN;HLA-DRB5;CDKN1A;SYK;DDX58;TAP1;TNFAIP3;NFKB1;HLA-E;GADD45G;HLA-DMA;HLA-DMB;HLA-DPB1;HLA-DRA;TAB2;HES1;CD58;HLA-DOA;MYD88;HLA-DPA1;TLR2
<b>Pathways in cancer</b>	<0.001	NOTCH2;CDKN1A;NOTCH4;FLT4;LEF1;SLC2A1;CXCR4;LPAR3;FGF1;FGF2;DLL1;GLI3;EGFR;GLI2;RASGRP3;MECOM;GNA12;HES1;RXRG;IL6R;PDGFRB;CDKN2B;TGFB2;LAMB2;FZD7;HGF;FN1;MITF;GNG12;GNG11;NFKB1;TGFB1;GADD45G;TGFB2;HEYL;SMO;ITGA6
<b>Systemic lupus erythematosus</b>	<0.001	C1QB;C1QA;HLA-DRB5;C4B;C4A;HLA-DMA;HLA-DMB;FCGR2A;HLA-DPB1;HIST1H4H;HLA-DRA;HLA-DOA;HIST1H2AC;HIST1H2BD;HLA-DPA1;C1QC
69 more..		
<b>Female Downregulated Genes (n= 398)</b>		
<b>Neuroactive ligand-receptor interaction</b>	0.001	CHRN2;GABRB2;GABRA1;CHRNA2;GABRA4;PTH2R;CKK;GRIK2;HTR5A;RXFP1;GABRG2;MCHR2;MAS1;GLRA3;GLRB;CNR1;SST;NPY;TAC1;VIP;GABRD
<b>Retrograde endocannabinoid signaling</b>	0.01	RIMS1;GABRB2;MAPK9;GABRA1;NDUFA5;CNR1;GABRA4;ITPR1;ADCY1;GABRD;GABRG2
<b>Synaptic vesicle cycle</b>	0.01	RIMS1;SLC6A7;ATP6V1G2;ATP6V1B2;ATP6V1H;ATP6V0E2;CPLX1;STX1A
<b>Aldosterone synthesis and secretion</b>	0.01	CAMK2D;STAR;PRKCE;CAMK4;CAMK2A;ITPR1;ADCY1;ATP1B1;CAMK1G
<b>Nicotine addiction</b>	0.02	CHRN2;GABRB2;GABRA1;GABRA4;GABRD;GABRG2
<b>GnRH signaling pathway</b>	0.03	MAPK9;EGR1;CAMK2D;PRKCD;CAMK2A;ITPR1;PTK2B;ADCY1
<b>Male Upregulated Genes (n= 415)</b>		
<b>PI3K-Akt signaling pathway</b>	0.009	PDGFRB;CSF3R;CSF1;ITGB5;IRS1;LAMB2;FLT4;FN1;LAMC1;GNG12;OSMR;FGF2;NFKB1;BCL2L11;ITGA10;KDR;SPP1;ITGA6;EPHA2;TLR2
<b>MAPK signaling pathway</b>	0.01	PDGFRB;TGFB2;CSF1;IL1R1;DUSP1;FLT4;GNG12;FGF2;NFKB1;TGFB2;MRAS;MECOM;MAPKAPK2;KDR;CD14;MYD88;EPHA2;MAP4K4
<b>NOD-like receptor signaling pathway</b>	0.03	NEK7;CARD6;ERBIN;ANTXR2;GBP2;ANTXR1;IKBKE;GBP1;NFKB1;MYD88;GBP4;GBP3
<b>ECM-receptor interaction</b>	0.03	ITGB5;LAMB2;ITGA10;SPP1;FN1;ITGA6;LAMC1;HSPG2
<b>Proteoglycans in cancer</b>	0.04	TGFB2;MRAS;ITGB5;FZD7;KDR;FN1;GAB1;HCLS1;WNT7A;HSPG2;FGF2;TLR2
<b>Focal adhesion</b>	0.04	VAV3;PDGFRB;ITGB5;LAMB2;ITGA10;FLT4;KDR;FN1;SPP1;CAPN2;ITGA6;LAMC1
<b>Fc gamma R-mediated phagocytosis</b>	0.04	VAV3;PTPRC;FCGR2A;MYO10;INPP5D;DOCK2;WASF2;PLPP1
<b>Amoebiasis</b>	0.04	TGFB2;IL1R1;LAMB2;FN1;LAMC1;CD14;NFKB1;TLR2
<b>Pathways in cancer</b>	0.04	NOTCH2;PDGFRB;TGFB2;CSF3R;LAMB2;FZD7;FLT4;LEF1;FN1;SLC2A1;WNT7A;CXCR4;MITF;LAMC1;GNG12;FGF2;NFKB1;GLI2;TGFB2;BCL2L11;MECOM;ITGA6;NFE2L2
<b>Cytokine-cytokine receptor interaction</b>	0.04	TGFB2;CSF3R;CSF1;IL1R1;CXCR4;LIFR;INHBB;OSMR;TNFRSF1B;IL17RB;TGFB2;IL1RL1;ACKR3;TNFRSF25;IL18R1
2 more..		



**Male Downregulated Genes (n= 98)**

No enriched pathways

## **Network Analysis in the Brain Identifies a Stronger Disease Signature in Females**

To assess transcriptomic changes on a gene network level, we utilized WGCNA. Gene networks were derived separately for male and female samples and compared using network preservation methods, as previously described<sup>48</sup>. We identified two AD-associated modules in males and 11 AD-associated modules in females (Figure 3A) that met the significance threshold (FDR < 0.05) and were either positively or negatively correlated with case/control status. Among the male modules, a 463-gene module (termed black) was upregulated in AD, and a 151-gene module (termed tan) was downregulated in AD. The black module in males had significant overlap with two modules in females (termed yellow and pink) ( $P < 0.001$ ; hypergeometric test) as indicated by asterisks in Figure 3B. The black module also had strong preservation in the female network (Z-summary score > 10). Among the female-specific disease associated modules, four modules (termed green, red, black and turquoise) were downregulated in AD, while seven were upregulated (Figure 3A).

Enrichment analysis of disease-associated modules using the 2019 KEGG Human pathway database revealed pathways relevant to AD that were consistent with those identified in the single gene analysis (Figure 3A). For example, in both males and females, an upregulated module was enriched for Akt signaling related pathways and downregulated modules were enriched for oxidative phosphorylation and thermogenesis related pathways, consistent with single gene level analyses.

Notably, several additional pathways not seen through single gene analysis were observed in the network analyses. An upregulated module in both males and females was highly enriched

for zinc finger nuclease genes related to Herpes simplex viral infection, consistent with recent work demonstrating Herpes virus infection in AD brains<sup>53</sup>.

Consistent with the single gene analysis, we observed greater number of disease associated modules in females with AD than in males. For example, an upregulated female module was enriched for cell structural processes related to adherens junctions, actin cytoskeleton and axonal guidance. An additional downregulated female module was enriched for neurological signaling pathways including synaptic vesicle exocytosis, aldosterone synthesis and secretion and morphine addiction. Interestingly, an additional female downregulated module was enriched for autophagy and proteolysis pathways, consistent with molecular studies demonstrating decreased autophagy in AD, particularly in females<sup>55</sup> (Figure 3A).

We also conducted an analysis identifying modules with apoE4:disease interactive effect to understand differential penetrance of the apoE  $\epsilon$ 4 allele in males and females. In the male gene network, we were unable to identify modules with significant apoE4:disease interactive effect. Interestingly, in the female network, we identified one module that was downregulated (2211 genes) in AD, and two modules (329 genes and 439 genes) that were upregulated in AD and exhibited a significant apoE4:disease interactive effect (Figure 3A). The two upregulated modules (termed pink and purple) were significantly enriched for several zinc finger nuclease genes related to Herpes simplex viral infection. The downregulated module was enriched for metabolic pathways including oxidative phosphorylation and the TCA cycle. Together these results suggest a female-specific network dysregulation involving zinc finger nucleases and metabolic alteration supporting differential apoE4 penetrance in males and females.

There were 102 hub genes among disease associated modules in the female network identified as module membership greater than 0.8, gene significance greater than 0.2, and

differentially expressed between AD and controls (Figure 3C; Supplementary Table S7). In contrast, zero hub genes were identified in the male gene network. Protein-protein interaction maps generated by STRING v11 suggest several  $\text{Ca}^{+2}$ - and G protein-dependent interconnected genes including ITPKB, PDGFRB, GNG12, and GNA12 among the female disease associated modules (Figure 3C). Among modules with apoE4:disease interactive effect in females, 35 hub genes were identified, including ITPKB as a highly connected regulator (Figure 3D). For a full list of genes in each module, including hub genes, please refer to Supplementary Table S7).

