

RESEARCH ARTICLE

Identification of Alzheimer's disease subtypes and biomarkers from human multi-omics data using subspace merging algorithm

Ziyan Song¹  | Xiaoqing Huang¹ | Asha Jacob Jannu^{2,3} | Travis S. Johnson¹ | Jie Zhang⁴ | Kun Huang¹ 

¹Department of Biostatistics and Health Data Science, Indiana University School of Medicine, Indianapolis, Indiana, USA

²Department of Bioengineering and Informatics, Indiana University School of Medicine, Indianapolis, Indiana, USA

³Department of Developmental Neurobiology, St. Jude Children's Research Hospital, Indianapolis, Indiana, USA

⁴Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana, USA

Correspondence

Kun Huang, Department of Biostatistics and Health Data Science, Indiana University School of Medicine, HITS 3000, 410 W 10th St, Indianapolis, IN 46202, USA.
Email: kunhuang@iu.edu

Jie Zhang, Department of Medical and Molecular Genetics, Indiana University School of Medicine, HITS 5015, 410 W 10th St, Indianapolis, IN 46202, USA.
Email: jizhan@iu.edu

Ziyan Song, Lead contact: zs8@iu.edu

Funding information

National Institutes of Health (NIH), Grant/Award Numbers: 5U54AG065181, R21AG075541

Abstract

INTRODUCTION: Alzheimer's disease (AD) is a heterogeneous disease with diverse disease progression trajectories and brain pathology. Identifying AD subtypes is essential for understanding AD etiology, heterogeneity, and developing precise treatment.

METHODS: We applied a subspace-merging algorithm to integrate multi-omics data from brain tissues of three large AD cohorts and identify data-driven AD subtypes. Within each cohort, we performed multiple analyses to characterize subtype-specific biology. A Phenome-wide Association Study (PheWAS) of expression quantitative trait loci (eQTLs) targeting differentially expressed genes (DEGs) was conducted to link molecular differences to disease phenotypes.

RESULTS: We identified AD subtypes that differed in cognitive and pathological phenotypes in three cohorts. Further analyses highlighted synaptic and neurotransmission pathways, and the PheWAS revealed significant associations with disease phenotypes.

DISCUSSION: Our developed integration algorithm successfully merged different data modalities into a common subspace for patient clustering and identified data-driven subtypes. The identified transcriptomic signatures provide valuable insights into the molecular mechanisms underlying AD heterogeneity, paving the way for personalized AD treatment.

KEYWORDS

AD subtyping, Alzheimer's disease, multi-omics, real-world data, subspace merging algorithm

Highlights

- We integrated brain transcriptomics, DNA methylation, and proteomics from three large Alzheimer's disease (AD) cohorts: ROSMAP, MSBB, and MayoRNAseq, using a subspace merging algorithm to embed these different data modalities onto a common subspace for patient-level presentation and clustering, and identified two

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.

© 2026 The Author(s). *Alzheimer's & Dementia* published by Wiley Periodicals LLC on behalf of Alzheimer's Association.

data-driven patient subtypes with significant cognitive and AD pathology differences in both studies, which are independent of apolipoprotein E (APOE) status and other demographic traits.

- Cross-tissue comparison identified nine overlapping differentially expressed genes (DEGs) between brain and blood: *RAB27B*, *XK*, *NWD2*, *DGAT2*, *ENO2*, *SYT2*, *NAPB*, *MLLT11*, and *SMYD2*, some of which may serve as potential blood biomarkers for subtype identification.
- Gene ontology (GO) enrichment revealed consistent altered pathways between the subtypes identified independently from ROSMAP and MSBB. In the better-cognition subtype, enriched molecular functions included neuropeptide hormone activity, anion activity, chloride channel activity, chloride transporter activity, hormone activity, and GTPase activity; the enriched biological processes included potassium ion transport, positive regulation of secretion by cell, regulation of catecholamine secretion, catecholamine secretion, positive regulation of secretion, regulation of amine transport, and amine transport.
- Validation in the MayoRNAseq cohort demonstrated high cross-cohort consistency with ROSMAP, identifying 439 shared DEGs that are consistently upregulated in the less-severe subtype. GO enrichment analysis revealed shared biological processes, including synapse organization, synaptic vesicle cycling, and cognition, and shared molecular functions, including gated ion channel activity, γ -aminobutyric acid A (GABA-A) receptor activity, and ion channel activity.
- Phenome-wide Association Study (PheWAS) revealed 16 significant associations between expression quantitative trait loci (eQTLs) targeting brain or blood subtype DEGs and clinical phenotypes, and broadened our understanding of AD heterogeneity and links to other disease phenotypes.

1 | BACKGROUND

Alzheimer's disease (AD) is characterized by a high degree of heterogeneity, encompassing diverse brain pathological changes, progression trajectories, and a multitude of risk factors among affected individuals.¹ Utilizing data-driven approaches for subtyping, AD holds the promise for elucidating its etiology and profoundly improving AD diagnosis, treatment, and disease management strategies. This project endeavors to delineate subtypes of AD by leveraging integrative analyses of multi-omics data derived from matched postmortem brain and blood samples of patients with AD.

Integration of multi-omics data for disease subtyping has been studied widely in diseases such as cancers. There are multiple hurdles with multi-omics data integration, including high-dimensionalities of the omics data, different numerical ranges between data types, and strong correlations between features (e.g., genes and proteins). These hurdles make it challenging to integrate the features from different omics data modalities for downstream subtyping analysis, which is often based on unsupervised learning methods.

In this study, we employ an advanced subspace merging algorithm² to conduct unsupervised clustering analysis on patients with AD, integrating gene expression, proteomics, and DNA methylation data from three large-scale AD studies: the Religious Orders Study and Memory and Aging Project (ROSMAP),³ the Mount Sinai Brain Bank (MSBB),⁴ and the Mayo RNAseq Study (MayoRNAseq).⁵

This design preserves modality-specific structure while capturing shared patient-level patterns, providing a flexible and mathematically grounded framework for discovering robust molecular subtypes.

We aim to translate the subtypes discovered using brain tissues to blood samples, which can lead to the discovery of blood markers in live patients with clinical significance for early AD subtype diagnosis. To achieve these goals, we analyzed RNA sequencing (RNAseq) gene expression data collected from the blood samples of matched patients from ROSMAP, expanding our analytical scope to peripheral biomarkers. We performed differential expression analysis, gene ontology (GO) enrichment analysis, and pathway analysis to establish potential connections between differentially expressed genes (DEGs) identified from brain samples and those identified from blood samples.

Furthermore, we investigate whether any clinical phenotypes are associated with the expression levels of the discovered brain and blood markers. Such phenotypes may serve as early indicators or markers to predict different AD risks and progression trajectories. Specifically, we conducted a phenome-wide association study (PheWAS)⁶ using comprehensive genomics data from the *All of Us* Research Project, which contains both genotype data and extensive clinical records for a large heterogeneous cohort of adults covering different sexes and races. We used established expression quantitative trait locus (eQTLs)⁷ to identify genetic variants associated with the expression levels of the DEGs and then explored the association between the genetic variants of these eQTLs and a wide variety of phenotypes in the *All of Us* cohort.

The identification of blood-based biomarkers holds significant promise for advancing AD research and clinical practice, offering potential advancement for enhancing diagnosis accuracy, treatment efficacy, and preventive strategies. By exploring the molecular underpinnings of AD heterogeneity and establishing connections between brain and peripheral biomarkers, our study contributes to precision medicine aimed at improving the understanding and management of AD.

2 | METHODS

2.1 | Previous work

With the accumulation of multi-omics data for diseases, several integrative approaches have been proposed. One strategy involves constructing patient similarity networks for each omics modality and merging them into a unified consensus network. Similarity Network Fusion (SNF) adopts this approach and has been widely applied to disease subtyping, including cancer and neurological disorders.⁸ Another representative method is Multi-Omics Factor Analysis (MOFA), which models shared and modality-specific latent factors using a Bayesian framework. MOFA is particularly effective in exploring the underlying sources of variability across heterogeneous omics layers.⁹ Similarly, iClusterPlus leverages a joint latent variable model and likelihood-based inference to simultaneously integrate multiple omics datasets with different distributions, such as Gaussian or binomial, and has been applied successfully to cancer subtype discovery.¹⁰ In this study we apply the subspace merging algorithm. The subspace merging algorithm was similar to the SNF algorithm, as it aims at integrating multiple patient networks generated from different types of data for downstream clustering. The algorithm is based on a rigorous graph-theoretical framework and has been demonstrated to be effective in discovering new disease subtypes in cancers with superior performance than SNF and other related algorithms.²

2.2 | Data used in this study

Three types of omics data, namely bulk tissue transcriptome, DNA methylome, and proteome data, were obtained from the Alzheimer's

RESEARCH IN CONTEXT

- 1. Systematic review:** We reviewed peer-reviewed (PubMed) and preprint literature and found limited studies on Alzheimer's disease (AD) subtyping with multi-omics data. We leveraged a subspace merging algorithm to generate representations that capture shared and omics-specific features. We performed differential expression, gene ontology (GO) enrichment, and phenome-wide association study (PheWAS) analyses to characterize the molecular differences distinguishing the discovered AD subtypes.
- 2. Interpretation:** We identified AD subtypes that differ in cognition and pathological phenotypes. Enrichment analysis pointed to synaptic and neurotransmission pathways. A cross-tissue analysis revealed nine overlapping differentially expressed genes, highlighting tissue-specific and context-dependent regulation that links peripheral biology with central transcriptional states. PheWAS demonstrated novel links between subtype-specific genetic markers and various phenotypes, broadening the biological scope of AD heterogeneity.
- 3. Future directions:** Future research will focus on validating subtype generalizability in independent and longitudinal cohorts. In addition, we will test peripheral markers for prospective subtype prediction in living patients and extend PheWAS analyses to prioritize causal targets.

Disease Knowledge Portal (<https://adknowledgeportal.synapse.org>) for both the ROSMAP and the MSBB study. For ROSMAP, additional monocyte RNA-seq data collected on peripheral blood mononuclear cells (PBMCs) from a subset of 28 matched patients with AD were also incorporated into the analysis. The resulting ROSMAP dataset included 79 AD patients with matched gene transcription,^{11,12} DNA methylation,¹³ and proteomics^{14–18} data from the dorsolateral prefrontal cortex (DLPFC) of autopsied brains and blood (Table 1).

For the MSBB study, multi-omics data from 83 AD patients were incorporated, including RNA-seq data collected from four distinct Brodmann areas: areas 22 and 36 from the temporal lobe and areas 10 and 44 from the frontal lobe of the brain (Table 1).⁴ Furthermore, Isobaric Multiplex Tandem Mass Tag (TMT)-labeled proteomics¹⁹ and DNA methylation array data from matched patients were included. For the MayoRNAseq Study, we analyzed three omics data modalities from 65 AD patients: label-free mass spectrometry-based proteomics and bulk RNA-seq data derived from two different brain regions, the cerebellum (CBE) and temporal cortex (TCX), according to data availability (Table 1). Clinical data for all patients were also included in the follow-up downstream analysis. We conducted integrative multi-omics data analyses with a subspace merging algorithm for patient subtype detection and downstream analysis. Those analyses were

TABLE 1 Summary of dataset characteristics: matched RNA-sequencing data, DNA methylation data, and proteomics data were obtained from the DLPFC of 79 patients with AD from the ROSMAP study, with an additional subset of 28 patients providing transcriptomics from PBMC. In addition, matched RNA-seq, DNA methylation, and proteomics data were collected from four distinct brain regions (Brodmann areas 10, 22, 36, and 44) among 83 patients in the MSBB dataset, and matched RNA-seq from CBE and TCX and proteomics among 65 patients in the MayoRNAseq study.

Study	Omics type	Number of samples	Number of features
ROSMAP	mRNA DLPFC	79	15,582
	mRNA PBMC	28	18,625
	DNA Methylation	79	54,596
	Proteomics	79	5211
MSBB	mRNA BM10	83	17,401
	mRNA BM22	83	17,401
	mRNA BM36	83	17,401
	mRNA BM44	83	17,401
	DNA Methylation	83	12,147
	Proteomics	83	2608
MayoRNAseq	mRNA TCX	65	12,758
	mRNA CBE	65	12,758
	Proteomics	65	3453

Abbreviations: DLPFC: dorsolateral prefrontal cortex; PBMC: peripheral blood mononuclear cell; BM: Brodmann area; BM10: anterior prefrontal cortex; BM 22: superior temporal gyrus; BM 36: parahippocampal gyrus; BM 44: pars opercularis; TCX: temporal cortex; CBE: cerebellum

conducted separately within each cohort to avoid the possible discrepancies that may exist among different brain regions and the need for further harmonization between datasets. For the ROSMAP cohort, because only data from the DLPFC were available, we focused on this region, integrating matched transcriptomics, DNA methylation, and proteomics data collected from the same brain region. For MSBB, we utilized RNA-seq data from four brain regions (BM10, BM22, BM36, and BM44), integrating these with available proteomics and methylation profiles. The data preprocessing and multi-omics data integration were performed within each cohort independently using anatomically and experimentally consistent data. In addition, we compared the patient subtypes we identified separately from either cohort to evaluate whether the integrative framework can capture consistent AD subtypes across independent cohorts, focusing on different brain regions, while capturing both the commonalities between cohorts and cohort-specific characteristics.

To enhance the robustness of downstream analyses, only feature variance among the top 75th percentile in each omics dataset including gene expression, DNA methylation, and proteomics collected from brain tissues were retained for each study. From the differential expression analysis and subtype analysis in the ROSMAP, MSBB, and MayoRNAseq cohorts, we identified distinct sets of DEGs representing molecular differences across AD subtypes. Once we obtained the

corresponding cis-eQTLs regulating these DEGs from the ROSMAP eQTL summary statistics, we wanted to further investigate whether these genetic variants were also associated with phenotypic traits related or unrelated to AD. Therefore, we conducted a PheWAS analysis using the whole-genome sequencing (WGS) dataset from the All of Us Research Project.²⁰ Our objective was to determine if eQTLs regulating the DEGs between subtypes, derived from ROSMAP cis-eQTL summary statistics (located within ± 1 Mb of each DEG), also showed associations with other phenotypes, particularly those related to cognition. Leveraging electronic health records (EHRs), we investigated the association between phenotypic information, condensed into International Classification of Diseases (ICD) codes, and genetic variations associated with gene expression, known as eQTLs. By focusing on eQTLs linked to DEGs identified in brain tissues,²¹ we explored the associations of the identified molecular markers with various clinical phenotypes to check if any potential link exists between AD subtypes and broader physiological conditions, to enhance our understanding of AD heterogeneity and gain deeper insights into the potential regulatory mechanisms underlying AD pathology. Data source and data pre-processing information can be found in Supplementary File 1.

Together, the integration of multi-omics data from the ROSMAP, MSBB, and MayoRNAseq, coupled with the comprehensive genomics-data from the All of Us Research Project, offers a robust framework for dissecting the molecular underpinnings of AD. By elucidating the intricate interplay between genetic, epigenetic, and proteomic factors, our study contributes to the broader effort of unraveling the complexity of AD and advancing personalized approaches to diagnosis, treatment, and disease management.

