


RESEARCH ARTICLE

Using blood transcriptome analysis for Alzheimer's disease diagnosis and patient stratification

Huan Zhong^{1,2} | Xiaopu Zhou^{1,2,3} | Hyebin Uhm¹ | Yuanbing Jiang^{1,2} | Han Cao¹ |
 Yu Chen^{1,3,4} | Tiffany T. W. Mak^{1,2} | Ronnie Ming Nok Lo¹ | Bonnie Wing Yan Wong^{1,2} |
 Elaine Yee Ling Cheng^{1,2} | Kin Y. Mok^{1,2,5} | Andrew Lung Tat Chan⁶ |
 Timothy C. Y. Kwok⁷ | Vincent C. T. Mok⁸ | Fanny C. F. Ip^{1,2,3} | John Hardy^{2,5,9} |
 Amy K. Y. Fu^{1,2,3} | Nancy Y. Ip^{1,2,3} 

¹Division of Life Science, State Key Laboratory of Molecular Neuroscience and Molecular Neuroscience Center, The Hong Kong University of Science and Technology, HKSAR, China

²Hong Kong Center for Neurodegenerative Diseases, InnoHK, HKSAR, China

³Guangdong Provincial Key Laboratory of Brain Science, Disease and Drug Development, HKUST Shenzhen Research Institute, Shenzhen, Guangdong, China

⁴The Brain Cognition and Brain Disease Institute, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen–Hong Kong Institute of Brain Science–Shenzhen Fundamental Research Institutions, Shenzhen, Guangdong, China

⁵Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK

⁶Department of Medicine, Queen Elizabeth Hospital, HKSAR, China

⁷Therese Pei Fong Chow Research Centre for Prevention of Dementia, Division of Geriatrics, Department of Medicine and Therapeutics, The Chinese University of Hong Kong, HKSAR, China

⁸Lau Tat-chuen Research Centre of Brain Degenerative Diseases in Chinese, Therese Pei Fong Chow Research Centre for Prevention of Dementia, Gerald Choa Neuroscience Institute, Li Ka Shing Institute of Health Sciences, Division of Neurology, Department of Medicine and Therapeutics, The Chinese University of Hong Kong, HKSAR, China

⁹Institute for Advanced Study, The Hong Kong University of Science and Technology, HKSAR, China

Correspondence

Nancy Y. Ip, Division of Life Science, State Key Laboratory of Molecular Neuroscience and Molecular Neuroscience Center, The Hong Kong University of Science and Technology, HKSAR 999077, China.

Email: boip@ust.hk

Abstract

INTRODUCTION: Blood protein biomarkers demonstrate potential for Alzheimer's disease (AD) diagnosis. Limited studies examine the molecular changes in AD blood cells.

METHODS: Bulk RNA-sequencing of blood cells was performed on AD patients of Chinese descent ($n = 214$ and 26 in the discovery and validation cohorts, respectively) with normal controls ($n = 208$ and 38 in the discovery and validation cohorts, respectively). Weighted gene co-expression network analysis (WGCNA) and deconvolution analysis identified AD-associated gene modules and blood cell types. Regression and unsupervised clustering analysis identified AD-associated genes, gene modules, cell types, and established AD classification models.

RESULTS: WGCNA on differentially expressed genes revealed 15 gene modules, with 6 accurately classifying AD (areas under the receiver operating characteristics curve

Huan Zhong and Xiaopu Zhou contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Authors. *Alzheimer's & Dementia* published by Wiley Periodicals LLC on behalf of Alzheimer's Association.

[auROCs] > 0.90). These modules stratified AD patients into subgroups with distinct disease states. Cell-type deconvolution analysis identified specific blood cell types potentially associated with AD pathogenesis.

DISCUSSION: This study highlights the potential of blood transcriptome for AD diagnosis, patient stratification, and mechanistic studies.

KEYWORDS

Alzheimer's disease, blood, co-expression, deconvolution, diagnosis, neutrophil, stratification, transcriptome

Highlights

- We comprehensively analyze the blood transcriptomes of a well-characterized Alzheimer's disease cohort to identify genes, gene modules, pathways, and specific blood cells associated with the disease.
- Blood transcriptome analysis accurately classifies and stratifies patients with Alzheimer's disease, with some gene modules achieving classification accuracy comparable to that of the plasma ATN biomarkers.
- Immune-associated pathways and immune cells, such as neutrophils, have potential roles in the pathogenesis and progression of Alzheimer's disease.

1 | BACKGROUND

Alzheimer's disease (AD), one of the most common aging-associated diseases, has become a leading cause of death worldwide.¹ Despite its huge socioeconomic burden, intervention strategies for the disease are limited. Current diagnosis mostly relies on physicians' subjective judgements based on family history and cognitive tests. However, those diagnostic methods are often insufficiently sensitive to detect the early stages of the disease or insufficiently specific to distinguish AD from other types of dementia. Meanwhile, the implementation of positron emission tomography (PET) and lumbar puncture for detecting AD hallmarks, including amyloid plaques, tau tangles, and neurodegeneration—collectively termed the “ATN” biomarker panel—can quantitatively assess disease severity, greatly improving the sensitivity and accuracy of AD diagnosis. In particular, the sequential changes in the ATN biomarkers enable us to examine the detailed molecular phenotypic changes during disease progression, which facilitates disease staging, intervention, and the development of tailored intervention strategies. Furthermore, the recent identification of blood-based biomarkers that reflect the respective changes of the ATN biomarkers in the brain, including amyloid-beta ($A\beta$), phosphorylated tau at threonine 181 (p-tau181), and neurofilament light polypeptide (NFL), has revolutionized the diagnosis and monitoring of AD.²⁻⁵ However, it is unclear if these biomarkers fully reflect the various disease states during AD progression. Moreover, the ATN biomarkers might not provide a detailed picture of the molecular mechanisms that underlie AD progression.

The clinical use of blood-based diagnostics for AD is being intensely investigated because of the ease of access to blood samples. Besides

the blood ATN biomarkers, the expressions of hundreds of plasma proteins—many of which are expressed by blood cells—are altered in patients with AD.⁶⁻⁸ Therefore, in parallel to the plasma ATN biomarkers, blood cells may also undergo pathophysiological changes along AD progression.

Transcriptome analysis is a widely used method for studying how genes behave in different biological contexts, including the brains of individuals with AD. Specifically, analyses of the brain transcriptome in AD provide a comprehensive view of the molecular changes in genes, pathways, gene modules, and cell types.⁹⁻¹⁵ For example, co-expression network analysis of transcriptomic data has revealed that the immune system is involved in AD,¹⁴ while cell-type deconvolution analysis has shown a decrease in the number of neurons and an increase in the numbers of microglia and astrocytes within the AD brain.¹⁵ Specific gene sets from transcriptomic data can also distinguish AD-affected brains from undemented brain tissues.¹⁶ Considering the changes in blood cells that occur upon AD onset and progression, transcriptome analysis may reveal molecular changes in the blood cells of patients with AD, providing critical insights into the disease's pathophysiological mechanisms. Furthermore, recent studies demonstrate that blood transcriptome analysis can aid the diagnosis of uncommon muscle and mitochondrial disorders, underscoring its potential utility for the diagnosis and characterization of diseases, including AD.¹⁷ Hence, applying blood transcriptome analysis to AD studies may also reveal new diagnostic biomarkers. Nevertheless, few studies involving transcriptome analysis have investigated AD blood cells or the application of blood transcriptomic data to disease diagnosis or stratification. Therefore, the present study aims to assess the potential usage of blood

transcriptome analysis for AD, including delineating the molecular phenotypic changes in blood cells as well as clinical diagnosis, stratification, and monitoring.

2 | METHODS

2.1 | Study cohorts

Two cohorts of Hong Kong Chinese individuals aged ≥ 60 years were included in this study. The discovery cohort ($n = 422$) comprised 214 patients with AD and 208 normal controls (NCs) who visited the Specialist Outpatient Department of the Prince of Wales Hospital of the Chinese University of Hong Kong from April 2013 to February 2018. The validation cohort ($n = 64$) included 38 NCs and 26 patients with AD with plasma p-tau181 levels ≤ 2.55 and > 2.55 pg/mL, respectively (the p-tau181 cutoff was based on our previous publication⁸). The validation cohort comprised 18 participants (14 patients with AD and 4 NCs) from the Specialist Outpatient Department of the Prince of Wales Hospital of the Chinese University of Hong Kong recruited from April 2013 to February 2018; 15 participants (12 patients with AD and 3 NCs) who visited Queen Elizabeth Hospital from February 2018 to March 2020; and 31 patients with AD who visited the Community CareAge Foundation or Haven of Hope Christian Service from October 2019 to January 2020 (Table S1).

All participants recruited from hospitals underwent medical history assessment, clinical assessment, and cognitive and functional assessment (i.e., the Montreal Cognitive Assessment [MoCA]).¹⁸ Participants with any psychiatric disorder or significant neurological disease other than AD were excluded. For participants recruited from the Specialist Outpatient Department of the Prince of Wales Hospital of the Chinese University of Hong Kong, the clinical diagnosis of AD was based on the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5).¹⁹ Participants recruited from Queen Elizabeth Hospital also underwent neuroimaging assessment by MRI.²⁰ For these participants, the clinical diagnosis of AD was based on the US National Institute on Aging and Alzheimer's Association (NIA-AA) workgroup 2011 revised criteria.^{21,22} In addition, the participants recruited from the Community CareAge Foundation or Haven of Hope Christian Service (representing population-level NCs), underwent medical history assessment as well as cognitive and functional assessment with the MoCA.¹⁸

This study was approved by the Clinical Research & Ethics Committees of the Joint Chinese University of Hong Kong-New Territories East Cluster for the Prince of Wales Hospital (CREC Ref. No. 2015.461), the Kowloon Central Cluster/Kowloon East Cluster for Queen Elizabeth Hospital (KC/KE-15-0024/FR-3), Ethics and Research Committee of Haven of Hope Christian Service and the Human Participants Research Panel of The Hong Kong University of Science and Technology (CRP#180 & CRP#225). All participants provided written informed consent for both study participation and sample collection.

RESEARCH IN CONTEXT

- 1. Systematic review:** The authors reviewed the literature using traditional sources (e.g., PubMed). While the utility of blood transcriptome analysis for investigating disease mechanisms, diagnosis, and stratification has not been extensively researched, a few publications describe its applications in Alzheimer's disease. These relevant citations are cited appropriately.
- 2. Interpretation:** Our findings suggest that there are molecular changes in the blood of patients with Alzheimer's disease, including alterations to genes, gene modules, and biological pathways—particularly immune-associated pathways and certain blood cell types. In addition, the results demonstrate the potential use of blood transcriptomic data to classify and stratify patients with Alzheimer's disease.
- 3. Future directions:** This study presents evidence of systemic alterations to the molecular phenotypes of blood cells in Alzheimer's disease. It also suggests the potential utility of blood transcriptomic data for investigating the disease's pathophysiological mechanisms and developing tools for diagnosis and stratification. Future studies might focus on the specific involvement of different blood cell types in the pathogenesis and progression of Alzheimer's disease as well as the use of blood transcriptome analysis for diagnosis and stratification in diverse ethnic groups and investigating other diseases.

