International Journal of General Medicine

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Identification of Plasma Cell Subsets and Molecular Markers of Alzheimer's Disease Based on Single-Cell Weighted Gene Co-Expression Network Analysis, Mendelian Randomization Analysis and Clinical Validation

Chao Xin[1,2], Hong-Wei Zhi[3], Da-Hua Wu[4], Peng-Li Ding[1], Zhong-Lin Wang[1,3], Ya-Han Wang[4]

[1]Shandong University of Traditional Chinese Medicine, Jinan, Shandong, People's Republic of China; [2]Affiliated Hospital of Shandong Academy of Traditional Chinese Medicine, Jinan, Shandong, People's Republic of China; [3]Shandong University of Traditional Chinese Medicine Affiliated Hospital, Jinan, Shandong, People's Republic of China; [4]Hunan Provincial Hospital of Integrated Traditional Chinese and Western Medicine, Changsha, Hunan, People's Republic of China

Correspondence: Zhong-Lin Wang, Affiliated Hospital of Shandong Academy of Traditional Chinese Medicine, Jingshi Road No. 16369, Lixia District, Jinan, Shandong, People's Republic of China, Tel +86-0531-6861-6039, Email doctorzhonglinwang@outlook.com; Ya-Han Wang, Hunan Provincial Hospital of Integrated Traditional Chinese and Western Medicine, Lushan Road No. 58, Yuelu District, Changsha, Hunan, People's Republic of China, Tel +86-15811105935, Email wyahan621@outlook.com

Introduction: Alzheimer's disease (AD) is a neurodegenerative disorder characterized by gene expression alterations and immune dysregulation.

Methods: In this study, we downloaded the GSE181279 single-cell dataset from the Gene Expression Omnibus (GEO) and applied bioinformatic analysis and clinical subject validation.

Results: After quality control and harmony integration, we identified 21 cell subsets, including T cells, B cells, plasma cells, and macrophages. Plasma cells were significantly elevated in AD patients, and six plasma cell subtypes were associated with AD. High-dimensional weighted gene co-expression network analysis (hdWGCNA) revealed two AD-related modules. Mendelian randomization identified *RGS1* as a key risk gene ($p = 0.0123$). Immune infiltration analysis showed *RGS1* negatively correlated with macrophages and positively with tumor-infiltrating lymphocytes. Functional enrichment indicated that *RGS1* is involved in JAK-STAT, NF-κB, and Wnt-β-catenin signaling pathways, suggesting a role in immune regulation and neuroinflammation. Furthermore, validation in AD patients confirmed that *RGS1* expression levels were higher than in controls ($p < 0.01$).

Discussion: This study identified the key gene *RGS1* related to AD and explored multiple signaling pathways associated with it, which provided important clues for the research on AD-related inflammation, gut microbiota, stretch-gated ion channel, and the evaluation of AD therapeutic targets.

Clinical Trial Registration Number: CTR20210477.

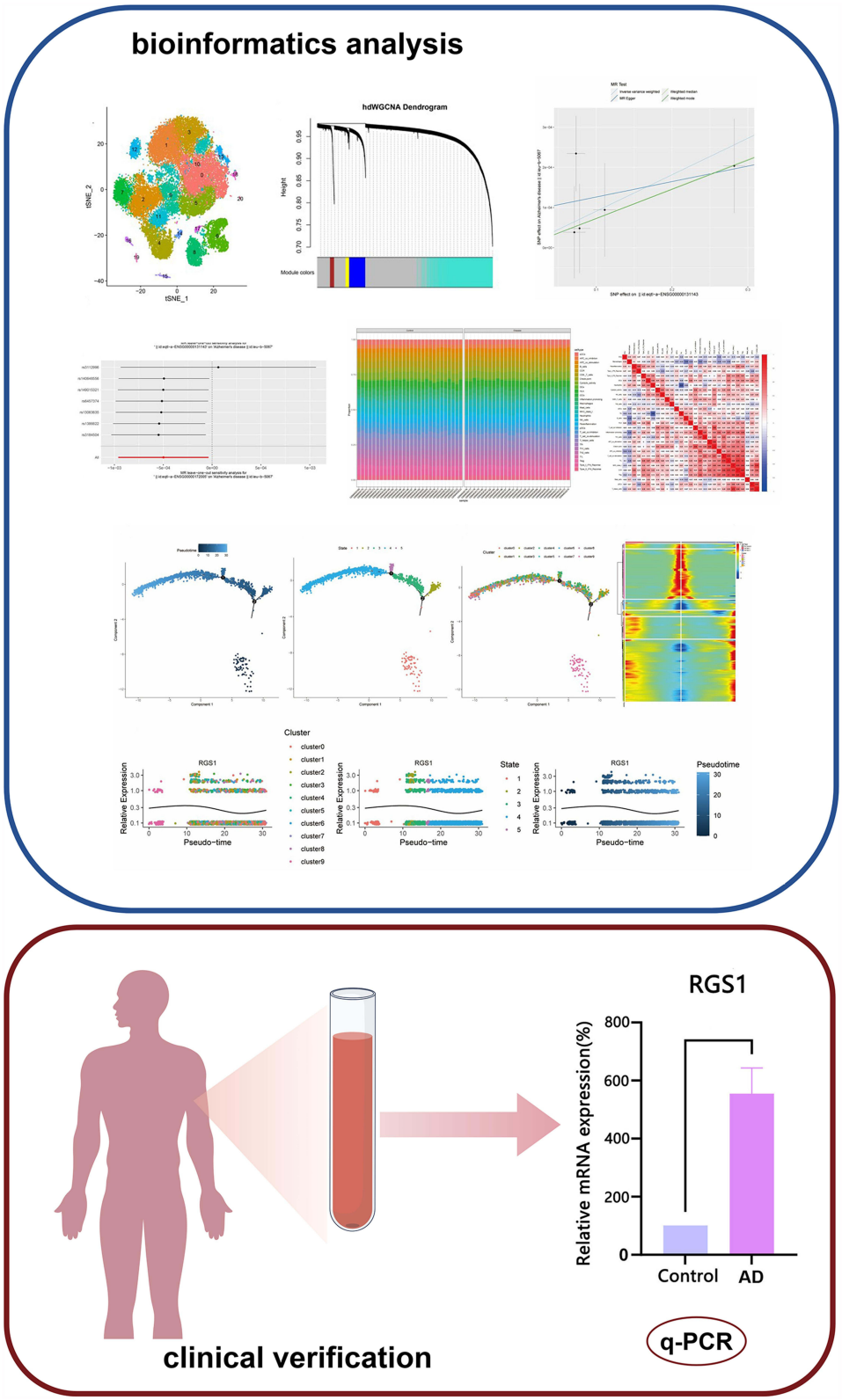
Keywords: alzheimer's disease, mendelian randomization, single-cell analysis, key genes, clinical validation

Background

Alzheimer's disease (AD) is a neurodegenerative disease characterized by memory loss as the first symptom. Its primary pathological features include amyloid beta (Aβ) plaque deposits and neurofibrillary tangles (NFTs) in the brain, both of which eventually lead to neuronal damage and death. Inflammatory responses play an important role in memory impairment. Activated macrophages and astrocytes produce inflammatory mediators that lead to inflammatory damage and neuronal death. The pathological manifestations of AD have been reported to appear decades before the onset of symptoms.[1–3] In recent years, the treatment of AD has shifted from single symptomatic intervention to multi-target

Received: 10 May 2025
Accepted: 9 September 2025
Published: 18 September 2025

Graphical Abstract



intervention, and new drugs such as Aβ monoclonal antibodies have brought breakthroughs. However, the current therapies cannot reverse the disease and they can only delay its progression. Moreover, the intervention timing is generally too late, and there is a lack of effective early diagnosis and prevention strategies. Therefore, it is important to explore the pathological mechanisms of AD to identify new targets for the development of novel therapeutic strategies.

Resolving the genetic and molecular mechanisms of complex diseases is a challenging but crucial task. Bioinformatics analysis of microarray data has been widely used to identify key genes for subsequent analyses.[4] With the advancement of technology, methods such as single-cell sequencing and Mendelian randomized analysis have gradually become important tools for revealing disease mechanisms. Recent evidence suggests that monocytes infiltrate the brain and differentiate into macrophage-like cells, further enhancing the inflammatory response. Monocytes, as key immune cells, have received much attention due to their involvement in neuroinflammatory processes and AD progression. Advances in single-cell sequencing technology have made it possible to identify cell subtypes associated with disease at high resolution, providing valuable insights into disease mechanisms. By examining gene expression at the single-cell level, it is possible to pinpoint genes that are specifically upregulated in monocytes and assess their role in AD pathology. Mendelian randomization (MR) is an analysis of genetic variables that follows the Mendelian laws of inheritance, which use single nucleotide polymorphisms (SNPs) as instrumental variables to infer the observed causal relationship between modifiable exposure and clinically relevant outcomes. Single-cell sequencing and Mendelian randomization have their respective advantages, it is of great innovative significance to combine the two in research. By identifying key genes that are highly expressed in specific cell subtypes through single-cell sequencing, we can further evaluate the causal relationship between these genes and disease using Mendelian randomization analysis. This study explored target genes in the pathological process of AD based on bioinformatics analysis and MR, and further verified their potential molecular mechanisms through in vivo experiments. A workflow diagram of this study is shown in Figure 1a.

Methods

Bioinformatics Analysis

Data Acquisition

The Series Matrix File of GSE153104 was downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) and annotated as GPL16791. The expression profile data of 55 patients were included, with 28 and 27 patients in the control and AD groups, respectively. The single-cell data file for GSE181279, which contains data from five samples with complete single-cell expression profiles, was downloaded. Expression quality loci (eQTL) data were obtained from the eQTLGen Consortium database (https://www.egtlgen.org) to serve as exposure data. The outcome data were from genome-wide association studies (GWAS) pooled from the MR Base database (ieu-b-5067). There were 954 cases in the AD group and 487,331 cases in control group, respectively. All participants in the GWAS selected for this study were of European descent. Data source is shown in Table 1.

Single-Cell Analysis

First, the expression profile was read using the Seurat package and low-expression genes were screened (nFeature_RNA > 200 and nFeature_RNA < 4000 and percent.mt < 8 and nCount_RNA < 20,000). The data were standardized, homogenized, and analyzed using principal component analysis (PCA). According to the elbow plot, the optimal number of principal components (pc) was 20. The location relationship between each cluster was obtained through t-distributed stochastic neighbor embedding (t-SNE). Each cluster was annotated using known cell markers that were assigned to cells important for disease occurrence.

High-Dimensional Weighted Gene Co-Expression Network Analysis (hdWGCNA)

SetupForWGCNA was used to construct a co-expression network of genes expressed in at least 5% of cells in Seurat objects, with a soft threshold of 6. The weighted gene co-expression network analysis (WGCNA) tree was visualized using a PlotDendrogram, and the module Eigengene (ME) level of each module was visualized using GetMEs. The n_hub was set to 100 to obtain the module genes.

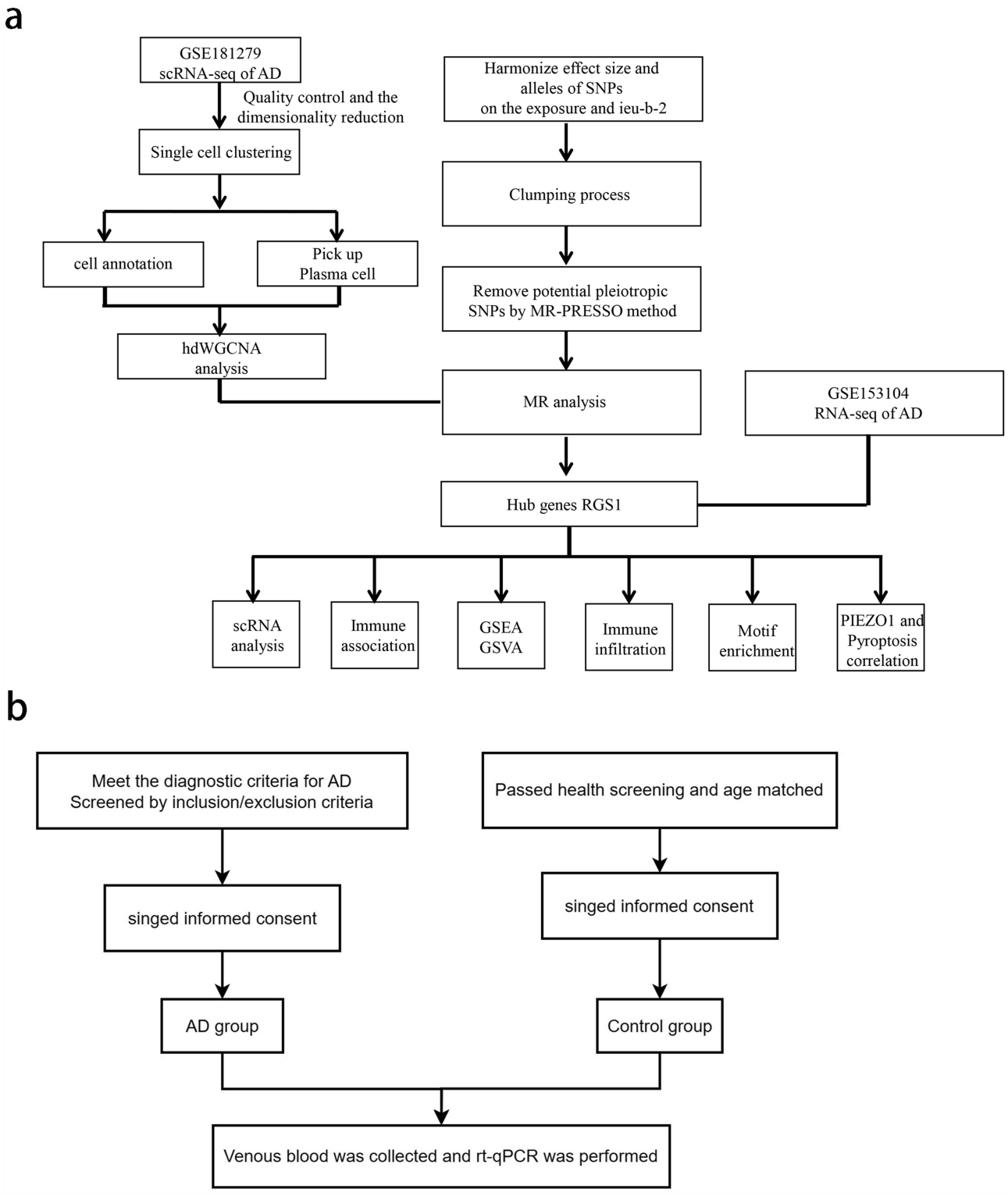


Figure 1 Flow charts for bioinformatics analysis and clinical validation. (**a**) Bioinformation analysis flow chart. (**b**) Clinical validation flow chart.

