Received: 6 May 2024 | Revised: 18 July 2024 | Accepted: 2 August 2024

DOI: 10.1002/alz.14241

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

Predicting Alzheimer's disease subtypes and understanding their molecular characteristics in living patients with transcriptomic trajectory profiling

Xiaoqing Huang[1] | **Asha Jacob Jannu**[2] | **Ziyan Song**[1] | **Nur Jury-Garfe**[3] | **Cristian A. Lasagna-Reeves**[3] | **Alzheimer's Disease Neuroimaging Initiative**[#] | **Travis S. Johnson**[1] | **Kun Huang**[1] | **Jie Zhang**[4]

[1]Department of Biostatistics & Health Data Science, Indiana University School of Medicine, Indianapolis, Indiana, USA

[2]Department of Biohealth Informatics, Indiana University School of Medicine, Indianapolis, Indiana, USA

[3]Department of Anatomy, Cell Biology & Physiology, Indiana University School of Medicine, Indianapolis, Indiana, USA

[4]Department of Medical & Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana, USA

Correspondence
Xiaoqing Huang, Department of Biostatistics & Health Data Science, Indiana University School of Medicine, Indianapolis, IN 46202, USA.
Email: huanxi@iu.edu

Jie Zhang, Department of Medical & Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN 46202, USA.
Email: jizhan@iu.edu

[#]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Funding information
Alzheimer's Disease Neuroimaging Initiative; National Institutes of Health, Grant/Award Number: U01 AG024904; Department of Defense, Grant/Award Number: W81XWH-12-2-0012; National Institute on Aging; National Institute of Biomedical Imaging and Bioengineering; AbbVie;

Abstract

INTRODUCTION: Deciphering the diverse molecular mechanisms in living Alzheimer's disease (AD) patients is a big challenge but is pivotal for disease prognosis and precision medicine development.

METHODS: Utilizing an optimal transport approach, we conducted graph-based mapping of transcriptomic profiles to transfer AD subtype labels from ROSMAP monocyte samples to ADNI and ANMerge peripheral blood mononuclear cells. Subsequently, differential expression followed by comparative pathway and diffusion pseudotime analysis were applied to each cohort to infer the progression trajectories. Survival analysis with real follow-up time was used to obtain potential biomarkers for AD prognosis.

RESULTS: AD subtype labels were accurately transferred onto the blood samples of ADNI and ANMerge living patients. Pathways and associated genes in neutrophil degranulation-like immune process, immune acute phase response, and IL-6 signaling were significantly associated with AD progression.

DISCUSSION: The work enhanced our understanding of AD progression in different subtypes, offering insights into potential biomarkers and personalized interventions for improved patient care.

KEYWORDS
AD prognosis, AD subtyping, ADNI, ANMerge, disease progression, optimal transport, ROSMAP, sex-specific difference, trajectory pseudotime

Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd; Indiana University Precision Health Initiative NIH TREAT-AD U54, Grant/Award Numbers: 5U54AG065181, 5R01NS119280

Highlights

- We applied an innovative optimal transport-based approach to map transcriptomic data from different Alzheimer's disease (AD) cohort studies and transfer known AD subtype labels from ROSMAP monocyte samples to peripheral blood mononuclear cell (PBMC) samples within ADNI and ANMerge cohorts.
- Through comprehensive trajectory and comparative analysis, we investigated the molecular mechanisms underlying different disease progression trajectories in AD.
- We validated the accuracy of our AD subtype label transfer and identified prognostic genetic markers associated with disease progression, facilitating personalized treatment strategies.
- By identifying and predicting distinctive AD subtypes and their associated pathways, our study contributes to a deeper understanding of AD heterogeneity.