2.3 | AD patient clustering with subspace merging algorithm

To integrate heterogeneous omics data for patient stratification in AD, we adopt a subspace merging algorithm originally developed for integrative cancer analysis,² with minor modifications to identify AD subtypes with matched three types of omics data collected from AD patients in ROSMAP. This method follows the intermediate data integration paradigm by converting each omics dataset into a patient-to-patient similarity graph, performing spectral embedding to obtain subspaces, and subsequently merging them on a Grassmann manifold. This framework captures both shared and omics-specific structure while preserving the local geometry of each data layer. An overview of this approach is shown in Figure 1.

Let $\{X\}_{m=1}^M$ be M omics data matrices, each of dimension $N \times p_m$, where N is the number of patients and p_m the number of features in the m_{th} omics layer. Features in each omics matrix X_m were independently standardized to have zero mean and unit variance prior to graph construction and subspace embedding by Equation (1), where f is any feature, $\hat{f}$ is the corresponding standardized feature, $E(f)$ is the sample mean, and the $Var(f)$ is the sample variance for the selected feature. This standardization ensured comparable scaling within each modality

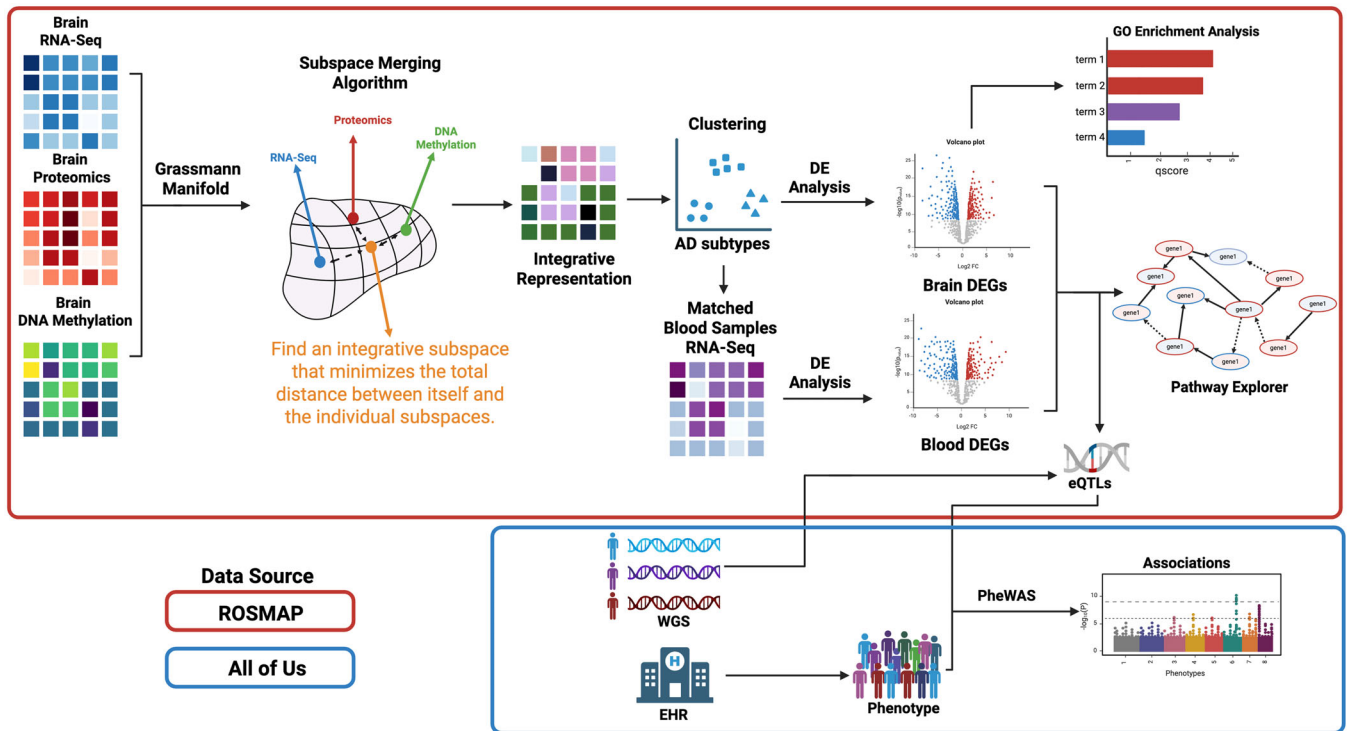


FIGURE 1 The workflow of the discovery phase in the study. The analysis begins with multi-omics data integration from ROSMAP studies, incorporating postmortem brain RNAseq, proteomics, and DNA methylation profiles from matched Alzheimer's disease (AD) patients separately. A subspace merging algorithm is applied to identify distinct AD molecular subtypes in both studies. In parallel, the ROSMAP study provides additional transcriptomic data from PBMCs for a subset of matched individuals, enabling cross-tissue analysis. Differential expression (DE) analysis is performed on both brain and blood transcriptomic data to identify subtype-associated DEGs. Gene ontology (GO) enrichment analysis is conducted on brain DEGs to characterize subtype-specific biological functions. The relationships between brain and blood differentially expressed genes (DEGs) are further explored using the Pathway Explorer module in QIAGEN Ingenuity Pathway Analysis (IPA) to identify converging molecular pathways. Finally, using genotypic and phenotypic data from the All of Us Research Program, a Phenome-wide Association Study (PheWAS) is conducted to assess the associations between expression quantitative trait loci (eQTLs) of DEGs between AD subtypes and a broad spectrum of human traits and clinical phenotypes.

and cohort while preserving cohort-specific biological variation

$$\hat{f} = \frac{f - E(f)}{\sqrt{\text{Var}(f)}} \quad (1)$$

For each omics modality, we construct a similarity graph $G^{(m)}$ by first computing the pairwise heat kernel similarity matrix with Equation (2),²² where σ_i presents the distance to 7th nearest neighbor of x_i .

We then define the pairwise similarity matrix $S^{(m)} \in \mathbb{R}^{N \times N}$ using a symmetric heat kernel. This kernel formulation guarantees that patients with similar profiles form locally dense neighborhoods, while preserving relative scaling across modalities. We then compute the normalized graph Laplacian $L^{(m)} \in \mathbb{R}^{N \times N}$ for each similarity matrix $S^{(m)}$ with Equation (3), where $D^{(m)}$ is a diagonal matrix with $D_{ii}^{(m)} = \sum_j S_{ij}^{(m)}$. This corresponds to the symmetric normalized Laplacian.

$$S_{ij}^{(m)} = \exp \left(-\frac{\|x_i^{(m)} - x_j^{(m)}\|^2}{\sigma_i \sigma_j} \right), \quad S_{ii} = 0 \quad (2)$$

$$L^{(m)} = I - D^{(m)-1/2} S^{(m)} D^{(m)-1/2} \quad (3)$$

We perform eigen decomposition of $L^{(m)}$, and obtain the k eigenvectors corresponding to the smallest eigenvalues. These eigenvectors $U^{(m)} \in \mathbb{R}^{N \times k}$ form the subspace representations for each modality. Each $U^{(m)}$ is a point on the Grassmann manifold $G(k, N)$, which denotes the collection of all k -dimensional linear subspaces in n dimensional Euclidean space $\mathbb{R}^N$.² The Grassmann manifold provides a principled geometric framework for modeling subspaces, enabling us to quantify the dissimilarity between modality-specific patient embeddings.

We measure the dissimilarity between two subspaces using the squared projection Frobenius norm as shown in Equation (4).

$$d_{proj}^2(U, U^{(m)}) = k - \text{tr}(UU^T U^{(m)} (U^{(m)})^T) \quad (4)$$

This projection metric quantifies the distance between the projection matrices of U and $U^{(m)}$, capturing how aligned the patient representations from different modalities are in terms of their subspace orientation. To construct a consensus representation U , we jointly minimize two objectives: the spectral loss from each omics graph and the projection distance between U and the modality-specific embeddings.

The resulting optimization problem is by solving Equation (5):

$$\min_{U \in \mathbb{R}^{N \times k}, U^T U = I} \sum_{m=1}^M \text{tr}(U^T L^{(m)} U) + \alpha \sum_{m=1}^M d_{\text{proj}}^2(U, U^{(m)}) \quad (5)$$

The first term in Equation (5) encourages U to preserve the local manifold structure encoded in each Laplacian $L^{(m)}$, whereas the second aligns the consensus with the individual subspaces. The hyperparameter α controls the balance between structure preservation and modality agreement.

$$L^{\text{mod}} = \sum_{m=1}^M L^{(m)} - \alpha \sum_{m=1}^M U^{(m)} (U^{(m)})^T \quad (6)$$

We then obtain the final merged subspace U by computing the top k eigenvectors of L^{mod} as given in Equation (6). Unlike iterative optimization methods, this eigen decomposition-based approach is computationally efficient and converges to a global solution. Our framework treats all omics layers with equal importance during integration. Both the Laplacians and the subspace projection terms are summed without scalar weighting. This reflects an assumption of equal biological contribution across gene expression, proteomics, and DNA methylation.

The rows of the merged subspace U provide patient-level embeddings that capture integrative structure across omics. We apply k -means clustering to U to define patient subtypes. To determine the optimal number of clusters k , we evaluate the average silhouette score for a range of candidate values (e.g., $k = 2-8$). For each value of k , we normalize the eigenvectors of L^{mod} , perform k -means clustering with multiple restarts ($n = 10$), and compute the silhouette score using the Euclidean distance. The value of k that yields the highest average silhouette score is selected as the optimal cluster number. This procedure ensures that the final clustering structure is both geometrically coherent and supported by internal validation metrics. After determining the optimal clustering solution for the AD subtypes, we compared demographic and clinical variables between the resulting groups. We specifically examined the distribution of apolipoprotein E (APOE) genotypes to verify that subtype membership was not confounded by APOE status. Thus, it ensures the differences observed in the omics profiles represent subtype-specific biological signals rather than effects driven by APOE. Continuous variables were tested using the Wilcoxon test, and categorical variables were evaluated with chi-square test with Monte Carlo-simulated p -value (based on 2000 replicates).²³

2.4 | Differential gene expression analysis

Differential gene expression and downstream pathway analyses were applied to establish the gene signatures between the subtypes (patient clusters) for ROSMAP, MSBB, and MayoRNAseq, as well as to the transcriptomic data of the PBMCs from the matching patients in ROSMAP in a search for possible blood markers. Differential expression anal-

ysis was conducted using the linear modeling approach implemented in the “limma” R package.^{24,25} The batch effect was corrected, and confounding variables such as age, race, and sex were adjusted if the processed data had not applied such adjustment. The empirical Bayes method implemented in “limma” was used to moderate standard errors and obtain more stable estimates of gene-wise variances. Comparative gene expression analysis was conducted to determine the differences between the two subgroups identified from the previous steps. The resulting differential expression statistics, including log-fold changes ($\log_2\text{FC}$) and moderated t -statistics, were used to identify genes that were significantly differentially expressed. To control the false discovery rate (FDR), the Benjamini-Hochberg (BH) procedure²⁵ was applied to adjust p -values for multiple testing. Genes with an adjusted p -value below a specified threshold ($\text{FDR} < 0.05$) were considered statistically significant.

Comparative analyses were conducted to determine the differences between the two subgroups identified from the previous steps. The resulting differential expression statistics, including $\log_2\text{FC}$ and moderated t -statistics, were used to identify genes that were significantly differentially expressed. To control the FDR, the BH procedure was applied to adjust p -values for multiple testing. Genes with an adjusted p -value below a specified threshold ($\text{FDR} < 0.05$) were considered statistically significant.

2.5 | GO enrichment analysis on subtype-associated DEGs

DEGs were further analyzed for biological relevance and functional enrichment using R Package “clusterProfiler,”²⁶ which performs GO and pathway enrichment analyses by applying a hypergeometric test to identify significantly enriched GO terms and pathways among a set of genes based on their annotations, identifying over-represented biological processes, molecular functions, or cellular components.

2.6 | PheWAS

PheWAS is the analysis that explores the association between genetic variants and a wide range of phenotypic traits or outcomes across a large population or cohort. Unlike traditional genome-wide association studies (GWAS), which typically focus on a specific disease or trait, PheWAS examines the genetic influences on diverse phenotypes, including diseases, clinical measurements, laboratory values, and other health-related characteristics.

This approach allows for a quick screening for the genetic variants that may influence various aspects of health and disease beyond those initially considered in the study. PheWAS can uncover unexpected potential relationships between genetic variants and phenotypes, provide insights into the pleiotropic effects of genes, and contribute to

a more comprehensive understanding of the genetic basis of complex traits and diseases.

2.6.1 | Genotypic data

To elucidate the complex relationships between genetic variation, gene expression, and phenotyping outcomes in our PheWAS study, we collected eQTL summary data from ROSMAP²¹ via the AD Knowledge Portal. This dataset included single nucleotide polymorphism (SNP)–gene associations with corresponding *p*-values and effect sizes. We selected ROSMAP because it contains both brain and matched blood transcriptomic data together with eQTL analysis summary statistics, whereas the latter two are not available in MSBB dataset. We focused on cis-eQTLs, defined as variants within ± 1 Mb of each DEG, and considered the union of cis-eQTLs targeting DEGs identified in brain and blood samples from ROSMAP between the two subtypes. For each DEG, we selected at most the top three significant eQTLs with an FDR less than 10^{-4} to prevent over-representation. This selection process yielded a total of 105 eQTLs derived from 37 DEGs, 49 from blood DEGs, and 56 from brain DEGs for the ROSMAP cohort. We employed genotypic data of these 105 eQTLs of interest from 11,024 unrelated participants from the All of Us study, 45 to 95 years of age, who were diagnosed with AD, diffuse Lewy body disease, behavioral disturbances concurrent with late-onset Alzheimer's dementia, or had a family history of dementia (parents, siblings, or self) as recorded in their EHR.

2.6.2 | Phenotypic data

Phenotype data used for the PheWAS were extracted from their EHR from All of Us, and mapped to *phcodes* using the R package “PheWAS.”²⁷ In the phenotype table, each participant was assigned a case/control label for each health outcome. A participant was classified as “TRUE” for a condition if they had at least two distinct instances of the corresponding *phcode*, indicating they were a case. Conversely, a “FALSE” classification indicated the absence of the health condition, classifying them as a control. *Phcodes* with fewer than 20 cases or controls were excluded, and sex-specific *phcodes* were assigned only to the corresponding sex. In total, 1858 phenotypes were extracted for the downstream association test.

2.6.3 | Statistical analysis of PheWAS

We performed logistic regression analyses for each *phcode*, where the binary case–control outcome was regressed on the genotype of each eQTL. The models were adjusted for age, sex at birth, and the first six principal components to correct for ancestry prediction. Multiple testing correction strategies were employed: We applied the BH method across all association tests and additionally conducted

an eQTL-specific correction for tests involving each respective eQTL. An arbitrary *p*-value threshold of 10^{-4} was utilized to capture any potential important associations.