2.2 | Blood transcriptome sequencing and analysis

RNA samples were extracted with PAXgene Blood RNA Tubes using PAXgene Blood miRNA Extraction kits (Qiagen, PreAnalytiX GmbH, Hilden, Germany). For each participant, 3 to 10 μg total RNA, quantified by a BioDrop DUO machine (BioDrop Ltd., UK), was treated with a GLOBINclear-Human Kit (Invitrogen, Waltham, MA, USA) to remove hemoglobin transcripts. Then, 2 to 5- μg cleaned RNA samples were subjected to RNA sequencing by Novogene (Beijing, China) using an Illumina NovaSeq 6000 platform. For each participant, 40 million 150-bp paired-end reads were produced.

The raw sequencing reads were first subjected to FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) to evaluate data quality and then to Trimmomatic²³ to trim and filter low-quality reads. The cleaned reads were then mapped to the GRCh37 human reference genome (GRCh37.87 GTF file, sourced from ENSEMBL) using the STAR aligner ("spliced transcripts alignment to a reference"; version 2.5.3a)²⁴ with the default settings. Gene and transcript abundance were then quantified by analyzing the mapped BAM files with Stringtie2 version 2.1.4.²⁵ A total of 63,677 genes from ENSEMBL

encompassing different categories were analyzed, including 22,810 protein-coding genes, pseudogenes, and long noncoding RNAs. There are 28,010 genes remaining after excluding genes with low expression (with at least one read count in each sample).

2.2.1 | Differential expression analysis

For the 28,010 genes that passed the quality control and low-abundance filtering, the raw read counts were subjected to DESeq2²⁶ to determine the genes that were differentially expressed in patients with AD compared to NCs, with age, sex, and population structure (represented by the top five principal components from the genomic data) included as covariates. A cutoff (i.e., adjusted $p < 0.1$ calculated using the Benjamini and Hochberg method) was applied to identify the 12,837 differentially expressed genes.

2.2.2 | Gene ontology enrichment analysis

The online tool PANTHER version 14.0²⁷ was utilized to perform Gene Ontology and pathway enrichment analyses on the differentially expressed genes (i.e., upregulated, downregulated, or from specific gene modules). Gene Ontology terms with a false discovery rate (FDR) < 0.05 were considered statistically significant.

2.2.3 | Co-expression network analysis

WGCNA²⁸ was applied to construct co-expression networks and identify genes that are co-regulated in human blood. Before analysis, the *lm()* function was applied to regress out the confounding effects of age and sex on gene expression matrix. The resultant clean gene expression matrix (measured as transcripts per million [TPM]) was then subjected to WGCNA, with a signed similarity matrix (i.e., Pearson's correlation coefficients) generated to represent the strength of the correlations among the gene expression profiles. Gene modules were identified using the default settings, and those with at least 30 genes were retained for downstream analysis. To determine the correlations between each gene module and phenotype or other AD-associated endophenotypes (e.g., plasma biomarkers), the first principal component of the module's gene expression (termed the "module eigengene") was used.

2.2.4 | Disease classification analysis

The clean gene expression matrix (quantified in TPM) mentioned above was subjected to a generalized linear model using the *glm* function in the R *Stats Package* version 3.6.0. The trained models were then applied to both the discovery and validation cohorts, and the model output—termed the "module score"—was used to evaluate the accuracy of AD classification. As an indicator of model performance, the area under the receiver operating characteristics curve (auROC) was calculated using the *roc* function in the R *pROC* package.²⁹

2.2.5 | Participant stratification analysis

The predicted score for each selected gene module—M01, M02, M05, M09, M13, and M15—was subjected to *k*-means clustering using the *kmeans()* function in the R *stats* package to determine the subgroups of participants in the discovery cohort. The optimal number of clusters was determined using the elbow method, which was implemented using the *fviz_nbclust()* function in the R *factoextra* package.³⁰ Then, for uniform manifold approximation and projection (UMAP) analysis, the module score was subjected to the *umap()* function in the *umap* package in R to project individual participants into a two-dimensional plane for visualization.

2.2.6 | Cell-type deconvolution analysis

The *lm()* function was applied to regress out the confounding effects of age and sex on gene expression before analysis. The clean gene expression matrix (measured in TPM) was subjected to CIBERSORTx software³¹ for cell-type deconvolution analysis using the LM22 signature matrix as a reference. One-tailed robust regression analysis was performed to examine the associations between gene modules (represented by predicted scores) and cell-type enrichment scores (obtained from CIBERSORTx) using data from the NCs in the discovery cohort. Association analysis was conducted to identify cell types that were altered in AD using multiple linear regression, taking all the analyzed cell types jointly and with age and sex included as covariates.

2.2.7 | Functions and tools for statistical analysis

Linear regression was performed using the *lm()* function in the R *stats* package, and robust regression analysis was performed using the *lmrob()* function in the R *robustbase* package.³²

2.3 | Calculation of population structure using whole-genome sequencing data

Whole-genome sequencing data for the studied participants were retrieved from previous studies.^{33,34} In brief, whole-genome sequencing (5× coverage with 150-bp paired-end reads) was performed by Novogene on an Illumina HiSeq × Ten and NovaSeq platform (San Diego, CA, USA). The GotCloud pipeline³⁵ was adopted to detect variants from our whole-genome sequencing data. In brief, the sequencing data were subjected to FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) for quality control and Trimmomatic³⁶ to trim and filter low-quality reads. The clean data were mapped to the GRCh37 reference genome containing decoy fragments using BWA-mem. The data were then subjected to the GotCloud pipeline for data processing and variant detection using the default settings.³⁷ The clean genotype files were subsequently to Beagle³⁸ and Thunder³⁹ for genotyping refinement. After passing the GATK Variant

Quality Score Recalibration and minor allele frequency filter (> 5%), the common variants underwent linkage disequilibrium pruning using PLINK software (version 1.9). The linkage disequilibrium-pruned variants were subsequently utilized for principal component analysis with PLINK to identify population structure.

2.4 | Measurement of plasma protein biomarker levels

Data for plasma biomarkers, including the $A\beta_{42/40}$ ratio and total tau, p-tau181, and NfL levels, were obtained from a previous publication.⁴⁰ Among the 422 participants in the discovery cohort, 184 including 100 patients with AD and 84 NCs had available plasma biomarker data. Plasma biomarker data were available for all 64 participants in the validation cohort.

For brain imaging analysis, T1-weighted magnetization-prepared rapid gradient-echo (MPRAGE) and fluid-attenuated inversion recovery (FLAIR) sequences were retrieved for 54 patients with AD and 85 NCs from the Prince of Wales Hospital in the discovery cohort. The raw imaging files were deidentified and sent to BrainNow Medical Technology (HKSAR, China) to analyze the volumes of different brain regions and white matter hyperintensity levels.

2.5 | Single-cell analysis

2.5.1 | Study participants

Five additional elderly Hong Kong Chinese individuals (all females aged 69 to 83 years) were recruited for the blood single-cell RNA sequencing analysis. Three patients with AD or mild cognitive impairment were recruited from the Specialist Outpatient Department of the Prince of Wales Hospital at the Chinese University of Hong Kong, and two individuals without a history of major disease were recruited from the Prince of Wales Hospital. Patients were diagnosed with AD or mild cognitive impairment according to the DSM-5 criteria.¹⁹ This study was approved by the Clinical Research & Ethics Committees of Joint Chinese University of Hong Kong-New Territories East Cluster for the Prince of Wales Hospital (CREC Ref. No. 2015.461) and the Human Participants Research Panel of the Hong Kong University of Science and Technology (CRP#180 and CRP#225). All participants provided written informed consent for both study participation and sample collection.

2.5.2 | Blood collection and leukocyte isolation

Whole-blood samples (2 mL) were collected from individuals using K3EDTA tubes (VACUETTE; Greiner Bio-One). To lyse red blood cells, the blood samples were treated with 40 mL ACK Lysing Buffer (A1049201, Thermo Fisher) for 5 min at room temperature. The remaining cells were then washed with 10 mL phosphate buffered

saline (PBS) and pelleted by centrifugation at 300 × g for 5 min at room temperature. For a second round of red blood cell lysis, the cell pellets were then resuspended in 15 mL ACK Lysing Buffer (A1049201) and incubated for 5 min at room temperature. After a second washing step with PBS and centrifugation, the cell pellets were resuspended in 2 mL PBS with 0.04% BSA and filtered with a 40- μ m cell strainer (352340, Corning). The cell suspensions were then assessed for cell viability and counted under a microscope. The concentration was adjusted to 1000 cells per μ L with PBS and 0.04% bovine serum albumin (BSA).

2.5.3 | Single-cell RNA sequencing library construction and sequencing

We prepared scRNA-seq libraries using a Chromium Next GEM Single Cell 3' Library & Gel Bead Kit v3.1 (1000121; 10x Genomics) according to the manufacturer's protocol. In brief, 10 μ L cell suspension (10,000 cells) was mixed with reverse-transcription reagents and loaded into a chip to partition single cells into droplets. The droplets were incubated on a thermocycler to generate barcoded cDNA from polyadenylated mRNA by reverse-transcription. Libraries were constructed from cDNA according to the manufacturer's instructions. The library concentration was measured using Qubit (Thermo Fisher Scientific), and fragment lengths were measured using a Fragment Analyzer (Advanced Analytical Technologies). The libraries were then sent to Novogene for sequencing on a NovaSeq 6000 system, which generated 150-bp paired-end reads for downstream analysis.

2.5.4 | Single-cell RNA sequencing preprocessing and quality control

Demultiplexed FASTQ files were aligned to the hg19 reference genome using CellRanger version 6.1.2 to obtain gene counts. Single cells with ≥ 200 unique molecular identifiers and genes detected in ≥ 5 cells were selected for analysis. For quality control, cell-free mRNA contamination was estimated and removed using the *SoupX* package.⁴¹ Potential dead cells (i.e., $\geq 10\%$ mitochondrial genes or ≤ 200 features) and multiplets (i.e., aggregation of two or more cells into single droplets; ≥ 5000 features) were excluded using Seurat version 4.0.6.⁴² After filtering, 19,958 cells remained in the dataset.