Functional Annotation

Functional annotation of the module genes was performed using the Metascape database (www.metascape.org) to explore their functional correlations. Gene Ontology (GO) pathway analysis was used to identify specific genes. Min overlap ≥ 3 and p values ≤0.01 were considered statistically significant.

Table 1 Data Source

Data Source	Setting	Participants	Genetic Variation Measurement, Quality Control, and Selection	Ethics Approval and Informed Consent
ieu-b-5067	Data from the IEU OpenGWAS project, with Alzheimer's disease (AD) data, collected in 2022.	954 AD cases and 487,331 controls.	SNPs extracted from GWAS summary data, used as instrumental variables for Mendelian randomization (MR) analysis.	Data obtained with ethics committee approval, all participants provided informed consent.
ebi-a-GCST90012877	Data from the EBI GWAS Catalog, with Alzheimer's disease data, collected in 2022.	472,868 participants.	Relevant SNPs selected from GWAS summary data for replication and validation in external cohorts.	Data obtained with ethics committee approval, all participants provided informed consent.
GSE153104	Data from Gene Expression Omnibus (GEO) database, including 55 AD patients (28 controls, 27 AD cases), data from 2020.	55 patients, divided into control and AD groups.	Data cleaning and quality control performed using standardized methods.	Data obtained with ethics committee approval, all participants provided informed consent.
eQTLGen	Data from eQTLGen Consortium database, with epigenetic data from 2019.	Mainly individuals of European ancestry.	Relevant genes selected from eQTL data, with quality control applied.	Data obtained with ethics committee approval, all participants provided informed consent.
Clinical Study	Conducted at a hospital health check center, participants aged 50 to 85, data collected in 2023.	10 AD patients and 10 healthy controls.	Total RNA extracted using TRIzol reagent, gene expression analyzed by RT-qPCR.	Study approved by ethics committee, all participants signed informed consent forms.

Mendelian Randomization (MR) Analysis

The outcome ID screened using the MRBase database (http://app.mrbase.org/) was used to extract relevant causality in the GWAS summary data (https://gwas.mrcieu.ac=/) in the eQTL. The SNPs associated with each gene in the whole locus significance threshold ($p < 1e{-}5$) were selected as potential instrumental variables. The linkage disequilibrium between SNPs was calculated and only SNPs with $p2 < 5e{-}5$ were retained among the SNPs with $R2 < 0.001$ (clumping window size = 10,000 kb). Inverse-variance weighted (IVW), MR-Egger, weighted median, and weighted mode (or Wald ratio if there was only one SNP in the causal relationship) were used to evaluate the reliability of the causal relationship and obtain an overall estimate of the effect of all cis isomers and cross-regional gene expression on AD using whole blood samples. Sensitivity analysis was performed on the MR data using the leave-one-out method to assess the effect of specific genetic variants on the risk of AD and to determine the robustness of the results.

Immune Cell Infiltration Analysis

The immune cells in the expression profile were quantified using a single-sample gene set enrichment analysis (ssGSEA) algorithm to infer the relative proportions of the 29 types of immune-infiltrating cells.

Gene Set Enrichment Analysis (GSEA)

Based on the level of gene expression, patients were divided into high- and low-expression groups, and differences in signaling pathways between the two groups were further analyzed using gene set enrichment analysis (GSEA). Background gene sets were downloaded from the Molecular Signatures Database (MSigDB) as annotated gene sets for subtype pathways. Differential expression analysis of pathways between subtypes was performed and significantly enriched gene sets (adjusted $p < 0.05$) were sorted according to their consistency scores.

Gene Set Variation Analysis (GSVA)

Gene sets were downloaded from MSigDB v7.0, and each gene set was comprehensively scored using the GSVA algorithm to evaluate potential biological and functional changes in different samples.

Analysis of Regulatory Networks of Important Genes

We used the RcisTarget package with the RcisTarget hg19 MotifDB CisbpOnly, a 500 bp database. This database contains motifs in the 500 bp range of the human genome (hg19) and includes only motifs in the cisBP database. The

area under the curve (AUC) was calculated for all the motifs. The AUC represents the degree of enrichment of a gene set under a particular motif. Next, we compared the AUC of all motifs in the gene set with random background distributions. From this comparison, we obtained a normalized enrichment score (NES) for each motif. NES is a standardized enrichment score that reflects the relative importance of a motif in a gene concentration. Finally, using NES values, we identified motifs that were significantly enriched in specific gene sets, and thus inferred possible transcription factors.

Clinical Validation

Diagnostic Criteria

The diagnosis of AD was based on the AD Core Criteria (National Institute on Aging-Alzheimer's Association [NIA-AA], 2011):[5] (1) latent onset with a clear history of progressive deterioration of cognitive function; (2) history and examination confirming one of the earliest and most significant categories of cognitive impairment (amnestic and non-amnestic symptoms); and (3) exclusion criteria.

Inclusion Criteria

The inclusion criteria were: (1) age $\geq$ 50 and $\leq$ 85 years; (2) met the NIA-AA 2011 diagnostic criteria for AD; (3) Clinical Dementia Rating (CDR) score > 0.5; (4) modified Hachinski Ischemia Scale (m-HIS) total score < 4; (5) Hamilton Depression Scale (HAMD) (24 items) < 8, Hamilton Anxiety Scale (HAMA) < 7; (6) head magnetic resonance imaging (MRI) showing hippocampal atrophy consistent with AD; (7) no obvious positive signs on nervous system examination; (8) availability of a stable and reliable caregiver to accompany the patient and assist the researchers with completing the HAMD and HAMA; and (9) provision of written informed consent by the patients and their families.

Exclusion Criteria

The exclusion criteria were as follows: (1) presence of other neurological diseases that can cause brain dysfunction; (2) presence of systemic diseases that can cause cognitive impairment; (3) presence of a condition that makes the patient unable to cooperate with the completion of cognitive tests; (4) presence of contraindications to MRI; (5) presence of mental retardation or neurological impairment; and (6) refusal to provide a blood sample.

Quality Control in Clinical Research

Inclusion and exclusion criteria were established to minimize clinical study bias. The healthy control group comprised patients screened at the Health Examination Center of our hospital. Patients in the control group were required to have a CDR = 0, and were similar to patients in the AD group in terms of age, sex, HAMA and HAMD score. Brain information was collected using 1.5T/3.0T magnetic resonance imaging (MRI), and MR reports were issued by two professional doctors in the radiology department of our hospital. Participants and their families were assessed by a qualified neuropsychologist in a quiet setting, followed by screening by two neurologists (including at least one lead physician) according to the inclusion and exclusion criteria.

After participants were enrolled, venous blood was collected by a nurse, and peripheral blood mononuclear cells (PBMCs) were extracted. This process is illustrated in Figure 1b.

This study was conducted in accordance with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of the First Clinical Medical School of Shandong University of Chinese Medicine (Date: 2024/01/23/No.AF/SC-08/03.0).

Quantitative Reverse Transcription Polymerase Chain Reaction (RT-qPCR)

Total RNA was extracted from 20 PBMC samples, from ten patients with AD and ten healthy controls using TRIzol reagent according to the manufacturer's instructions.[6] RNA (1 μL) was detected with a nanometer photometer N50. Reverse transcription of mRNA was performed using the SureScript First-Strand cDNA Synthesis Kit (GeneCopoeia Inc., Rockville, MD, USA).[6] Amplification and dissolution curves were drawn, and cycle threshold (Ct) values were determined. The primer sequences are listed in Table 2.

Finally, the relative expression was calculated using the $2^{-\Delta\Delta}$Ct method. Finally, the $2^{-\Delta\Delta}$Ct value was calculated and the p-value was calculated using GraphPad Prism 9 (GraphPad Software, San Diego, CA, USA).

Table 2 Primers

Primers	Sequence
H-RGS1-239 F	GAGTTCTGGCTGGCTTGTGA
H-RGS1-239 R	GTGAGGAACCTGGGATAAGAGTC
GAPDH F	CGAAGGTGGAGTCAACGGATTT
GAPDH R	ATGGGTGGAATCATATTGGAAC

Statistical Analysis

The biological information analysis was conducted in R language (version 4.2). In order to control the false positive rate caused by multiple hypothesis tests, the mr_heterogeneity() function was used to directly calculate the Q value of the heterogeneity test. The p. adjust method was used to control the false discovery rate (FDR). For functional annotation and differential gene analysis, a q value of <0.05 was defined as statistically significant. In addition, in gene enrichment analysis (such as GSEA, GSVA) and hdWGCNA analysis, we used a p value <0.05 as a significant threshold to ensure the robustness and reliability of the results. SPSS 19.0 (IBM Corp., Armonk, NY, USA) and GraphPad Prism (version 9.4.1; GraphPad Software, San Diego, CA, USA) were also used to perform statistical analyses and data mapping. The statistical significance of the differences in sex between groups was assessed with the Fisher's exact test. Age, CDR score, HAMA score, HAMD score and relative expression of mRNA were normally distributed, and an independent-sample *t*-test was used. P* < 0.05 were considered statistically significant, with p** < 0.01 flagged as highly significant.

Results

Gene Subpopulation Annotation

Single-cell data from GSE181279 included five samples. In this analysis, nFeature_RNA and nCount_RNA were used to screen the data samples (nFeature_RNA > 200, nFeature_RNA < 4000, percentage mt < 8, and nCount_RNA < 20,000) (Supplementary Figure S1A and B), and the ten genes with the highest standard deviations are shown (Supplementary Figure S1C). The data were successively analyzed using standardization, homogenization, PCA, and harmony (Supplementary Figure S1D–F), and 21 subgroups were identified using t-SNE analysis (Figure 2a). Twenty-one clusters were annotated as T cells, $CD8^+$ T cells, $CD4^+$ T cells, natural killer (NK) cells, B cells, plasma cells, monocyte-macrophages, and megakaryocyte progenitor cells (Figure 2b). In addition, we analyzed the bubble diagram (Figure 2c) of the classical markers of these eight cell types and the cell ratio histogram of the two groups (Figure 2d) and found that the proportion of plasma cell subtypes was higher in the AD samples than in the control samples. Subsets of plasma cell subtypes were successively analyzed using PCA, Harmony, ElbowPlot, and FindClusters to obtain ten subtypes (Figure 3a–d). The proportions of subtypes C0, C2, C4, C5, C6, and C7 were higher in the AD samples than in the control samples (Figure 3e). Thus, these six subtypes were defined as the disease subtypes.

High-Dimensional Weighted Gene Co-expression Network Analysis (hdWGCNA) of Plasma Cell Subsets

To determine the co-expression networks of genes in plasma cell subsets, hdWGCNA was performed. The groups were set as the cluster subtype and custom subtype, which were used to construct the co-expression network and explore biomarkers for AD development. The soft threshold was set to six (Figure 4a). Four gene modules were detected (yellow, turquoise, brown, and blue) (Figure 4b). Further analysis of the four modules and ME levels (Figure 4c and d) showed that the ME levels in the brown and blue modules were higher than those in the AD cell subtypes (Figure 4e). Therefore, the module genes obtained from analysis of the brown and blue modules were selected as candidate genes for subsequent studies.