3 | RESULTS

We performed the subspace merging algorithm and downstream analyses as demonstrated in Figure 1 on three independent cohorts of ROSMAP, MSBB, and MayoRNAseq multi-omics data in order to see if the subtype(s) we identified is a universal observation or is dataset/cohort-specific. In the following sections, we describe the results separately for each cohort. The complete list of DEGs between subtypes for each cohort can be found in supplementary tables (ROSMAP in Table S1, MSBB in Table S2, and MayoRNAseq in Table S3).

3.1 | Analysis of the ROSMAP dataset

3.1.1 | Patient k-means clustering results and their associations with clinical phenotypes

Based on their similarities computed from the integration of the three types of omics data from brains: gene transcriptome, proteome, and DNA methylome, we clustered the 79 patients from the ROSMAP cohort into two clusters (33 patients in Cluster 1 and 46 patients in Cluster 2 via k-means (Figure S1A). The number of clusters is determined by maximizing the silhouette score, which indicates the separating stage when clusters are most compact internally and most distinct from one another.

Demographic and clinical variables including age at death (age); sex; APOE genotype²⁸; clinical diagnosis of cognitive status (dcfdx)^{29,30}; clinical consensus diagnosis of cognitive status at time of death (cogdx)³¹; Braak stage score (braaksc),^{32–34} which is a semi-quantitative measure of neurofibrillary triangle pathology; the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) score (ceradsc),^{34,35} which is a semi-quantitative measure of neuritic plaques, the Mini-Mental State Examination (MMSE) score³⁶; and MMSE, which assesses cognitive impairment, were compared between the two clusters. Among them, dcfdx (χ^2 -test *p*-value = 0.025), MMSE (Wilcoxon test *p*-value = 0.04), and CERAD score (*p*-value = 0.035) were significantly different between the two clusters, indicating that there were mild differences in disease progression in terms of cognition. Overall, patients in Cluster 1 (referred as Subtype 1) presented relatively higher cognition score compared with Cluster 2 (referred as Subtype 2). However, there was no statistically significant difference between age at death, sex at birth, APOE genotype, or Braak stage. Clinical consensus diagnosis of cognitive status (cogdx) is marginally significant (*p* = 0.087). The clustering result suggests that the observed multi-omics difference cannot be explained by age, sex, or APOE status (Table 2) and was only revealed by integrating the multi-omics data-driven approach.

TABLE 2 Summary statistics (mean, standard deviation (SD) for continuous variables; counts and proportion for categorical variables) and *p*-values from Wilcoxon or chi-square tests between AD subtypes in ROSMAP, MSBB, and MayoRNAseq cohorts.

Variable	ROSMAP			MSBB			MayoRNAseq		
	Subtype 1	Subtype 2	<i>p</i> value	Subtype 1	Subtype 2	<i>p</i> value	Subtype 1	Subtype 2	<i>p</i> value
Age	86.19 (5.21)	86.68 (3.86)	0.789	82.298 (7.8322)	83.056 (7.566)	0.607	85.38 (4.84)	79.81 (7.78)	0.002
APOE genotype									
ε2/ε3	2 (6.1%)	3 (6.5%)	0.732	5 (19.2%)	3 (14.3%)	0.906	2 (3%)	0	0.255
ε2/ε4	0 (0%)	1 (2.2%)		16 (61.5%)	13 (61.9%)				
ε3/ε3	24 (72.7%)	28 (60.9%)		5 (19.2%)	4 (19.0%)		14 (21.6%)	15 (23.1%)	
ε3/ε4	7 (21.2%)	14 (30.4%)		0 (0%)	1 (4.7%)		12 (18.5%)	17 (26.2%)	
ε4/ε4							1 (1.5%)	4 (6.2%)	
Sex									
Male	8 (24.2%)	12 (26.1%)	1	12 (44.7%)	14 (38.9%)	0.653	12 (18.5%)	15 (23.1%)	0.999
Female	25 (75.8%)	34 (73.9%)		26 (55.3%)	22 (61.1%)		17 (26.2%)	21 (32.3%)	
Race									
Black				4 (8.5%)	6 (16.7%)	0.491			
Hispanic				4 (8.5%)	3 (8.3%)				
White				39 (83%)	27 (75%)				
Braaksc									
Low	3 (9.1%)	3 (6.5%)	0.862	24 (51.1%)	5 (13.9%)	<0.001	4 (6.2%)	2 (3.1%)	0.400
Intermediate	24 (72.7%)	33 (71.7%)		10 (21.1%)	9 (25.0%)				
High	6 (18.2%)	10 (21.7%)		13 (27.7%)	22 (61.1%)		25 (38.5%)	34 (52.3%)	
Dcfdx									
No cognitive impairment	13 (39.4%)	12 (26.1%)	0.025	–	–	–			
Mild cognitive impairment	13 (39.4%)	10 (21.7%)		–	–				
Alzheimer's dementia	7 (21.2%)	23 (50.0%)		–	–				
Other dementia	0 (0%)	1 (2.2%)		–	–				
Cogdx									
No cognitive impairment	15 (45.5%)	12 (26.1%)	0.087	–	–	–			
Mild cognitive impairment	9 (27.3%)	9 (19.6%)		–	–				
Alzheimer's dementia	9 (27.3%)	23 (50.0%)		–	–				
Other dementia	0 (0%)	2 (4.3%)		–	–				
Mmse	24.909 (6.156)	19.867 (8.656)	0.004	–	–	–			
Ceradsc									
Definite AD	9 (27.3%)	24 (52.2%)	0.035	–	–	–			
Probable AD	24 (72.3%)	22 (47.8%)		–	–				
Cdr									
Non- or mild dementia	–	–	–	22 (46.8%)	6 (16.7%)	0.006			
Dementia	–	–	–	25 (53.2%)	30 (83.3%)				
Plaque	–	–	–	6.170 (9.13)	9.954 (9.15)	0.024			

For continuous variables, the *p*-value is calculated based on the Wilcoxon rank-sum test with continuity correction, and for categorical variables, the *p* value is calculated by Pearson's χ^2 -test with Monte Carlo simulated *p* value (based on 2000 replicates).²³

Abbreviations: Braaksc, Braak stage score; Dcfdx, clinical diagnosis of cognitive status; Cogdx, clinical consensus diagnosis of cognitive status at time of death; Mmse, mini-mental state examination; Ceradsc, the consortium to establish a registry for Alzheimer's disease score; Cdr, clinical dementia rating; Plaque, mean of the density of neuritic plaques

3.1.2 | Differential analysis for each data type in brain samples

Differential expression analyses were conducted independently for the three omics data types generated from brain samples between the two subtypes of patients. In proteomics data, 87 proteins were

found to be differentially expressed at a significance level ($FDR < 0.05$). However, the effect size in terms of $\log_2 FC$ between the two clusters was less than 0.2, indicating that the differences were not substantial (Figure 2A). In contrast, in the DNA methylation analysis, we observed that the average magnitude of foldchange between subtypes was minimal (Figure 2B) and that no sites passed the significance threshold

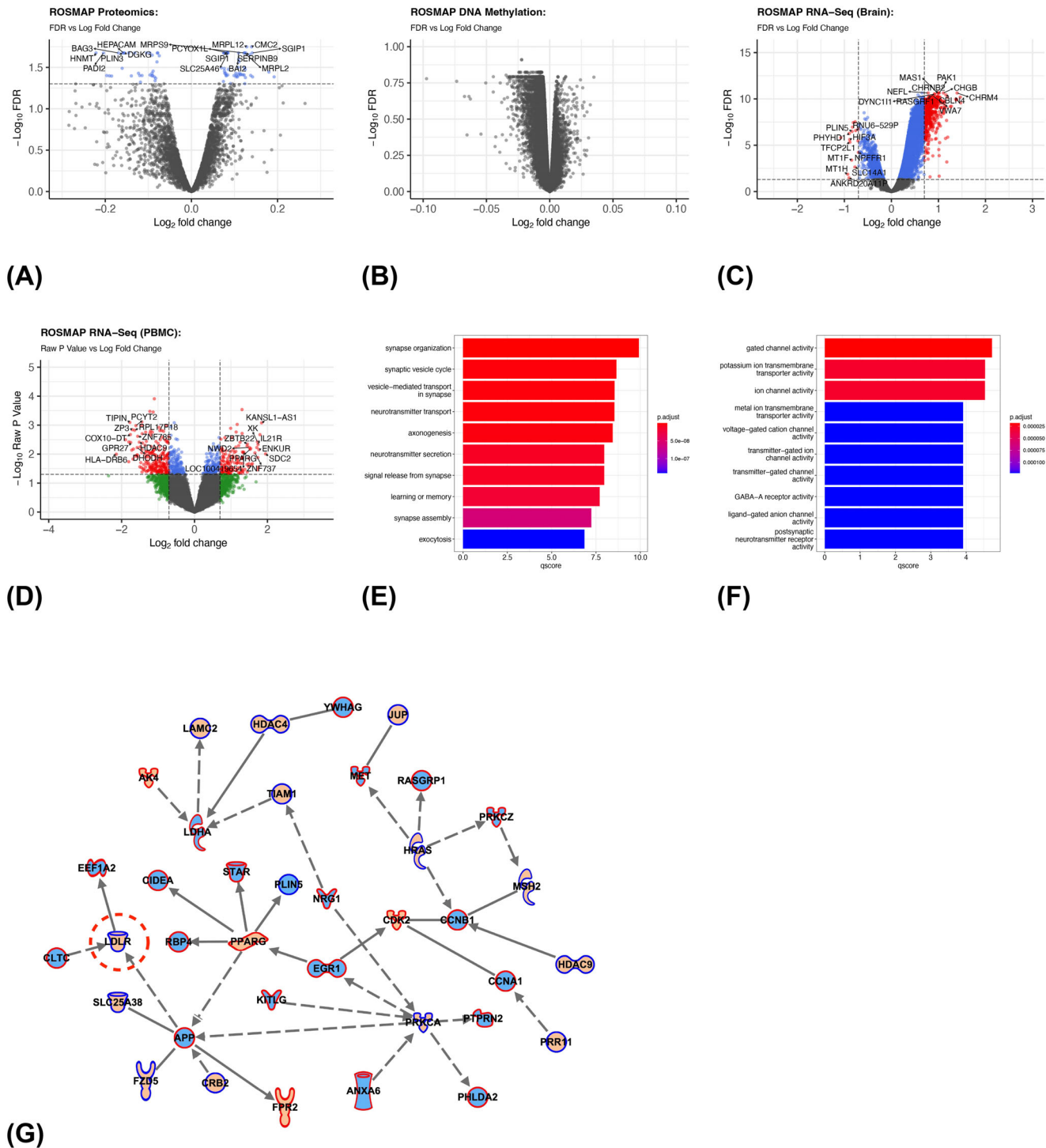


FIGURE 2 Volcano plots for differentially expressed genes (DEGs) in ROSMAP. (A) Proteomics data; (B) DNA methylation data; (C) bulk brain RNAseq; and (D) Peripheral blood mononuclear cell (PBMC) RNAseq data. There were no significantly expressed genes or probes for the DNA methylation and proteomics dataset. There were 524 genes that were significantly expressed between the two clusters in the brain tissue. There were 508 upregulated (Subtypes 1–2) and 16 downregulated genes with the magnitude of $\log_2 FC$ greater than 0.7 and FDR less than 0.05. In addition, in the blood samples, 655 genes were significant with a less strict threshold at the magnitude of $\log_2 FC$ greater than 0.26 and raw p -value less than 0.05. The top 10 upregulated and downregulated genes were labeled with gene symbols in the volcano plot. (E) The top 10 significantly enriched biological processes on DEGs that are upregulated in Subtype 1 compared with Subtype 2; and (F) the top 10 significantly enriched molecular functions for DEGs that are upregulated in Subtype 2 from brain samples in the ROSMAP study. The p -values were corrected with the Benjamini-Hochberg (BH) method, and the q -score was calculated by $-\log_{10} p.adjust$; (G) the network plot displays the shortest pathways between DEGs in brain tissues and DEGs in blood samples. The shape of the node indicates the type of gene, the edge indicates the direction and type of the relationship between the two connected genes, the fill color indicates whether the gene was identified from the brain or blood RNAseq data, and the outline color indicates whether the gene is upregulated or downregulated.

(FDR < 0.05). For brain differential gene expression analysis, 524 genes exhibited significant differential expression between the two subtypes, including 508 upregulated genes in Subtype 1 (higher cognition score) compared to Subtype 2 (lower cognition score), and 16 downregulated genes (Figure 2C). For RNA-seq data collected from PBMC samples, we identified 371 DEGs, including 147 upregulated and 224 downregulated genes, using a threshold of $|\log_2 FC| > 0.7$ and raw p -value < 0.05. Because the transcriptional signal in blood is generally weaker than that in brain tissue, we relaxed the cutoff criteria to capture additional potentially relevant DEGs between Subtype 1 and Subtype 2. With a smaller foldchange threshold, $|\log_2 FC| > 0.26$, a total of 655 genes were identified, of which 374 were downregulated and 281 were upregulated in Subtype 1 compared to Subtype 2 (Figure 2D). Among genes differentially expressed between Subtype 1 and Subtype 2 in both brain tissue and blood, nine genes overlap. Of these DEGs, three genes (*NAPB*, *MLLT11*, and *SMYD2*) were found to have an opposite direction of expression (Figures S2A–C, S3A–C). Specifically, they were upregulated in Subtype 1 in the brain but downregulated in the blood of this group, compared with Subtype 2. The other six genes (*RAB27B*, *XK*, *NWD2*, *DGAT2*, *ENO2*, and *SYT2*) were both upregulated in Subtype 1 for both brain and blood samples (Figures S2D–I, S3D–I).

The genes with consistent upregulation in Subtype 1 are consistent with a robust neuroprotection for individuals with higher cognitive scores for this group. *RAB27B* gene encodes a small GTPase (Ras-related protein Rab-27B) that is critical for the endosomal–lysosomal pathway, regulating vesicular trafficking and the autophagic clearance of toxic protein aggregates. Experimental models demonstrate that the knockdown (loss of function) of *RAB27B* impairs autophagic flux, resulting in the accumulation of insoluble protein and increased neuronal toxicity, suggesting a neuroprotective role.³⁷ This is consistent with our observed upregulation of the gene in the cognitively better subtype. The *SYT2* gene (Synaptotagmin 2), encodes a synaptic vesicle protein, which drives efficient neurotransmitter release. Its concentration has been shown to be decreased in the brains of patients with Parkinson's disease dementia and dementia with Lewy bodies, reflecting significant synaptic loss and dysfunction in those neurodegenerative conditions,³⁸ which is consistent with the observed upregulation of *SYT2* in our cognitively better subtype, suggesting a preservation or compensatory effort to maintain synaptic integrity and efficient neurotransmission. In addition, the observation of upregulated *DGAT* (diacylglycerol acyltransferase-2) in both the brain and blood samples of the cognitively better Subtype 1 highlights the context-dependent nature of lipid metabolism in AD heterogeneity. *DGAT2* is a key enzyme driving the conversion of neurotoxic free fatty acids (FFAs) into inert triglycerides (TGs) and promoting lipid droplet (LD) formation. Previous research has demonstrated that amyloid beta ($A\beta$) exposure elevates *DGAT2*, specifically in microglia, leading to excessive LD buildup that subsequently impairs $A\beta$ phagocytosis and causes neuronal damage.³⁹ These results suggest a context-dependent role for *DGAT2* in AD, where increased expression may be adaptive in systemic and non-microglial compartments and align with preserved cognition.