2.5.5 | Single-cell RNA sequencing integration, cell clustering, and cell-type annotation

The gene count matrices of each sample were normalized and integrated with the reciprocal principal component analysis and SCTransform-based workflows in Seurat. In brief, variable features were identified and normalized using the *SCTransform* function. The first 30 principal components from reciprocal principal component analysis were selected to identify integration anchors between samples using the *FindIntegrationAnchors* function. The samples were then

integrated into a single Seurat object using the anchors via the *IntegrateData* function. We performed principal component analysis on the integrated data and selected the first 30 principal components for cluster identification using *k*-nearest neighbor algorithm. We identified the differentially expressed genes of each cluster using the Wilcoxon rank-sum test with the *FindAllMarkers* function. Only genes with a $\log_2(\text{fold-change}) > 0.25$ that were expressed in $\geq 10\%$ of cells in the cluster were analyzed. Known cell-type marker genes among the differentially expressed genes were used to assign clusters to 13 cell types based on the literature.^{43,44}

2.6 | Data visualization

A heatmap of the differentially expressed genes was generated using the *heatmap.2()* function in the *ggplot* package in R. The heatmap of the co-regulation of gene expression was generated using the *TOMplot* function in the *WGCNA* package in R. Bar charts, heatmaps of the association results between gene modules and disease-associated endophenotypes, and receiver operating characteristic curves were generated using GraphPad Prism version 8.0.2. Radar plots were generated using the *ggradar()* function using the R *ggradar* package, and Ridgeline plots were generated using the *geom_ridgeline()* function in the R *ggridges* package.

3 | RESULTS

3.1 | Blood transcriptome analysis reveals dysregulated genes and pathways in patients with AD

To understand the molecular changes in the blood cells of patients with AD, we performed bulk RNA-seq on blood cells from an AD cohort recruited in Hong Kong,^{45,46} comprising 422 participants (patients with AD: $n = 214$, normal controls [NCs]: $n = 208$), hereafter referred to as the “discovery cohort” (Table S1; see Figure 1 for study design). Based on these RNA-seq data, we first identified genes that were dysregulated in the blood in AD by using a generalized linear model with age, sex, and population structure (i.e., the top five principal components estimated from genomic data) as covariates. Out of 28,010 genes detected in the blood, 12,837 were differentially expressed between the patients with AD and NCs, including 7736 upregulated and 5101 downregulated genes (adjusted $p < 0.1$; Figure 2A, Table S2). We subsequently conducted Gene Ontology enrichment analysis to examine the biological processes associated with those differentially expressed genes. The upregulated genes are involved in the metabolic process (*false discovery rate* [FDR] = 1.22×10^{-16}), translation (FDR = 3.99×10^{-10}) and oxidative phosphorylation (FDR = 5.31×10^{-4} ; Figure 2B, Table S3), whereas the downregulated genes are involved in the cell cycle (FDR = 1.77×10^{-17}), histone modification (FDR = 3.18×10^{-8}), and immune-associated pathways such as autophagy (FDR = 4.59×10^{-5}) and lymphocyte activation (FDR = 6.20×10^{-4} ; Figure 2B,

Table S4). Hence, by identifying which genes and pathways are altered in blood cells in AD, our results suggest the existence of molecular changes in the blood cells of patients with AD.

To validate the observed gene expression changes in AD blood, we repeated the RNA-seq and differential expression analyses on blood cells from an independent cohort ($n = 64$; comprising 26 p-tau-positive patients with AD and 38 p-tau-negative NCs)—hereafter referred to as the “validation cohort”—with a stricter diagnosis based on the examination of plasma p-tau181 levels. The changes in transcript levels of those differentially expressed genes (i.e., the fold-change between the AD and NC groups) between the discovery and validation cohorts were strongly correlated ($R^2 = 0.89$, $p < 2.2 \times 10^{-16}$). This finding validates the observed molecular changes in the blood cells of patients with AD (Figure 2C, Table S5). Furthermore, we examined the expression changes of key AD risk genes implicated in a genome-wide association study (GWAS) of both the discovery and validation cohorts.⁴⁷ Many of these key AD GWAS risk genes, such as *apolipoprotein E* (APOE), *ABCA7*, and *CR1*, which are associated with inflammation and immune function, were differentially expressed at the gene level in the blood of patients with AD in the discovery cohort, with a consistent trend of dysregulation in the validation cohort (Figure 2D). Indeed, single-cell RNA-seq studies have provided compelling evidence suggesting that APOE is expressed in specific blood cell types, particularly in human macrophages and mouse leukocytes.^{48,49} We extended the analysis to all genes that resided in AD GWAS risk loci based on the annotations from Kunkle et al.⁵⁰ Notably, for specific AD risk loci, multiple genes are significantly dysregulated in both the discovery and validation cohorts (e.g., *ABCA7*, *CR1*, *HLA-DRB1*, *MS4A2*, *MYAP1*, and *SPI1*; Figure S1).⁵⁰ These findings support the findings of previous Gene Ontology analyses that highlight the dysregulation of immune-associated pathways in AD, further indicating that the molecular changes in blood cells are associated with AD pathogenesis.

3.2 | Gene modules in blood cells are involved in diverse biological processes in AD

Given the involvement of the above-mentioned differentially expressed genes in diverse biological pathways, it is important to identify gene modules to focus on understanding the key biological processes in AD. As genes related to a given biological process likely have overlapping regulatory mechanisms and similar expression patterns, we performed WGCNA to identify modules of co-expressed genes that are altered in the blood from patients with AD and whose expression levels are highly correlated. In brief, we constructed a co-expression network from the 12,837 differentially expressed genes in the blood samples from the discovery cohort and identified 15 AD-associated gene modules, which we designated M01–M15 (Figure 3A, B; Tables S6–19). Interestingly, Gene Ontology analysis revealed that many of these gene modules are involved in immune-associated pathways. For example, M07 is involved in B cell activation (FDR = 1.04×10^{-5}), M09 in adaptive immune response (FDR = 3.96×10^{-11}), M11 in innate immune response (FDR = 5.82×10^{-24}), and

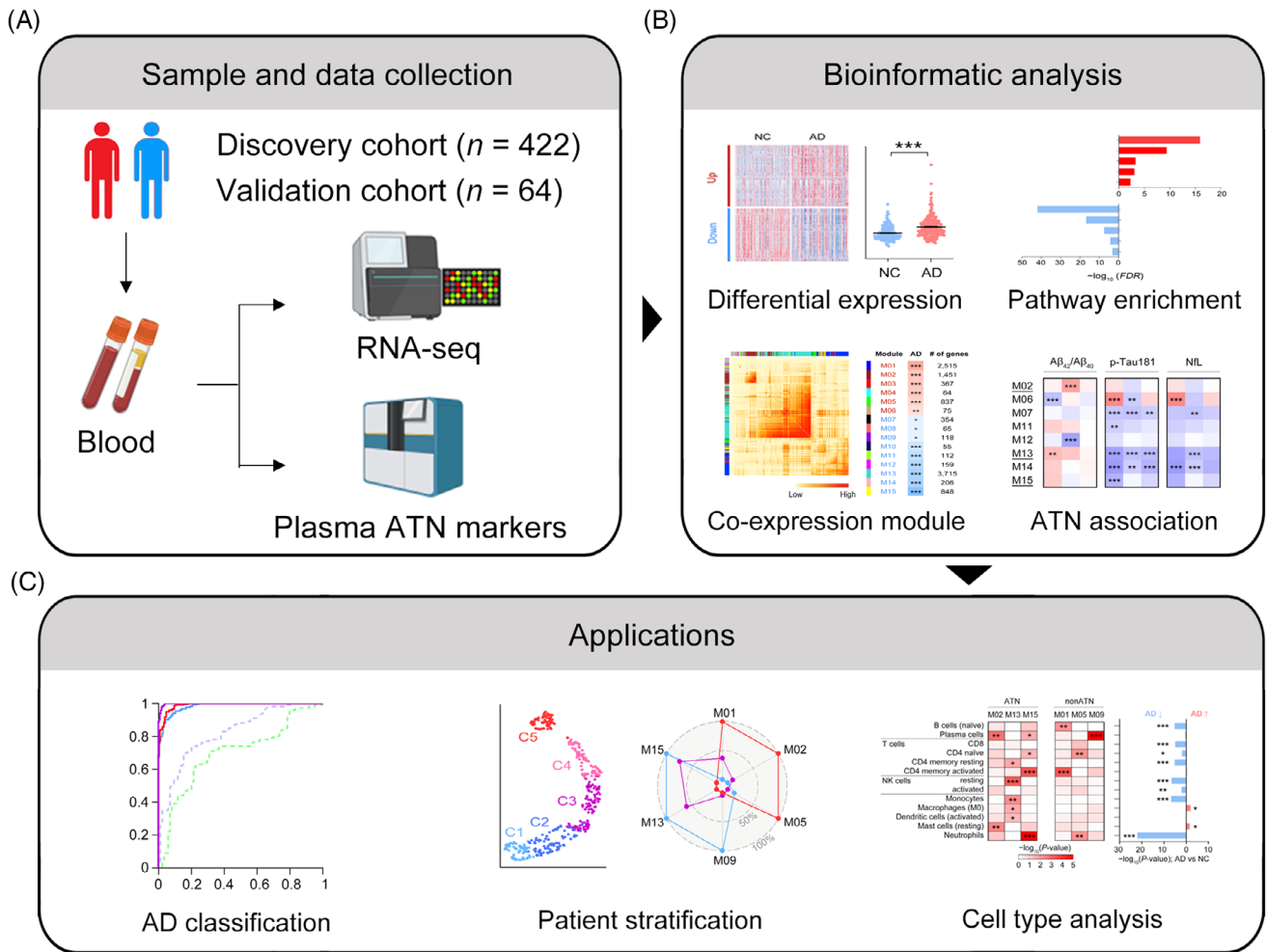


FIGURE 1 Schematic diagram of the study design. (A) Samples and data collected. (B) Types of analyses performed. (C) Proposed applications of blood transcriptomic data for Alzheimer’s disease classification, patient stratification, and cell-type analysis. AD, Alzheimer’s disease; ATN, amyloid/tau/neurodegeneration; RNA-seq, RNA sequencing

M15 in myeloid leukocyte activation ($FDR = 1.53 \times 10^{-5}$). These results further corroborate the dysregulation of immune pathways in blood cells in AD and suggest the parallel involvement of diverse immune-associated pathways during AD pathogenesis and progression. In addition, M06 is associated with synaptic transmission ($FDR = 3.42 \times 10^{-2}$; Table S8), suggesting a link between the molecular changes in blood cells and the brain during AD pathogenesis and progression.

3.3 | Associations between gene modules and AD

We subsequently evaluated whether these gene modules contribute equally to AD or not. Accordingly, we used lasso logistic regression to condense the gene expression data from each module into a single numeric score—termed the “module score”—that reflects the relative overall expression changes of genes within the module. We trained the model using data from the discovery cohort and determined if it can distinguish patients with AD from NCs in both cohorts. We then evaluated the ability of the identified gene modules to identify AD by

calculating the auROCs in both cohorts (as a measure of the strength of the association between the identified gene modules and AD). Interestingly, all 15 AD-associated gene modules were associated with AD ($auROC > 0.7$), and 6 modules—M01, M02, M05, M09, M13, and M15—exhibited consistently strong associations with AD ($auROC > 0.9$) in both cohorts (Figure 4A; Figures S2, S3; Table S20). Among these six modules, M09 and M15 are related to immune functions: adaptive immune response and myeloid leukocyte activation, respectively. These results indicate that several distinct biological processes, especially those related to immune cells or pathways, are dysregulated in blood cells in AD, highlighting the roles of these processes in AD.