Mendelian Randomization (MR) Analysis

We further analyzed the pathways of brown and blue MEs, and the results showed that blue MEs were mainly enriched in the intrinsic apoptotic signaling pathway and negative regulation of mRNA metabolic processes, whereas brown MEs

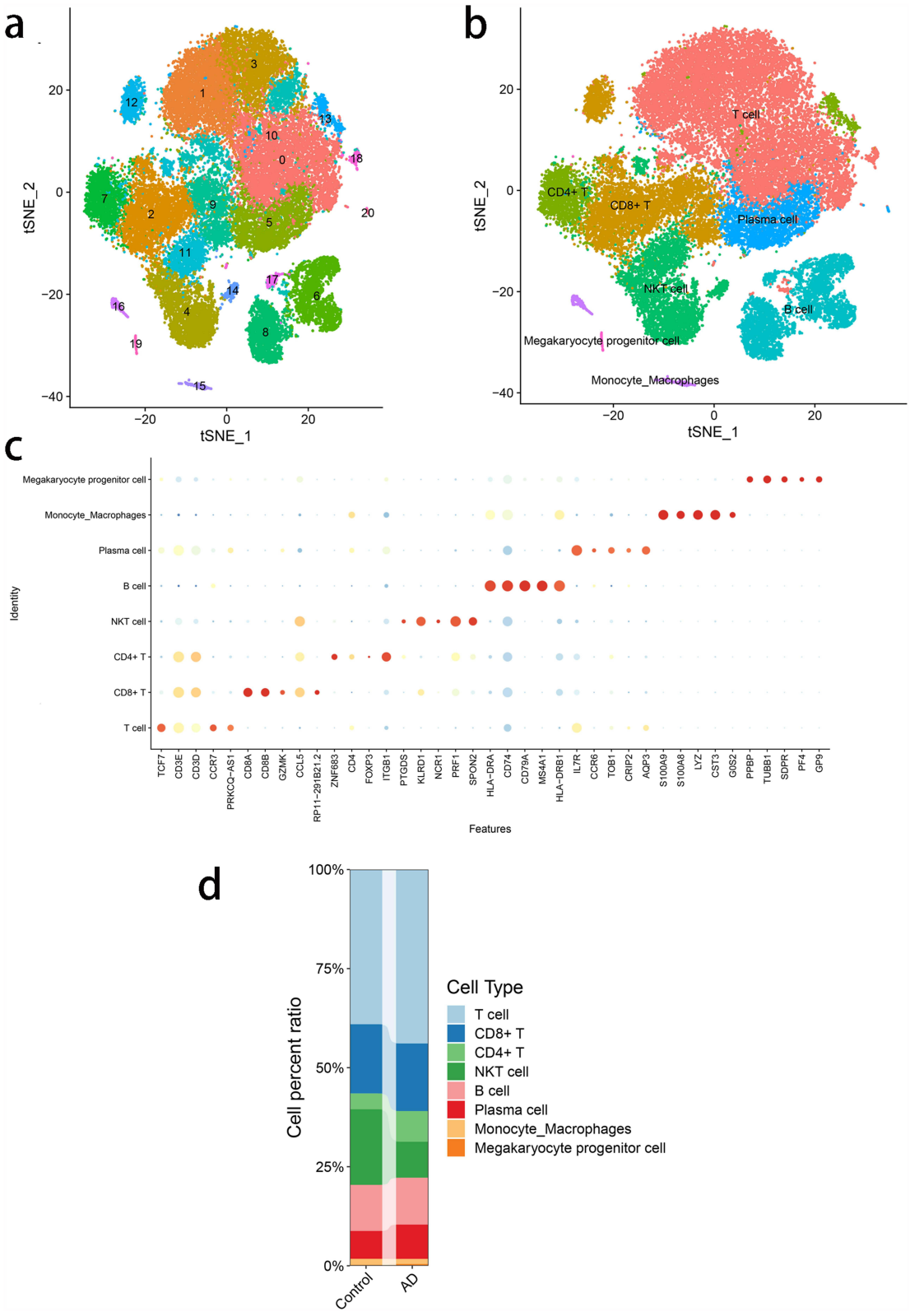


Figure 2 Cell annotation (**a**) We divided the cells into 21 clusters with the t-distributed stochastic neighbor embedding (t-SNE) algorithm based on the significant components available in the principal component analysis (PCA). X-axis/Y-axis: coordinates of tSNE dimensionality reduction, used to show the cluster distribution of different subtypes after dimensionality reduction. (**b**) Cell annotation of 21 clusters. The 21 clusters were classified into eight cell types: megakaryocyte progenitor cell, monocyte macrophages, plasma cell, B cell, NK T cell, $CD4^+$ T cell, $CD8^+$ T cell, and T cell. X-axis/Y-axis: coordinates of tSNE dimensionality reduction, used to show the clustering distribution of different cells after dimensionality reduction. (**c**) Dot-plot of the eight cell types. X-axis: cell markers; Y axis: Cell type, bubble color depth indicates the amount of gene expression, size indicates the degree of significance. (**d**) Difference in the proportion of the eight cell types in the Alzheimer's disease (AD) group and the control group. X-axis: Grouping; Y-axis: cell content percentage.

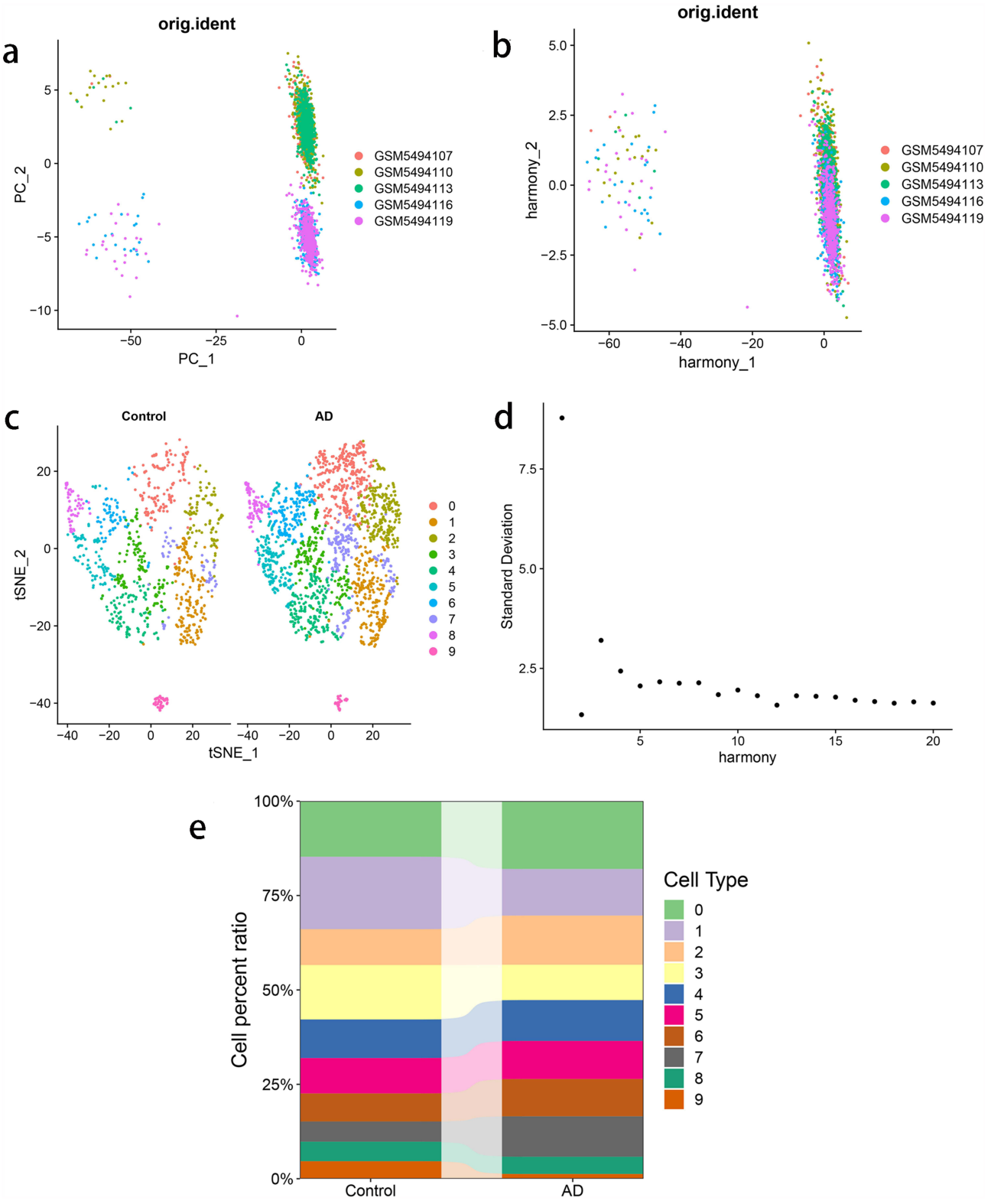


Figure 3 Secondary clustering of cells (**a** and **b**) Display of the principal component analysis (PCA) and the distribution of the principal components (PC). The points represent cells, and the colors represent samples. X /Y axis of A: The two principal components (PC1 and PC2) after PCA reduction to show the position of the cell in the principal component space projected in the data. X/Y axis of B: The two main components of HarmonY after dimensionality reduction to show the distribution of cells after integration. (**c**) We divided the cells into ten clusters using the t-distributed stochastic neighbor embedding (t-SNE) algorithm based on the significant components available in the PCA. X-axis/Y-axis: coordinates of tSNE dimensionality reduction, used to show the cluster distribution of different subtypes after dimensionality reduction. (**d**) Variance ordering plot for each PC. X-axis: Principal component (PC), each of which explains a different part of the variance of the data; Y-axis: Standard deviation of each principal component. (**e**) Difference in the proportion of the ten cell types in the samples of Alzheimer's disease (AD) group and control group. X-axis: Grouping; Y-axis: cell content percentage.

https://doi.org/10.2147/IJGM.S539547

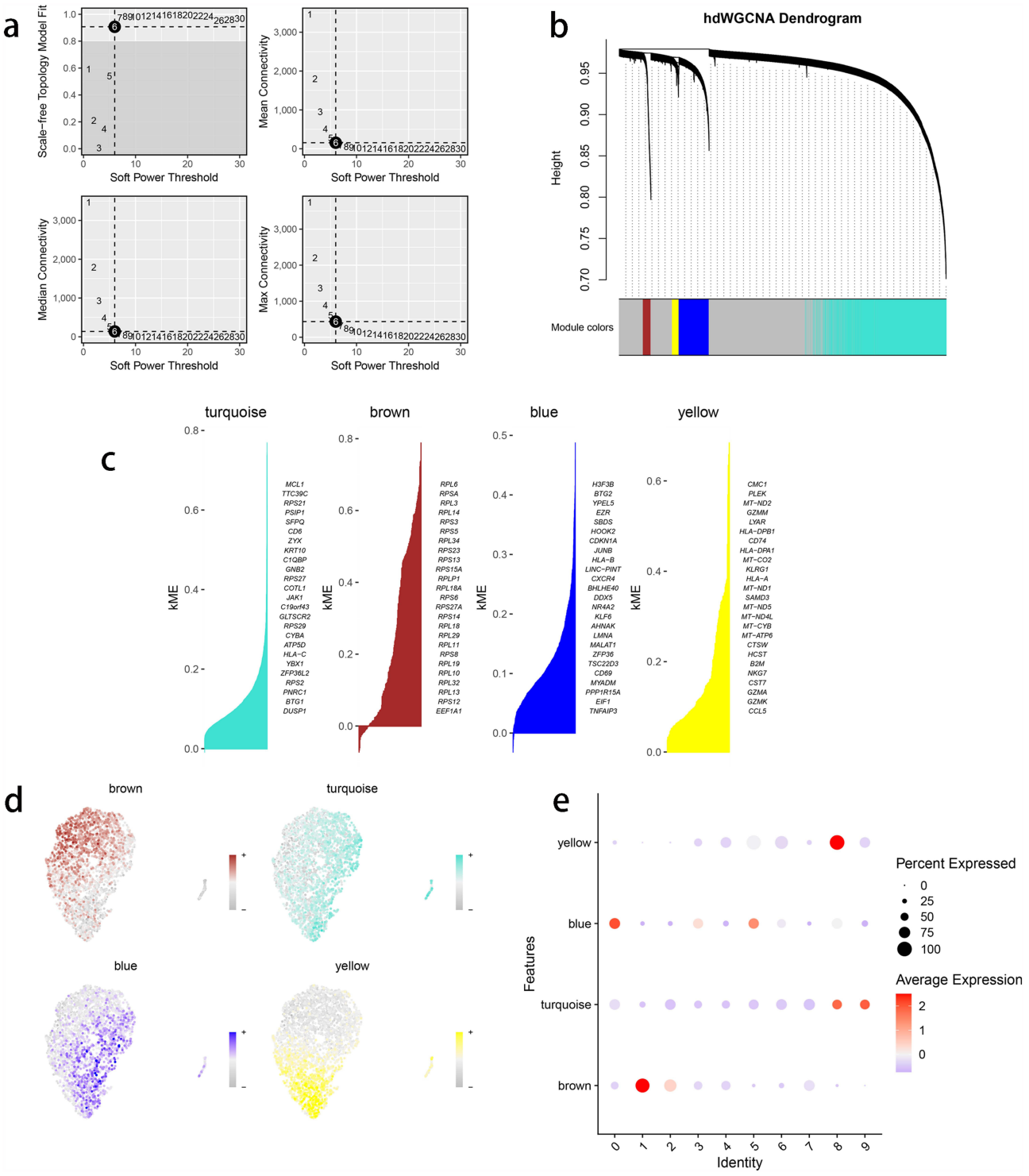


Figure 4 Single-cell weighted gene co-expression network analysis (WGCNA) (**a**) Scale-free index and average connectivity of each soft threshold. X axis (Soft Thresholding Power): indicates the selected soft threshold. Y-axis: Indicates whether the network conforms to the fit of the scale-free topology. (**b**) Tree of gene clusters, with different colors representing different modules. (**c** and **d**) Feature maps of top-ranking gene and module scores in different modules. (**e**) Correlation between module characteristic genes.

were mainly concentrated in the regulation of translation and peptidase activator activity involved in the apoptotic process (Supplementary Figure S2A and B).