1 | BACKGROUND

Although beta-amyloid (Aβ) plaque and neural fibrillary tangles (NFTs) that result from tau protein aggregation have been the generally accepted Alzheimer's disease (AD) hallmark pathological changes, a subset of AD patients demonstrates a disproportionate cognitive decline compared to the severity of their pathologies. These atypical patients remain relatively understudied, and there is a critical need to stratify patients with different disease progression subtypes, identify biomarkers for each subtype for early diagnosis, and develop personalized therapy.[1–4] Here, we focus on the uncoupled tauopathy and cognitive decline and have identified asymptomatic (Asym) AD (high NFTs, low or no cognitive decline) and Low-NFT AD (low NFTs, severe cognitive decline) subtypes in various postmortem AD brain studies.[1,3] However, detecting such atypical disease progression subtypes in living patients poses significant challenges, and current non-invasive methods for measuring tauopathies, such as plasma assays and positron emission tomography (PET) imaging, lack sufficient accuracy for diagnosing such subtypes brain-wide.[5–8] Moreover, the current AD field assesses AD progression using clinical data of observable symptoms and cognitive assessments, which may not fully capture the underlying molecular and neuropathological changes occurring in the brain.[9,10] This can result in a limited understanding of disease progression and hinder the identification of early biomarkers and therapeutic targets. Lastly, many previous studies have focused on static snapshots of gene expression, overlooking the temporal dynamics inherent in AD pathogenesis.[3,11–14] Given the slowly developing nature of AD, it is imperative to employ methodologies that can effectively capture the evolving molecular signatures over time. Investigating the disease's pseudo-trajectory through transcriptomic profiling sheds light on the molecular traits associated with disease progression and variability, an area briefly explored but not extensively studied in previous research, particularly in AD subtypes differentiated by tauopathy and cognitive decline.[15–19] As the heterogeneous nature of AD phenotypes demands methodologies capable of capturing the intrinsic molecular signatures present in distinct subgroups of patients, we aim to enhance our understanding of atypical AD (ie, the subtypes with tauopathy and cognitive decline uncoupling) by mapping the blood transcriptomics from subtype-known postmortem cohorts' to living patients' peripheral blood mononuclear cell (PBMC) samples in the new cohorts and by forecasting such atypical subtypes in the PBMC transcriptomic data for the new cohorts.

In our study, we first propose a novel graph-based approach for estimating molecular characteristics in living patients of AD subgroups using optimal transport (OT),[20] transferring the known subtype label from the Religious Orders Study and Rush Memory and Aging Project (ROSMAP) onto the Alzheimer's Disease Neuroimaging Initiative (ADNI) and AddNeuroMed consortium (ANMerge) cohorts. Leveraging this methodology in the context of AD allows for the precise identification of molecular features unique to specific patient subgroups while maintaining the integrity of the original data structure. This approach facilitates the transfer of AD patient classifications from the postmortem ROSMAP cohort to living patient cohorts such as ADNI and ANMerge, ensuring consistency in labeling across different study groups. Subsequently, in order to validate our prediction and to elucidate the molecular characteristics possessed within each predicted subtype, we undertake a comparative analysis of gene expression between the Asym AD subtype versus Typical AD to uncover underlying genes and pathways associated with Asym AD samples which can be of neuroprotective value. Additionally, through the integration of inferred patient labels with longitudinal clinical parameters, we are capable of further probing into the temporal dynamics of AD-related gene alterations and their correlations with disease subtypes and clinical outcomes. Finally, we investigate these potential gene markers' association with clinical outcomes to evaluate their clinical significance as prognostic biomarkers. Grasping the link between genetic marker expression and cognitive decline can provide valuable insights into the underlying mechanisms of AD. Our approach holds great potential for early prognosis and effective personalized treatments for AD and related tauopathies by providing clinicians with actionable insights into the dynamic molecular processes underlying disease progression.

2 | METHODS

2.1 | AD subtype definition and patient characteristics

In our study, ROSMAP monocyte transcriptomic data was obtained from the Accelerating Medicines Partnership (AMP)-AD consortium Synapse (https://www.synapse.org). Four subgroups are classified based on their Braak & Braak stage (BBS), clinical consensus diagnosis of cognitive status at the time of death (COGDX), and semiquantitative estimates of neuritic plaque density as recommended by the Consortium to Establish a Registry for Alzheimer's Disease (CERAD), modified to be implemented without adjustment for age and clinical diagnosis. We defined our AD subtypes as Asym AD (BBS ≥4, COGDX ≤2, $N = 86$ cases), which has high tau protein deposition of NFTs but low or no AD cognitive decline, Low-NFT AD (BBS ≤2, COGDX = 4 or 5, $N = 5$ cases), which has Low NFTs but with AD symptoms, normal Control (BBS ≤2, COGDX = 1, CERAD = 4, $N = 18$ cases), and Typical AD (BBS ≥4, COGDX = 4 or 5, CERAD = 1, $N = 45$ cases). For ADNI, with the inferred BBS using OT, discussed in Section 2.2, in combination with clinical diagnosis (Clinical Dementia Rating [CDR]), we assigned the patients to four groups: Asym AD (inferred BBS ≥4, CDR ≤4, $N = 341$ cases), Low-NFT AD (inferred BBS ≤2, CDR ≤2, $N = 20$ cases), normal Control (inferred BBS ≤2, CDR ≤0.5, N = 32 cases), and Typical AD (Inferred BBS ≥4, CDR > 4, $N = 70$ cases). Similarly, for ANMerge, we assigned the patients to four groups, Asym AD (inferred BBS ≥4, CDR ≤4, $N = 250$ cases), Low- NFT AD (inferred BBS ≤2, CDR ≤2, $N = 38$ cases), normal Control (inferred BBS ≤2, CDR ≤0.5, $N = 27$ cases), and Typical AD (inferred BBS ≥4, CDR > 4, $N = 141$ cases). The demographics for these three datasets are shown in Supplementary Table S1. All source data analyzed in the study are publicly available.