3.4 | Associations between gene modules and plasma ATN biomarkers

During AD progression, the plasma ATN biomarkers are correlated with key pathological hallmarks in the brain.^{51,52} Given that the above-mentioned gene modules can identify AD based on the blood cell

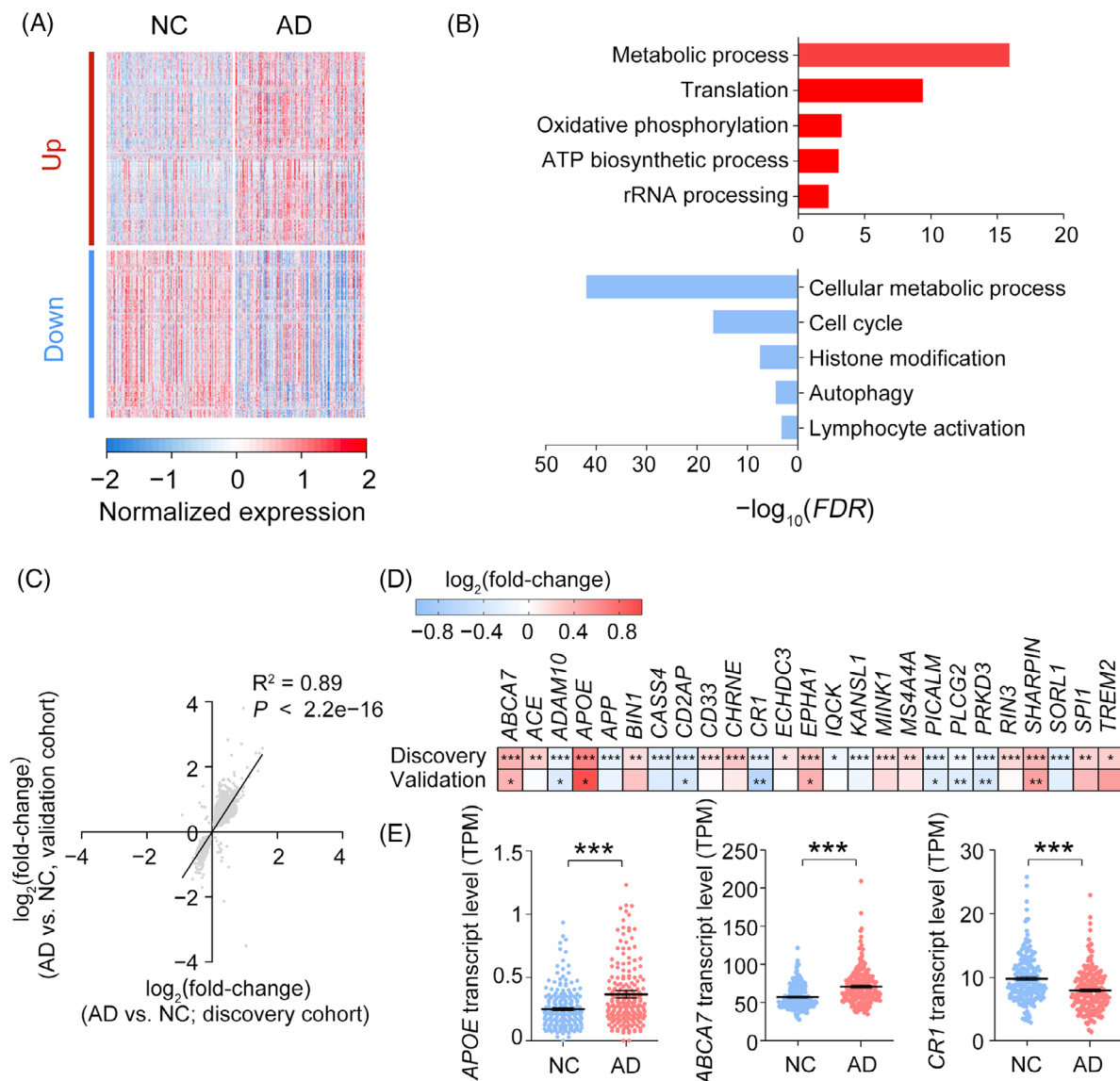


FIGURE 2 Dysregulation of genes and pathways in the blood of patients with AD. (A) Heatmap of the normalized expressions of genes that were differentially expressed between patients with AD (n = 214) and NCs (n = 208) in the discovery cohort (FDR < 0.1). Rows represent individual genes (n = 7736 and 5101 up- and downregulated genes, respectively), and columns represent individual participants. Colors denote the levels of gene expression (TPM). (B) Representative gene ontology terms enriched by genes that were up- (red) or downregulated (blue) in patients with AD versus NCs. (C) Correlations between the $\log_2(\text{fold-change})$ of differentially expressed genes in the discovery cohort (n = 422) and validation cohort (n = 64). (D) Heatmap showing the differential expressions of AD genes identified by GWAS in the discovery and validation cohorts. (E) Dot plot of the transcript levels of representative AD genes that were differentially expressed in the blood of patients with AD versus NCs in the discovery cohort. Data are mean \pm SEM. (D, E) $\log_2(\text{fold-change})$ calculated by negative binomial generalized linear models from DESeq2; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. AD, Alzheimer's disease; FDR, false discovery rate; GWAS, genome-wide association study; NC, normal control; Norm., normalized; SEM, standard error of the mean; TPM, transcripts per million

transcriptome, we subsequently investigated if these modules are associated with pathological changes of AD. Accordingly, we evaluated the plasma ATN biomarkers using the ultrasensitive SIMOA ("single-molecule array") platform and determined their correlations with the module scores of each gene module. Interestingly, in the discovery cohort, only 9 of the 15 gene modules were significantly correlated with the plasma ATN biomarkers ($p < 0.05$; Figure S4). Among the six modules that were closely associated with AD (i.e., M01, M02, M05, M09, M13, and M15), only M02, M13, and M15 were correlated with

the plasma ATN biomarkers ("ATN modules" hereafter; Figure 4B). We further compared the auROCs for AD classification between the three ATN modules and the three that were not ("non-ATN modules" hereafter). Interestingly, there was no significant difference between the auROCs of the ATN and non-ATN modules (Figure 4C). Hence, our results suggest that there are specific AD-associated molecular changes that occur in blood cells but are not correlated with the plasma ATN biomarkers. Accordingly, these findings indicate that blood transcriptome analysis could supplement the ATN biomarkers

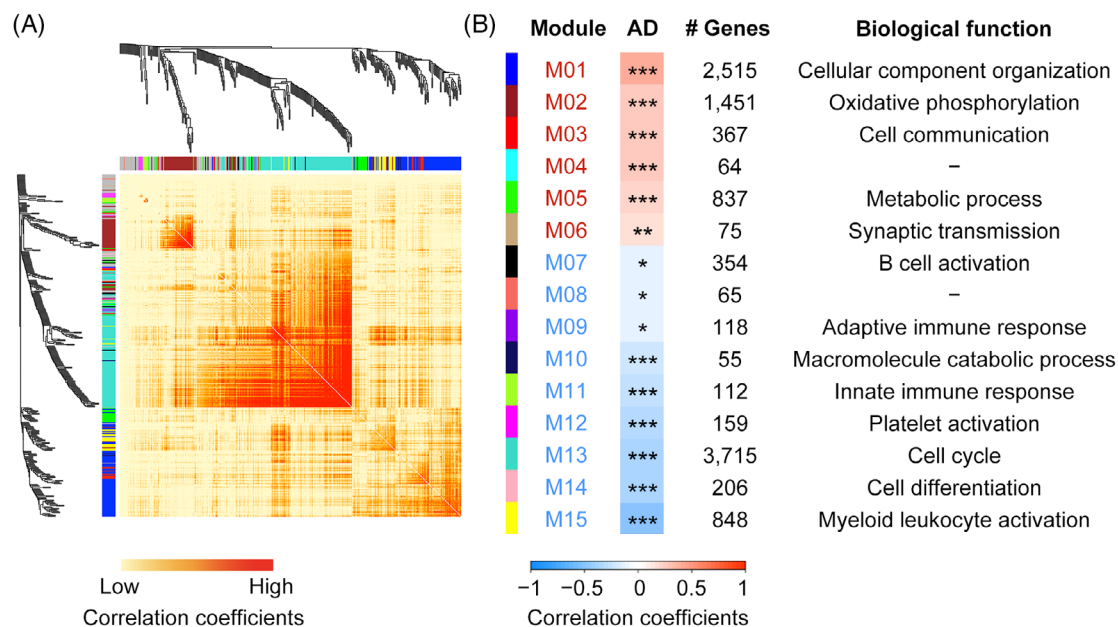


FIGURE 3 Gene modules with distinct biological functions identified by co-expression network analysis. (A) Heatmap showing the co-expression (i.e., correlation coefficients) among differentially expressed genes in the blood of patients with AD. Colors along the left side and top represent the 15 gene modules defined by weighted correlation network analysis. (B) Detailed information about the 15 identified gene modules, including (from left to right) the correlations between the modules and the presence of AD, gene count, and associated biological functions inferred from Gene Ontology analysis (Pearson correlation analysis; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). AD, Alzheimer's disease

for the diagnosis of AD. Of note, these non-ATN modules might still be associated with brain functions.

Next, to examine whether the six modules that are closely associated with AD are associated with the brain structural and cognitive changes in AD, we determined the correlations between the module scores and cognitive performance measured by the MoCA as well as brain volume (i.e., the amygdala and hippocampus) measured by MRI among participants from the discovery cohort. Interestingly, both the ATN and non-ATN modules were correlated with cognitive performance as well as volumetric changes in the amygdala and hippocampus in all participants (Figures S5, S6). This suggests that the molecular changes in blood cells captured by transcriptome analysis are correlated with changes in cognitive function and brain structure, which might provide additional information for determining disease status beyond the existing ATN biomarkers.

3.5 | Accuracy of the identified gene modules for AD classification

We subsequently investigated whether blood transcriptomic data can assist AD classification. Accordingly, we trained logistic regression models that include the module scores from the six modules closely associated with AD—M01, M02, M05, M09, M13, and M15—and output a single numeric score representing an individual's disease status. We again calculated the auROCs to evaluate the performance of the AD classification models. Interestingly, the models based on the ATN modules (i.e., M02, M03, and M15) were highly accurate in both the

discovery (auROC = 0.990) and validation cohorts (auROC = 0.897) as were the non-ATN modules (i.e., M01, M05, and M09) (auROC = 0.983 and 0.927 in the discovery and validation cohorts, respectively). Furthermore, when combining all six modules, the classification model exhibited similarly high accuracy (auROC = 0.998 and 0.920 in the discovery and validation cohorts, respectively; Figure 4D, E). Thus, our results indicate that blood transcriptomic data can be used to classify AD with relatively high accuracy. Furthermore, the plasma ATN biomarkers and blood gene modules exhibited similar accuracy of AD classification. In the discovery cohort, the models based on the blood gene modules (auROC = 0.98–1.00) classified AD more accurately than those based on the plasma ATN biomarkers (auROC = 0.70–0.82; Figure 4D, Table S21). We obtained similar results in the validation cohort: the models developed using the blood gene modules performed better (auROC = 0.90–0.92) than that based on plasma A $\beta_{42/40}$ ratio (auROC = 0.76) and comparably to that based on plasma NfL level (auROC = 0.95; Figure 4E, Table S21; data for p-tau181 are not presented, because plasma p-tau181 was used as a selection criterion for the validation cohort). Taken together, these results suggest that blood transcriptome analysis can accurately distinguish patients with AD from NCs.