To further identify the key genes that influence AD, we used 200 genes in the brown and blue modules to analyze 488,285 records from 954 patients with AD and 487,331 controls. Extract_instruments and extract_outcome_data were

used to extract 171 causal associations between genes and outcomes (using ieu-b-5067 from the MR Base database) (Supplementary Table S1). Furthermore, the causal relationship between the four genes, *COX4I1, MAL, RGS1*, and *RPS5*, corresponding to eQTL positive outcomes was screened using MR analysis (Figure 5a–d, IVW $p < 0.05$). *MAL* (0.9995; 95% confidence interval [CI]: 0.9990−1.0000; $p = 0.0347$) and *RPS5* (0.9995; 95% CI: 0.9991−0.9999; $p = 0.0230$) were associated with a lower risk of AD, whereas *COX4I1* (1.0009; 95% CI: 1.0002−1.0016; $p = 0.0095$) and *RGS1* (1.0020; 95% CI: 1.0004−1.0037; $p = 0.0123$) were associated with a higher risk of AD. A sensitivity analysis of the causal relationships between the four genes was performed to determine the reliability of the MR analysis. The results showed that the effect on the overall error line was not significant after the removal of any SNP, indicating that the four selected causality pairs were robust (Figure 6a–d). Then, with the aggregated statistical data of 472,868 samples related to AD in

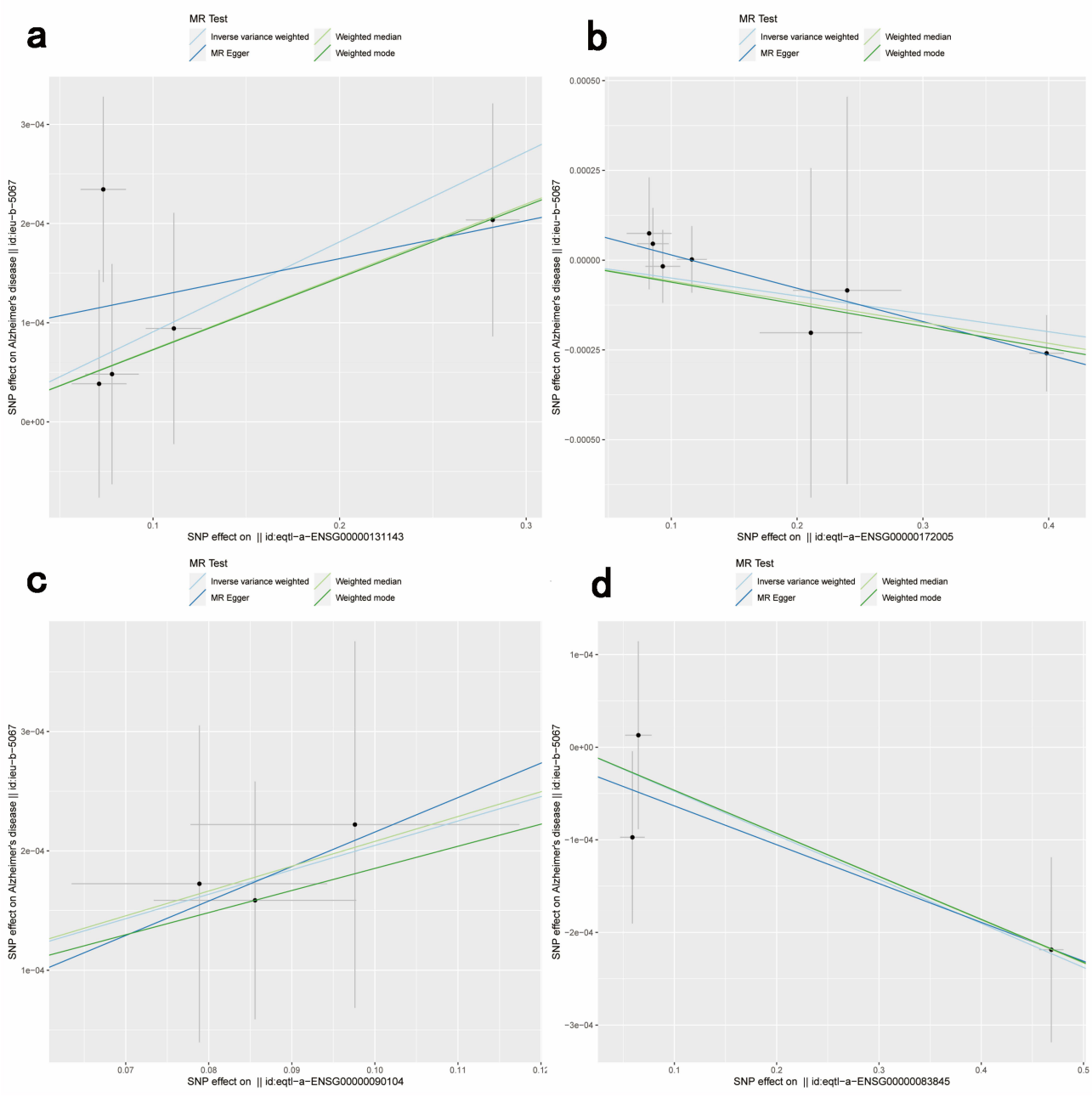


Figure 5 Mendelian randomization (MR) analysis (**a–d**) Scatter plots of the MR analysis of key genes. Different colors indicate different statistical methods, and the slope of the lines indicates the causal effect of each method. The X-axis represents the beta value of the instrumental variable in exposure, and the Y-axis represents the beta value of the instrumental variable in outcome.

https://doi.org/10.2147/IJGM.S539547

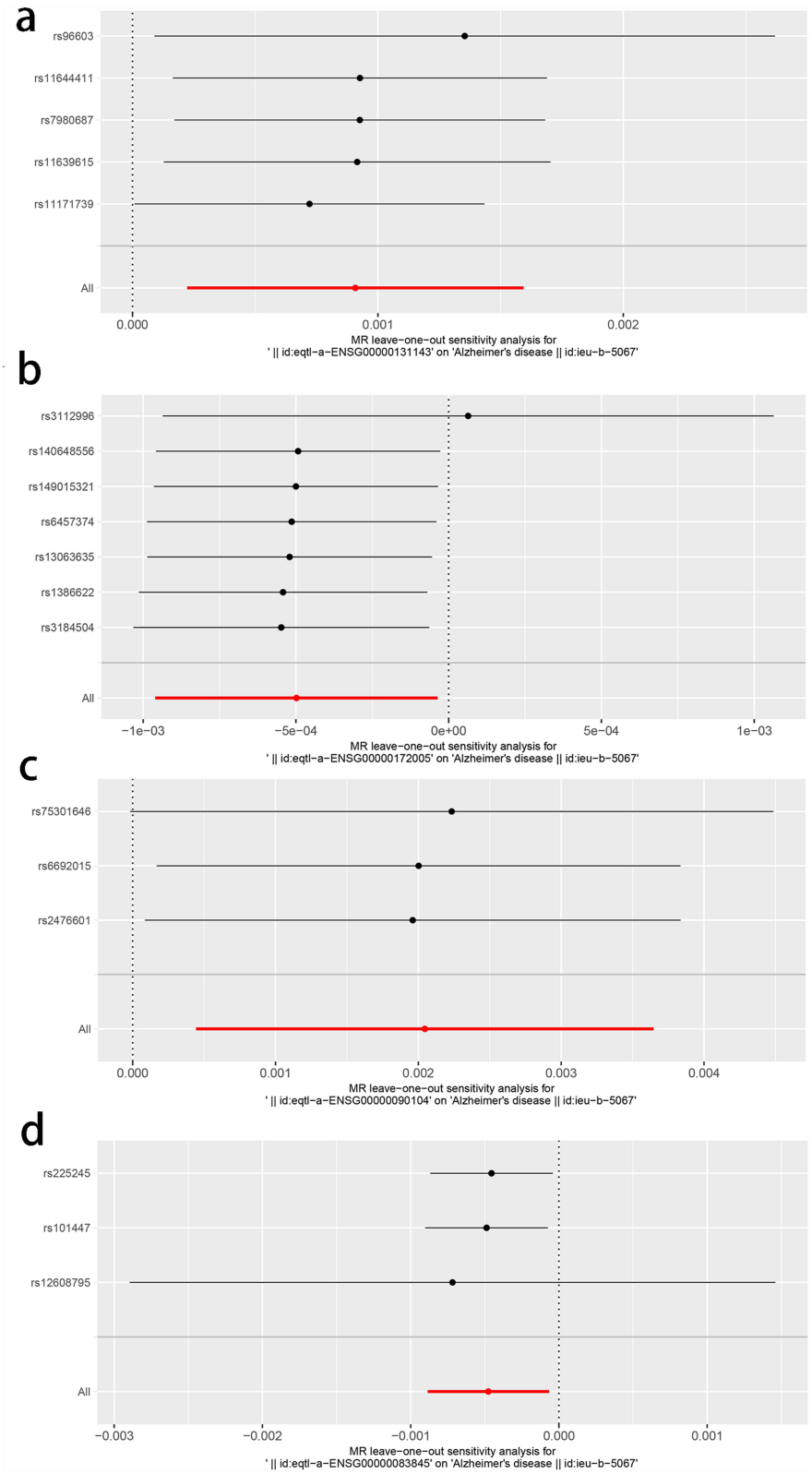


Figure 6 Leave-one-out test (**a–d**) Forest map of leave-out tests for single-nucleotide polymorphisms (SNPs) corresponding to key genes. The X axis represents beta values and the y axis represents tool variables.

the validation set, the outcome id was ebi-a-GCST90012877. Extract_instruments and extract_outcome_data were used to extract the causal relationship between genes and outcome, and the causal relationship corresponding to eQTL positive outcome was further screened by Mendelian randomization analysis (IVW pval < 0.05) for 1 pair of genes (Figure 7). Gene *RGS1* (1.2211; 1.0338–1.4424; p= 0.0187) may be associated with a higher risk of Alzheimer's disease. Therefore, *RGS1* will be a key gene for subsequent analysis in this study.