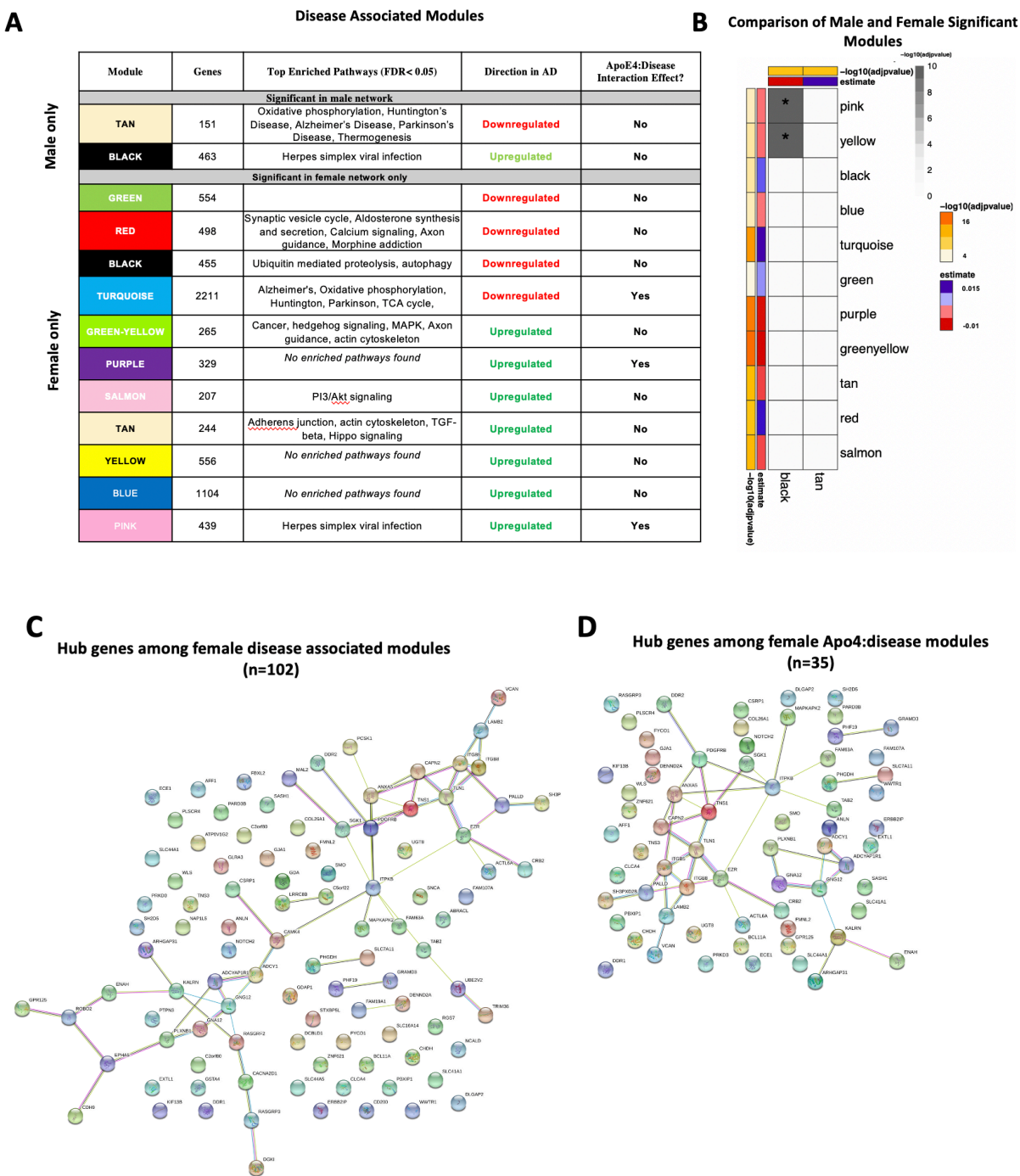


Figure 3. Network Analysis in the Brain.

## **Differential Gene Expression in Whole Blood Identifies Stronger Disease Signatures in Females with AD in Comparison to Males**

Similar to the brain, we observed distinct AD-associated transcriptomic signatures between males and females with AD in whole blood. We observed a total of 599 differentially expressed genes in females with AD, including 294 upregulated genes and 305 downregulated genes ( $q < 0.05$ ; Figures 2C-D; Supplementary Table 8). In males, 98 genes were differentially expressed in AD, including 38 upregulated genes and 50 downregulated genes ( $q < 0.05$ ; Figures 2C-D; Supplementary Table 8). Altogether, 542 genes were uniquely dysregulated in females, including 271 upregulated genes and 271 downregulated genes. In males, 31 genes were uniquely dysregulated, including 15 upregulated genes and 16 downregulated genes. There was a significant overlap of dysregulated genes across males and females with AD ( $P < 0.05$ ; hypergeometric test).

Next, we characterized the transcriptomic signatures observed in the blood of male and female AD patients. Among upregulated genes in female AD patients, we found 14 enriched pathways, many of them relating to components of the innate and adaptive immune system (Table 3; Supplementary Table S9). Several cytokine response elements including STAT5B, STAT6, and IL10RB contributed to enrichment of a number of pathways relating to response to infection (Table 3). Similar to the brain, components of actin cytoskeleton regulation were also dysregulated in females. (Table 3; Supplementary Table S9). Downregulated genes in female AD patients were enriched for a number of metabolism related processes including oxidative phosphorylation and thermogenesis, consistent with the single-gene and network analysis in the brain (Supplementary Table S10).

Similar to the brain analysis, we observed dramatically fewer enriched pathways in males with AD. Among upregulated genes in male AD patients, we did not identify any enriched pathways. Among downregulated genes in male AD patients, components of the proteasome were enriched including PSMD4 and PSMC3 (Table 3; Supplementary Table S11). For a full list of enriched pathways, refer to Supplementary Tables S9-S11.

Lastly, we performed a non-stratified analysis comparing gene expression between AD and control samples irrespective of sex in whole blood. Analyses were adjusted for sex, apoE4 carrier status, age and education. A total of 339 genes were upregulated and 360 genes were downregulated in patients with AD compared to controls (Figure S3B, Supplementary Table S8). Upregulated genes were enriched for several pathways previously implicated in AD, including MAPK signaling, autophagy and NFkB signaling (Supplementary Table S12). In addition, a number of immune related pathways were enriched including tuberculosis, Escherichia coli infection, salmonella infection, and inflammatory bowel disease. Several components of the NFkB cascade and antigen presentation system including NFKBIA, ITGAM, STAT5B, TLR5, TLR4, CD14 and C4A, contributed to this enrichment (Supplementary Table S12). Among downregulated genes, pathways related to protein synthesis and metabolism, including ribosome, proteasome, protein export, thermogenesis, and oxidative phosphorylation were enriched. Included in these pathways were several oxidation phosphorylation related genes including NDUFA9, NDUFA8, COX4I2 (Supplementary Table S13).

**Table 3: Enriched Pathways in Blood**

Term	Adjusted P	Genes
<b>Female Upregulated Genes (n= 294)</b>		
<b>Tuberculosis</b>	<0.001	ATP6V0B;CEBPB;ITGAM;IL10RB;IFNGR2;TCIRG1;CTSS; CREB1;IRAK1;LAMP2;ITGAX;RAF1;CAMK2G
<b>Necroptosis</b>	0.004	PYCARD;STAT5B;MLKL;H2AFJ;IFNGR2;STAT6;TYK2;CF LAR;CAMK2G;HIST1H2AC;HIST2H2AC