Among DEGs that are upregulated in the brain and downregulated in the blood in Subtype 1 versus Subtype 2, *NAPB* (SNAP-BETA), which is expressed primarily in the brain and plays a role in synaptic vesicle recycling, yields higher expression in the better-cognition subtype (Subtype 1) in the brain but lower expression in the blood. The loss of *NAPB* function has been shown in previous literature to result in impaired neuronal development.⁴⁰ Therefore, the increased expression of *NAPB* in the brain of Subtype 1 likely represents a successful local compensation to preserve synaptic integrity and maintain the functional network connectivity required for higher cognitive function, directly opposing the synaptic dysfunction characteristic of AD. However, the different expression patterns between the central nervous system (CNS) and blood between the two subtypes indicates that *NAPB* expression levels could vary among tissue types. Another gene with the same opposite expression pattern between brain and blood is *MLLT11* (*MLLT11* transcription factor 7 cofactor). The upregulation of *MLLT11* in the brain of Subtype 1 suggests a protective role in neuronal structural integrity. Previous studies have reported that its expression is restricted to neurons, where it is known to regulate neurite outgrowth, neuronal migration, and the stability of axonal projections.⁴¹ This maintenance of crucial neuronal connectivity and morphology is consistent with the preservation of cognitive function observed in our subtype. However, in blood, *MLLT11* is an established oncogene in leukemia,⁴² and its reduced expression in the higher cognition group may point to a healthier, non-malignant hematopoietic system that supports better overall health and neurological outcomes.

3.1.3 | GO enrichment analysis on DEGs

The GO enrichment was conducted separately for up- and downregulated DEGs for brain samples; however, as the downregulated gene set contained fewer than 15 mapped genes, only the enrichment results for upregulated DEGs are reported. There were 154 biological process (BP) terms in GO enriched in the gene set of DEGs identified from the previous section. The examination revealed notable enrichment in categories associated with "synapse organization" (adj. p -value < 0.001) and "cognition" (adj. p -value < 0.001). Further examination showed that the 41 DEGs involved in the synapse organization pathway are all consistently upregulated in Subtype 1 relative to Subtype 2. This suggests that Subtype 1 exhibits a heightened level of synaptic activity and cellular processes compared to Subtype 2. Moreover, the analysis found that the 30 DEGs related to the cognition pathway are all consistently upregulated in Subtype 1 compared to Subtype 2. This molecular finding aligns directly with the clinical observation that patients in Subtype 1 have better cognitive performance scores than those in Subtype 2, suggesting that the heightened expression of these genes may be a key contributor to the improved cognitive function. Figure 2E shows a compilation of the top enriched GO BP terms, offering more insight into the key molecular processes associated with DEGs between the two subtypes identified by the subspace merging algorithm. These findings imply that the selected genes are pivotal in brain synaptic activities, cognition, and learning, offering a deeper understanding on

the molecular mechanisms underlying AD progression. Similarly, the analysis also revealed significant enrichment in 62 molecular function (MF) terms. For instance, the gene set demonstrated a notable overrepresentation in terms related to “gated channel activity” (adj. p -value < 0.001) and “potassium ion trans-membrane transporter activity” (adj. p -value < 0.001), and “GABA (γ -aminobutyric acid) receptor A activity” (adj. p -value < 0.001), which are all upregulated in Subtype 1. These findings suggest that the genes that are differentially expressed between the two subtypes play crucial roles in the transport of ions across cell membranes and forming channels with the ability to open and close in response to specific signals, which are all essential functions for the nervous system, providing valuable insights into the potential roles of those genes in cellular processes involving ion transport and signaling. Further details and statistical significance measures for each pathway can be referenced in Figure 2F, and the full GO enrichment results for ROSMAP can be found in Table S4.

3.1.4 | Network analysis linking DEGs in brain and blood samples

Despite the absence of genes exhibiting highly statistically significant expression changes with the FDR < 0.05, a subset of 371 genes characterized by a relatively substantial effect size ($|\log_2 FC| > 0.7$) has been identified from the PBMC samples. We performed pathway explorer analysis using QIAGEN Ingenuity Pathway Analysis (IPA)⁴³ to identify the shortest pathways between all DEGs identified within two clusters from the brain region (comprising 524 genes) and genes exhibiting a large foldchange in blood samples (totaling 371 genes). The results revealed networking between multiple genes from blood samples and DEGs within the brain sample.

In the network plot generated by IPA, we want to see if there is any connection between the DEGs identified in the brain samples and those identified from the blood (PBMC) samples. As shown in Figure 2G, our attention was particularly drawn to the *LDLR* gene (low density lipoprotein receptor), as we focused on blood DEGs that connect to multiple brain DEGs in the shortest-path network, and among these, *LDLR* (Figures S2J, S3J) emerged as a key candidate due to its role in lipid metabolism and its multiple connections to brain DEGs, suggesting a potential cross-tissue regulatory relationship. *LDLR* identified as a DEG from the PBMC samples, is a member of the low-density lipoprotein (LDL) family. Its hub role in this pathway indicates this gene is a potential key regulator or mediator for the blood–brain connection as well as for brain transcriptomic changes.

Notably, the *APOE* ϵ 4 (*APOE*4) allele stands as the most common genetic risk factor for AD. Previous studies revealed that apoE plays a crucial role in cholesterol metabolism by binding to various receptors, with *LDLR* exhibiting a notably high affinity for apoE.⁴⁴ It was noteworthy that *LDLR* was the sole member of its receptor family to demonstrate an isoform-specific binding affinity (E4 binds more strongly than E3, which in turn binds more strongly than E2). The *LDLR* gene regulates apoE levels in both the periphery and the CNS in mouse. It has been identified in astrocytes for this function and was shown

to modulate amyloid deposition in AD transgenic mice.⁴⁵ Although previous studies hypothesized that *LDLR* overexpression reduced A β aggregation and enhanced A β clearance from the brain extracellular space,⁴⁶ the observed downregulation of *LDLR* in the blood of individuals in Subtype 1, which is the group with higher cognitive function, challenged *LDLR*'s protective role. Thus, it could serve as a key blood biomarker for identifying distinct AD subtypes and may offer insights into the different underlying mechanisms of cognitive functions. In addition, we identified two other DEGs with established links to AD: *APP* (β -amyloid precursor protein) gene and *PPAR* γ (peroxisome proliferator-activated receptor gamma). The *APP* gene, which is cleaved to form A β plays a crucial role in the aggregation of A β in the brain, a hallmark of AD.^{47,48} Previous literature also reported that *APP* mRNA is highly expressed in neurons and is upregulated in AD brains, with expression patterns and regulatory transcription mechanisms changing progressively with age and increasingly shifting toward dysfunction.⁴⁸ In contrast, our analysis revealed *APP* upregulation in Subtype 1, comprising individuals with higher cognitive scores relative to Subtype 2. This finding contradicts prior reports and further underscores the importance of AD subtyping for disentangling disease heterogeneity, which further emphasizes the necessity and importance of AD subtyping. Furthermore, the nuclear receptor *PPAR* γ (Figure S3K) has emerged as a promising therapeutic target.⁴⁹ *PPAR* γ agonists, which regulate glucose and lipid metabolism and suppress inflammatory gene expression, have shown significant improvements in memory and cognition in patients with AD, highlighting their therapeutic potential in AD treatment.⁵⁰ In line with these findings, we observed *PPAR* γ upregulation in Subtype 1, whereas Subtype 2 showed lower expression. This differential pattern supports that Subtype 1 may retain a protective metabolic and anti-inflammatory signature, thereby distinguishing it from Subtype 2.

3.2 | Analysis of the MSBB dataset

3.2.1 | Patient k-means clustering results and their associations with clinical phenotypes

Similar to the ROSMAP data, the subspace merging algorithm clustered these 83 patients from the MSBB study into two clusters (36 patients in Cluster 1 and 47 patients in Cluster 2; Figure S1B) according to their similarities based on the following omics data: gene expression for BM10, BM22, BM36, BM44, proteomics from BM36, and DNA methylation from the parahippocampal gyrus data.

We also examined the differences in terms of demographic and clinical diagnostic variables between the two clusters in the MSBB study (Table 2). Consistent with what we have seen in ROSMAP, we found that age at death, race, sex, and *APOE* genotype did not show significant differences between the two clusters. Instead, between the two clusters, there were significant differences in A β plaque levels (p -value = 0.024), Clinical Dementia Rating (CDR) (p -value = 0.006), and Braak stage (p -value < 0.001), with Cluster 1 (Subtype 1) exhibited lower Braak State, lower CDR, and reduced plaque accumulation

compared to Cluster 2 (Subtype 2), suggesting a less-advanced disease state. This result demonstrated that at least two major cognitively different subtypes commonly exist in AD, which is not relevant to age, race, sex, and APOE genotype. However, unlike ROSMAP, where the two clusters did not differ by tauopathy (Braak stage), the MSBB clusters showed significant differences in both tauopathy (Braak stage) and amyloid pathology (A β plaque burden) as well as dementia severity (CDR).

3.2.2 | Differential analysis for each data type in MSBB

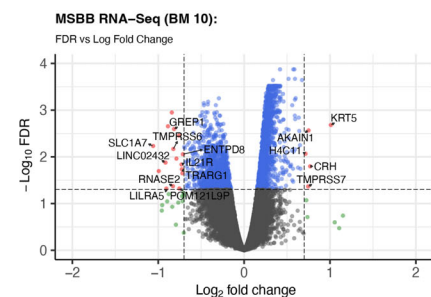
The analysis of brain gene expression patterns revealed prominent distinctions between the two subtypes (Subtype 1 vs Subtype 2). In the BM10 brain region, 21 genes (5 upregulated and 16 downregulated) in Figure 3A exhibited statistically significant differential expression level between the identified clusters with FDR < 0.05 and magnitude of $\log_2FC > 0.7$. Similarly, the BM22 region displayed significant expression differences in 134 genes (24 upregulated and 110 downregulated) in Figure 3B. In BM36, 169 genes (96 upregulated and 73 downregulated) in Figure 3C showed significant expression changes, and in the BM44 region, 110 genes (32 upregulated and 78 downregulated) in Figure 3D demonstrated differential expression patterns, indicating discerning regional variations. However, no differentially expressed features were identified in DNA methylation data (Figure 3E). In addition, we identified 145 differentially expressed proteins ($|\log_2FC| > 0.26$ and FDR < 0.05) with 138 downregulated and seven upregulated. The three downregulated genes with the largest foldchange were SAA2, SAA1, and A4_{A β 42} (Figure 3F). A β protein 42 (or A β ₄₂) is a 42-amino acid peptide derived from the A β precursor protein. Notably, the APP gene was differentially expressed in brain tissue from ROSMAP and found to be upregulated in individuals with higher cognitive scores.

In contrast, A β ₄₂ expression levels were also elevated in Subtype 2 of MSBB cohort Subtype 2, which is consistent with this group's higher A β plaque burden and more advanced Braak stage, which all point to the worse cognitive outcomes. Furthermore, the check of gene intersections between these four regions unveiled shared and region-specific expression profiles (Figure 3). Notably, only the SLC1A7 gene displayed differential expression consistently across BM10, BM22, BM36, and BM44. It is a member of the solute carrier (SLC) superfamily that specifically encodes the EAAT5 protein, which influences the levels of intracellular chloride within neurons and membrane potential.⁵¹ The gene is downregulated in Subtype 1 consistently in all four regions, 19 genes across three regions, and 54 genes across two regions. Of note, all shared DEGs exhibited consistent directions of expression change across regions. The number of DEGs in each region can be found in Figure 3G. Among the DEGs in BM22, GAD2, which encodes glutamate decarboxylase 65, was significantly upregulated in the cognitively preserved Subtype 1. A previous study reported that GAD2 plays a key role in the synthesis of GABA and the maintenance of inhibitory neurotransmission, and reduced GAD1 and GAD2 expression were observed in the brains of patients with AD.⁵² The

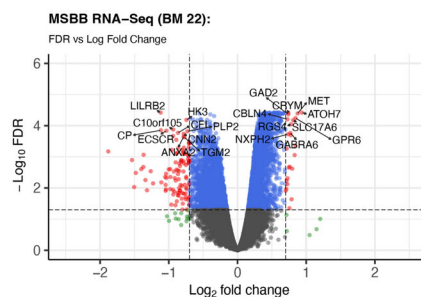
elevated GAD2 expression observed in our Subtype 1 may reflect preserved inhibitory function and reduced excitotoxic stress, consistent with better cognitive performance. In addition, we observed that STAT4, another DEG from BM36, is upregulated in Subtype 1. Previous preclinical work demonstrates that myeloid lineage deficiency of STAT4 protects against Western diet-induced cognitive impairment, type 2 diabetes, and neuroinflammation, suggesting that suppression of STAT4 activation may reduce the risk of AD and other neurodegenerative disorders.⁵³ By contrast, our results show elevated STAT4 expression in Subtype 1, where the patients have better cognition scores, indicating its local protective adaptation in BM36 region.