Immunoinfiltration Analysis

The microenvironment is mainly composed of immune cells, the extracellular matrix, various growth factors, inflammatory factors, and special physicochemical characteristics that significantly affect the diagnosis and clinical treatment sensitivity of

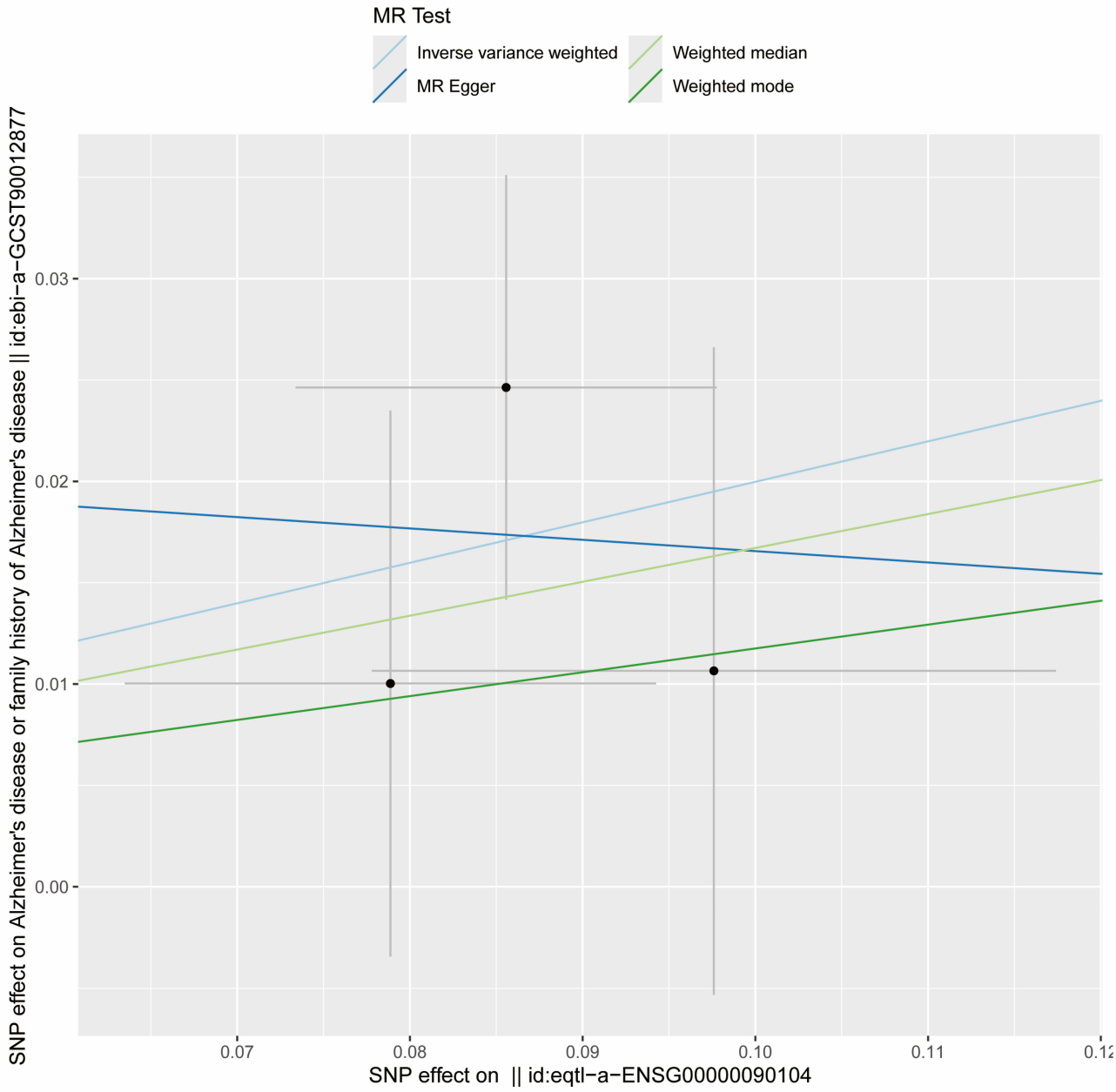


Figure 7 Mendelian randomization analysis *RGS1* was associated with a higher risk of Alzheimer's disease (p= 0.0187). The X-axis represents the beta value of the instrumental variable in exposure, and the Y-axis represents the beta value of the instrumental variable in outcome.

https://doi.org/10.2147/IJGM.S539547

diseases. By analyzing the relationship between key genes and immune infiltration in the AD dataset, we further explored the potential molecular mechanisms by which key genes influence AD progression. This study showed the proportion of immune cells in each patient and the correlation between different types of immune cells (Figure 8a and b). In addition, the results showed that APC_co_stimulation and $CD8^+$ T cells were significantly increased in the control and AD groups (Figure 8c). We further explored the relationship between key gene and immune cells and found that key gene was highly correlated with immune cells (Figure 8d). *RGS1* was negatively correlated with macrophages and positively correlated with tumor infiltrates lymphocytes (TIL). In this study, the correlation between this key gene and different immune factors, including immunomodulators, chemokines and cell receptors, was obtained from the TISIDB database (Figure 9a–e). These analyses suggest that key genes are closely related to the level of immune cell infiltration and play an important role in the immune microenvironment.

Gene Set Enrichment Analysis (GSEA) and Gene Set Variation Analysis (GSVA)

Next, the specific signaling pathways involving the key gene were studied to explore the potential molecular mechanisms by which these genes affect AD progression. The GSEA results showed that the pathways enriched by *RGS1* included the JAK-

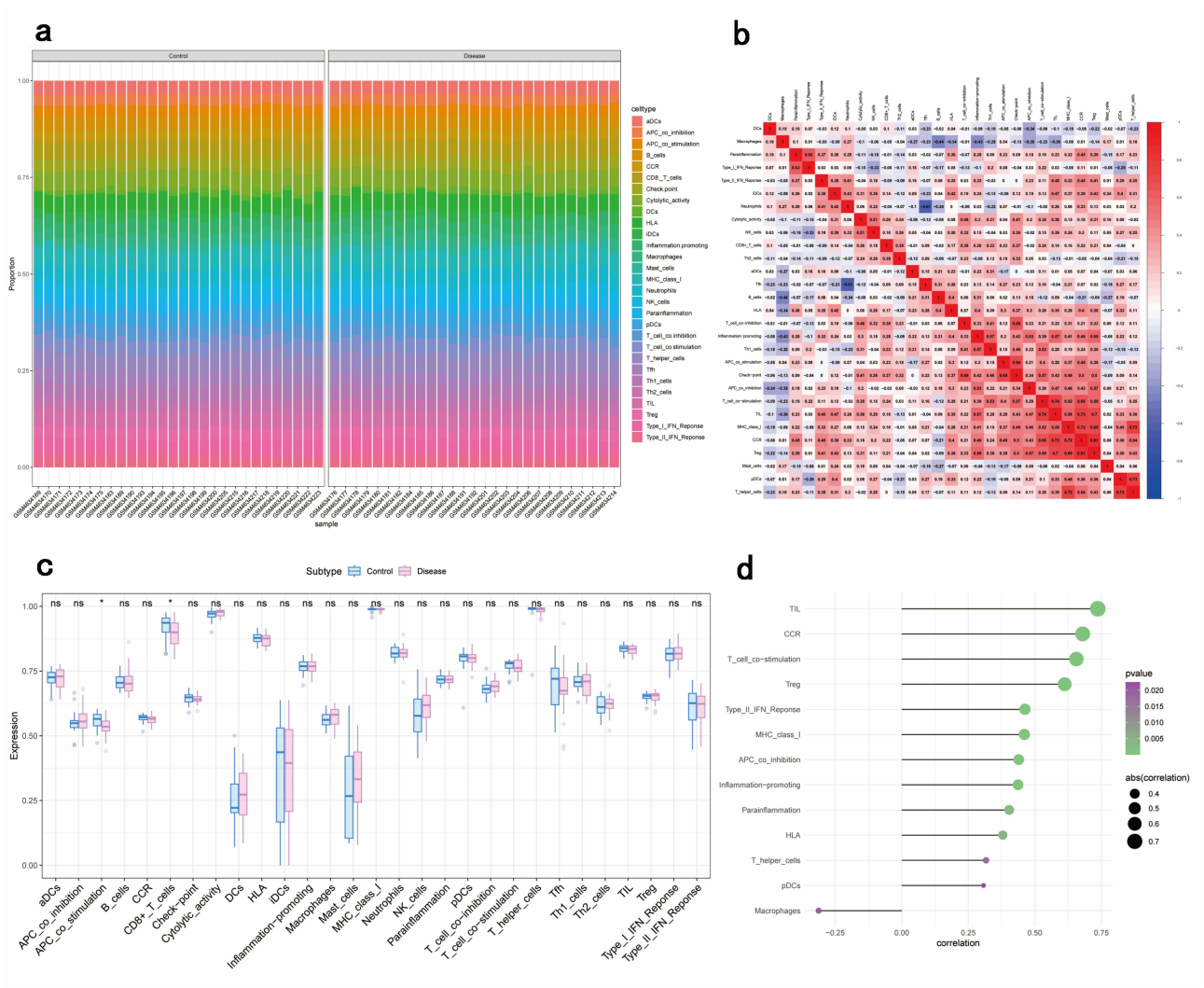


Figure 8 Immunoinfiltration analysis (**a**) Relative percentage of 29 immune cell subpopulations. The X-axis represents the name of the sample; the Y-axis represents the relative proportion of different cell types in each sample. (**b**) Pearson correlations among 29 immune cell subpopulations. Negative correlations are shown in blue and positive correlations are shown in red. The X and Y axes represent immune cell types. (**c**) Differences in immune cell content between control and disease samples. The X-axis shows the type of immune cells; the Y-axis represents the quantity of expression. (*p < 0.05) (**d**) The correlation between key gene and immune cells. The X-axis represents the correlation coefficient; the Y-axis represents the quantity of expression.

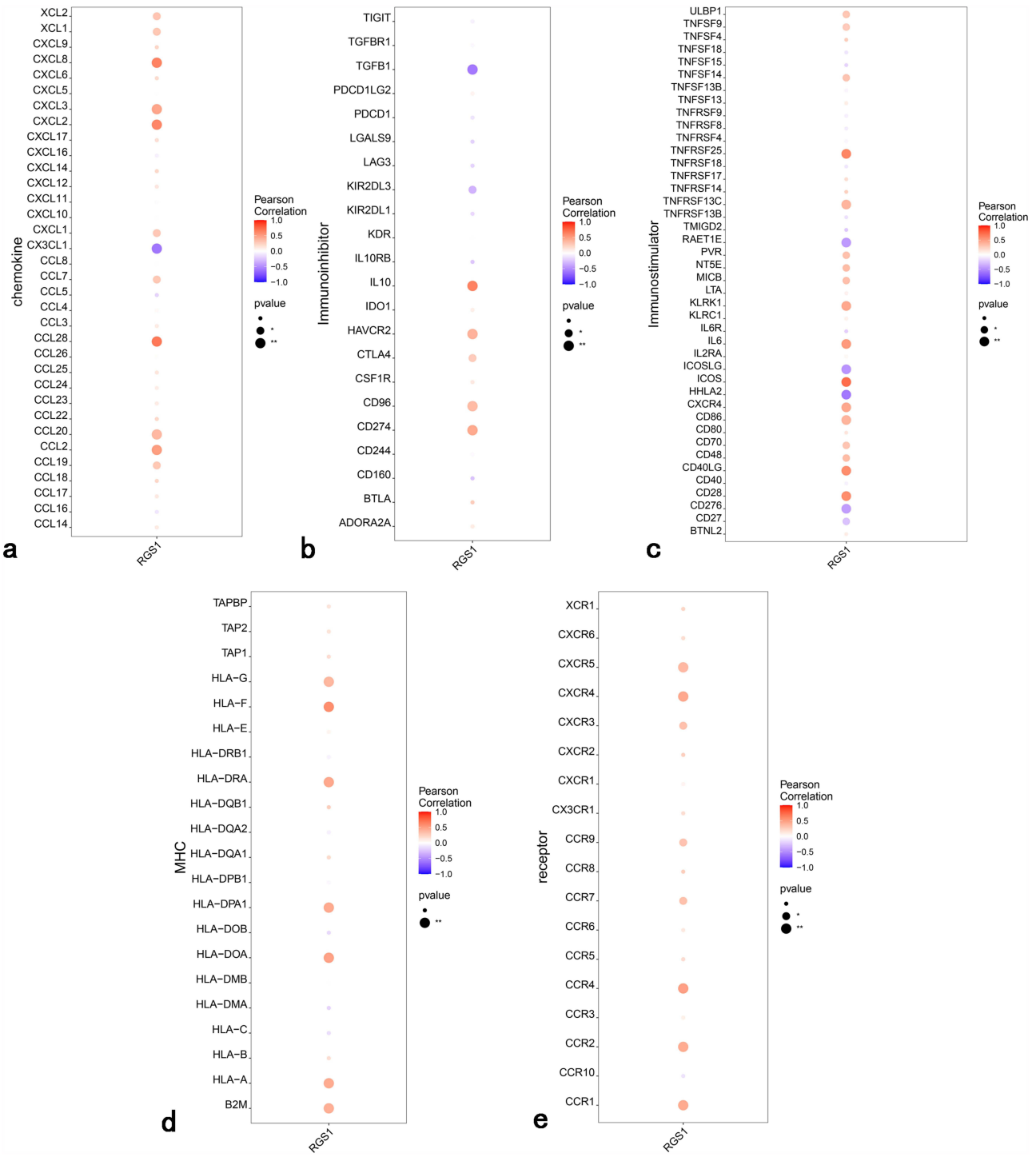


Figure 9 Relationship between key gene and immune factors (**a**–**e**) Correlation of key gene with chemokines, immunoinhibitors, immunostimulators, major histocompatibility complexes (MHC) and receptors. X-axis: Key genes; Y-axis: immune factors, bubble color indicates correlation coefficient, size indicates degree of significance.

STAT and NF-kappa B signaling pathways (Figure 10a and b). GSVA showed high expression of *RGS1* enriched the Wnt-β-catenin and mTORC1 signaling pathways (Figure 10c).