<b>Fc gamma R-mediated phagocytosis</b>	0.006	HCK;PTPRC;ARPC1A;PRKCD;RAC2;ASAP1;ARPC5;RAF1
<b>Pathogenic Escherichia coli infection</b>	0.01	ARPC1A;NCK2;ARHGEF2;ARPC5;TLR5;TUBA4A
<b>TNF signaling pathway</b>	0.01	CEBPB;RPS6KA5;CREB1;MLKL;MAP3K8;FOS;CFLAR;CREB5
<b>Regulation of actin cytoskeleton</b>	0.02	FGD3;ITGAM;SPATA13;ARPC1A;RAC2;ITGAX;IQGAP1;ARPC5;RAF1;SSH2;PAK2
<b>Lysosome</b>	0.02	GNPTG;CD63;ATP6V0B;LAMP2;IDS;TCIRG1;GNS;CTSS
<b>Phagosome</b>	0.02	ATP6V0B;ITGAM;LAMP2;CANX;TAP1;TCIRG1;TUBA4A;CTSS;ATP6V1F
<b>JAK-STAT signaling pathway</b>	0.02	STAT5B;CCND3;CSF3R;IL10RB;IFNGR2;STAT6;TYK2;RAF1;MCL1
<b>Estrogen signaling pathway</b>	0.03	CREB1;PRKCD;FOS;KRT10;RAF1;ADCY7;FKBP5;CREB5
4 more..		
<b>Female Downregulated Genes (n= 305)</b>		
<b>Ribosome</b>	<0.001	RPL4;RPL5;RPL30;RPL41;RPL32;RPL12;RPL22;RPL11;RPL35A;MRPL36;MRPL24;RPL6;MRPL33;RPS25;RPL36AL;RPL35;RPL24;RPS20;RPL26;RPS27A;RPL39;RPS24;RPS12
<b>Proteasome</b>	<0.001	PSMB6;PSMA5;PSMB7;PSMA3;PSMD4;PSMC3;PSMC1;POMP;PSMB1;PSMC2;PSMD1;PSMF1
<b>Spliceosome</b>	<0.001	ISY1;HSPA8;SF3B5;CCDC12;BUD31;DDX42;PLRG1;PQB1;SNRPD2;ZMAT2;SYF2;SNRPG;PPIH;SNRPA1;SNRPB2;SLU7;CTNBL1
<b>Protein export</b>	<0.001	SRP19;SEC61G;SRPRB;SRP68;SRP14;SEC11A
<b>Oxidative phosphorylation</b>	<0.001	NDUFA9;NDUFA8;NDUFS5;COX17;NDUFB2;NDUFA1;COX6A1;ATP6V1E1;NDUFV2;COX6C;ATP6V1D;UQCRH
<b>Huntington disease</b>	<0.001	NDUFA9;NDUFA8;NDUFB2;NDUFA1;CLTA;COX6C;COX6A1;UQCRH;SOD1;SIN3A;NDUFS5;VDAC3;BAX;NDUFV2
<b>Non-alcoholic fatty liver disease (NAFLD)</b>	<0.001	NDUFA9;NDUFA8;NDUFS5;NDUFB2;NDUFA1;BAX;PIK3R1;COX6A1;NDUFV2;COX6C;ADIPOR2;UQCRH
<b>Protein processing in endoplasmic reticulum</b>	0.002	DNAJA1;ATXN3;HSPA8;HSP90AA1;HSPH1;HSP90AB1;EIF2AK1;SEC61G;ERP29;BAX;UBXN6
<b>Parkinson disease</b>	0.002	NDUFA9;NDUFA8;NDUFS5;VDAC3;NDUFB2;NDUFA1;COX6A1;NDUFV2;COX6C;UQCRH
<b>Thermogenesis</b>	0.007	NDUFA9;COA3;NDUFA8;SMARCC1;NDUFS5;COX17;NDUFB2;NDUFA1;COX6C;COX6A1;NDUFV2;UQCRH
3 more...		
<b>Male Upregulated Genes (n=38)</b>		
<b>No enriched pathways</b>		
<b>Male Downregulated Genes (n=50)</b>		
<b>Proteasome</b>	0.06	PSMD4;PSMC3;POMP

## Network Analysis in Whole Blood Identifies a Stronger Disease Signature in Females

We identified five AD-associated modules in females and zero AD-associated modules in males (Figure 4) that met the significance threshold ( $FDR < 0.05$ ) and were either positively or negatively correlated with case/control status. Among the modules in female samples, three modules including a 483-gene module (termed turquoise), a 129-gene module (termed pink) and 153-gene module (termed black) were upregulated in AD. Two modules including a 270-gene module (termed blue) and 119-gene module (termed magenta) were downregulated in AD (Figure 4A). No modules with significant apoE4:disease interaction effect were found in female or male network analyses from the blood datasets.

Enrichment analysis of disease-associated modules using the 2019 KEGG Human pathway database revealed pathways relevant to AD that were consistent with those identified in the single gene analysis (Figures 4A and 3A). For example, upregulated modules in females were strongly enriched for innate immune system activity, neutrophil degranulation, CSF signaling, IL2 signaling, and cytokine signaling. Consistent with single gene analyses, downregulated modules in females were enriched for metabolic processes including metabolism of RNA and metabolism of amino acids (Figure 4A).

There were 35 hub genes among disease associated modules in the female-specific network identified as module membership greater than 0.8, gene significance greater than 0.2 and differentially expressed between AD and controls (Figure 4B). In contrast, zero hub genes were identified in the male-specific gene network. Protein-protein interaction maps generated by STRING v11 suggest several interconnected genes including the B cell development related protein, IGLL1, and ribosomal proteins RPS20, RPS25, RPL4, and RPL35A (Figure 4B).



For a full list of genes in each module, including hub genes, please refer to Supplementary Table S14).

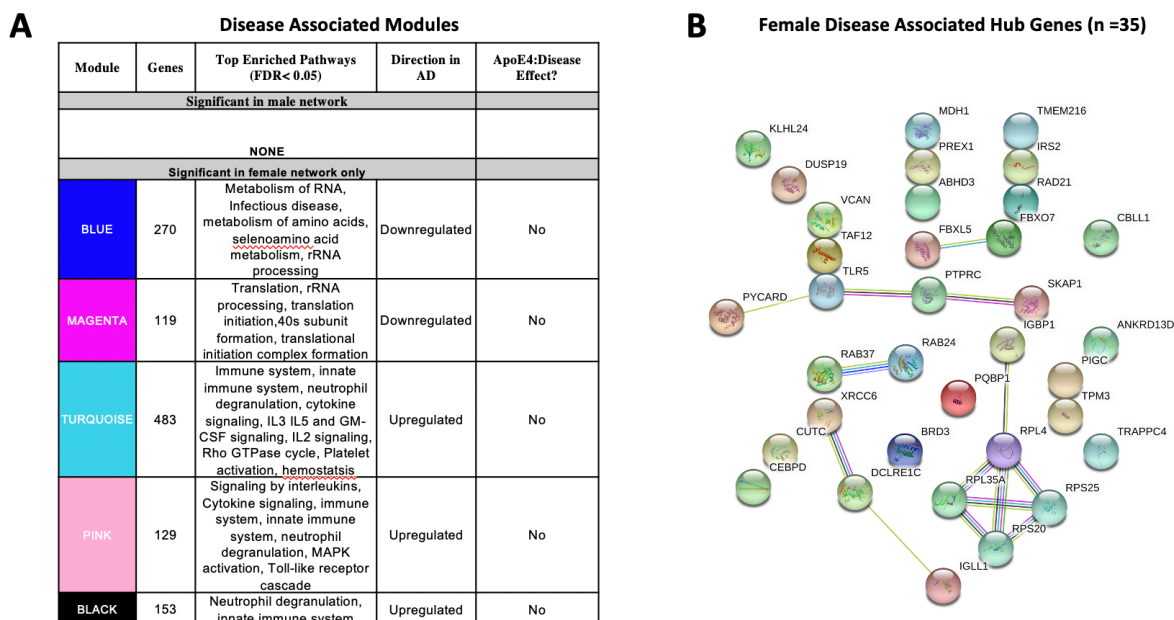


Figure 4. Network Analysis in Whole Blood.

## Comparison of Brain and Blood Transcriptomic Signatures Reveals Common Immune Related Signals in Females

We next identified genes that were commonly dysregulated in both blood and brain (Figure 2E). In females, a total of 23 genes were dysregulated in the brain and blood in the same direction (two downregulated and 21 upregulated). Several genes among the commonly upregulated genes have roles in antigen presentation including TAP1, CTSS, and PTPRC. Enrichment analysis of commonly upregulated genes revealed an enrichment of the KEGG terms primary immunodeficiency, phagosome, and cell adhesion molecules (adjusted  $P < 0.1$ ; Figure 2F). In addition, eight genes were dysregulated but in different directions in the brain and blood including PRKCD, VAMP8, GIMAP7, LAPTM5, HLA-DOA, TNS1, DBI, GIMAP7, TUBA4A

(Figure 2E). In contrast, in males we found one upregulated gene, VCAN encoding vesican, dysregulated in both the blood and brain (Figure 2E).