3.2.3 | GO enrichment analysis on DEGs

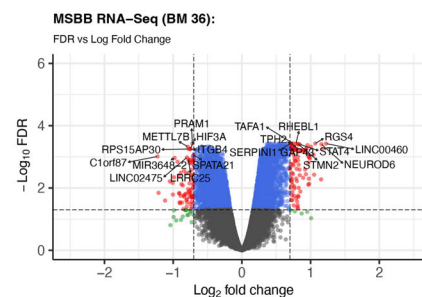
We then conducted enrichment analysis on DEGs (up- and down-regulated DEGs separately) for each of the four above distinct brain regions, where multi-testing was adjusted with BH method. The outcomes reveal significant enrichment in diverse biological processes, with 122 statistically significant GO terms and pathways, including the acute inflammatory response (adj. *p*-value < 0.001), intrinsic apoptotic signaling pathway (adj. *p*-value < 0.001), cellular transition metal ion homeostasis (adj. *p*-value < 0.001), and leukocyte migration (adj. *p*-value < 0.001) within the BM22 region (Figure 3H). And all these enriched GO terms were downregulated in Subtype 1. Moreover, 27 significant molecular functions such as calcium-dependent protein binding (adj. *p*-value < 0.001), cytokine receptor activity (adj. *p*-value < 0.001), RAGE receptor binding (adj. *p*-value < 0.001), and immune receptor activity (adj. *p*-value < 0.001) were identified as enriched in BM22 (Figure 3I). Consistently, all these enriched GO terms were downregulated in Subtype 1, indicating a more active immune response process going on in the more AD-severe Subtype 2 as compared to the higher cognition Subtype 1. Furthermore, two biological process terms are downregulated in Subtype 1 compared to Subtype 2 in both the BM22 and BM44 regions, namely acute inflammatory response (adj. *p*-value = 0.005 in BM22 and adj. *p*-value = 0.013 in BM44), and acute-phase response (adj. *p*-value = 0.004 in BM22 and adj. *p*-value = 0.017 in BM44). Moreover, previous studies indicated that extracellular copper ions can bind directly to A β and promote the aggregation of A β monomers and the formation of more toxic A β fibrils and plaques.⁵⁴ Consistent with this mechanism, we observed upregulation of copper ion binding (adj. *p*-value = 0.006) in Subtype 2 with worse cognitive performance compared with Subtype 1 in BM22. Furthermore, the upregulated DEGs in Subtype 1 were also enriched in nine significant biological process terms (Figure 3J), including positive regulation of secretion by cell (adj. *p*-value = 0.032), potassium ion transport (adj. *p*-value = 0.042), and neuropeptide signaling pathway (adj. *p*-value = 0.032). In addition, the BM36 (Figure 3K) region exhibited enrichment in eight statistically significant molecular functions. Noteworthy functions included neuropeptide hormone activity (adj. *p*-value < 0.001) and peptide hormone receptor binding (adj. *p*-value < 0.001), which were all upregulated in Subtype 1, consistent with this group's higher neuronal function phenotype.



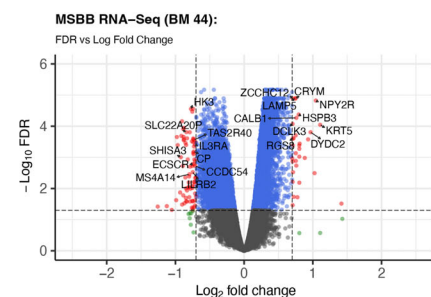
(A)



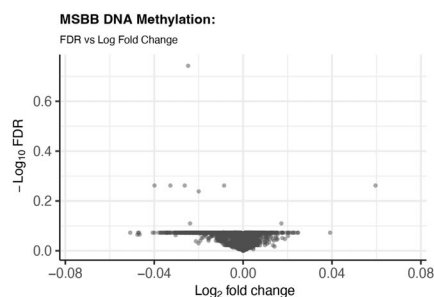
(B)



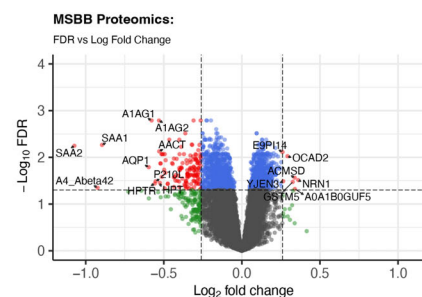
(C)



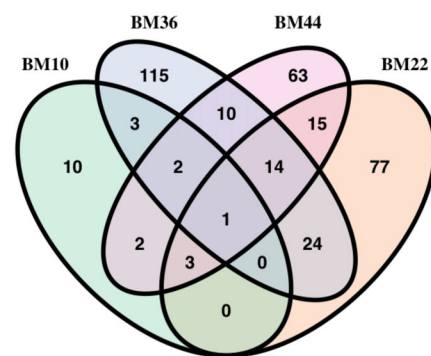
(D)



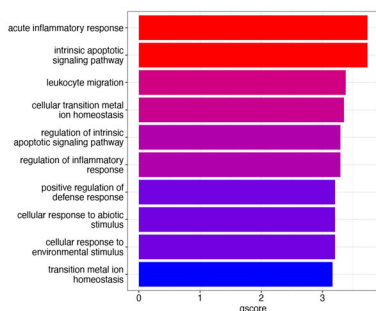
(E)



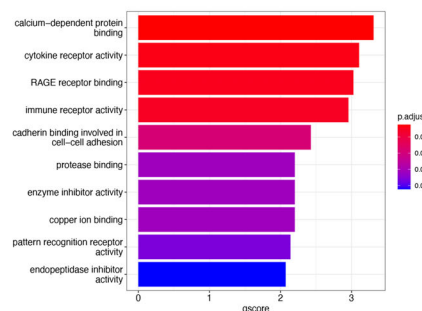
(F)



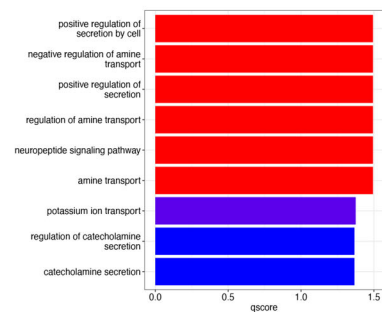
(G)



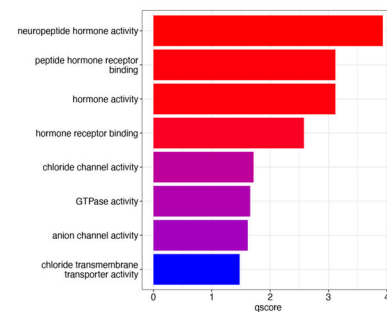
(H)



(I)



(J)



(K)

All the above results showed consistency with the prior knowledge between the gene expression differences and the matched cognitive differences between the two groups, and it enhanced our comprehension of the intricate molecular profile specific to the BM22, BM36, and BM44 regions, and can emphasize commonalities in biological process aspects shared between BM22 and BM44 (Figure 3G). The full GO enrichment results for MSBB can be found in Table S5.

3.3 | Analysis of the MayoRNAseq dataset

3.3.1 | Patient k-means clustering results and their associations with clinical phenotypes

The subspace merging algorithm was applied to cluster the 65 patients from the MayoRNAseq study into two clusters (29 patients in Cluster 1 and 36 patients in Cluster 2; Figure S1C) according to their similarities based on the following omics data: gene transcriptomics data for CBE and TCX and proteomic data. We examined the differences in terms of demographic and clinical diagnostic variables between the two clusters in the MayoRNAseq study (Table 2). There were no significant differences between the two AD subtypes in sex (p -value = 0.999), APOE genotype (p -value = 0.255), or Braak stage (p -value = 0.400). Age differed significantly between subtypes, with Subtype 1 being notably older (p -value = 0.002). Although the Braak stage did not reach statistical significance, Subtype 2 exhibited a visible enrichment of patients with highest Braak scores, Braak score at VI, (20 of 36 in Cluster 2 vs 10 of 29 in Cluster 1). Unlike the ROSMAP and MSBB data, clinical variables besides the Braak stage are not available through the AD Knowledge Portal for the MayoRNAseq dataset, and thus they were not compared.

3.3.2 | Differential analysis for each data type in brain samples

The differential expression analysis between the two AD subtypes identified using the subspace merging algorithm on multi-omics MayoRNAseq data revealed a total of 210 differentially expressed proteins between the two AD subtypes (Figure 4A), including 96 proteins upregulated in Subtype 1, which had fewer patients with higher Braak stages, and 114 proteins downregulated in Subtype 1 ($|\log_2 FC| > 0.26$,

$FDR < 0.05$). In the CBE (Figure 4B), no genes met the threshold for differential expression ($|\log_2 FC| > 0.7$, $FDR < 0.05$). In contrast, 2328 genes were differentially expressed between the subtypes in the TCX region using the same criteria, with 1534 genes upregulated in Subtype 1 and 794 downregulated in Subtype 1 (Figure 4C). Among the DEGs distinguishing the two AD subtypes, *FGR*, a member of the Src family tyrosine kinases, was downregulated in the subtype with fewer individuals at Braak Stage 6, suggesting a relatively earlier or less-advanced pathological state. Previous work⁵⁵ has shown that *FGR* expression increases along the AD pseudotime trajectory and is involved in pathways associated with the transition from early to late stages of disease progression, including neutrophil degranulation-related immune processes. This reported pattern is consistent with the expression differences observed in the MayoRNAseq dataset. Another downregulated DEG in Subtype 1, *SERPINA5*, a gene whose expression is known to increase with worsening tau pathology as previous studies,^{56,57} identified this gene as a novel tau-binding partner, co-localizing with neurofibrillary tangles and neuritic plaques, and elevated *SERPINA5* levels have been linked to hippocampal vulnerability in AD brains. Therefore, the lower *SERPINA5* expression in our less-severe subtype aligns with the reduced tau burden and supports the biological plausibility of the molecular differences captured by our clustering.

3.3.3 | GO enrichment analysis on DEGs

To further characterize the biological differences underlying the two AD subtypes, we conducted GO enrichment analysis using the top 500 upregulated DEGs and top 500 downregulated DEGs between the two subtypes, ranked by log foldchange (Subtype 1 – Subtype 2). We assessed the enrichment of molecular function and biological process categories to determine which pathways were overrepresented among the most differentially expressed genes in the TCX. We identified 458 enriched biological process items among the downregulated DEGs in Subtype 1 (Figure 4D), including regulation of vasculature development (adj. p -value < 0.001) and regulation of angiogenesis (adj. p -value < 0.001), copper detoxification (adj. p -value < 0.001), and zinc ion homeostasis (adj. p -value < 0.001). These pathways collectively suggest that the more severe subtype exhibits greater neurovascular remodeling and altered metal ion regulation, which contributes to A β aggregation and remodeling of the blood–brain barrier.^{58–61}

FIGURE 3 Volcano plots for differentially expressed genes (DEGs) in MSBB. (A) Brodmann area (BM)10 RNAseq; (B) BM22 RNAseq; (C) BM36 RNAseq; (D) BM44 RNAseq; (E) DNA methylation; and (F) Proteomics data. In BM10, 21 genes (5 upregulated and 16 downregulated) were differentially expressed between Subtypes 1 and 2 (false discovery rate (FDR) < 0.05 and $|\log_2 FC| > 0.7$). BM22, BM36, and BM44 showed 134 (24 upregulated and 110 downregulated), 169 (96 upregulated and 73 downregulated), and 110 (32 upregulated and 78 downregulated) differentially expressed genes, respectively, under the same threshold. No differentially expressed features were detected in the DNA methylation data. In the proteomics data, 145 differentially expressed proteins were identified ($|\log_2 FC| > 0.26$ and $FDR < 0.05$), of which 138 were downregulated and 7 were upregulated in Subtype 1. (G) The number of DEGs in each region. The p -values were corrected with the BH method. Pathway analysis for MSBB RNA-seq data from different brain regions. (H) Top 10 significant biological processes of DEGs downregulated in Subtype 1 in BM22; (I) top 10 significant molecular functions of DEGs downregulated in Subtype 1 in BM22; (J) top 10 significant biological processes of DEGs upregulated in Subtype 1 in BM36; and (K) top 10 significant molecular functions of DEGs upregulated in Subtype 1 in BM36.

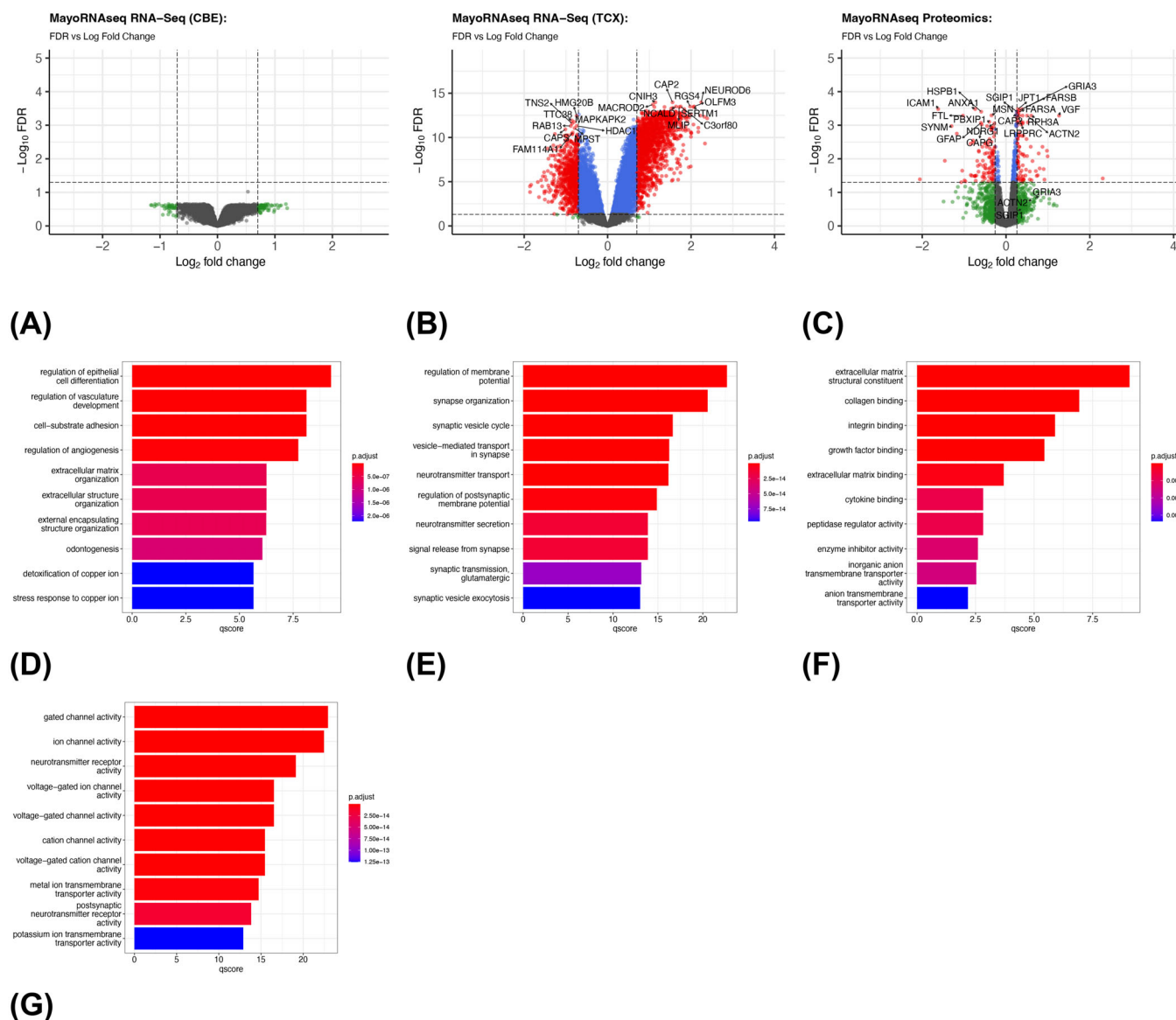


FIGURE 4 Volcano plots for differentially expressed genes (DEGs) in MayoRNAseq. (A) Cerebellum (CBE) RNAseq; (B) Temporal Cortex (TCX) RNAseq; (C) and proteomics data. In CBE region, no genes were differentially expressed between Subtype 1 and Subtype 2 (false discovery rate (FDR) < 0.05 and $|\log_2 FC| > 0.7$). TCX region showed 2328 (1678 upregulated and 794 downregulated) differentially expressed genes, under the same threshold. In the proteomics data, 210 differentially expressed proteins were identified ($|\log_2 FC| > 0.26$ and FDR < 0.05), of which 114 were downregulated and 96 were upregulated in Subtype 1. (D) Top 10 significant biological processes of DEGs downregulated in Subtype 1 in TCX; (E) top 10 significant biological processes of DEGs upregulated in Subtype 1 in TCX; (F) top 10 significant molecular functions of DEGs downregulated in Subtype 1 in TCX; and (G) top 10 significant molecular functions of DEGs upregulated in Subtype 1 in TCX.