3.6 | Stratification of participants based on blood transcriptomic data

Given the progressive nature and complex multifactorial etiology of AD, accurate staging is crucial for effective disease management.

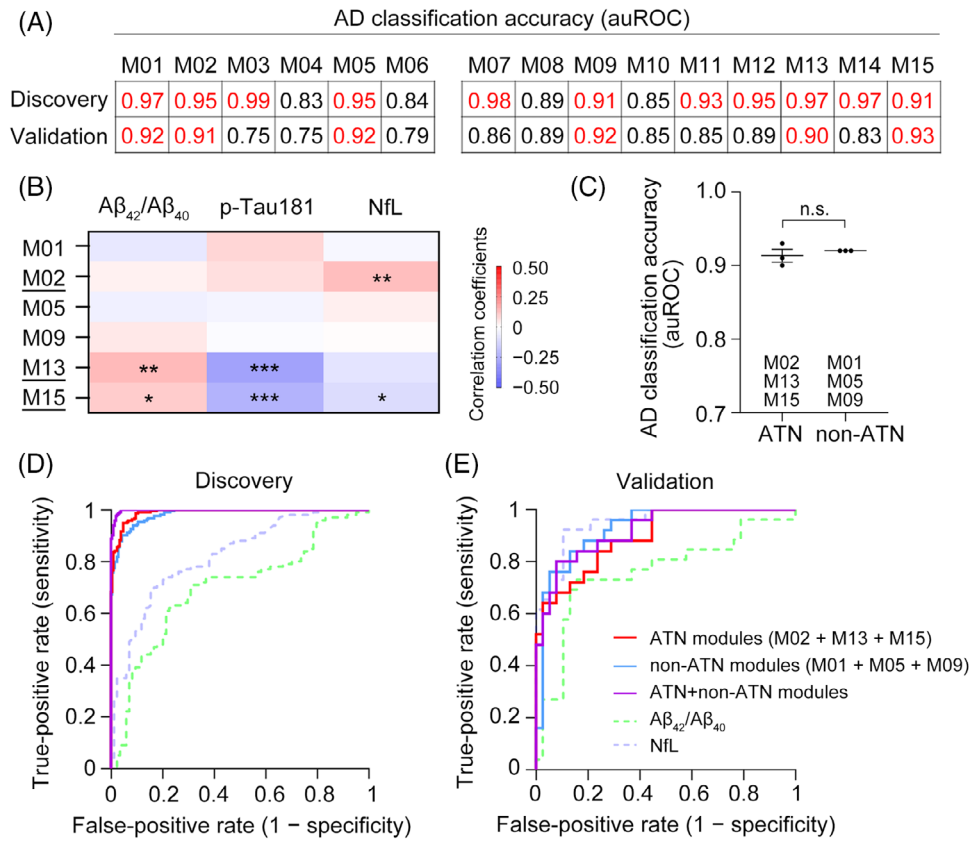


FIGURE 4 Classification of AD status according to the ATN and non-ATN gene modules. (A) AD classification accuracy of individual gene modules in the discovery and validation cohorts. The auROCs > 0.9 are highlighted in red. (B) The correlations between modules with a auROC > 0.9 in the two analyzed cohorts and the levels of the ATN biomarkers in all participants were examined. Modules correlated with any ATN biomarker are underlined (Pearson correlation analysis; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). (C) AD classification accuracy for selected modules grouped according to their association with the ATN biomarkers. Data are the mean ± SEM auROCs obtained using each module score within the group (one-sample t -test). ROC curves showing the discriminatory accuracy of selected modules and the plasma ATN biomarkers in (D) the discovery cohort and (E) validation cohort. AD, Alzheimer's disease; ATN, amyloid/tau/neurodegeneration; auROC, area under the receiver operating characteristic curve; NC, normal control

Despite recent studies suggesting the existence of multiple AD subtypes, there is limited information about the molecular mechanisms associated with these subtypes or stages. Of note, the AD-associated blood gene modules identified herein involve distinct biological processes that may be altered along with disease progression. Hence, the dysregulation of these gene modules may represent individuals' specific status through AD progression.

Accordingly, to determine if the identified gene modules from blood cells can be used to stratify participants into subgroups, we performed unsupervised clustering analysis using the module scores from M01, M02, M03, M05, M09, and M15. Accordingly, the participants in the discovery cohort clustered into five subgroups, designated C1–C5 (Figure 5A), with increasing proportions of patients with AD indicating that the current stratification is associated with disease progression (Figure 57). Therefore, we investigated whether individuals from different subgroups had different statuses by comparing the AD endophenotypes among subgroups. Notably, there were significant differences in plasma ATN biomarker levels (Figure 5B), cognitive performance (Figure 58a), and the volumes of the amygdala (Figure

58b) and hippocampus (Figure 58c) among subgroups ($p < 0.05$). Thus, besides classifying AD risk, blood transcriptomic data can also stratify individuals into subgroups corresponding to different disease stages or states.

Of note, the module score used for stratification analysis was calculated based on the expression levels of genes within each module, which are associated with specific biological processes (Figure 3). Therefore, alterations to these biological processes can be represented by changes in the associated module score. Accordingly, to investigate whether specific biological processes associated with the six gene modules are altered among the identified subgroups, we compared the module scores from these modules among subgroups C1–C5. Interestingly, our analysis revealed significant alterations in the gene module scores among these subgroups. For instance, the module scores of M01, whose genes are involved in cellular component organization, were significantly higher in C2–C5 than in C1 ($p < 1 \times 10^{-3}$; Figure 5C, D). Meanwhile, the module scores of M15, whose genes are involved in myeloid leukocyte activation, were significantly lower in C2–C5 than in C1 ($p < 1 \times 10^{-3}$; Figure 5C, D). These results indicate

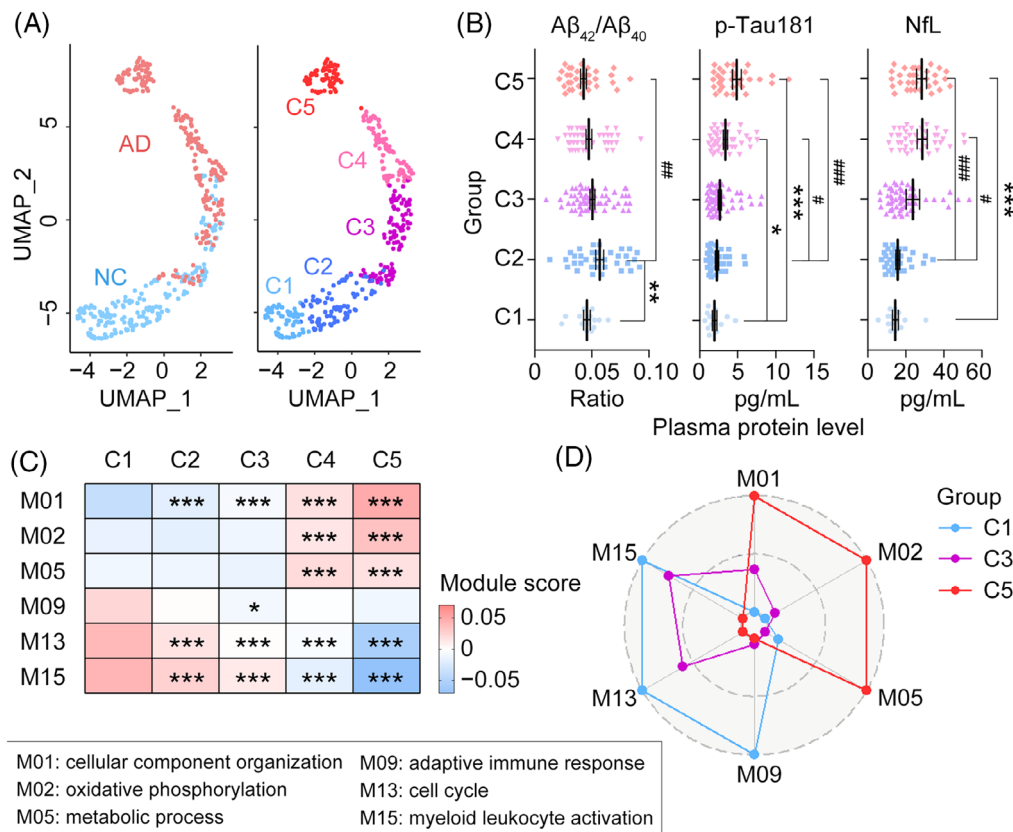


FIGURE 5 Stratification of participants based on the ATN and non-ATN gene modules (A) UMAP plots of participants stratified according to phenotype group (left panel) and subgroups (right panel: C1–C5). Colors denote the indicated group phenotypes (AD and NC, left panel) or indicated subgroups classified by *k*-means unsupervised clustering (right panel). (B) Comparison of plasma ATN biomarker levels among subgroups (generalized linear regression; **p* < 0.05, ***p* < 0.01, ****p* < 0.001 vs. C1; #*p* < 0.05, ##*p* < 0.01, ###*p* < 0.001 vs. C2). Data are mean ± SEM. (C) Heatmap comparing module scores from six modules that closely associated with AD among subgroups (generalized linear regression; **p* < 0.05, ****p* < 0.001 vs. C1). (D) Radar plot showing the relative differences in average module scores among C1, C3, and C5. *Aβ*, amyloid-beta; AD, Alzheimer's disease; ATN, amyloid/tau/neurodegeneration; NC, normal control; NfL, neurofilament light polypeptide; p-tau181, tau phosphorylated at threonine 181; UMAP, uniform manifold approximation and projection

that blood transcriptome analysis can further elucidate the molecular changes that occur during AD progression, providing insights into possible pathophysiological mechanisms associated with the disease pathogenesis and progression of AD.