### **Cell-type Deconvolution Identifies Sex-specific Immune Cell Dysregulation in Females with AD**

Differences in 22 immune blood cell types (Figures 5A-B) were evaluated by deconvolving the transcriptomic signature obtained via meta-analysis of blood studies. Analysis of cell type proportions adjusting for age, sex, and apoE4 status revealed an increase in neutrophils and naïve B cells, and a decrease in M2 macrophages and CD8<sup>+</sup> T cells in AD patients compared to controls in pooled male and female samples (Figure 5C, FDR  $P < 0.05$ ). Among females with AD, relative to controls, we observed an increase in neutrophils and naïve B cells and a decrease in M2 macrophages, memory B cells, and CD8<sup>+</sup> T cells in AD samples (Figure 5C, FDR  $P < 0.05$ ). Interestingly, among males with AD, we did not find any significant differences in immune cell proportions compared to controls.

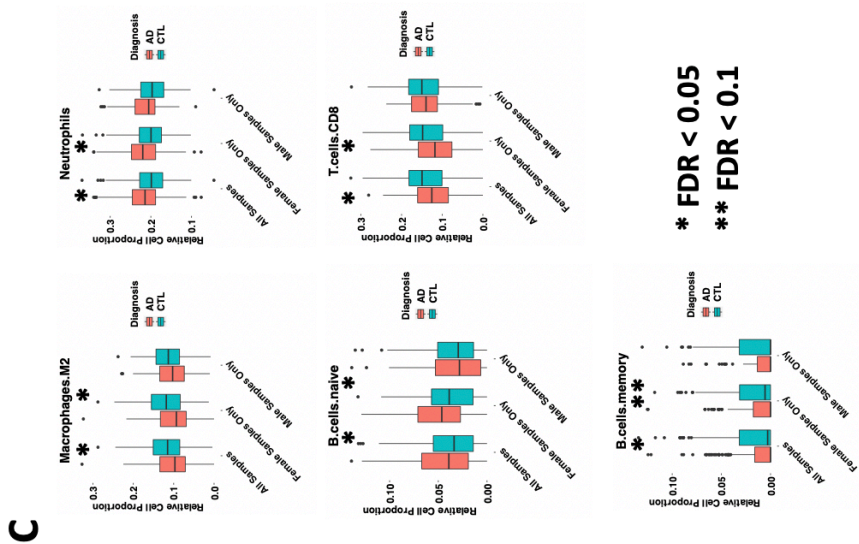
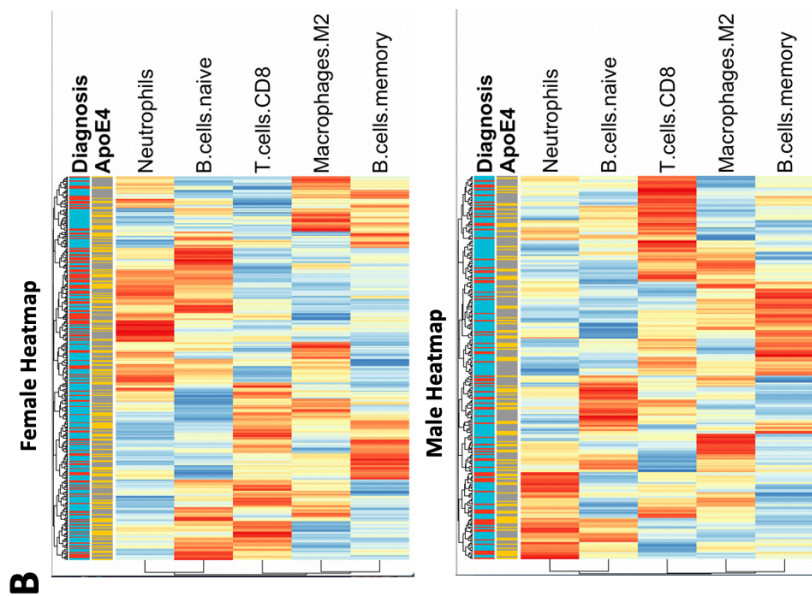
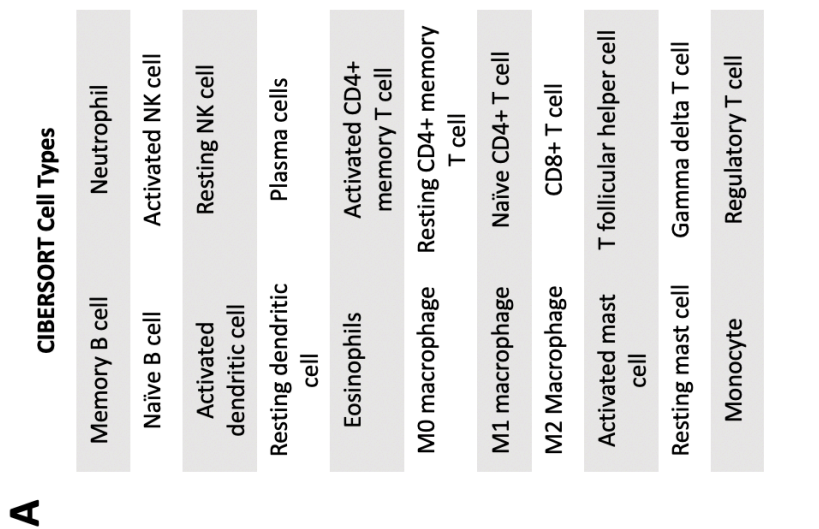


Figure 5. Cell Type Analysis in Whole Blood.

## Sex-specific Transcriptomic Data Improves AD Classification Accuracy

To assess the value of sex-specific transcriptomic data in developing a blood-based classifier in AD, we trained a linear SVM model to classify AD patients controls using the transcriptomic signature obtained via meta-analysis of blood studies. We trained a ‘clinical model’ with age, sex, education, and apoE4 status and a ‘clinical + molecular model’ with age, sex, education, apoE4 status, and blood transcriptomic data. Using pooled male and female samples, the ‘clinical + molecular model’ achieved a higher AUROC compared to the ‘clinical model’ (AUROC = 0.88 for ‘clinical + molecular model’; AUROC = 0.77 for ‘clinical model’) on a test set composed of 25% of samples (Figures 6A and S4A).

Interestingly, a model trained with only female data achieved a higher AUROC (‘clinical + molecular model’: 0.90 and ‘clinical model’: 0.86; Figures 6B and S4B) than the pooled male and female model. In contrast, a model trained with only male data obtained a lower AUROC (‘clinical + molecular’ model 0.81 and ‘clinical model’ 0.83; Figures 6C and S4C) than the pooled male and female model.

Figures 6G-H summarizes shared features between models. In all simple models (pooled male and female, female only, and male only), age and apoE4 status had a positive feature importance while education had a negative feature importance. A positive feature importance means that the expression of that feature increases the likelihood of being classified as AD (termed risk factor). A negative feature importance means that expression of the feature expression reduces the likelihood of being classified as AD (termed protective factor). In the female ‘clinical + molecular model’, 57 features, including known risk factors including apoE4 and age, had a positive feature importance (Supplementary Table S15). In addition, 50 features had negative feature importance. Among these were education and previously implicated AD risk

genes including CETN2 (Supplementary Table S15). In the male ‘clinical + molecular model’, 103 features, including apoE4, had positive feature importance. (Supplementary Table S16). In addition, 105 features, including education, had negative feature importance (Supplementary Table S16).

Altogether, we observed a significant overlap ( $P < 0.001$ , hypergeometric test) in features with non-zero feature importance between the pooled male and female ‘clinical + molecular model’ and female ‘clinical + molecular model’; female ‘clinical + molecular model’ and male ‘clinical + molecular model’; and pooled male and female ‘clinical + molecular model’ and male ‘clinical + molecular model’ (Figure 6G).

Functional annotation of features with a non-zero feature importance was performed via enrichment analysis using the 2019 KEGG database of human pathways. Among features with non-zero feature importance, we did not identify any enriched biological pathways in the male only and female only complex models. In the male and female pooled complex model, features with positive feature importance (risk factors), were enriched for staphylococcus aureus infection, graft-vs-host disease, and antigen presentation and processing KEGG pathways (adjusted  $P < 0.05$ ; Figure 6H). The HLA genes HLA-DRB4 and HLA-DQA1 contributed to this enrichment. In addition, the P-selection glycoprotein ligand-1 gene (SELPLG) and killer cell immunoglobulin-like receptor (KIR2DL3) also contributed to enrichment, suggesting a role for leukocyte recruitment and natural killer cell activity in AD pathology.