Among the top 500 upregulated DEGs between the two subtypes, 357 biological process items were significantly enriched (Figure 4E), including pathways related to neurotransmission and synaptic signaling such as regulation of membrane potential (adj. p -value < 0.001), synapse organization (adj. p -value < 0.001), synaptic vesicle cycle (adj. p -value < 0.001), and neurotransmitter transport (adj. p -value < 0.001). In addition, these genes that are overexpressed in the subtype with fewer Braak stage VI patients were enriched in learning or memory (adj. p -value < 0.001) neuron cognition (adj. p -value = 0.025), and cognition (adj. p -value < 0.001), consistent with our observation

that Subtype 1 exhibits relatively preserved neuronal and cognitive functions.

In addition, molecular functions related to the extracellular matrix (ECM), such as ECM structural constituent (adj. p -value < 0.001) and ECM binding (adj. p -value < 0.001), were downregulated in Subtype 1 (Figure 4F). These ECM-related processes are known to play key roles in inflammation response and tissue structural changes in neurodegenerative diseases, particularly AD.⁶² Furthermore, top upregulated DEGs in Subtype 1 were enriched in key neuronal and synaptic molecular functions, including ion channel activity (adj. p -value < 0.001),

voltage-gated channel activity (adj. p -value < 0.001), neurotransmitter receptor activity (adj. p -value < 0.001), and GABA receptor activity (adj. p -value < 0.001). These enrichments point to preserved synaptic signaling and neuronal function in this less-severe subtype (Figure 4G). The full GO enrichment results for MayoRNAseq can be found in Table S6.

3.4 | Comparison of results from ROSMAP and MSBB datasets

In our studies on both cohorts, we identified two distinct AD subtypes when applying the subspace merging algorithm to multi-omics data. Demographic variables, including age at death and sex at birth, as well as the *APOE* genotype (a well-established genetic risk factor for AD), did not differ significantly between subtypes in either cohort. This indicates that the observed clinical differences between subtypes are purely data-driven and unlikely to be confounded by demographic or genetic variation and are more likely driven by differences in their molecular profiles. As mentioned above, we observed phenotypic consistency among the two subtypes identified separately from ROSMAP and MSBB. At the molecular level, we also found common DEGs between the two subtypes across the two cohorts. One gene, *CRH* encodes corticotropin-releasing hormone, was identified as differentially expressed in both ROSMAP brain tissue and the BM10 brain region from MSBB. Moreover, *CRH* was identified as an upregulated DEG in Subtype 1 of ROSMAP and also an upregulated DEG in Subtype 1 of MSBB. Across both cohorts, *CRH* expression was consistently higher in the patient subtype with better cognitive scores. In addition, we found nine DEGs from BM22, 35 from BM36, and 11 from BM44 that are common to ROSMAP brain DEGs, all of which showed consistent directions of differential expression between the two cohorts. This cross-cohort concordance not only reinforces the biological and clinical relevance of the subtype-specific transcriptomic signatures but also demonstrates the robustness and generalizability of our approach and findings across independent datasets and brain regions.

Furthermore, GO enrichment analysis revealed that several molecular functions were consistently upregulated in the AD subtype associated with better cognitive outcomes across both ROSMAP and MSBB (BM36) cohorts, where six molecular functions were significantly enriched including neuropeptide hormone activity, anion channel activity, chloride channel activity, chloride transmembrane transporter activity, hormone activity, and GTPase activity, suggesting that enhanced hormonal, ion channel activities, and ion transportation may characterize a protective role against cognitive decline. In addition, we also identified seven biological processes that upregulated in the AD subtype of better cognition in both cohorts, including potassium ion transport, positive regulation of secretion by cell, regulation of catecholamine secretion, catecholamine secretion, positive regulation of secretion, regulation of amine transport, and amine transport, indicating that the increased activity in these pathways involving neuronal signaling and synaptic communications suggests stronger communi-

cations between brain cells, which is consistent with the phenotypic characteristics of Subtype 1 in both cohorts.

3.5 | Comparison of results from the ROSMAP and MayoRNAseq datasets

In both cohorts, we applied the proposed subspace merging algorithm and generated *APOE* genotype-independent AD subtypes. Although in MayoRNAseq data, the clinical diagnosis variable Braak stage is not significantly different between the two AD subtypes, Subtype 1 contains fewer patients with a Braak score of VI, suggesting a comparatively less-pathological group. At the molecular level, among the 1534 overexpressed DEGs in Subtype 1 of the MayoRNAseq dataset, 439 were also differentially expressed between the two AD subtypes in the ROSMAP brain samples. Of note, all of these genes were consistently upregulated in the subtype characterized by better cognitive performance. Notably, several genes were differentially expressed in both brain and blood tissues in ROSMAP, including *NAPB*, *NWD2*, *RAB27B*, *MLLT11*, *SMYD2*, *DGAT2*, *ENO2*, and *XK*, highlighting their potential as robust, cross-tissue biomarkers associated with this less-severe AD subtype. In addition, we also identified several DEGs that were consistently downregulated in the subtype with better cognition or less pathological compared with the other subtype across both ROSMAP and MayoRNAseq datasets, including *PHYHD1*, *HIF3A*, *MT1F*, *SLC14A1*, and *MT1H*. The concordant downregulation of these genes in two independent cohorts further supports the robustness of the molecular differences that distinguish the AD subtype.

Furthermore, across the two cohorts, the DEGs distinguishing the AD subtypes showed strong functional concordance. Specifically, 140 GO biological process terms were enriched in both datasets, of which 135 exhibited the same direction of change in the subtype with better cognition and lower pathological burden. These included key neuronal and cognitive pathways such as synapse organization, synaptic vesicle cycling, learning or memory, cognition, and hormone transport. Only five biological process terms showed opposite directions between cohorts, including cell junction assembly, chloride transport, cell-cell adhesion via plasma membrane adhesion molecules, chloride transmembrane transport, and inorganic anion transmembrane transport.

A similar pattern was observed at the molecular function category. In total, 48 GO MF terms were shared between cohorts and demonstrated the same direction of regulation, including major neuronal functions such as gated channel activity, ion channel activity, potassium ion transmembrane transport, voltage-gated cation channel activity, and hormone binding. Only seven molecular function terms displayed opposite directions, including anion channel activity, chloride channel activity, chloride transmembrane transporter activity, transmembrane receptor protein kinase activity, transmembrane receptor protein tyrosine kinase activity, glycosaminoglycan binding, and inorganic anion transmembrane transporter activity. In conclusion, these findings indicate a high degree of cross-cohort consistency, with the vast majority of molecular pathways showing the same directional regulation.

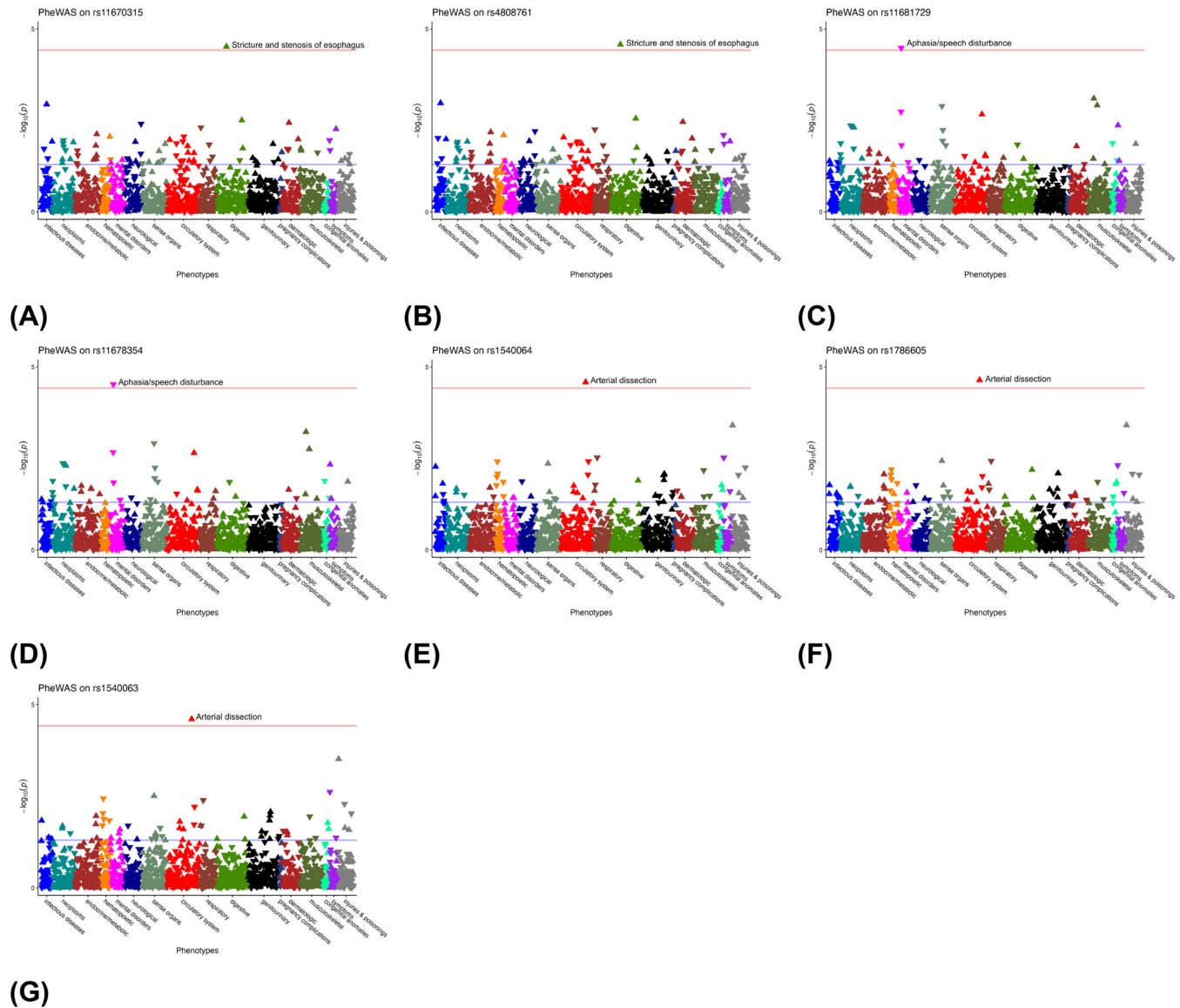


FIGURE 5 Manhattan plot of expression quantitative trait loci (eQTLs). The y-axis presents the $-\log_{10}p$ - value and the x-axis presents the phenotype category. The arrow presents the direction of the odds ratio. The blue line suggests the raw p -value at level 0.05, and the red line presents the Bonferroni corrected p -value using only all association tests involving the selected eQTL. (A) rs11670315, (B) rs4808761, (C) rs11681729, (D) rs11678354, (E) rs1540064, (F) rs1786605, and (G) rs1540063.

3.6 | PheWAS analysis

Although no significant association was identified with a stringent FDR correction of 0.05 on all association tests, we were still able to identify 16 associations at the raw p -value threshold of 10^{-4} , which helps to reveal the potential association between pre-selected eQTLs that are identified from previous eQTL analysis and the phenotypical data available in the EHR. In addition, among these 16 associations mentioned, seven were significant with the eQTL-specific multi-testing correction strategy. The details of the PheWAS results can be found in Table 3. Notably, two eQTLs of *RAB3A*, which was a DEG from the brain tissues: rs11670315 (odds ratio (OR) = 1.756, $p = 3.02 \times 10^{-5}$, Figure 5A) and rs4808761 (OR = 1.764, $p = 2.66 \times 10^{-5}$, Figure 5B) on chromosome 19 were strongly associated with stricture and stenosis of the esoph-

agus. In addition, the two eQTLs targeting *MEIS1* identified from the blood sample: rs11681729 (OR = 0.596, $p = 3.30 \times 10^{-5}$, Figure 5C) and rs11678354 (OR = 0.595, $p = 2.98 \times 10^{-5}$, Figure 5D) on chromosome 2 were strongly associated with aphasia/speech disturbance. Furthermore, the three eQTLs of *DTNA*, which was a DEG from blood sample: rs1540064 (OR = 3.543, $p = 2.60 \times 10^{-5}$, Figure 5E), rs1786605 (OR = 3.570, $p = 2.27 \times 10^{-5}$, Figure 5F), and rs1540063 (OR = 3.552, $p = 2.48 \times 10^{-5}$, Figure 5G) on chromosome 18 were strongly associated with arterial dissection.