2.2 | Graph-based label transfer approaches

As mentioned in Section 2.1, since the whole brain BBS measurement is not available in living patients, we need to infer the BBS and assign subtype labels to them. The overview of the label transfer workflow is shown in Figure 1. In the following sections, we describe in detail how the graph-based OT method was used to transfer BBS labels from ROSMAP patients to those in ADNI and ANMerge, then the BBS inferred by our OT method was used to define AD subtypes in the ADNI and ANMerge cohorts. Note that we pre-validated the workflow by transferring ROSMAP postmortem brain RNA sequencing (RNA-seq) to the matched ROSMAP monocyte RNA-seq data to obtain the label transfer model prototype, which is not included in this paper. Furthermore, we compared our OT method with other state-of-the-art label transfer methods in terms of label transfer accuracy. Clinical pTau and cognitive measurements were also compared among groups to verify the subtype accuracy.

RESEARCH IN CONTEXT

1. **Systematic review**: Examination of peer-reviewed (PubMed) and preprint literature found limited studies of AD subtyping based on uncoupled tauopathy and cognitive decline. We started with predefined AD subtypes, then leveraged optimization techniques to infer the AD subtype labels in living patients. Subsequently, we performed differentially expressed, trajectory, survival, and comparative analyses to identify the underlying drivers for disease progression.
2. **Interpretation**: Neutrophil degranulation, acute phase response, and IL-6 signaling pathways were found to be significantly associated with the pathological progression of AD in living patients, including 41 key genes with temporal features. Moreover, the study also investigated sex-specific differences related to key genes.
3. **Future directions**: Our research underscores the importance of utilizing trajectory profiling on transcriptomic data to elucidate pertinent disease progression mechanisms, emphasizing the role of chronic drivers in AD risk. Future endeavors will involve conducting association analyses on larger datasets and undertaking biological and clinical studies to further validate their implication in AD progression.

2.2.1 | Co-embed transcriptomic data by optimal transport

Label transfer refers to transferring the labels in reference data to a set of target data. Many label transfer methods have been developed recently.[20–23] The overall principles of these label transfer methods are similar. First, the reference data and the target data are co-embedded in the same low-dimensional space. Then, the labels of the samples in the target data are assigned to the most frequently represented labels around the samples by the classic k-nearest neighbors (KNN) algorithm (Figure 1A). Here, we propose using OT to transfer BBS from ROSMAP to ADNI and ANMerge. OT is a mathematical tool that aims to find the most efficient way to move probability mass between two distributions. Using label transfer from ROSMAP to ADNI as an example, we describe the whole process, which includes two steps: (1) co-embed ROSMAP RNA-seq data and ADNI microarray data to the same low-dimension space; and (2) label propagation in the co-embedded space using the KNN algorithm. We elaborate on the details as follows.

We first used OT to co-embed ROSMAP RNA-seq data and ADNI microarray data. We let $\{x_i \in \mathbb{R}^{genes}\}_{i=1}^{n}$ and $\{y_j \in \mathbb{R}^{genes}\}_{j=1}^{m}$ represent the collection of transcriptomic data for n patients in ROSMAP data and m patients in ADNI data, respectively. Please note that x_i and y_j are transcriptomic data with the same set of genes. Furthermore, we