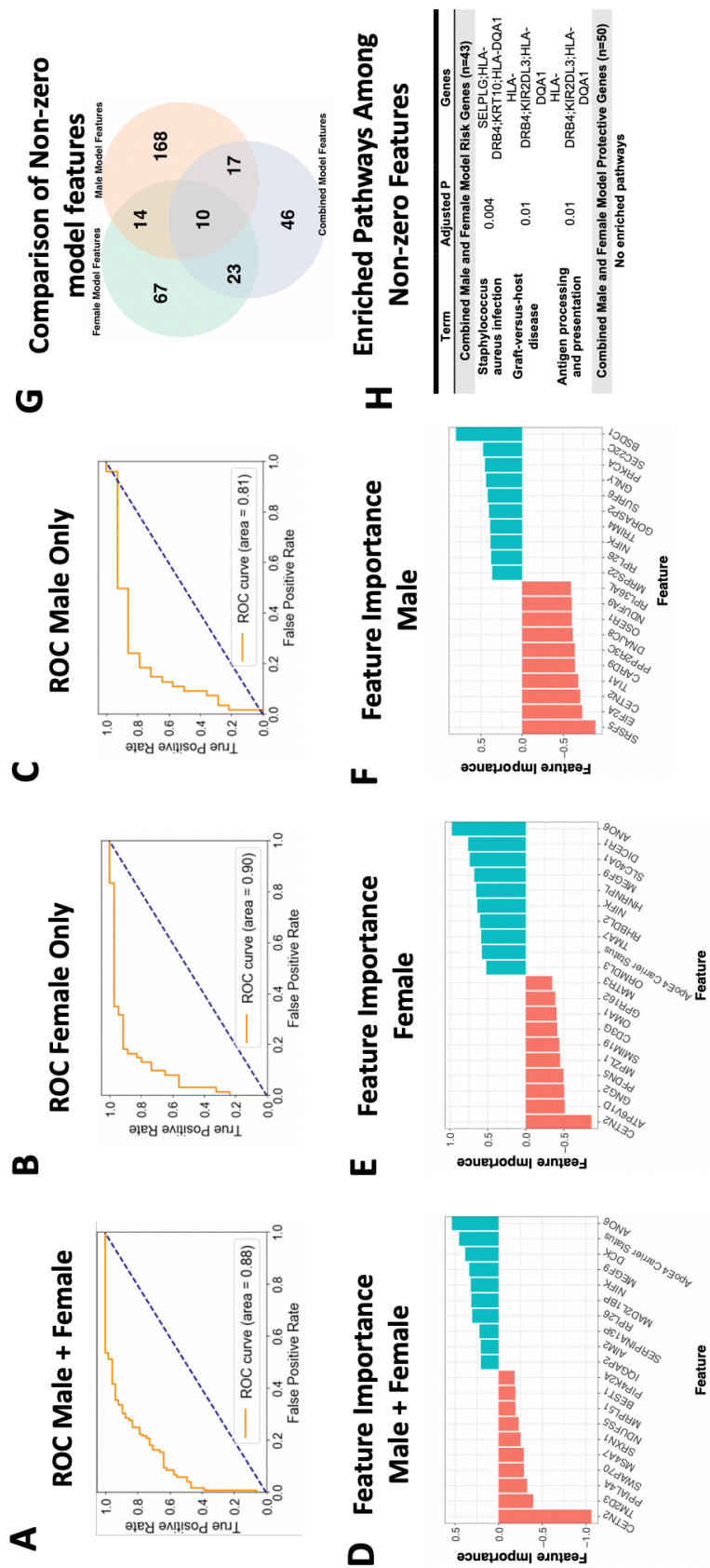


Figure 6. Linear SVM Molecular Model on Whole Blood Data.



## DISCUSSION

In this study, through computational analysis of publicly available gene expression datasets from brain and blood samples, we evaluated AD at the transcriptome level using single gene and network approaches to gain insight into the mechanisms underlying sex and apoE4-genotype based differences in AD. We also evaluated how including sex-specific transcriptomic data from blood samples with clinical data would affect the performance of a machine learning classifier for AD diagnostics.

Our characterization of brain transcriptomic signatures revealed, among upregulated genes in the brains of both females and males with AD, an enrichment of pathways related to components of the innate and adaptive immune systems as well as the MAPK signaling pathway. This result is consistent with past findings where the brain's immune system has been indicated as a major component of AD pathogenesis<sup>60,61</sup>. Additionally, MAPKs, enzymes that play critical roles in cellular signaling, have also been implicated as accelerators of AD development<sup>62</sup>. Overall, findings from our brain transcriptome analysis provide supporting evidence for therapeutics currently being explored for AD, such as p38 MAPK inhibitors<sup>67</sup>, and suggest that possible treatments targeting the MAPK pathway may have a greater effect in females with AD.

Interestingly, from our differential expression analysis, we found a nearly two-fold greater total number of dysregulated genes in the brain transcriptome that met our significance cutoff for females with AD compared to males with AD (974 vs 509, respectively). Many of these genes are in pathways related to antigen presentation and processing, complement activation, suggesting a female-specific role of neuroinflammation in the pathogenesis of AD. Additionally, for downregulated genes in AD patients, we observed enrichment of neurological signaling pathways in females only and no enriched pathways in males.

Through network analysis, we identified more AD-associated modules in the brain transcriptome of females than males. Enrichment analysis of AD-associated modules also revealed some pathways that were enriched in both sexes, including an upregulated module for a PI3/Akt signaling related pathway and downregulated modules for oxidative phosphorylation and thermogenesis related pathways. Unique to females, we observed upregulated modules associated with cell structural processes (adherens junctions, actin cytoskeleton and axonal guidance) and HSV infection-related zinc finger nuclease genes, as well as a downregulated module for neurological signaling pathways, autophagy and proteolysis.

Upon performing hub gene analysis, we identified hub genes in female disease-associated modules but were unable to identify male disease associated hub genes. These female hub genes consisted of several potentially interconnected genes including ITPKB, PDGFRB, GNG12, and GNA12. In our subsequent analysis to assess an apoE4:disease interaction effect, we identified three modules, one of which was significantly enriched for HSV infection-related zinc finger nuclease genes as well as containing the ITPKB hub gene as a highly connected regulator. These results suggest zinc finger nucleases as a potential mechanism underlying sex-associated differential penetrance of apoE4 in AD.

Our findings suggest a neuroinflammatory model of AD pathogenesis in females with dysregulation in components of the adaptive and innate immune system including antigen presentation and processing and complement activation and genes including MAPK and ITPKB. It has been postulated that accumulation of damage from HSV infection and major neuroinflammatory effects can lead to the development of AD, and that apoE4 carriers suffer either greater viral damage or have poorer repair of such damage<sup>54</sup>. Previous studies have demonstrated that ITPKB expression is increased in human AD brains and exacerbates AD



pathology in an animal model<sup>64</sup>. Our brain transcriptome findings for females with AD, including downregulation of autophagy and proteolysis pathways, upregulation of pathways related to the immune system and HSV infection, as well as ITPKB as a hub gene, particularly in female apoE4 carriers, highlight specific gene-encoded processes in the brain that may be more involved in AD for women than for men.

Similar to our brain findings, in analysis of blood transcriptomes, we observed more dysregulated genes in the blood of females with AD than in males with AD. Further characterization of these transcriptomic signatures revealed, among upregulated genes, enrichment in only females with AD of pathways related to components of the innate and adaptive immune systems as well as actin cytoskeleton regulation; however, for downregulated AD genes, we observed enriched metabolic pathways (oxidative phosphorylation and thermogenesis) in females and enriched pathways for protein homeostasis in males.

Through network analysis, we identified AD-associated modules and hub genes in the female blood transcriptome but not in males. In the blood of females with AD, upregulated modules were strongly enriched for innate immune system activity (neutrophil degranulation, CSF signaling, IL2 signaling, and cytokine signaling). Consistent with single gene analyses, female downregulated modules were enriched for metabolic processes (e.g. metabolism of RNA and amino acids). Hub genes identified in the blood of females with AD include those related to immunity (the B cell development related protein, IGLL1) and viral RNA translation (ribosomal proteins RPS20, RPS25, RPL4, and RPL35A).