These findings highlighted potential genetic links between AD subtype differences and other diseases, warranting further investigation into their biological mechanisms and potential clinical implications. For instance, the alternative allele of both rs11678354 (T/A) and rs11681729 (A/G) was associated with an increase in the gene

TABLE 3 PheWAS results.

eQTL	Chr	Gene	Phecode	Phenotype description	Category	OR	p value	adj.p value	eQTL.P value
rs1786605	18	DTNA	442.4	Arterial dissection	Circulatory system	3.570	2.273×10^{-5}	0.567	0.031
rs1540063	18	DTNA	442.4	Arterial dissection	Circulatory system	3.552	2.483×10^{-5}	0.567	0.033
rs1540064	18	DTNA	442.4	Arterial dissection	Circulatory system	3.543	2.597×10^{-5}	0.567	0.035
rs4808761	19	RAB3A	530.3	Stricture and stenosis of esophagus	Digestive	1.764	2.656×10^{-5}	0.567	0.036
rs11678354	2	MEIS1	292.1	Aphasia/speech disturbance	Mental Disorders	0.595	2.980×10^{-5}	0.567	0.040
rs11670315	19	RAB3A	530.3	Stricture and stenosis of esophagus	Digestive	1.756	3.023×10^{-5}	0.567	0.041
rs11681729	2	MEIS1	292.1	Aphasia/speech disturbance	Mental Disorders	0.596	3.299×10^{-5}	0.567	0.044
rs9489197	6	ROS1	963.1	Antineoplastic and immunosuppressive drugs causing adverse effects	Injuries and Poisonings	3.167	4.153×10^{-5}	0.567	0.056
rs9320603	6	ROS1	963.1	Antineoplastic and immunosuppressive drugs causing adverse effects	Injuries and Poisonings	3.167	4.162×10^{-5}	0.567	0.056
rs9320602	6	ROS1	963.1	Antineoplastic and immunosuppressive drugs causing adverse effects	Injuries and Poisonings	3.166	4.185×10^{-5}	0.567	0.056
rs11086095	19	RAB3A	530.3	Stricture and stenosis of esophagus	Digestive	1.736	4.424×10^{-5}	0.567	0.060
rs6707386	2	PAX8-AS1	870.3	Other open wound of head and face	Digestive	2.856	6.910×10^{-5}	0.686	0.090
rs9789607	2	MEIS1	292.1	Aphasia/speech disturbance	Mental Disorders	0.612	7.120×10^{-5}	0.686	0.096
rs75271568	7	PTPRN2	288	Diseases of white blood cells	Hematopoietic	0.503	7.575×10^{-5}	0.686	0.102
rs4849177	2	PAX8-AS1	870.3	Other open wound of head and face	Injuries and Poisonings	2.806	9.480×10^{-5}	0.686	0.125
rs6755040	2	PAX8-AS1	870.3	Other open wound of head and face	Injuries and Poisonings	2.805	9.520×10^{-5}	0.686	0.124

We identified 16 important associations at an arbitrary *p*-value threshold of 10^{-5} suggesting a potential association between 105 pre-selected expression quantitative trait loci (eQTLs) and the Phecodes available in the electronic health records (EHRs). The adj. *p*-values were adjusted with Benjamini-Hochberg (BH) procedure, and the eQTL *P* values were adjusted with BH method using only associations involving the selected eQTL.

expression level of *MEIS1*, a gene that has been identified as a risk factor for sudden cardiac death, and previous mouse models suggested that *MEIS1* interacts with various trophic factors signaling pathways during postmitotic neurons differentiation.⁶³ Furthermore, both these two eQTLs were negatively associated with the risk of aphasia/speech disturbance, indicating that as the number of alternative alleles appeared in a patient's genotypic profile, the risk of getting the disease would decrease.

Furthermore, the alternative allele of both rs11670315 (G/A) and rs4808761 (A/C) was associated with an increase in the gene expression level of *RAB3A*, which was upregulated in Subtype 1 of ROSMAP. Previous studies identified that this gene could regulate A β production and was deregulated in AD brains.^{64–67} In addition, the PheWAS results also showed that the frequency of alternative alleles in the three eQTLs targeting *DTNA* positively correlates with the risk of arterial dissection. Previous studies revealed that the expression profile of *DTNA* in the hippocampus (HIP) increased significantly among subjects with dementia and associated levels of phosphorylated tau (p-tau) in the temporal cortex.⁶⁸ A meta-analysis of gene expression data also revealed that this gene was upregulated in three studies on Cornu Ammonis 1 region (CA1) of the HIP of brains with AD.⁶⁹

4 | DISCUSSION

In this study, we demonstrated that the Subspace Merging Algorithm effectively integrates multi-omics data using graph-based subspace analysis, generating informative clusters by preserving complementary omic-specific representations. The repeatable results in three independent large AD cohort studies showed that it is a powerful tool to integrate multiple types of omics data and clustering disease patients into subtypes with distinctive phenotypes, and in our AD cases, the distinctive cognitive and brain pathology. The resulting clusters are not significantly associated with age (except for the MayoRNAseq dataset), sex, or *APOE* status. The clusters are revealed only by data-driven integrative multi-omics analysis.

Among genes differentially expressed between Subtype 1 and Subtype 2 in both the brain tissue and blood in ROSMAP, which contains the most comprehensive multi-modal data, nine genes overlap in total. Of these DEGs, three (*NAPB*, *MLLT11* and *SMYD2*) were found to have an opposite direction of expression (Figure S2). Specifically, they were upregulated in Subtype 1 in the brain but downregulated in the blood of this group. The upregulations in the brain of Subtype 1 showed consistent neuroprotective roles of those genes and the subtype's superior cognition condition, whereas the downregulation in blood may be due to the interference with physiological conditions from elsewhere of the body. The other six genes that overlap between two tissues: *RAB27B*, *XK*, *NWD2*, *DGAT2*, *ENO2* and *SYT2* were both upregulated in Subtype 1 for both brain and blood samples (Figure S2). The observed differences in DEGs patterns between Subtype 1 and Subtype 2 in ROSMAP between brain and blood samples reveal critical insights into the tissue-specific and context-dependent functions of key genes. This seeming inconsistency can be attributed to the different cellular environments

and biological processes occurring in the CNS versus the peripheral immune system.

Through pathway analysis, we identified *LDLR* (Figure S2) as one of the key genes linking multiple brain DEGs and blood DEGs. The *LDLR* showed a larger fold-change expression level in PBMCs between the two subtypes. Notably, increased *LDLR* expression level has been associated with reduced amyloid burden and enhanced glial response, suggesting its potential as a therapeutic target.⁷⁰ Previous pre-clinical studies have strongly suggested that *LDLR* enhances the clearance of brain A β thereby modulating A β metabolism.⁷¹ *LDLR* emerges as a pivotal pathway influencing A β metabolism, rendering it an attractive therapeutic target for AD.⁴⁶ A comprehensive understanding of the underlying biological mechanisms of this functional interaction can contribute to more effective treatment design and management strategies for different subtypes of AD. Of interest, our analysis revealed that *LDLR* expression in the blood sample was lower in Subtype 1, which exhibited better cognitive outcomes. It seems to be contrary to the purported protective role of *LDLR* against AD pathology in the aforementioned pre-clinical study. We hypothesize that it could be either a feedback mechanism related to AD pathology or the effects of *LDLR* regulation from liver or other internal organs. In the periphery, *LDLR* is key to LDL clearance, and defects leading to hypercholesterolemia have been correlated with cognitive impairment in mouse models.⁷² In addition, we postulated that the high *LDLR* expression in PBMC in Subtype 2 could be interpreted as an ineffective compensatory response by the peripheral immune cells to systemic metabolic distress, where increased quantity of the receptor fails to mitigate the functional consequences of the underlying severe dyslipidemia and inflammation. Future follow-up work may further clarify the exact role of *LDLR*. In addition, *APP* was upregulated in Subtype 1 in the ROSMAP dataset, while the A β ₄₂ level is upregulated in Subtype 2, the patient group with worse cognitive score in the MSBB study. Another lipid metabolism player, *PPAR γ* , is also identified to be upregulated in Subtype 1. It is a key regulator of lipid metabolism and inflammation whose agonists improve cognition,⁵⁰ consistent with Subtype 1's better cognition phenotype.

The study further reveals that the DEGs between the two subtypes identified from the brain sample in ROSMAP are enriched in pathways governing synaptic plasticity, cognitive function, and neuronal signaling, which suggests that impaired neuronal communication is a central mechanism underlying their distinct pathologies. The consistent upregulation of synapse- and cognition-related genes (such as *BDNF*, *GRM5*, *NPTN*, and *CHRNA2*) in Subtype 1 may underlie the better cognitive outcomes observed in this group. This work suggests that future therapeutic interventions may be most effective if tailored to the specific pathway disruptions characterizing each patient subtype.

Our enrichment analysis also highlighted molecular function pathways, with significant overrepresentation of terms related to gated channel activity, potassium ion transmembrane transport, and GABA-A receptor activity. These pathways were enriched in Subtype 1, the subgroup with relatively better cognitive performance. Therefore, the enrichment of ion transport and receptor activity pathways in Subtype 1 may indicate preserved synaptic signaling capacity, providing

a molecular basis for the comparatively better cognitive outcomes observed in this subgroup. Together, these findings offer strong evidence for the reproducibility, robustness, and biological relevance of our subtyping framework.

One challenge of validating the study is the lack of datasets that exactly matches the data modality, brain region, and clinical information. Nevertheless, we further applied the method on two related large cohort studies of AD besides ROSMAP to test if similar subgrouping of patients can be detected. Although these datasets may not fully validate each other due to different choices of samples and data modalities, potential concordances among the subgroups can shed light on key genes and processes that are associated with AD clinical presentations.

In the MSBB cohort, among the DEGs between Subtype 1 and Subtype 2 in the four brain regions, we observed significant upregulation of copper ion-binding pathways in the BM22 region in Subtype 2, which was associated with worse cognitive outcomes compared with Subtype 1. This finding is consistent with previous studies showing that extracellular copper can directly bind to A β , accelerating the aggregation of monomers into toxic fibrils and plaques. The enrichment of copper-binding genes in the worse cognition group, therefore, supports the role of copper ions in promoting amyloid pathology and cognitive decline.

In the MayoRNAseq cohort, we observed a high degree of cross-cohort consistency between itself and ROSMAP, with 439 shared DEGs consistently upregulated in the less-severe subtype in both cohorts. We identified 135 biological process terms, including synapse organization, synaptic vesicle cycling, and cognition, and 41 molecular function terms such as gated ion channel activity and voltage-gated channel activity showed the same directional regulation across both datasets. The observation of these shared directionally consistent GO terms across independent cohorts further illustrated the robustness of the AD subtypes and yielded a reproducible molecular framework underlying AD heterogeneity.

Through PheWAS, we identified novel associations between eQTLs of subtype-specific DEGs and common disease phenotypes in the EHR data of the All of Us study. The PheWAS results broadened our understanding of how specific genetic variations could influence multiple health conditions.

To summarize, in this study, we applied an advanced graph-based unsupervised subspace merging algorithm to integrate multi-omics data for clustering patients and identifying potential subtypes in AD from postmortem brain samples. We successfully identified different groups of AD patients with distinct clinical presentations only from this data-driven approach and discovered potential blood-based markers. We further explored the associations of the genetic variants of these marker genes of other disease phenotypes with AD subtypes.

Although our subspace merging framework provides a powerful approach to integrating multi-omics data for identifying AD subtypes, several limitations exist in this study. First, due to the data availability, we focused only on postmortem AD brain tissue, which may not fully capture the dynamic and temporal nature of AD progression.

Additional studies incorporating longitudinal data, such as imaging or blood-based biomarkers would further shed light on the heterogeneity of AD progression, which is not reflected in this study. Second, the MSBB cohort may not be an ideal validation cohort for the findings from ROSMAP due to the anatomic differences in the brain regions and lack of blood data in this cohort, limiting the test for generalizability of the identified subtypes. Nevertheless, we compared the clustering findings from the two cohorts in parallel and discussed the common characteristics of the subtypes as well as their differences. Third, the integrative model assumes equal contribution from each omics layer. Although this assumption simplifies computation and reduces overfitting risk, it may obscure modality-specific signals that are biologically relevant to certain pathways or patient subgroups. Adaptive fusion strategies, such as modality-specific weighting or attention mechanisms, can enhance the model's ability to prioritize more informative data layers. Although we initiated joint modeling of brain and peripheral tissues, extending this approach through transfer learning or multitask learning may help identify more clinically actionable subtypes. With the rapid progress in machine learning, we will also further explore advanced graph-learning methods to refine subtype discovery and representation.

ACKNOWLEDGMENTS

We extend our sincere gratitude to the participants of the All of Us Research Program. We also acknowledge the National Institutes of Health's (NIH's) All of Us Research Program for providing access to the participant data examined in this research.

The data available in the AD Knowledge Portal would not be possible without the participation of research volunteers and the contribution of data by collaborating researchers. The results published here are in whole or in part based on data obtained from the AD Knowledge Portal (<https://adknowledgeportal.org>).

The Religious Orders Study (ROS) and Memory and Aging Project (MAP) data were provided by the Rush Alzheimer's Disease (AD) Center, Rush University Medical Center, Chicago. Data collection was supported through funding by National Institute on Aging (NIA) grants P30AG10161 (ROS), R01AG15819 (ROSMAP; genomics and RNAseq), R01AG17917 (MAP), R01AG30146, R01AG36042 (5hC methylation, Assay for Transposase-Accessible Chromatin using sequencing (ATACseq)), RC2AG036547 (H3K9Ac), R01AG36836 (RNAseq), R01AG48015 (monocyte RNAseq) RF1AG57473 (single nucleus RNAseq), U01AG32984 (genomic and whole exome sequencing), U01AG46152 (ROSMAP Accelerating Medicines Partnership Program for Alzheimer's Disease (AMP-AD), targeted proteomics), U01AG46161 (Tandem Mass Tag (TMT) proteomics), U01AG61356 (whole genome sequencing, targeted proteomics, ROSMAP AMP-AD), the Illinois Department of Public Health (ROSMAP), and the Translational Genomics Research Institute (genomic). Additional phenotypic data can be requested at www.radc.rush.edu.

The Mount Sinai Brain Bank (MSBB) data were generated from post-mortem brain tissue collected through the Mount Sinai VA Medical Center Brain Bank and were provided by Dr. Eric Schadt from Mount Sinai School of Medicine.

The MSBB proteomics data were provided by Dr. Levey from Emory University based on postmortem brain tissue collected through the Mount Sinai VA Medical Center Brain Bank provided by Dr. Eric Schadt from Mount Sinai School of Medicine.

The ROSMAP TMT proteomics data were provided through the Accelerating Medicine Partnership for AD (U01AG046161 and U01AG061357) based on samples provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984, U01AG46152, the Illinois Department of Public Health, and the Translational Genomics Research Institute.

Data generation was supported by the following NIH grants P30AG10161, P30AG72975, R01AG15819, R01AG17917, R01AG036836, U01AG46152, U01AG61356, U01AG046139, P50 AG016574, R01 AG032990, U01AG046139, R01AG018023, U01AG006576, U01AG006786, R01AG025711, R01AG017216, R01AG003949, R01NS080820, U24NS072026, P30AG19610, U01AG046170, RF1AG057440, and U24AG061340, and the Cure PSP, Mayo and Michael J Fox foundations, Arizona Department of Health Services, and the Arizona Biomedical Research Commission.

We thank the participants of the ROSMAP projects for the generous donation, the Sun Health Research Institute Brain and Body Donation Program, the Mayo Clinic Brain Bank, and the Mount Sinai/JJ Peters VA Medical Center NIH Brain and Tissue Repository. Data and analysis contributing investigators include Nilüfer Ertekin-Taner, Steven Younkin (Mayo Clinic, Jacksonville, FL), Todd Golde (University of Florida), Nathan Price (Institute for Systems Biology), David Bennett, Christopher Gaiteri (Rush University), Philip De Jager (Columbia University), Bin Zhang, Eric Schadt, Michelle Ehrlich, Vahram Haroutunian, Sam Gandy (Icahn School of Medicine at Mount Sinai), Koichi Iijima (National Center for Geriatrics and Gerontology, Japan), Scott Noggle (New York Stem Cell Foundation), Lara Mangravite (Sage Bionetworks).

The Mayo RNAseq study (MayoRNAseq) data was led by Dr. Nilüfer Ertekin-Taner, Mayo Clinic, Jacksonville, FL as part of the multi-PI U01 AG046139 (MPIs Golde, Ertekin-Taner, Younkin, Price). Samples were provided from the following sources: The Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, and R01 AG003949, National Institute of Neurological Disorders and Stroke (NINDS) grant R01 NS080820, CurePSP Foundation, and support from Mayo Foundation. Study data include samples collected through the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona. The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026 National Brain and Tissue Resource for Parkinsons Disease and Related Disorders), the NIA (P30 AG19610 Arizona Alzheimers Disease Core Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimers Research Center), the Arizona Biomedical Research Commission (con-

tracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson's Disease Consortium), and the Michael J. Fox Foundation for Parkinsons Research.