Prediction and Expression Patterns of Key Gene Transcription Factors

In this study, key genes were used for the gene set in this analysis, and they were found to be regulated by common mechanisms such as multiple transcription factors. Therefore, these transcription factors were enriched by cumulative

https://doi.org/10.2147/IJGM.S539547

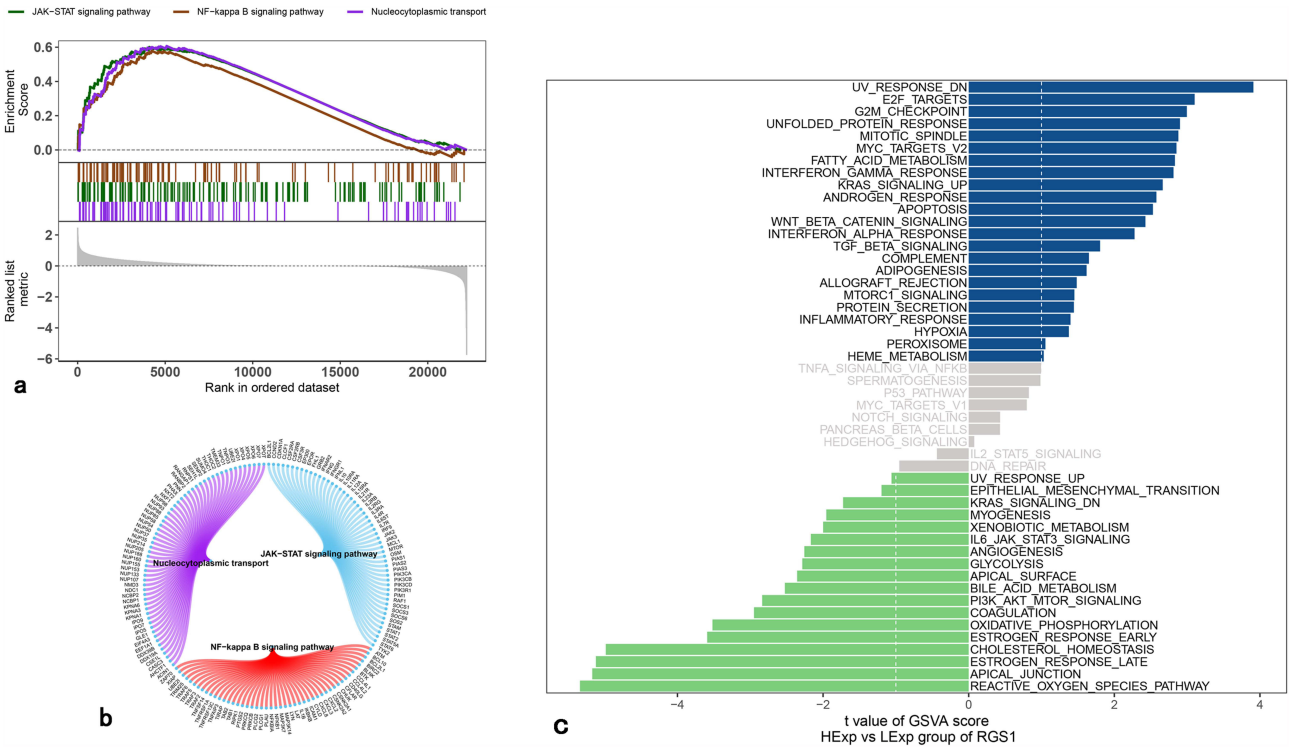


Figure 10 Gene set enrichment analysis (GSEA) analysis and gene set variation analysis (GSVA) analysis of key genes (**a** and **b**) Kyoto Encyclopedia of Genes and Genomes (KEGG) signaling pathways involving key gene, pathway regulation, and gene involved. X-axis: Indicates the position of the gene in the overall gene set. Y-axis: represents enrichment score. (**c**) GSVA analysis of key genes. The signaling pathways involved in high and low gene expression are shown in blue and green, respectively. The hallmark gene set was used as the background gene set. The X-axis represents enrichment fraction and the Y-axis represents enrichment pathway.

recovery curves. Motif-tf annotation and selection analysis of important genes showed that the Motif Motif with the highest standardized enrichment score (NES: 4.74) was cisbp__M1404. In this study, all the enriched motifs of key genes and corresponding transcription factors were shown (Figure 11a and b).

With GeneCards database (https://www.genecards.org/), we acquired Pyroptosis related genes, and the expression of *GZMB* and *IFI27* showed significant difference between AD patients and control group (Figure 11c). In addition, correlation analysis results of key genes and Pyroptosis related genes showed a significant correlation, in which *RGS1* and *NLRP3* were significantly positively correlated (r= 0.785), and *RGS1* and *PYCARD* were significantly negatively correlated (r=−0.71) (Figure 11d).

In addition, we analyzed the key gene in T cells, $CD8^+$ T cells, $CD4^+$ T cells, NK T cells, B cells, plasma cells, monocyte-macrophages, and megakaryocyte progenitor cells in a single cell (Figure 12a and b). Next, we quantified the immune and metabolic pathways using ssGSEA and analyzed and demonstrated the correlation between the key gene and the immune and metabolic pathways (Figure 12c). We also visualized the co-expression of AD-related regulatory genes (*APOE, APP*, and *PSEN1*) and key gene in the eight cell types (Supplementary Fig. S3). Finally, we analyzed the correlation between key genes and *PIEZO1*, and found that *RGS1* had a significant correlation with *PIEZO1* (Figure 13).

In order to obtain the developmental trajectories of key cell subtypes, we calculated the similarity between cells and constructed the cell differentiation trajectories. Then, with visualizing the tracks, we generated a cell differentiation track map constructed in pseudo-time to show the cell development process, which was used to study the process of cell differentiation and gene expression patterns at different time points. Images of cells colored with pseudotime values, celltype, and state were output respectively (Figure 14a–c). Genes were differentially expressed in different branches, but this could not be shown in the overall heat map. We selected all the branch points, calculated the genes with large differences in cell expression before and after the branch points, and visualized them as the branch heat map (Figure 14d). Then we looked at the expression changes of key genes with cell state, etc. (Figure 14e).

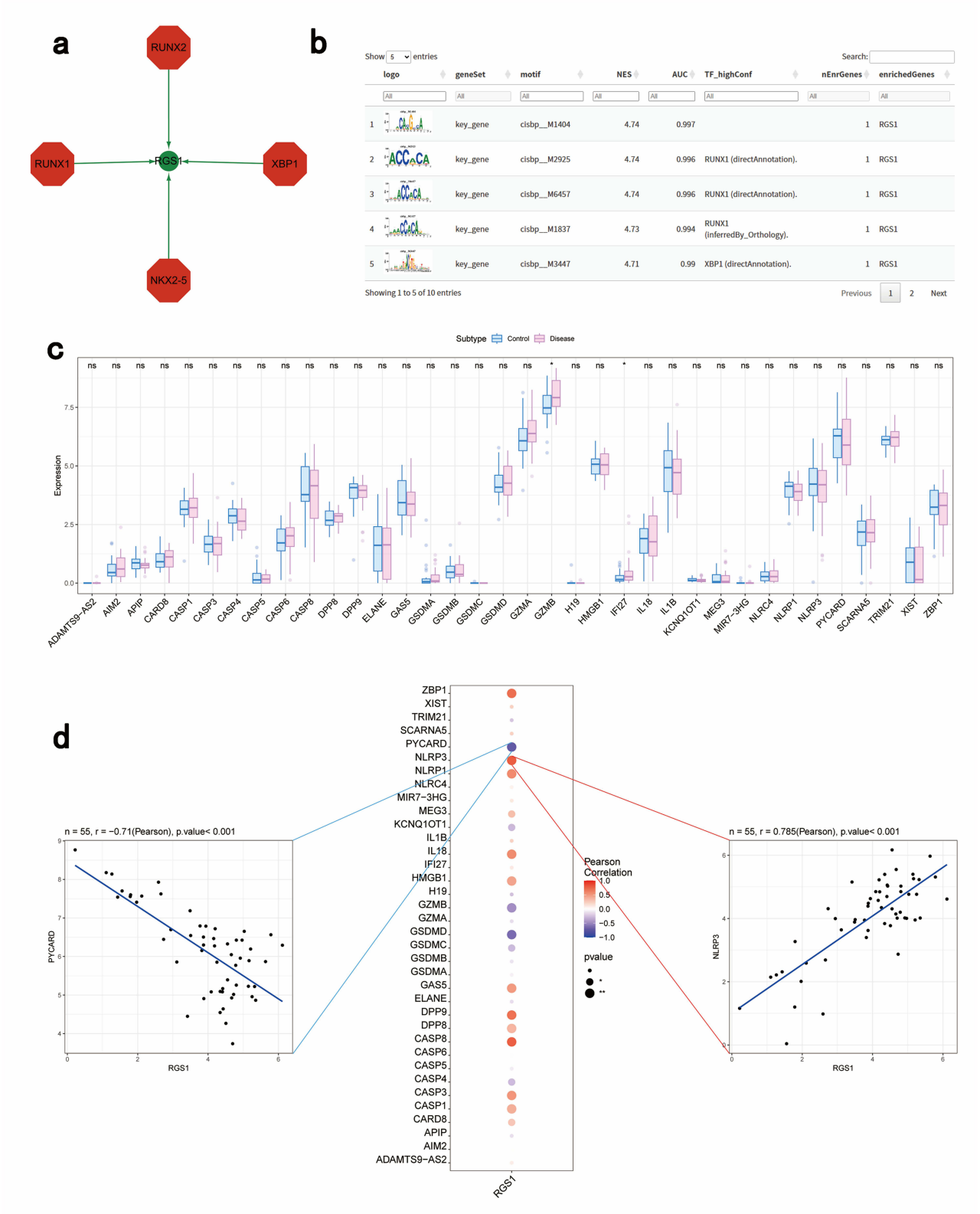


Figure 11 Correlation between key gene-related transcriptional regulation and disease progression genes (**a**) Transcriptional regulatory networks of key gene, with green representing key gene and red representing transcription factors. (**b**) All the enriched motifs and corresponding transcription factors of key gene are displayed. (**c**) Differences in the expression of disease-regulating gene, with blue representing control patients and pink representing patients with Alzheimer's disease (AD). The X-axis represents genes and the Y-axis represents gene expression values. (*p < 0.05) (**d**) Correlation between key gene and AD regulatory genes. Bubble plot and scatter plot. Red bubbles indicate a positive correlation and blue bubbles indicate a negative correlation. The X-axis represents key genes, the Y-axis represents PYroptosis genes, the bubble size represents significance, and the bubble color depth represents the correlation coefficient.

https://doi.org/10.2147/IJGM.S539547

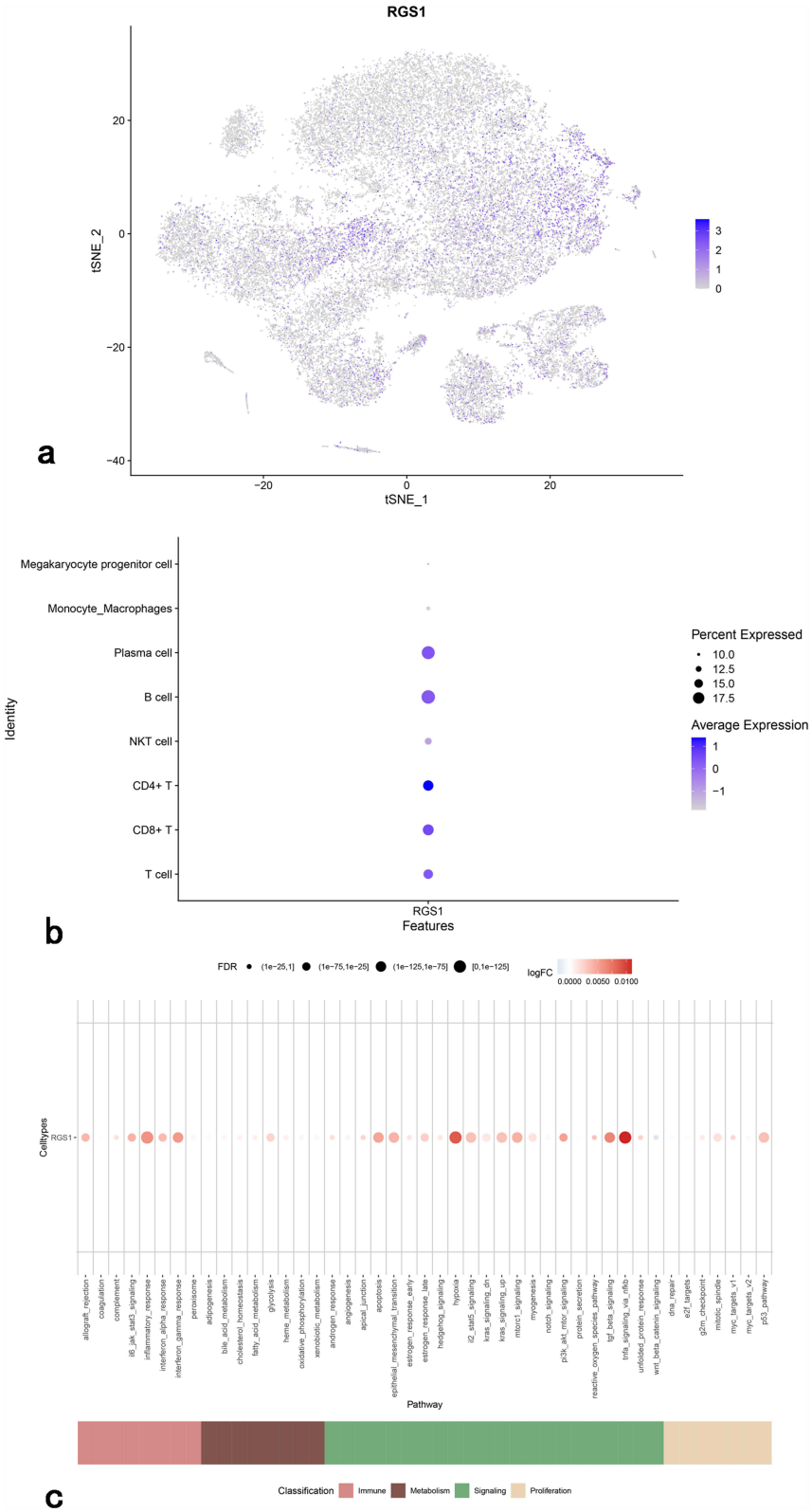


Figure 12 Expression profile of single cells (**a** and **b**) Overview of key gene expression in single cells. X /Y axis of (**a**) Coordinates of tSNE dimensionality reduction, used to show the cluster distribution of key cells after dimensionality reduction. The X axis of B represents key genes, the Y axis represents cells, the size of the bubble represents significance, and the depth of the bubble color represents the amount of gene expression. (**c**) Correlation of key gene with immune metabolic pathways. The X-axis represents immune metabolic pathways, the Y-axis represents key genes, bubble size represents significance, and bubble color depth represents activity.