In addition to neuroinflammation's role in AD, dysregulation of the immune system outside of the brain has also been noted to be a factor in AD<sup>65</sup>. Our findings feature specific gene-encoded processes in peripheral blood cells that may be more involved in AD for women than for

men. Furthermore, our cell-type deconvolution analysis revealed dysregulation of peripheral immune cells uniquely in females with AD and not males with AD.

When including blood transcriptomic features with clinical features (age, sex, education, and apoE4 status) to train a machine learning prediction model of AD, our model performed better with these additional molecular features than without (AUROC: 0.88 vs 0.77, respectively). The performance of this model also improved when trained with only female data (clinical + molecular model AUROC: 0.90 and clinical model AUROC: 0.86) and worsened when trained with only male data (molecular model AUROC: 0.81 and clinical model AUROC: 0.83) than with pooled male and female model. This finding suggests that the molecular changes in females compared to males are better able to model AD-related changes. Further, given the distinct transcriptomic signature observed in males and females, stratifying by sex may aid future efforts to identify biomarkers in AD.

Diagnostic tests currently available for AD, including  $A\beta$  position emission tomography (PET), lack accuracy or are implemented through invasive and painful procedures such as lumbar puncture<sup>56-59</sup>. Diagnostic tests for AD that are more accurate and less invasive are worthwhile for preventing undue uncertainty and physical discomfort experienced by patients. Our machine learning AD prediction model based on clinical and blood transcriptomic features has the potential to complement currently available clinical AD diagnostic tests, and improve the accuracy of these tests, particularly for women, with minimal additional discomfort for patients.

Based on the nature of our analyses, there are a number of limitations to note. We analyzed publicly available datasets, which were limited in sample size and contained annotation differences. This provided challenges in selecting cases from controls and restricted our ability to answer certain questions. For instance, the Allen Brain Atlas dataset provided only a binary

classification for apoE (apoE4: Y/N). This confined our analysis to only look at the presence of apoE4, instead of looking at difference across different genotype combinations. Next, we did not stratify our analysis by age or disease stage, so we cannot describe whether these transcriptomic signatures differ with age or disease severity. Additionally, since we aggregated bulk tissue from different brain regions in our analysis, we cannot infer sex differences across brain region. Consequently, using bulk tissue transcriptomics reduces our resolution of the more complex interactions and contributions of different brain cell types in AD. Future approaches to better characterize sex-specific changes in AD would involve stratification by brain regions, age and disease stage, apoE genotype, as well as an analysis of single cell AD datasets.

In conclusion, the major finding of this study is a distinct, sex-specific transcriptomic signature in the brains and whole blood of patients with AD. Gene expression meta-analysis and network-based analyses revealed an immune signature in the brains and whole blood of females with AD that was absent in males. Our analyses also revealed more pronounced neurosignaling and metabolism signatures in the brains whole blood of females with AD than in males with AD. Stratification by sex improved machine-learned based classification of AD using whole-blood transcriptomic data. Results from this work will help to better understand molecular etiologies underlying sex differences in AD and pave the way for sex-specific biomarker and therapeutic development in AD.

## References

1. Alzheimer, Association. 2018 Alzheimer's disease facts and figure. *Alzheimer's Dement.* 2018;14(3):367-429. doi:10.1016/J.JALZ.2018.02.001
2. Bekris LM, Yu CE, Bird TD, Tsuang DW. Review article: Genetics of Alzheimer disease. *J Geriatr Psychiatry Neurol.* 2010;23(4):213-227. doi:10.1177/0891988710383571

3. Karch CM, Goate AM. Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol Psychiatry*. 2015;77(1):43-51. doi:10.1016/j.biopsych.2014.05.006
4. Hardy J, Selkoe DJ. The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science*. 2002;297(5580):353-356. doi:10.1126/science.1072994
5. Holtzman DM, John CM, Goate A. Alzheimer's Disease: The Challenge of the Second Century. *Sci Transl Med*. 2011;3(77):77sr1. doi:10.1126/scitranslmed.3002369
6. Medeiros ADM, Silva RH. Sex Differences in Alzheimer's Disease: Where Do We Stand? *J Alzheimers Dis*. 2019;67(1):35-60. doi:10.3233/JAD-180213
7. Torrão AS, Café-Mendes CC, Real CC, et al. Different Approaches, One Target: Understanding Cellular Mechanisms of Parkinson's and Alzheimer's Diseases. *Rev Bras Psiquiatr*. 2012;34:194-218. doi:10.1016/j.rbp.2012.08.004
8. Riedel BC, Thompson PM, Brinton RD. Age, APOE and sex: Triad of risk of Alzheimer's disease. *J Steroid Biochem Mol Biol*. 2016;160:134-147. doi:10.1016/j.jsbmb.2016.03.012
9. Hebert LE, Weuve J, Scherr PA, Evans DA. Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. *Neurology*. 2013;80(19):1778-1783. doi:10.1212/WNL.0b013e31828726f5
10. Goedert M, Spillantini MG. A century of Alzheimer's disease. *Science*. 2006;314(5800):777-781. doi:10.1126/science.1132814
11. Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: Few candidates, frequent failures. *Alzheimers Res Ther*. 2014;6(4). doi:10.1186/alzrt269
12. World Population Prospects 2019: Highlights | Multimedia Library - United Nations Department of Economic and Social Affairs. <https://www.un.org/development/desa/publications/world-population-prospects-2019-highlights.html>. Accessed November 12, 2019.
13. Guerreiro RJ, Gustafson DR, Hardy J. The genetic architecture of Alzheimer's disease: beyond APP, PSENs and APOE. *Neurobiol Aging*. 2012;33(3):437-456. doi:10.1016/j.neurobiolaging.2010.03.025
14. Holtzman DM, Herz J, Bu G. Apolipoprotein E and apolipoprotein E receptors: Normal biology and roles in Alzheimer disease. *Cold Spring Harb Perspect Med*. 2012;2(3). doi:10.1101/cshperspect.a006312
15. Mahley RW, Weisgraber KH, Huang Y. Apolipoprotein E4: A causative factor and therapeutic target in neuropathology, including Alzheimer's disease. *Proc Natl Acad Sci U S A*. 2006;103(15):5644-5651. doi:10.1073/pnas.0600549103

16. Paranjpe MD, Chen X, Liu M, et al. The effect of ApoE  $\epsilon$ 4 on longitudinal brain region-specific glucose metabolism in patients with mild cognitive impairment: a FDG-PET study. *NeuroImage Clin*. 2019. doi:10.1016/j.nicl.2019.101795
17. Liu M, Paranjpe MD, Zhou X, et al. Sex modulates the ApoE  $\epsilon$ 4 effect on brain tau deposition measured by 18F-AV-1451 PET in individuals with mild cognitive impairment. *Theranostics*. 2019;9(17):4959-4970. doi:10.7150/thno.35366
18. Leduc V, Jasmin-Bélanger S, Poirier J. APOE and cholesterol homeostasis in Alzheimer's disease. *Trends Mol Med*. 2010;16(10):469-477. doi:10.1016/j.molmed.2010.07.008
19. Bu G. Apolipoprotein e and its receptors in Alzheimer's disease: Pathways, pathogenesis and therapy. *Nat Rev Neurosci*. 2009;10(5):333-344. doi:10.1038/nrn2620
20. Saunders AM, Strittmatter WJ, Schmechel D, et al. Association of apolipoprotein E allele  $\epsilon$ 4 with late-onset familial and sporadic Alzheimer's disease. :7.
21. Corder EH, Saunders AM, Risch NJ, et al. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet*. 1994;7(2):180-184. doi:10.1038/ng0694-180
22. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*. 1993;261(5123):921-923. doi:10.1126/science.8346443
23. Andersen K, Launer LJ, Dewey ME, et al. Gender differences in the incidence of AD and vascular dementia: The EURODEM Studies. EURODEM Incidence Research Group. *Neurology*. 1999;53(9):1992-1997. doi:10.1212/wnl.53.9.1992
24. Irvine K, Laws KR, Gale TM, Kondel TK. Greater cognitive deterioration in women than men with Alzheimer's disease: a meta analysis. *J Clin Exp Neuropsychol*. 2012;34(9):989-998. doi:10.1080/13803395.2012.712676
25. Barnes LL, Wilson RS, Bienias JL, Schneider JA, Evans DA, Bennett DA. Sex differences in the clinical manifestations of Alzheimer disease pathology. *Arch Gen Psychiatry*. 2005;62(6):685-691. doi:10.1001/archpsyc.62.6.685
26. Carter CL, Resnick EM, Mallampalli M, Kalbarczyk A. Sex and Gender Differences in Alzheimer's Disease: Recommendations for Future Research. *J Womens Health*. 2012;21(10):1018-1023. doi:10.1089/jwh.2012.3789
27. Altmann A, Tian L, Henderson VW, Greicius MD. Sex modifies the APOE-related risk of developing Alzheimer disease. *Ann Neurol*. 2014;75(4):563-573. doi:10.1002/ana.24135
28. Bretsky PM, Buckwalter JG, Seeman TE, et al. Evidence for an interaction between apolipoprotein E genotype, gender, and Alzheimer disease. *Alzheimer Dis Assoc Disord*. 1999;13(4):216-221. doi:10.1097/00002093-199910000-00007