3.7 | Changes in blood cell subtypes are implicated in AD

The transcriptomic changes described herein are likely due to molecular changes occurring within individual blood cell subtypes. Therefore, we identified which gene modules are associated with the roles of specific blood cell types (Figure S9a). Accordingly, to examine how the molecular changes in individual blood cell subtypes are associated with AD or the identified gene modules, we conducted cell-type deconvolution analysis using the blood transcriptomic data from the discovery cohort. The enrichment scores estimated from the deconvolution analysis represent the contribution of each blood cell subtype to the transcriptomic data. We subsequently performed multiple regres-

sion analysis to determine the associations of the enrichment scores with the module scores and AD phenotypes. Notably, the key ATN and non-ATN gene modules were significantly associated with the expression profiles of various blood cell types (Figure 6A, Figure S9a, Table S22). In particular, some associations were closely aligned with the annotated biological functions of the individual gene modules. For instance, M09 is involved in adaptive immune response and associated with plasma cells ($p = 6.82 \times 10^{-130}$; Figure 6A, Table S22), whereas M15 is involved in myeloid leukocyte activation and associated with neutrophils ($p = 8.13 \times 10^{-7}$; Figure 6A, Table S22). Of note, among the analyzed cell types, neutrophils were most closely associated with AD ($p = 1.09 \times 10^{-22}$; Figure 6B, Figure S9b, Table S23). Therefore, the dysregulation of genes in M15 is likely linked to molecular changes in neutrophils during AD pathogenesis and progression.

To further examine the relationship between M15 and neutrophils, we performed single-cell RNA-seq analysis of the blood cells from three patients with AD and 2 NCs. We estimated the module scores of M15 for individual cell types and then visualized the distribution of

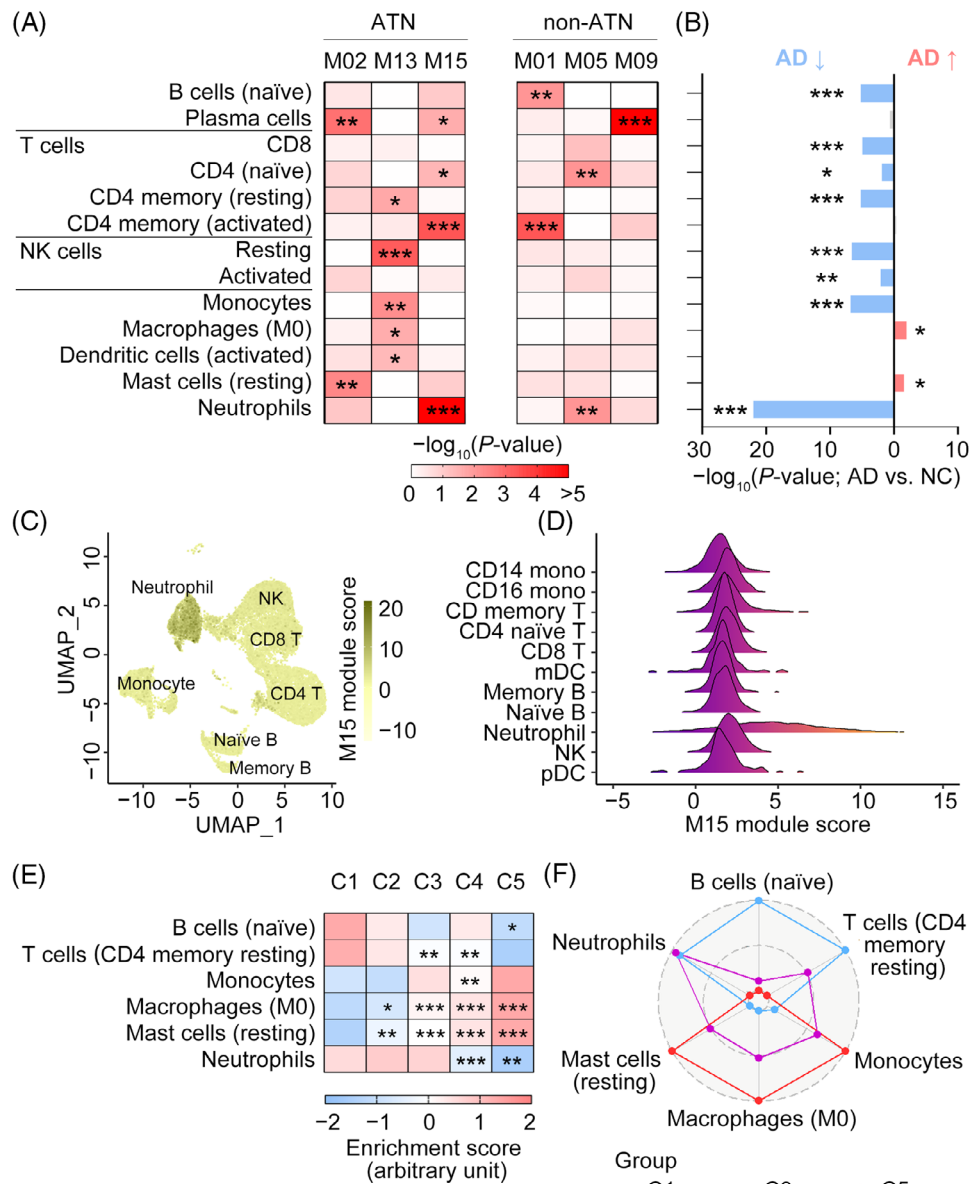


FIGURE 6 Changes in blood cell subtypes are implicated in AD. (A) Heatmap showing the associations among the selected gene modules and enrichment scores for individual cell types estimated by cell-type deconvolution analysis (NCs; one-tailed robust regression; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). (B) Associations between disease phenotypes and enrichment scores for individual cell types from cell-type deconvolution analysis (generalized liner regression; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). (C) UMAP plot showing the scores of the M15 module in the blood single-cell dataset. (D) Module scores for M15 in distinct blood cell types classified according to the blood single-cell RNA sequencing data. (E) Heatmap displaying cell-type enrichment scores among subgroups of participants (i.e., C1–C5) (generalized liner regression; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ vs. C1). (F) Radar plot displaying the relative differences in cell-type enrichment scores among C1, C3, and C5. AD, Alzheimer's disease; NC, normal control; UMAP, uniform manifold approximation and projection

M15 module scores across different cell types. The module score of M15 was significantly higher for neutrophils than for other cell types (Figure 6C, D), which is corroborated by the correlation between M15 module scores and neutrophil enrichment scores in our cell deconvolution analysis (Figure 6A, Table S22). Thus, these results collectively suggest that the dysregulation of M15 in AD is predominantly driven by neutrophils.

To better understand changes in blood cell types during AD progression or in different disease stages, we compared the cell-type

enrichment scores across different subgroups classified according to gene modules from the blood transcriptomic data. Accordingly, we observed significant alterations in key blood cell subtypes among these subgroups. For instance, the enrichment scores of M0 macrophages and resting mast cells were significantly higher in C3–C5 than in C1 ($p < 1 \times 10^{-3}$; Figure 6E, F). Meanwhile, the enrichment scores of neutrophils were significantly lower in C4 and C5 than in C1 ($p < 1 \times 10^{-2}$; Figure 6E, F). These results show that, besides diagnosis and patient stratification, blood transcriptome profiling can reveal

molecular changes in specific blood cell types that play critical roles in AD pathogenesis and progression.

4 | DISCUSSION

Recent advancements in “omics” research have revolutionized the prediction of AD risk and diagnosis.^{53–56} Notably, transcriptome analysis is now widely used to accurately quantify the levels of all genes expressed in tissues of interest and is extensively applied in disease studies. Therefore, blood transcriptome analysis is expected to comprehensively profile disease-associated molecular changes in blood cells during AD pathogenesis and progression. Accordingly, as one of the first comprehensive analyses of the AD blood transcriptome, we assayed the blood transcriptome of individuals from well-defined AD cohorts with available data on plasma ATN biomarkers, brain volumetric data, and cognitive function to identify molecular signatures in blood cells that are associated with AD (i.e., dysregulated genes, modules, and biological pathways). In particular, we identified specific immune pathways and immune-associated blood cell types that are altered in the blood of patients with AD. We verified some of these alterations using single-cell transcriptomic data in independent participants. Of note, the gene modules identified from blood transcriptomes also classified patients with AD with accuracy comparable to that with the plasma ATN biomarkers. Furthermore, these gene modules could also stratify participants into distinct subgroups corresponding to different disease stages or states. Thus, our findings collectively suggest that there are changes in the molecular phenotypes of blood cells during the onset and development of AD, highlighting the potential utility of blood transcriptomic data for both mechanistic and drug development studies in AD as well as patient classification and stratification.

Importantly, our findings support the potential of using blood transcriptome analysis to identify presymptomatic AD patients who already show molecular changes. Specifically, by using blood transcriptome analysis, we can stratify AD patients and NC into five subgroups; specifically, C1 consists entirely of undemented participants, and C2 primarily consists of undemented individuals (87 out of 94; see Figure S7). Of note, compared to individuals in C1, individuals in C2 have a lower plasma A β 42/40 ratio but no obvious differences in cognitive performance, or volume of the hippocampus or amygdala. This suggests that, despite being classified as NCs according to other analyses, individuals in C2 may already exhibit amyloid pathology (Figure 5 and Figure S8). Therefore, analyzing the blood transcriptome could also aid in the early detection of individuals at risk of developing AD before symptoms manifest.

Numerous studies demonstrate alterations in the blood transcriptome of AD patients, although the overlap in the dysregulated genes among different published studies is small.^{57–60} This discrepancy may be attributable to the small sample sizes of those studies (Figure S10 and Table S24). Nonetheless, it is worth highlighting that approximately half of the dysregulated genes reported in those studies were differentially regulated in AD patients in the present study. Thus, com-

prehensive AD-associated blood transcriptome profiling can be used to detect biomarkers of AD and provide insights into the pathological mechanisms of the disease.