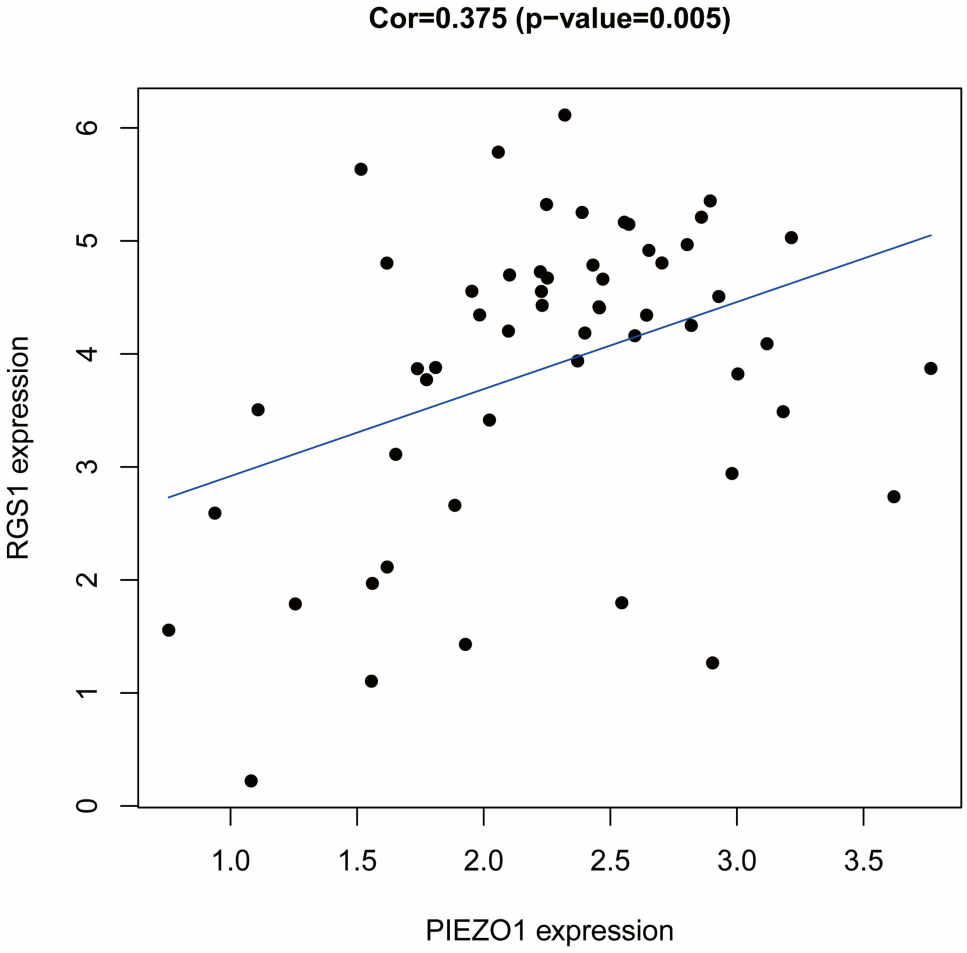


Figure 13 Correlation analysis of key gene Association of key gene with *PIEZO1*. X-axis (gene1): represents the expression of the gene"'gene'". Y-axis (gene2): represents the expression of the gene"'gene'".

Validation of Key Gene Expression in Patients with AD

To further validate the results of the biometric analysis, we screened and tested patients who visited our hospital. Ten patients with AD and ten healthy controls were screened. Baseline information is presented in Table 3. There were no significant differences in age, sex, HAMA score or HAMD score between the two groups. The CDR scores of patients with AD were higher than those of the control group ($p < 0.01$). The RT-qPCR results (Figure 15) showed that the relative mRNA expression levels of *RGS1* was higher in patients with AD than in the control group ($p < 0.01$), consistent with the results of the bioinformatics analysis.

Discussion

With the development of an aging society, the incidence of AD is increasing exponentially and has become a burden for many families and societies. The development of AD is related to genetics, brain damage and environmental factors. Monoclonal antibody drugs recently approved for the treatment of AD, aducanumab and lecanemab, primarily act to clear Aβ deposits. The formation of amyloid plaques and neurofibrillary tangles resulting from tau hyperphosphorylation are characteristic pathological manifestations of AD. In clinical diagnosis, FDG-PET and the detection of amyloid and p-tau protein levels in cerebrospinal fluid are meaningful.[7] However, these tests are difficult to popularize in the clinic due to their invasive and high cost. With advances in high-throughput sequencing and microarray technology, bioinformatics is increasingly being used to analyze genetic alterations in the nervous system. Currently, researchers are working

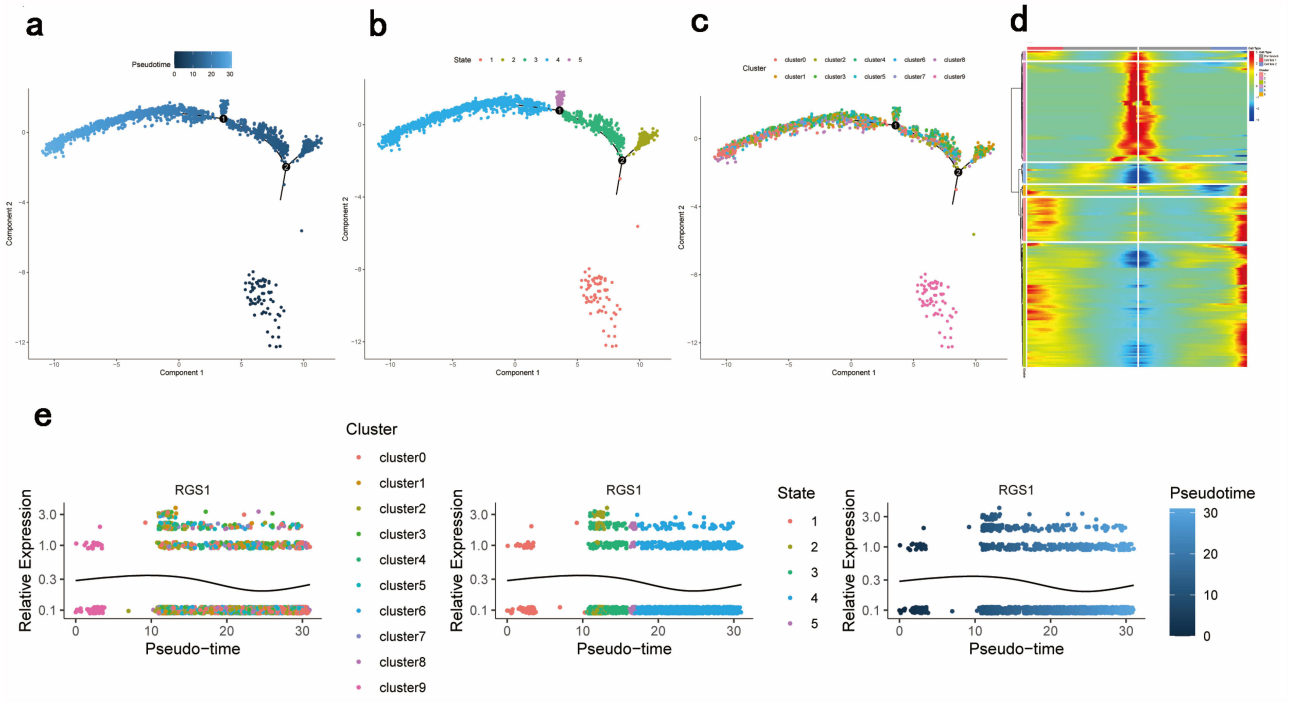


Figure 14 Quasi-temporal analysis of plasma cells (**a**) Pseudo time tracks of plasma cells generated by Monocle. Color gradients indicate progress along pseudo-time, with light blue indicating an early state and dark blue indicating a late state. X/Y-axis represent the cell position calculated by dimensionality reduction algorithm, and the color represents the continuous change of the cell along the pseudo-time sequence from early state to late state. (**b**) Cell states along a pseudo-time trajectory. Plasma cells are divided into five different states, marked with different colors, indicating the progression of cell differentiation or functional changes. X/Y-axis represent the cell position calculated by dimensionality reduction algorithm, and the color represents each state, usually representing a key stage or branch point along the cell development trajectory. (**c**) Cluster distribution of plasma cells. Plasma cells are divided into different clusters according to gene expression profiles, and each cluster is labeled with a different color. X/Y axes represent the cell positions calculated by the dimensionality reduction algorithm, and the colors represent different cell populations. (**d**) Heat maps of gene expression over supposed time. The heat map shows the expression dynamics of key genes over a pseudo-time period, with red indicating higher expression and blue indicating lower expression. Genes are divided into different modules according to their expression patterns. (**e**) Expression dynamics of *RGS1*. Relative expression of *RGS1* in relation to pseudotime, showing different cell clusters (left) and cell states (right). The black lines represent trends in gene expression over time. The X-axis represents the progression of pseudo-timing, and the Y-axis represents the amount of cell expression.

on early diagnosis and therapeutic targets for Alzheimer's disease, and many biomarkers and targets are primarily identified through computational methods aimed at minimizing the large investment required for drug development.

The role of immune disorders in AD has received increasing attention.[8] Inflammatory immune response is the factor that runs through the whole pathological mechanism of AD.[9] The researchers found several AD-related genes involved in neuroinflammation and immune activation in the test model.[10] Some researchers identified 18 immune-associated DEGs

Table 3 Baseline Information Analysis

	AD Group (n=10)	Control Group (n=10)	P-value
Male	4	5	0.5
Female	5	4	0.5
Age (mean ± standard deviation)	65.9±8.25	62.7±7.78	0.38
HAMA	2.3±0.36	2.5±0.34	0.70
HAMD	2.5±0.34	2.7±0.36	0.81
CDR (mean ± standard deviation)	2.05±0.83	0.00	0.00

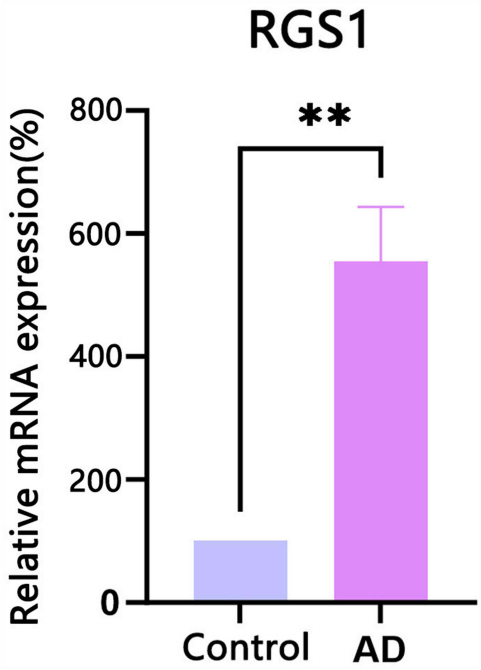


Figure 15 Assessment of the expression level of key gene in participants with AD and healthy controlsQuantitative reverse transcription (RT-q) PCR results show that the relative mRNA expression levels of *RGS1* was higher in the AD group than in the control group (** p<0.01).

with PPI network analysis and generated AD immune-associated ceRNA networks with tools such as StarBase, dianalnbase, and the Human MicroRNA Disease Database (HMDD).[11] Researchers have also identified differentially infiltrating immune cells (DIICs) using the CIBERSORT algorithm.[12] Weighted gene Co-expression Network Analysis (WGCNA) is a powerful screening tool that constructs a scale-free gene co-expression network to explore relationships between genes with similar expression patterns and external clinical information. Multiple studies have identified biomarkers associated with AD immune infiltration using WGCNA and machine learning algorithms. Brianne M. Bettcher et al suggested that peripheral inflammatory markers in asymptomatic elderly people were related to clinical outcomes, and peripheral inflammatory events were correlated with central inflammatory responses.[13]