29. Ungar L, Altmann A, Greicius MD. Apolipoprotein E, gender, and Alzheimer's disease: An overlooked, but potent and promising interaction. *Brain Imaging Behav.* 2014;8(2):262-273. doi:10.1007/s11682-013-9272-x
30. Farrer LA. Effects of Age, Sex, and Ethnicity on the Association Between Apolipoprotein E Genotype and Alzheimer Disease. *JAMA.* 1997;278(16):1349-1349. doi:10.1001/jama.1997.03550160069041
31. Paranjpe MD, Taubes A, Sirota M. Insights into Computational Drug Repurposing for Neurodegenerative Disease. *Trends Pharmacol Sci.* 2019. doi:10.1016/j.tips.2019.06.003
32. Huang L-K, Chao S-P, Hu C-J. Clinical trials of new drugs for Alzheimer disease. *J Biomed Sci.* 2020;27(1):18. doi:10.1186/s12929-019-0609-7
33. Cummings J, Lee G, Ritter A, Sabbagh M, Zhong K. Alzheimer's disease drug development pipeline: 2019. *Alzheimers Dement Transl Res Clin Interv.* 2019;5(1):272-293. doi:10.1016/j.trci.2019.05.008
34. Cummings J, Feldman HH, Scheltens P. The "rights" of precision drug development for Alzheimer's disease. *Alzheimers Res Ther.* 2019;11(1):76. doi:10.1186/s13195-019-0529-5
35. Wang M, Beckmann ND, Roussos P, et al. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci Data.* 2018;5(1):1-16. doi:10.1038/sdata.2018.185
36. Allen M, Carrasquillo MM, Funk C, et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci Data.* 2016;3(1):1-10. doi:10.1038/sdata.2016.89
37. Bennett DA, Schneider JA, Buchman AS, Barnes LL, Boyle PA, Wilson RS. Overview and Findings from the Rush Memory and Aging Project. *Curr Alzheimer Res.* 2012;9(6):646-663.
38. Hodes RJ, Buckholtz N. Accelerating Medicines Partnership: Alzheimer's Disease (AMP-AD) Knowledge Portal Aids Alzheimer's Drug Discovery through Open Data Sharing. *Expert Opin Ther Targets.* 2015. doi:10.1517/14728222.2016.1135132
39. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635
40. Picard. Picard. <http://broadinstitute.github.io/picard/>. Accessed January 15, 2019.
41. Download. Aging, Dementia and TBI Study. <https://aging.brain-map.org/>. Accessed January 15, 2019.
42. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics.* 2012;28(6):882-883. doi:10.1093/bioinformatics/bts034

43. Hodes RJ, Buckholtz N. Accelerating Medicines Partnership: Alzheimer's Disease (AMP-AD) Knowledge Portal Aids Alzheimer's Drug Discovery through Open Data Sharing. *Expert Opin Ther Targets*. 2016;20(4):389-391. doi:10.1517/14728222.2016.1135132
44. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology*. 2010;74(3):201-209. doi:10.1212/WNL.0b013e3181cb3e25
45. Zhang Z. Missing data imputation: focusing on single imputation. *Ann Transl Med*. 2016;4(1):9. doi:10.3978/j.issn.2305-5839.2015.12.38
46. Greenfield B, Kegerreis S, Walsworth M, et al. limma. *R-Package*. 2010;40(10):668-673.
47. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28(1):27-30.
48. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559. doi:10.1186/1471-2105-9-559
49. Langfelder P, Luo R, Oldham MC, Horvath S. Is My Network Module Preserved and Reproducible? Bourne PE, ed. *PLoS Comput Biol*. 2011;7(1):e1001057. doi:10.1371/journal.pcbi.1001057
50. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci*. 2019;28(11):1947-1951. doi:10.1002/pro.3715
51. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res*. 2019;47(Database issue):D590-D595. doi:10.1093/nar/gky962
52. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(Database issue):D607-D613. doi:10.1093/nar/gky1131
53. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453-457. doi:10.1038/nmeth.3337
54. Itzhaki RF. Corroboration of a Major Role for Herpes Simplex Virus Type 1 in Alzheimer's Disease. *Front Aging Neurosci*. 2018;10. doi:10.3389/fnagi.2018.00324
55. Congdon EE. Sex Differences in Autophagy Contribute to Female Vulnerability in Alzheimer's Disease. *Front Neurosci*. 2018;12. doi:10.3389/fnins.2018.00372
56. Elahi FM, Miller BL. A clinicopathological approach to the diagnosis of dementia. *Nat Rev Neurol*. 2017;13(8):457-476. doi:10.1038/nrneurol.2017.96



57. Ewers M, Mattsson N, Minthon L, et al. CSF biomarkers for the differential diagnosis of Alzheimer's disease: A large-scale international multicenter study. *Alzheimers Dement.* 2015;11(11):1306-1315. doi:10.1016/j.jalz.2014.12.006
58. Ritchie C, Smailagic N, Noel-Storr AH, Ukoumunne O, Ladds EC, Martin S. CSF tau and the CSF tau/ABeta ratio for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev.* 2017;3:CD010803. doi:10.1002/14651858.CD010803.pub2
59. Bergeron D, Ossenkoppele R, Jr Laforce R. Evidence-based Interpretation of Amyloid- $\beta$  PET Results: A Clinician's Tool. *Alzheimer Dis Assoc Disord.* 2018;32(1):28-34. doi:10.1097/WAD.0000000000000239
60. Frost GR, Jonas LA, Li Y-M. Friend, Foe or Both? Immune Activity in Alzheimer's Disease. *Front Aging Neurosci.* 2019;11. doi:10.3389/fnagi.2019.00337
61. Heneka MT, Golenbock DT, Latz E. Innate immunity in Alzheimer's disease. *Nat Immunol.* 2015;16(3):229-236. doi:10.1038/ni.3102
62. Du Y, Du Y, Zhang Y, et al. MKP-1 reduces A $\beta$  generation and alleviates cognitive impairments in Alzheimer's disease models. *Signal Transduct Target Ther.* 2019;4(1):1-12. doi:10.1038/s41392-019-0091-4
63. Heppner FL, Ransohoff RM, Becher B. Immune attack: the role of inflammation in Alzheimer disease. *Nat Rev Neurosci.* 2015;16(6):358-372. doi:10.1038/nrn3880
64. Stygelbout V, Leroy K, Pouillon V, et al. Inositol trisphosphate 3-kinase B is increased in human Alzheimer brain and exacerbates mouse Alzheimer pathology. *Brain.* 2014;137(2):537-552. doi:10.1093/brain/awt344
65. Cao W, Zheng H. Peripheral immune system in aging and Alzheimer's disease. *Mol Neurodegener.* 2018;13(1):51. doi:10.1186/s13024-018-0284-2
66. Peskind E, Nordberg A, Darreh-Shori T, Soininen H. Safety of Lumbar Puncture Procedures in Patients with Alzheimer's Disease. *Curr Alzheimer Res.* 2009;6(3):290-292.
67. Lee JK, Kim N-J. Recent Advances in the Inhibition of p38 MAPK as a Potential Strategy for the Treatment of Alzheimer's Disease. *Molecules.* 2017;22(8):1287. doi:10.3390/molecules22081287



### **Acknowledgements:**

This work was in part supported by NIA 1R0AG057683 and R01AG060393. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

### **Author Contributions:**

MDP, MS conceived and designed the study. MDP, AT, and BG extracted data from publicly available data sources used in the study. MDP, JKW, IP, AG completed all analyses and produced figures. MDP, MS, TO, and SB wrote the manuscript. MDP, MS, TO, SB, AT, KZ, BG, and YH critically revised the manuscript for important intellectual content. All authors commented and approved the manuscript.

### **Competing Interests:**

Y.H. is a co-founder and scientific advisory board member of E-Scape Bio, Inc. and GABAeron, Inc. M.S. is on the scientific advisory board for twoXAR. The other authors declare no competing interests.