Recent genetics and functional studies suggest the involvement of immune-associated signaling pathways, including the peripheral immune system,^{61,62} in the pathogenesis of AD.^{37,61,62,63} Concordantly, our blood transcriptome and gene ontology analyses demonstrate the potential involvement of immune-associated pathways (i.e., innate and adaptive immune-associated pathways) and specific immune cell types (i.e., B cells, neutrophils, and leukocytes) in AD. Given that several studies suggest that neutrophils are involved in AD,^{64–69} we further examined the genes associated with myeloid leukocyte activation from module M15, which is associated with neutrophils. Accordingly, we identified that LYN, a member of the Src family of nonreceptor protein tyrosine kinases, is involved in essential neutrophil functions such as phagocytosis, chemotaxis, and respiratory burst response (Figure S11a). LYN is mainly expressed in blood cells (Figure S11b), and our previous high-throughput proteomic analysis highlights LYN as one of the most differentially regulated proteins in AD,⁴⁶ both its blood transcript and plasma protein levels are consistently lower in patients with AD than in NCs. Our analysis of single-cell blood data confirms the high expression of LYN in neutrophils (Figure S11c). Moreover, the fraction of LYN-positive neutrophils was significantly lower in individuals with high p-tau181 levels than in those with low levels (Figure S11d). Meanwhile, although LYN plasma protein and blood transcript levels were lower in patients with AD, they were only correlated in NCs and not in patients with AD (Figures S11e–g). These observations again highlight the involvement of the peripheral immune system, including neutrophil functioning, in AD pathogenesis and progression. Notably, there is a recent report on the success of a phase three clinical trial of masitinib, a tyrosine kinase inhibitor that blocks LYN activity.^{70,71} This corroborates the feasibility of targeting LYN or its associated pathways as an intervention strategy for AD.^{72–74}

In summary, our comprehensive analysis of the blood transcriptome in patients with AD reveals key molecular phenotypes, including genes, modules, pathways, and subtypes of blood cells, that may be closely associated with AD pathogenesis and progression. Our findings also suggest the potential utility of blood transcriptome analysis for disease classification and stratification. Given the ease of blood sampling, blood transcriptome analysis could provide insights into human diseases, aiding the development of technologies for disease diagnosis, monitoring, and patient stratification. This could ultimately facilitate both early intervention and precision medicine for AD and other human diseases.⁷⁵

AUTHOR CONTRIBUTIONS

Huan Zhong, Xiaopu Zhou, Amy K. Y. Fu, and Nancy Y. Ip conceived of the project; Andrew Lung Tat Chan, Timothy C. Y. Kwok, Vincent C. T. Mok, and Fanny C. F. Ip organized patient recruitment and sample collection; Ronnie Ming Nok Lo, Bonnie Wing Yan Wong, and Elaine Yee Ling Cheng organized the clinical data and prepared the samples for sequencing analysis; Huan Zhong and Xiaopu Zhou set up the data-processing pipelines; Tiffany T. W. Mak coordinated the data

organization; Huan Zhong, Xiaopu Zhou, Hyebin Uhm, Yuanbing Jiang, Han Cao, Yu Chen, Kin Y. Mok, John Hardy, Amy K. Y. Fu, and Nancy Y. Ip analyzed the data; and Huan Zhong, Xiaopu Zhou, Amy K. Y. Fu, and Nancy Y. Ip wrote the manuscript with the input from all the co-authors.

ACKNOWLEDGMENTS

The authors thank Pauline Kwan, Hazel Mok, Dr Phillip Y.C. Chan, Choi Ying Ling, and Estella P S Tong for coordinating the collection of clinical data. The authors also thank Ka Chun Lok and Cara Wing Si Kwong for their excellent technical assistance as well as other members of the Ip Laboratory for many helpful discussions. John Hardy has served as a consultant for Eli Lilly and Eisai. This work was supported in part by the National Key R&D Program of China (2021YFE0203000); the Research Grants Council of Hong Kong (the Collaborative Research Fund [C6027-19GF], the Theme-Based Research Scheme [T13-605/18 W], and the General Research Fund [HKUST16103122]); the Areas of Excellence Scheme of the University Grants Committee (AoE/M-604/16); the Innovation and Technology Commission (InnoHK, ITCPD/17-9, ITS/207/18FP, MRP/042/18X and MRP/097/20X); the Guangdong Provincial Key S&T Program Grant (2018B030336001); the Guangdong Provincial Fund for Basic and Applied Basic Research (2019B1515130004), the NSFC-RGC Joint Research Scheme (32061160472), the Shenzhen Knowledge Innovation Program Grants (JCYJ20180507183642005, JCYJ20200109115631248, and ZDSYS20200828154800001); the Guangdong-Hong Kong-Macao Greater Bay Area Center for Brain Science and Brain-Inspired Intelligence Fund (2019001 and 2019003); and the Fundamental Research Program of Shenzhen Virtual University Park (2021Szvup137).

CONFLICT OF INTEREST STATEMENT

Y.J., F.C.I., A.K.F., and N.Y.I. are inventors of the protein biomarker-related technology licensed to Cognitact. Y.J. and F.C.I. are co-founders of Cognitact. The remaining authors declare no competing interests. Author disclosures are available in the [supporting information](#).

DATA AVAILABILITY STATEMENT

The data and results of the genetic and Alzheimer's disease-associated endophenotypic analysis are provided in the *Supplementary Information*. For raw data, the consent form signed by each participant states that the research content will remain private under the supervision of the hospital and research team. Therefore, these data will be made available and shared only in the context of a formal collaboration; applications for data sharing and project collaboration will be processed and reviewed by a Review Panel hosted at the Hong Kong University of Science and Technology. Researchers may contact sklneurosci@ust.hk for further details on project collaboration and the sharing of the data from this study.

CONSENT STATEMENT

This study was approved by the Clinical Research & Ethics Committees of the Joint Chinese University of Hong Kong-New Territories East Cluster for the Prince of Wales Hospital (CREC Ref. No.

2015.461), the Kowloon Central Cluster/Kowloon East Cluster for Queen Elizabeth Hospital (KC/KE-15-0024/FR-3), Ethics and Research Committee of Haven of Hope Christian Service and the Human Participants Research Panel of The Hong Kong University of Science and Technology (CRP#180 & CRP#225). All participants provided written informed consent for both study participation and sample collection.

CODE AVAILABILITY

The code used in this manuscript is available upon request.

Figure legends

ORCID

Nancy Y. Ip  <https://orcid.org/0000-0002-2763-8907>

REFERENCES

- 2021 Alzheimer's disease facts and figures. *Alzheimers Dement.* 2021;17(3):327-406. doi: [10.1002/alz.12328](https://doi.org/10.1002/alz.12328)
- Hampel H, O'Bryant SE, Molinuevo JL, et al. Blood-based biomarkers for Alzheimer disease: mapping the road to the clinic. *Nat Rev Neurol.* 2018;14(11):639-652. [10.1038/s41582-018-0079-7](https://doi.org/10.1038/s41582-018-0079-7)
- Zetterberg H. Blood-based biomarkers for Alzheimer's disease-An update. *J Neurosci Methods.* 2019;319:2-6. [10.1016/j.jneumeth.2018.10.025](https://doi.org/10.1016/j.jneumeth.2018.10.025)
- Teunissen CE, Verberk IMW, Thijssen EH, et al. Blood-based biomarkers for Alzheimer's disease: towards clinical implementation. *Lancet Neurol.* 2022;21(1):66-77. [10.1016/S1474-4422\(21\)00361-6](https://doi.org/10.1016/S1474-4422(21)00361-6)
- Hansson O, Edelmayer RM, Boxer AL, et al. The Alzheimer's Association appropriate use recommendations for blood biomarkers in Alzheimer's disease. *Alzheimers Dement.* [10.1002/ALZ.12756](https://doi.org/10.1002/ALZ.12756). Published online 2022.
- Whelan CD, Mattsson N, Nagle MW, et al. Multiplex proteomics identifies novel CSF and plasma biomarkers of early Alzheimer's disease. *Acta Neuropathol Commun.* 2019;7(1):169. [10.1186/s40478-019-0795-2](https://doi.org/10.1186/s40478-019-0795-2)
- Shi L, Westwood S, Baird AL, et al. Discovery and validation of plasma proteomic biomarkers relating to brain amyloid burden by SOMAscan assay. *Alzheimers Dement.* 2019;15(11):1478. [10.1016/J.JALZ.2019.06.4951](https://doi.org/10.1016/J.JALZ.2019.06.4951)
- Jiang Y, Zhou X, Ip FC, et al. Large-scale plasma proteomic profiling identifies a high-performance biomarker panel for Alzheimer's disease screening and staging. *Alzheimers Dement.* 2022;18(1):88-102. [10.1002/alz.12369](https://doi.org/10.1002/alz.12369)
- Miller JA, Horvath S, Geschwind DH. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proc Natl Acad Sci U S A.* 2010;107(28):12698-12703. [10.1073/pnas.0914257107](https://doi.org/10.1073/pnas.0914257107)
- Wan YW, Al-Ouran R, Mangleburg CG, et al. Meta-analysis of the Alzheimer's disease human brain transcriptome and functional dissection in mouse models. *Cell Rep.* 2020;32(2):107908. [10.1016/j.celrep.2020.107908](https://doi.org/10.1016/j.celrep.2020.107908)
- Panitch R, Hu J, Xia W, et al. Blood and brain transcriptome analysis reveals APOE genotype-mediated and immune-related pathways involved in Alzheimer disease. *Alzheimers Res Ther.* 2022;14(1):1-14. [10.1186/S13195-022-00975-Z/FIGURES/3](https://doi.org/10.1186/S13195-022-00975-Z/FIGURES/3)
- Twine NA, Janitz K, Wilkins MR, Janitz M. Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. *PLoS One.* 2011;6(1):e16266. [10.1371/journal.pone.0016266](https://doi.org/10.1371/journal.pone.0016266)
- Xu M, Zhang DF, Luo R, et al. A systematic integrated analysis of brain expression profiles reveals YAP1 and other prioritized hub genes