Members of the G protein signaling (RGS) family are important regulators of the G protein-coupled receptor (GPCR) signaling pathway in the brain, which can respond to a variety of extracellular stimuli, and interference in its function is a potential cause of AD.[14] The brain-gut axis or gut-brain axis is one of the potential mechanisms. Previous studies showed that mice with *RGS1* knocked out showed an accelerated return of B cells to the lymph nodes, accompanied by reduced B cell retention at the site of inflammation.[15] In another study, *RGS1* expression was elevated in intestinal T cells of patients with colitis.[16] These results suggest an association between *RGS1* and intestinal inflammation. *PIEZO1* is a stretch-gated ion channel required for mechanosensation in many organ systems. It acts as a sensor in the gut that regulates gut flora and inflammation.[17] Our research suggested a strong link between *RGS1* and *PIEZO1*. In our study, the role of *RGS1* in the regulation of CD4+ T cell immune responses further supports its critical role in the brain-gut axis, suggesting that the treatment of AD may require a combination of central and peripheral inflammatory interactions.[16] *RGS1* expression is elevated in both macrophages and microglia in patients with myocardial infarction and AD.[18] The mechanism of action of *RGS1* in AD is unclear; however, but it *RGS1* is considered a potential biomarker of AD[19] and may be related to neuroinflammation and microglial overactivation regulated by the Janus kinase/signaling and transcriptional activator JAK-STAT pathways.[20] Although microglia protect neuronal cells against external stimuli, the activated state can be transformed into a neurotoxic state, intensifying the aggregation of Aβ and tau.[21] The JAK/STAT signaling pathway plays an important role in regulating the fate of glial cells. In the context of microglial overactivation, the JAK/STAT signaling pathway initiates innate immunity, regulates adaptive immunity, and inhibits neuroinflammation in AD.[22] Previous studies have suggested that *RGS1* has a macrophage-like regulatory mechanism on microglia, which

leads to local accumulation of microglia by desensitizing chemokine receptors, leading to chronic neuroinflammation.[18] High expression of *RGS1* can be enriched in MTORC1 signaling, which is associated with increased mitochondrial calcium uptake and mitochondrial activity, and its activation can accelerate neurodegenerative changes.[23] In this study, we found that *RGS1* is highly expressed in monocytes of AD patients and may play an important role in neuroinflammation and mitochondrial dysfunction through enrichment in the JAK/STAT signaling pathway and MTORC1 signaling pathway. However, due to ethical restrictions, we did not directly detect the expression of *RGS1* in the brain tissue of patients with AD. The increase of *RGS1* in peripheral blood can serve as a hint.

RSG1, which has been screened by bioinformatics analysis, is regarded as a key marker for the diagnosis and treatment of AD and verified in clinical subjects. The limitations of this study include reliance on a single master dataset, the lack of direct mechanism validation, and the inherent limitations of cross-sectional studies. Although the sample size included in this study is limited and there are still limitations in the promotion of the universal significance of *RSG1*, the increased expression of this gene in AD patients prompts us to pay attention to it, hoping to further explore its value in future studies. Future studies should focus on the specific function of *RGS1* in monocytes and microglia, and explore whether the GPCR signaling pathway regulated by *RGS1* can be used as a new target to interfere with the inflammatory response of AD. In addition, the relationship between *RGS1* and JAK/STAT signaling and mitochondrial dysfunction suggests that *RGS1* may play an indirect role in Aβ metabolism and tau protein abnormal aggregation, which provides A way to further study its therapeutic potential. The discovery of the *RGS1* surpasses the traditional Aβ and Tau hypotheses and can inhibit the phagocytic function of microglia by regulating multiple signaling pathways, thereby exacerbating neuroinflammation and AD pathological deposition. More strikingly, *RGS1* has a potential connection with PIEZO1. The *RGS1*-PIEZO1- neuroinflammatory axis represents a brand-new mechanism and offers a very promising target for future treatments. By developing *RGS1* inhibitors, it is expected to restore the function of microglia and suppress neurotoxic inflammation, thereby achieving precise intervention for AD from the perspective of neuroimmunity.

Conclusions

Although the high expression of *RGS1* in AD patients has been verified by rt-qPCR, we still need more in-depth functional studies to explain its significance in the pathogenesis, diagnosis and treatment of AD. The limitation of this study is that chronic diseases such as diabetes and hypertension cannot be ruled out in clinical validation, the clinical sample size is limited, and there are significant differences in lifestyle and dietary patterns among subjects, all of which may affect gene expression. We plan to conduct further studies to clarify the regulatory roles of *RGS1* in different signaling pathways and identify possible intervention targets to provide a basis for the diagnosis and treatment of AD.

Abbreviations

Aβ, amyloid beta;AD, Alzheimer's disease; AUC, area under the curve; CDR, Clinical Dementia Rating; eQTL, expression quality locus; GEO, Gene Expression Omnibus; GO, Gene Ontology; GPCR, G protein-coupled receptor; GSEA, gene set enrichment analysis; GSVA, gene set variation analysis; GWAS, genome-wide association study/studies; HAMA, Hamilton Anxiety Scale; HAMD, Hamilton Depression Scale; hdWGCNA, high-dimensional weighted gene co-expression network analysis; IVW, inverse-variance weighted; JAK, Janus kinase; ME, module Eigengene; m-HIS, Modified Hachinski Ischemia Scale; MR, Mendelian randomization; MRI, magnetic resonance imaging; MSigDB, Molecular Signatures Database; NCBI, National Center for Biotechnology Information; NES, normalized enrichment score; NFT, neurofibrillary tangle; NIA-AA, National Institute on Aging-Alzheimer's Association; NK, natural killer; PBMC, peripheral blood mononuclear cell; PCA, principal component analysis; RGS, regulators of G protein signaling; RT-qPCR, quantitative reverse transcription polymerase chain reaction; SNP, single-nucleotide polymorphisms; ssGSEA, single-sample gene set enrichment analysis; STAT, signaling and transcriptional activator; TF, transcription factor; t-SNE, t-distributed stochastic neighbor embedding; WGCNA, weighted gene co-expression network analysis.

Data Sharing Statement

The datasets generated and/or analyzed during the current study are not publicly available because they protect patient privacy but are available from the corresponding author upon reasonable request.

Ethics Approval and Consent to Participate

Human experiments were approved by the Ethics Committee of our hospital, and all enrolled subjects signed an informed consent form (Date:2024/01/23/ version.4.2).

This study was conducted in accordance with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of the First Clinical Medical School of Shandong University of Chinese Medicine (Date: 2024/01/23/No. AF/SC-08/03.0).

Acknowledgments

We thank the Experimental Center of Shandong University of Traditional Chinese Medicine, Jinan 250355, China. This paper is available on ResearchSquare as a preprint https://doi.org/10.21203/rs.3.rs-4609987/v1.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Consent for Publication

Written informed consent was obtained from the clinical participants.

Funding

This work was supported by the [Natural Science Foundation of China] under Grant [number 82205064]; [Natural Science Foundation of Shandong Province] under Grant [number ZR2021QH110] and the [TCM Science and Technology Project of Shandong Province] under Grant [number Z-2022060].

Disclosure

The authors declare no competing interests in this work.

References

1. Flannagan K, Stopperan JA, Hauger BM, et al. Cell type and sex specific mitochondrial phenotypes in iPSC derived models of Alzheimer's disease. *Front Mol Neurosci*. 2023;16:1201015. doi:10.3389/fnmol.2023.1201015
2. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet*. 2015;16:197–212. doi:10.1038/nrg3891
3. Wilkins HM, Koppel SJ, Bothwell R, Mahnken J, Burns JM, Swerdlow RH. Platelet cytochrome oxidase and citrate synthase activities in APOE epsilon4 carrier and non-carrier Alzheimer's disease patients. *Redox Biol*. 2017;12:828–832. doi:10.1016/j.redox.2017.04.010
4. Zhang Z, Liu X, Zhang S, Song Z, Lu K, Yang W. A review and analysis of key biomarkers in Alzheimer's disease. *Front Neurosci*. 2024;18:1358998. doi:10.3389/fnins.2024.1358998
5. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7:280–292. doi:10.1016/j.jalz.2011.03.003
6. Qin G, Wang Y, Liu Z, Mana L, Huang S, Wang P. Shenzhiling oral solution promotes myelin repair through PI3K/Akt-mTOR pathway in STZ-induced SAD mice. *3 Biotech*. 2021;11:361. doi:10.1007/s13205-021-02900-x
7. Zhu M, Hou T, Jia L, Tan Q, Qiu C, Yifeng D; Alzheimer's Disease Neuroimaging Initiative. Development and validation of a 13-gene signature associated with immune function for the detection of Alzheimer's disease. *Neurobiol Aging*. 2023;125:62–73. doi:10.1016/j.neurobiolaging.2022.12.014
8. Rai SN, Singh S, Singh SK. *Neurodegenerative Diseases: Translational Models, Mechanisms, and Therapeutics*. 1st ed. CRC Press; 2024.
9. Rai S, Singh S, Singh S. Neurodegenerative diseases: translational model, mechanism and therapeutics. 2024.
10. Shen M, Shang M, Tian R, et al. Single cell molecular alterations reveal target cells and pathways of conditioned fear memory. *Brain Res*. 2023;1807:148309. doi:10.1016/j.brainres.2023.148309
11. Sirkis DW, Solsberg CW, Johnson TP, et al. Expansion of interferon signaling-associated gene (ISAG) hi T cells in early-onset Alzheimer's disease. *bioRxiv*. 2023. 2023–09.
12. Chen S, Chang Y, Li L, et al. Spatially resolved transcriptomics reveals genes associated with the vulnerability of middle temporal gyrus in Alzheimer's disease. *Acta Neuropathol Commun*. 2022;10:188. doi:10.1186/s40478-022-01494-6
13. Brianne MB, T MG, Guillaume D, Michael TH. Peripheral and central immune system crosstalk in Alzheimer disease - a research prospectus.*Nat. Rev Neurol*. 2022;17:689–701.

14. Squires KE, Montãnez-Miranda C, Pandya RR, Torres MP, Hepler JR. Genetic analysis of rare human variants of regulators of G protein signaling proteins and their role in human physiology and disease. *Pharmacol Rev*. 2018;70:446–474. doi:10.1124/pr.117.015354
15. Han S, Moratz C, Huang N, et al. Rgs1 and Gnai2 regulate the entrance of B lymphocytes into lymph nodes and B cell motility within lymph node follicles. *Immunity*. 2005;22:343–354. doi:10.1016/j.immuni.2005.01.017
16. Gibbons DL, Abeler-Dörner L, Raine T, et al. Cutting edge: regulator of G protein signaling-1 selectively regulates gut T cell trafficking and colitic potential. *J Immunol*. 2011;187:2067–2071. doi:10.4049/jimmunol.1100833
17. Yan X, Xiong Y, Liu Y, et al. Activation of goblet cell Piezo1 alleviates mucus barrier damage in mice exposed to WAS by inhibiting H3K9me3 modification. *Cell Biosci*. 2023;13:7. doi:10.1186/s13578-023-00952-5
18. Xue W, He W, Yan M, Zhao H, Pi J. Exploring shared biomarkers of myocardial infarction and Alzheimer's disease via single-cell/nucleus sequencing and bioinformatics analysis. *J Alzheimers Dis*. 2023;96:705–723. doi:10.3233/JAD-230559
19. Leandro GS, Evangelista AF, Lobo RR, Xavier DJ, Moriguti JC, Sakamoto-Hojo ET. Changes in expression profiles revealed by transcriptomic analysis in peripheral blood mononuclear cells of Alzheimer's disease patients. *J Alzheimers Dis*. 2018;66:1483–1495. doi:10.3233/JAD-170205
20. Nevado-Holgado AJ, Ribe E, Thei L, Furlong L, Mayer M-A. Simon Lovestone (2019) Genetic and real-world clinical data, combined with empirical validation, nominate Jak-Stat signaling as a target for Alzheimer's disease therapeutic development. *Cells*. 2019;8:425. doi:10.3390/cells8050425
21. d'Errico P, Ziegler-Waldkirch S, Aires V, et al. Microglia contribute to the propagation of Aβ into unaffected brain tissue. *Nat Neurosci*. 2022;25:20–25. doi:10.1038/s41593-021-00951-0
22. Rusek M, Smith J, El-Khatib K, Aikins K, Czuczwar SJ, Pluta. Role of the JAK/STAT signaling pathway in the pathogenesis of Alzheimer's disease: new potential treatment target. *Int J Mol Sci*. 2023;25:24. doi:10.3390/ijms25010024
23. Ryan KC, Ashkavand Z, Sarasija S, Laboy JT, Samarakoon R, Norman KR. Increased mitochondrial calcium uptake and concomitant mitochondrial activity by presenilin loss promotes mTORC1 signaling to drive neurodegeneration. *Aging Cell*. 2021;20:e13472.