## **Figure Legends**

### **Figure 1: Meta-analysis Overview**

Diagram depicting the study overview including all datasets used and analyses performed. Data sets were obtained via searching GEO or PubMed for the keyword Alzheimer's Disease. Samples with neurological conditions other Alzheimer's, including Parkinson's Disease and Huntington's Disease, and single cell preparations were excluded from analysis. Datasets were merged using the ComBat package in R. WGCNA was used for network analyses. CIBERSORT was used for cell type deconvolution. The linear SVM was trained to classify AD and control patients using the transcriptomic signature obtained via meta-analysis of blood studies. The performance of a molecular model consisting of gene expression, age, sex and apoE4 status was compared to that clinical model with age, sex and apoE4 status as features.

### **Figure 2: Cross-Tissue Sex Specific Differential Gene Expression**

**A.** Four-way plot with fold change in males vs fold change in females depicting differentially expressed genes in the brain. Differential expression was defined using a fold change  $> 1.2$  and FDR  $P < 0.05$ . Covariates of age and sex were included in statistical analyses. **B.** In the brain, a total of 631 genes were uniquely dysregulated in females with AD while 166 genes were uniquely dysregulated in males with AD. Common to both males and females in the brain were 343 genes. **C.** Four-way plot with fold change in males vs fold change in females depicting differentially expressed genes in the blood. Differential expression was defined using a fold change  $> 1.2$ . Covariates of age, sex, and education were included in statistical analyses **D.** A total of 542 genes were uniquely dysregulated in females with AD while 31 genes were uniquely dysregulated in males with AD in blood. Common to both males and females in the brain were 55 genes. **E** Fold change plot depicting genes that are dysregulated in both blood and brain tissues. Genes are colored by sex indicating if the gene is dysregulated in male samples (1 gene; red) or female samples (31 genes; blue). **F.** Enrichment analysis of the commonly dysregulated genes depicted in **E**. An adjusted  $P$ -value cutoff of 0.1 was used for significance to increase power.

### **Figure 3: Network Analysis in the Brain.**

WGCNA was used to construct gene network separately for males and females in the brain. Networks were randomly assigned colors. **A.** A description of the disease-associated gene networks (termed modules) produced using WGCNA. Significant disease-associated modules were identified by associating module eigengene to case/control status adjusting for age and apoE4 status ( $P < 0.05$ ). KEGG enrichment analysis of significant was conducted using an adjusted  $P$  value threshold of 0.05. The direction in AD is computed using the case/control coefficient of the model associating module eigengene to case/control status. Modules with significant apoE4:disease interaction effect were identified by adding the interaction term apoE4:disease to the previous model ( $P < 0.05$ ) **B.** Heatmap depicting the degree of module overlap assessed using a hypergeometric test between male and female disease-associated modules. The black module in males had significant overlap ( $P < 0.05$ ) with the pink and yellow modules. Estimate refers to the case/control coefficient in the model module eigengene  $\sim$  age + apoE4 + case/control status. **C.** Hub genes from female disease-associated modules. Hub genes were defined as genes with gene significance (the correlation between the gene expression and case/control status) greater than 0.2 and module membership (the correlation between gene expression and module eigengene) greater than 0.8. Hub genes were restricted to those that were

differentially expressed in AD vs control. Protein-protein interactions between hub gene visualization was performed using the STRING v11 database. Edge color represents the type of interaction evidence for protein-protein interaction (cyan: known interaction from curated databases; turquoise: experimentally determined; green: gene-neighborhood predicted interaction; red: gene-fusions predicted interaction; blue: gene co-occurrence predicted interaction; green-yellow: text mining; black: co-expression; light purple: protein homology). **D.** Hub genes among modules with significant apoE4:disease interaction effect. Protein-protein interaction between hub genes was visualized using STRING v11 with edge colors representing the same as in C.

#### **Figure 4: Network Analysis in Whole Blood.**

WGCNA was used to construct gene network separately for males and females in whole blood. Networks were randomly assigned colors. **A.** A description of the disease-associated gene networks (termed modules) produced using WGCNA. Significant disease-associated modules were identified by associating module eigengene to case/control status adjusting for age and apoE4 status and education ( $P < 0.05$ ). KEGG enrichment analysis of significant was conducted using an adjusted  $P$  value threshold of 0.05. The direction in AD is computed using the case/control coefficient of the model associating module eigengene to case/control status. Modules with significant apoE4:disease interaction effect were identified by adding the interaction term apoE4:disease to the previous model ( $P < 0.05$ ). **B.** Hub genes from female disease-associated modules. Hub genes were defined as genes with gene significance (the correlation between the gene expression and case/control status) greater than 0.2 and module membership (the correlation between gene expression and module eigengene) greater than 0.8. Hub genes were restricted to those that were differentially expressed in AD vs control. Protein-protein interactions between hub gene visualization was performed using the STRING v11 database. Edge color represents the type of interaction evidence for protein-protein interaction (cyan: known interaction from curated databases; turquoise: experimentally determined; green: gene-neighborhood predicted interaction; red: gene-fusions predicted interaction; blue: gene co-occurrence predicted interaction; green-yellow: text mining; black: co-expression; light purple: protein homology).

#### **Figure 5: Cell Type Analysis in Whole Blood.**

**A.** Cell types included in the panel of 22 reference cell types in CIBERSORT. **B.** Heatmap depicting cell type expression between cases and controls. apoE4 carrier status, sex, and case/control status is annotated for each sample. Only cell types that are significantly different between cases and controls in pooled male and female, male-only or female-only analyses are shown. **C.** Bar charts depicting cell type expression for individual cell types that are significantly different between cases and controls in pooled male and female, male-only or female-only analyses are shown. Significance was assessed by associated cell type proportion to case/control status, adjusting for age, sex (in the pooled male and female model) and apoE4 status.  $P < 0.05$  was deemed significant.

#### **Figure 6: Linear SVM Clinical + Molecular Model in Whole Blood.**

**A-C.** Receiver operating characteristic (ROC) curves depicting performance of each linear SVM model on a test set composed of 25% of samples. Features include gene expression data obtained via meta-analysis, age, sex, education, and apoE4 status. Three models were fit for male and female pooled samples (**A**), female samples only (**B**), and male samples only (**C**).

**D-F.** Feature importance plots for features with non-zero importance in the combined male and female model (**D**), female model (**E**), and male model (**F**). A positive feature importance means that the expression of that feature increases the likelihood of being classified as AD (risk factor). A negative feature importance means that expression of the feature expression reduces the likelihood of being classified as AD (protective factor). **G.** Comparison of non-zero features between combined male and female model, female model and male model.

### **Supplementary Figure Legends:**

#### **Figure S1: Brain Data PCA**

Principal component plots (PCA) of brain samples before (top) and after batch (bottom) correction. PCA plots depict principal component 1 and 2 and are colored by study (left), sex (middle) and apoE4 status (right).

#### **Figure S2: Blood Data PCA**

Principal component plots (PCA) of blood samples before (top) and after batch (bottom) correction. PCA plots depict principal component 1 and 2 and are colored by study (left), sex (middle) and apoE4 status (right).

#### **Figure S3: Non-stratified Differential Gene Expression**

**A.** Volcano plot depicting fold changes and  $p$  values from samples in the brain. Analyses were adjusted for sex, apoE4 status, and age. An adjusted  $P$  value  $< 0.05$  and  $\log FC > 1.2$  was deemed significant. In the brain, a total of 662 genes were upregulated and 430 genes were downregulated in patients with AD compared to controls. **B.** Volcano plot depicting fold changes and  $P$  values from samples in the blood. Analyses were adjusted for sex, apoE4 status, age, and education. An adjusted  $P$  value  $< 0.05$  was deemed significant. In blood, 339 genes were upregulated and 360 genes were downregulated in patients with AD compared to controls.

#### **Figure S4: Linear SVM Clinical Model**

**A-C.** Receiver operating characteristic (ROC) curves depicting performance of each linear SVM model on a test set composed of 25% of samples. Features include age, sex, education, and apoE4 status. Three models were fit for male and female pooled samples (**A**), female samples only (**B**), and male samples only (**C**). **D-F.** Feature importance plots for features with non-zero importance in the combined male and female model (**D**), female model (**E**), and male model (**F**). A positive feature importance means that the expression of that feature increases the likelihood of being classified as AD (risk factor). A negative feature importance means that expression of the feature expression reduces the likelihood of being classified as AD (protective factor).