- as important upstream regulators in Alzheimer's disease. *Alzheimers Dement*. 2018;14(2):215-229. [10.1016/j.jalz.2017.08.012](https://doi.org/10.1016/j.jalz.2017.08.012)
14. Wang M, Roussos P, McKenzie A, et al. Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease. *Genome Med*. 2016;8(1):104. [10.1186/s13073-016-0355-3](https://doi.org/10.1186/s13073-016-0355-3)
 15. Wang X, Allen M, Li S, et al. Deciphering cellular transcriptional alterations in Alzheimer's disease brains. *Mol Neurodegener*. 2020;15(1):38. [10.1186/s13024-020-00392-6](https://doi.org/10.1186/s13024-020-00392-6)
 16. Zhao Y, Xie YZ, Liu YS, Accelerated aging-related transcriptome alterations in neurovascular unit cells in the brain of Alzheimer's disease. *Front Aging Neurosci*. 2022;14:949074. doi: [10.3389/fnagi.2022.949074](https://doi.org/10.3389/fnagi.2022.949074)
 17. Frésard L, Smail C, Ferraro NM, et al. Identification of rare-disease genes using blood transcriptome sequencing and Irge control cohorts. *Nature Medicine*. 2019;25(6):911-919. [10.1038/s41591-019-0457-8](https://doi.org/10.1038/s41591-019-0457-8)
 18. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695-699. [10.1111/j.1532-5415.2005.53221.x](https://doi.org/10.1111/j.1532-5415.2005.53221.x)
 19. Arlington VA. Association, AP diagnostic and statistical manual of mental disorders. *Am Psychiatr Assoc*. 2013;5:612-613.
 20. Pangman VC, Sloan J, Guse L. An examination of psychometric properties of the mini-mental state examination and the standardized mini-mental state examination: implications for clinical practice. *Appl Nurs Res*. 2000;13(4):209-213.
 21. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):263-269.
 22. Albert MS, DeKosky ST, Dickson D, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):270-279.
 23. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120. [10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170)
 24. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
 25. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol*. 2019;20(1):1-13.
 26. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
 27. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res*. 2019;47(D1):D419-D426.
 28. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559. [10.1186/1471-2105-9-559](https://doi.org/10.1186/1471-2105-9-559)
 29. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12(1):1-8.
 30. Kassambara A, Mundt F. Package 'factoextra.' *Extract and visualize the results of multivariate data analyses*. 2017;76(2).
 31. Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*. 2019;37:773-782. <https://doi.org/10.1038/s41587-019-0114-2>
 32. Maechler M, Rousseeuw P, Croux C, et al. Package 'robustbase.' *Basic Robust Stat Available online: https://cran.r-project.org/web/packages/robustbase/index.html* (accessed on 10 October 2021). Published online 2022.
 33. Zhou X, Chen Y, Mok KY, et al. Non-coding variability at the APOE locus contributes to the Alzheimer's risk. *Nat Commun*. 2019;10(1):3310. [10.1038/s41467-019-10945-z](https://doi.org/10.1038/s41467-019-10945-z)
 34. Zhou X, Chen Y, Ip FCF, et al. Deep learning-based polygenic risk analysis for Alzheimer's disease prediction. *Commun Med*. 2023;3(1):49. doi:[10.1038/s43856-023-00269-x](https://doi.org/10.1038/s43856-023-00269-x)
 35. Jun G, Wing MK, Abecasis GR, Kang HM. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res*. 2015;25(6):918-925. doi:[10.1101/gr.176552.114](https://doi.org/10.1101/gr.176552.114)
 36. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120. doi:[10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170)
 37. Zhou X, Chen Y, Mok KY, et al. Identification of genetic risk factors in the Chinese population implicates a role of immune system in Alzheimer's disease pathogenesis. *PNAS*. 2018;115(8):1697-1706. doi:[10.1073/pnas.1715554115](https://doi.org/10.1073/pnas.1715554115)
 38. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*. 2007;81(5):1084-1097. doi:[10.1086/521987](https://doi.org/10.1086/521987)
 39. Li Y, Sidore C, Kang HM, Boehnke M, Abecasis GR. Low-coverage sequencing: implications for design of complex trait association studies. *Genome Res*. 2011;21(6):940-951. doi:[10.1101/gr.117259.110](https://doi.org/10.1101/gr.117259.110)
 40. Jiang Y, Zhou X, Ip FC, et al. Large-scale plasma proteomic profiling identifies a high-performance biomarker panel for Alzheimer's disease screening and staging. *Alzheimer Dement*. 2022;18(1):88-102.
 41. Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience*. 2020;9(12):giaa151.
 42. Butler A, Hoffman P, Smibert P, Papalexis E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*. 2018;36(5):411-420.
 43. Cano-Gamez E, Soskic B, Roumeliotis TI, et al. Single-cell transcriptomics identifies an effectorness gradient shaping the response of CD4+ T cells to cytokines. *Nat Commun*. 2020;11(1):1801. doi:[10.1038/s41467-020-15543-y](https://doi.org/10.1038/s41467-020-15543-y)
 44. Wigerblad G, Cao Q, Brooks S, et al. Single-cell analysis reveals the range of transcriptional states of circulating human neutrophils. *J Immunol*. 2022;209(4):772-782. doi:[10.4049/jimmunol.2200154](https://doi.org/10.4049/jimmunol.2200154)
 45. Zhou X, Chen Y, Ip FCF, et al. Genetic and polygenic risk score analysis for Alzheimer's disease in the Chinese population. *Alzheimers Dement*. 2020;12(1):e12074. doi:[10.1002/dad2.12074](https://doi.org/10.1002/dad2.12074)
 46. Jiang Y, Zhou X, Ip FC, et al. Large-scale plasma proteomic profiling identifies a high-performance biomarker panel for Alzheimer's disease screening and staging. *Alzheimers Dement*. 2022;18(1):88-102. doi:[10.1002/alz.12369](https://doi.org/10.1002/alz.12369)
 47. de Rojas I, Moreno-Grau S, Tesi N, et al. Common variants in Alzheimer's disease and risk stratification by polygenic risk scores. *Nat Commun*. 2021;12(1):3417. doi:[10.1038/s41467-021-22491-8](https://doi.org/10.1038/s41467-021-22491-8)
 48. Ren X, Wen W, Fan X, et al. COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell*. 2021;184(7):1895-1913. doi:[10.1016/j.cell.2021.01.053](https://doi.org/10.1016/j.cell.2021.01.053). e19.
 49. Zhou X, Fu AK, Ip NY. APOE signaling in neurodegenerative diseases: an integrative approach targeting APOE coding and noncoding variants for disease intervention. *Curr Opin Neurobiol*. 2021;69:58-67. doi:[10.1016/j.conb.2021.02.001](https://doi.org/10.1016/j.conb.2021.02.001)
 50. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A β , tau, immunity and lipid processing. *Nat Genet*. 2019;51(3):414-430. doi:[10.1038/s41588-019-0358-2](https://doi.org/10.1038/s41588-019-0358-2)
 51. Zicha S, Bateman RJ, Shaw LM, et al. Comparative analytical performance of multiple plasma A β 42 and A β 40 assays and their ability to

- predict positron emission tomography amyloid positivity. *Alzheimer's Dement.* 2023;19(3):956-966. doi:10.1002/alz.12697
52. Rauchmann BS, Schneider-Axmann T, Perneczky R, Alzheimer's Disease Neuroimaging Initiative (ADNI). Associations of longitudinal plasma p-tau181 and NfL with tau-PET, A β -PET and cognition. *J Neurol Neurosurg Psychiatry.* 2021;92(12):1289-1295. doi:10.1136/jnnp-2020-325537
 53. Hampel H, Nisticò R, Seyfried NT, et al. Omics sciences for systems biology in Alzheimer's disease: state-of-the-art of the evidence. *Ageing Res Rev.* 2021;69:101346. doi:10.1016/j.arr.2021.101346
 54. Peña-Bautista C, Baquero M, Vento M, Cháfer-Pericás C. Omics-based biomarkers for the early Alzheimer disease diagnosis and reliable therapeutic targets development. *Curr Neuropharmacol.* 2019;17(7):630-647. doi:10.2174/1570159X16666180926123722
 55. Sancesario GM, Bernardini S. Alzheimer's disease in the omics era. *Clin Biochem.* 2018;59:9-16. doi:10.1016/j.clinbiochem.2018.06.011
 56. Nativio R, Lan Y, Donahue G, et al. An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease. *Nat Genet.* 2020;52(10):1024-1035. doi:10.1038/s41588-020-0696-0
 57. Sood S, Gallagher IJ, Lunnon K, et al. A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status. *Genome Biol.* 2015;16(1):185. doi:10.1186/s13059-015-0750-x
 58. Maes OC, Xu S, Yu B, Chertkow HM, Wang E, Schipper HM. Transcriptional profiling of Alzheimer blood mononuclear cells by microarray. *Neurobiol Aging.* 2007;28(12):1795-1809. doi:10.1016/j.neurobiolaging.2006.08.004
 59. Panitch R, Hu J, Xia W, et al. Blood and brain transcriptome analysis reveals APOE genotype-mediated and immune-related pathways involved in Alzheimer disease. *Alzheimers Res Ther.* 2022;14(1):30. doi:10.1186/s13195-022-00975-z
 60. Shigemizu D, Mori T, Akiyama S, et al. Identification of potential blood biomarkers for early diagnosis of Alzheimer's disease through RNA sequencing analysis. *Alzheimers Res Ther.* 2020;12(1):87. doi:10.1186/s13195-020-00654-x
 61. 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
 62. Bettcher BM, Tansey MG, Dorothée G, Heneka MT. Peripheral and central immune system crosstalk in Alzheimer disease—a research prospectus. *Nat Rev Neurol.* 2021;17(11):689-701. doi:10.1038/s41582-021-00549-x
 63. Haage V, De Jager PL. Neuroimmune contributions to Alzheimer's disease: a focus on human data. *Mol Psychiatry* 2022;1-18. doi:10.1038/s41380-022-01637-0. Published online June 6, 2022.
 64. Kuyumcu ME, Yesil Y, Oztürk ZA, et al. The evaluation of neutrophil-lymphocyte ratio in Alzheimer's disease. *Dement Geriatr Cogn Disord.* 2012;34(2):69-74. doi:10.1159/000341583
 65. Zenaro E, Pietronigro E, Della Bianca V, et al. Neutrophils promote Alzheimer's disease-like pathology and cognitive decline via LFA-1 integrin. *Nat Med.* 2015;21(8):880-886. doi:10.1038/NM.3913
 66. Dong Y, Lagarde J, Xicota L, et al. Neutrophil hyperactivation correlates with Alzheimer's disease progression. *Ann Neurol.* 2018;83(2):387-405. doi:10.1002/ana.25159
 67. Shigemizu D, Mori T, Akiyama S, et al. Identification of potential blood biomarkers for early diagnosis of Alzheimer's disease through RNA sequencing analysis. *Alzheimers Res Ther.* 2020;12(1):1-12. doi:10.1186/s13195-020-00654-X/FIGURES/7
 68. Sayed A, Bahbah EI, Kamel S, Barreto GE, Ashraf GM, Elfil M. The neutrophil-to-lymphocyte ratio in Alzheimer's disease: current understanding and potential applications. *J Neuroimmunol.* 2020;349. doi:10.1016/J.JNEUROIM.2020.577398
 69. Smyth LCD, Murray HC, Hill M, et al. Neutrophil-vascular interactions drive myeloperoxidase accumulation in the brain in Alzheimer's disease. *Acta Neuropathol Commun.* 2022;10(1):38. doi:10.1186/s40478-022-01347-2
 70. Lortholary O, Chandesris MO, Bulai Livideanu C, et al. Masitinib for treatment of severely symptomatic indolent systemic mastocytosis: a randomised, placebo-controlled, phase 3 study. *Lancet.* 2017;389(10069):612-620. doi:10.1016/S0140-6736(16)31403-9
 71. Dubois B, López-Arrieta J, Lipschitz S, et al. Masitinib for mild-to-moderate Alzheimer's disease: results from a randomized, placebo-controlled, phase 3, clinical trial. *Alzheimers Res Ther.* 2023;15(1):39. doi:10.1186/s13195-023-01169-x
 72. Zieneldien T, Kim J, Sawmiller D, Cao C. The immune system as a therapeutic target for Alzheimer's disease. *Life.* 2022;12(9). doi:10.3390/LIFE12091440
 73. Bettcher BM, Tansey MG, Dorothée G, Heneka MT. Peripheral and central immune system crosstalk in Alzheimer disease — a research prospectus. *Nature Reviews Neurology.* 2021;17(11):689-701. doi:10.1038/s41582-021-00549-x
 74. Fu AKY, Hung KW, Yuen MYF, et al. IL-33 ameliorates Alzheimer's disease-like pathology and cognitive decline. *Proc Natl Acad Sci U S A.* 2016;113(19):E2705-E2713. doi:10.1073/PNAS.1604032113/-DCSUPPLEMENTAL
 75. Hampel H, Caraci F, Cuello AC, et al. A path toward precision medicine for neuroinflammatory mechanisms in Alzheimer's disease. *Front Immunol.* 2020;11:456. doi:10.3389/FIMMU.2020.00456/BIBTEX

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhong H, Zhou X, Uhm H, et al. Using blood transcriptome analysis for Alzheimer's disease diagnosis and patient stratification. *Alzheimer's Dement.* 2024;1-16. <https://doi.org/10.1002/alz.13691>