The MayoRNAseq proteomics data were provided by Dr. Levey from Emory University based on postmortem brain tissue provided by The Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01 AG003949, NINDS grant R01 NS080820, CurePSP Foundation, and support from Mayo Foundation.

This project is partially supported by the National Institutes of Health (NIH) 5U54AG065181 (Indiana TREAT-AD Center) and NIH R21AG075541 (Indiana University Precision Health Initiative).

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests. Author disclosures are available in the [Supporting Information](#).

CONSENT STATEMENT

This study did not involve human participants or the use of identifiable human data.

ORCID

Ziyan Song  <https://orcid.org/0009-0000-5679-1751>

Kun Huang  <https://orcid.org/0000-0002-8530-370X>

REFERENCES

1. Komarova NL, Thalhauser CJ. High degree of heterogeneity in Alzheimer's disease progression patterns. *PLoS Comput Biol*. 2011;7(11):e1002251.
2. Ding H, Sharpnack M, Wang C, Huang K, Machiraju R. Integrative cancer patient stratification via subspace merging. *Bioinformatics*. 2018;35(10):1653-1659. doi:10.1093/bioinformatics/bty866
3. A Bennett D, A Schneider J, Arvanitakis Z, S Wilson R. Overview and findings from the religious orders study. *Curr Alzheimer Res*. 2012;9(6):628-645.
4. Wang M, Beckmann ND, Roussos P, et al. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci Data*. 2018;5(1):180185.
5. Allen M, Carrasquillo MM, Funk C, et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci Data*. 2016;3(1):160089. doi:10.1038/sdata.2016.89
6. Denny JC, Ritchie MD, Basford MA, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*. 2010;26(9):1205-1210.
7. Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci*. 2013;368(1620):20120362.
8. Wang B, Mezlini AM, Demir F, et al. Brudno, and A. Haibe-Kains Benjamin and Goldenberg, Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods*. 2014;11(3):333-337.
9. Argelaguet R, Velten B, Arnol D, et al. Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol*. 2018;14(6):e8124.
10. Mo Q, Shen R, Guo C, Vannucci M, Chan KS, Hilsenbeck SG. A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data. *Biostatistics*. 2018;19(1):71-86.

11. Mostafavi S, Gaiteri C, Sullivan SE, et al. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat Neurosci*. 2018;21(6):811-819.
12. De Jager PL, Ma Y, McCabe C, et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data*. 2018;5(1):180142.
13. De Jager PL, Srivastava G, Lunnon K, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBD2 and other loci. *Nat Neurosci*. 2014;17(9):1156-1163.
14. Higginbotham L, Ping L, Dammer EB, et al. Integrated proteomics reveals brain-based cerebrospinal fluid biomarkers in asymptomatic and symptomatic Alzheimer's disease. *Sci Adv*. 2019;6(43):eaaz9360.
15. Johnson ECB, Dammer EB, Duong DM, et al. Deep proteomic network analysis of Alzheimer's disease brain reveals alterations in RNA binding proteins and RNA splicing associated with disease. *Mol Neurodegener*. 2018;13(1):52.
16. Ping L, Kundinger SR, Duong DM, et al. Global quantitative analysis of the human brain proteome and phosphoproteome in Alzheimer's disease. *Sci Data*. 2020;7(1):315.
17. Ping L, Duong DM, Yin L, et al. Global quantitative analysis of the human brain proteome in Alzheimer's and Parkinson's Disease. *Sci Data*. 2018;5(1):180036.
18. Johnson ECB, Dammer EB, Duong DM, et al. Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nat Med*. 2020;26(5):769-780. doi:10.1038/s41591-020-0815-6
19. Bai B, Wang X, Li Y, et al. Deep Multilayer Brain Proteomics Identifies Molecular Networks in Alzheimer's Disease Progression. *Neuron*. 2020;105(6):975-991.e7. doi:10.1016/j.neuron.2019.12.015
20. The All of Us Research Program Genomics Investigators. "Genomic data in the All of Us Research Program," *Nature* 627(8003):340-346, 2024.
21. Sieberts SK, Perumal TM, Carrasquillo MM, et al. Large eQTL meta-analysis reveals differing patterns between cerebral cortical and cerebellar brain regions. *Sci Data*. 2020;7(1):340.
22. Zelnik-manor L, Perona P. Self-Tuning Spectral Clustering. In: Saul L, Weiss Y, Bottou L, eds. *Advances in Neural Information Processing Systems*. MIT Press; 2004. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2004/file/40173ea48d9567f1f393b20c855bb40b-Paper.pdf
23. Hope ACA. A simplified Monte Carlo significance test procedure. *J R Stat Soc Series B Stat Methodol*. 1968;30(3):582-598.
24. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
25. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995;57(1):289-300.
26. Yu G, Wang L, Han Y, He Q. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-287.
27. Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for genome-wide association studies in the R environment. *Bioinformatics*. 2014;30(16):2375-2376.
28. Yu L, Lutz MW, Wilson RS, et al. TOMM40'523 variant and cognitive decline in older persons with APOE ε3/ε3 genotype. *Neurology*. 2017;88(7):661-668. doi:10.1212/WNL.0000000000003614
29. Bennett DA, Wilson RS, Schneider JA, et al. Natural history of mild cognitive impairment in older persons. *Neurology*. 2002;59(2):198-205.
30. Bennett DA, Schneider JA, Aggarwal NT, et al. Decision rules guiding the clinical diagnosis of Alzheimer's disease in two community-based cohort studies compared to standard practice in a clinic-based cohort study. *Neuroepidemiology*. 2006;27(3):169-176.
31. Schneider JA, Arvanitakis Z, Bang W, Bennett DA. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. *Neurology*. 2007;69(24):2197-2204.
32. Mufson EJ, Malek-Ahmadi M, Perez SE, Chen K. Braak staging, plaque pathology, and APOE status in elderly persons without cognitive impairment. *Neurobiol Aging*. 2016;37:147-153. doi:10.1016/j.neurobiolaging.2015.10.012
33. Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol*. 1991;82(4):239-259.
34. Bennett DA, Schneider JA, Arvanitakis Z, et al. Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology*. 2006;66(12):1837-1844.
35. Mirra SS, Heyman A, McKeel D, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology*. 1991;41(4):479-486.
36. Folstein MF, Folstein SE, McHugh PR. Mini-Mental State Examination. *American Psychological Association (APA)*. 2014.
37. Underwood R, Wang B, Carico C, Whitaker RH, Placzek WJ, Yacoubian TA. The GTPase Rab27b regulates the release, autophagic clearance, and toxicity of α-synuclein. *J Biol Chem*. 2020;295(23):8005-8016.
38. Bereczki E, Francis PT, Howlett D, et al. Synaptic proteins predict cognitive decline in Alzheimer's disease and Lewy body dementia. *Alzheimers Dement*. 2016;12(11):1149-1158.
39. Prakash P, Manchanda P, Paouri E, et al. Amyloid-β induces lipid droplet-mediated microglial dysfunction via the enzyme DGAT2 in Alzheimer's disease. *Immunity*. 2025;58(6):1536-1552.e8.
40. Ali G, Shin KC, Habbab W, et al. Characterization of a loss-of-function NSF attachment protein beta mutation in monozygotic triplets affected with epilepsy and autism using cortical neurons from proband-derived and CRISPR-corrected induced pluripotent stem cell lines. *Front Neurosci*. 2023;17:1302470.
41. Stanton-Turcotte D, Hsu K, Moore SA, Yamada M, Fawcett JP, Iulianella A. Mllt11 regulates migration and neurite outgrowth of cortical projection neurons during development. *J Neurosci*. 2022;42(19):3931-3948.
42. Xiong Y, Li Z, Ji M, et al. MIR29B regulates expression of MLLT11 (AF1Q), an MLL fusion partner, and low MIR29B expression associates with adverse cytogenetics and poor overall survival in AML. *Br J Haematol*. 2011;153(6):753-757.
43. Krämer A, Green J, Pollard J Jr, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*. 2014;30(4):523-530.
44. Johnson LA, Olsen RH, Merckens LS, et al. Apolipoprotein E-low density lipoprotein receptor interaction affects spatial memory retention and brain ApoE levels in an isoform-dependent manner. *Neurobiol Dis*. 2014;64:150-162.
45. Katsouri L, Georgopoulos S. Lack of LDL receptor enhances amyloid deposition and decreases glial response in an Alzheimer's disease mouse model. *PLoS One*. 2011;6(7):e21880.
46. Kim J, Castellano JM, Jiang H, et al. Overexpression of low-density lipoprotein receptor in the brain markedly inhibits amyloid deposition and increases extracellular Aβ clearance. *Neuron*. 2009;64(5):632-644.
47. Tcw J, Goate AM. Genetics of β-amyloid precursor protein in Alzheimer's disease. *Cold Spring Harb Perspect Med*. 2017;7(6):a024539.
48. Hampel H, Hardy J, Blennow K, et al. The amyloid-β pathway in Alzheimer's disease. *Mol Psychiatry*. 2021;26(10):5481-5503.
49. Auwerx J, Cock T, Knouff C. PPAR-γ: a thrifty transcription factor. *Nucl Recept Signal*. 2003;1(1):e006.
50. Landreth G, Jiang Q, Mandrekar S, Heneka M. PPARγ agonists as therapeutics for the treatment of Alzheimer's disease. *Neurotherapeutics*. 2008;5(3):481-489.

51. Lee A, Balcar VJ, McCombe, Pow DV. Human brain neurons express a novel splice variant of excitatory amino acid transporter 5 (hEAAT5v). *J Comp Neurol*. 2020;528(17):3134-3142.
52. Hill MA, Gammie SC. Alzheimer's disease large-scale gene expression portrait identifies exercise as the top theoretical treatment. *Sci Rep*. 2022;12(1):17189.
53. Zhang X, Hollander CM, Khan MY, et al. Myeloid cell deficiency of the inflammatory transcription factor Stat4 protects long-term synaptic plasticity from the effects of a high-fat, high-cholesterol diet. *Commun Biol*. 2023;6(1):967.
54. Li Y, Tan S, Wu D, Zhang Q, Wang T, Zheng G. The role of intracellular and extracellular copper compartmentalization in Alzheimer's disease pathology and its implications for diagnosis and therapy. *Front Neurosci*. 2025;19:1553064.
55. Huang X, Jannu AJ, Song Z, et al. Predicting Alzheimer's disease subtypes and understanding their molecular characteristics in living patients with transcriptomic trajectory profiling. *Alzheimer's & Dementia*. 2025;21(1):e14241. doi:10.1002/alz.14241
56. Crist AM, Hinkle KM, Wang X, et al. Transcriptomic analysis to identify genes associated with selective hippocampal vulnerability in Alzheimer's disease. *Nat Commun*. 2021;12(1):2311. doi:10.1038/s41467-021-22399-3
57. Matchett BJ, Lincoln SJ, Baker M, et al. The SERPINA5 coding variant E228Q does not contribute to clinicopathologic characteristics in Alzheimer's disease: a cross-sectional study. *Medicine*. 2023;102(24):e34017. doi:10.1097/MD.00000000000034017
58. Szewczyk B. Zinc homeostasis and neurodegenerative disorders. *Front Aging Neurosci*. 2013;5:33. doi:10.3389/fnagi.2013.00033
59. Qi Z, Liu KJ. The interaction of zinc and the blood-brain barrier under physiological and ischemic conditions. *Toxicol Appl Pharmacol*. 2019;364:114-119. doi:10.1016/j.taap.2018.12.018
60. Gao F, Zhan Y, Wang Q, Zhang M, Dai L, Shen Y. Pathological angiogenesis was associated with cerebrovascular lesion and neurodegeneration in Alzheimer's disease. *Alzheimer's & Dementia*. 2025;21(2):e14521. doi:10.1002/alz.14521
61. Eisenmenger LB, Peret A, Famakin BM, et al. Vascular contributions to Alzheimer's disease. *Transl Res*. 2023;254:41-53. doi:10.1016/j.trsl.2022.12.003
62. Sun Y, Xu S, Jiang M, et al. Role of the Extracellular Matrix in Alzheimer's Disease. *Front Aging Neurosci*. 2021;13:707466. doi:10.3389/fnagi.2021.707466
63. Bouilloux F, Thireau J, Ventéo S, et al. Loss of the transcription factor Meis1 prevents sympathetic neurons target-field innervation and increases susceptibility to sudden cardiac death. *Elife*. 2016;5:e11627.
64. Udayar V, Buggia-Prévot V, Guerreiro R, et al. A paired RNAi and RabGAP Overexpression screen identifies Rab11 as a regulator of β -amyloid production. *Cell Rep*. 2013;5(6):1536-1551.
65. Jordan KL, Koss DJ, Outeiro TF, Giorgini F. Therapeutic targeting of Rab GTPases: relevance for Alzheimer's disease. *Biomedicine*. 2022;10(5):1141.
66. Reddy PH, Mani G, Park BS, et al. Differential loss of synaptic proteins in Alzheimer's disease: implications for synaptic dysfunction. *J Alzheimers Dis*. 2005;7(2):103-117; discussion 173-80.
67. Zhang X, Huang TY, Yancey J, Luo H, Zhang Y. Role of Rab GTPases in Alzheimer's disease. *ACS Chem Neurosci*. 2019;10(2):828-838.
68. Simon MJ, Wang MX, Murchison CF, et al. Transcriptional network analysis of human astrocytic endfoot genes reveals region-specific associations with dementia status and tau pathology. *Sci Rep*. 2018;8(1):12389.
69. Hosseini S, Arefian E, Rakhsh-Khorshid H, et al. A meta-analysis of gene expression data highlights synaptic dysfunction in the hippocampus of brains with Alzheimer's disease. *Sci Rep*. 2020;10(1):8384.
70. Wood H. LDL receptor upregulation reduces tau pathology. *Nat Rev Neurol*. 2021;17(8):462.
71. Castellano JM, Deane R, Gottesdiener AJ, et al. Low-density lipoprotein receptor overexpression enhances the rate of brain-to-blood A β clearance in a mouse model of β -amyloidosis. *Proc Natl Acad Sci U S A*. 2012;109(38):15502-15507.
72. de Oliveira J, Engel DF, de Paula GC, et al. High cholesterol diet exacerbates blood-brain barrier disruption in LDLr^{-/-} mice: impact on cognitive function. *J Alzheimers Dis*. 2020;78(1):97-115.

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: Song Z, Huang X, Jannu AJ, Johnson TS, Zhang J, Huang K. Identification of Alzheimer's disease subtypes and biomarkers from human multi-omics data using subspace merging algorithm. *Alzheimer's Dement*. 2026;22:e71292. <https://doi.org/10.1002/alz.71292>