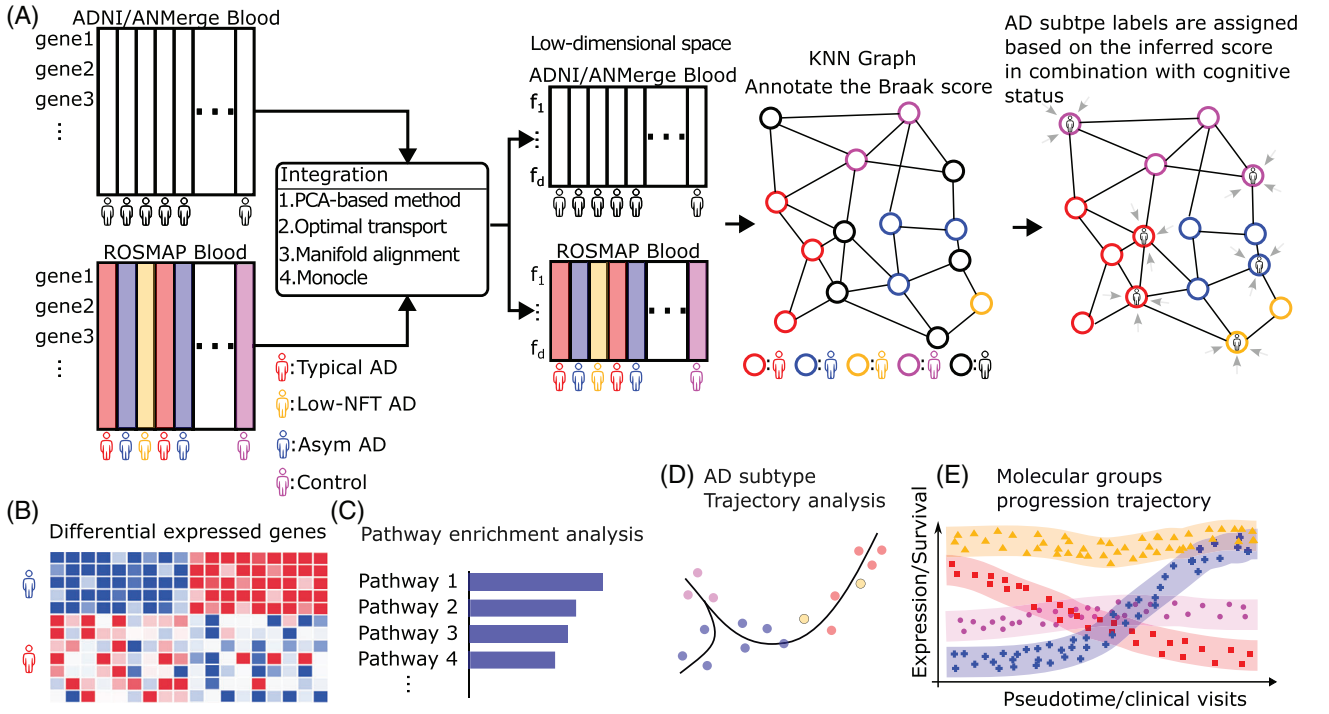

FIGURE 1 Overview of transfer learning to infer patients' subtypes from ROSMAP to ADNI and ANMerge. (A) An illustration of Braak & Braak staging (BBS) and AD subtype transfer learning steps from ROSMAP (monocyte RNA-seq) to ADNI/ANMerge (PMBCs in living patients). (B) Differential gene expression analysis and pathway enrichment annotation. (C) Pathway enrichment analysis using IPA. (D) Trajectory analysis and pseudotime inference. (E) Branch-specific characterization for disease progression. AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; ANMerge, AddNeuroMed consortium; Asym, asymptomatic; IPA, Ingenuity Pathway Analysis; KNN, k-nearest neighbors; NFT, neural fibrillary tangle; PCA, principal component analysis; PBMC, peripheral blood mononuclear cell; ROSMAP, Religious Orders Study and Rush Memory and Aging Project.

computed the pairwise distance (such as Euclidean distance or Manhattan distance) M_{ij} *between* x_i and y_j, which can be interpreted as the penalty of moving x_i and y_j together. Then the optimal transport that moves the two distributions of $\{x_i \in \mathbb{R}^{genes}\}_{i=1}^{n}$ and $\{y_j \in \mathbb{R}^{genes}\}_{j=1}^{m}$ could be found by solving

$$\min_{\Gamma \in \mathbb{R}^{n \times m}} \sum_{i,j} \Gamma_{ij} M_{ij}$$

$$s.t.\ \Gamma 1 = p_x\ ; \Gamma^T 1 = p_y\ ;\ \Gamma \geq 0,$$

where p_x and p_y are the empirical distribution of $\{x_i \in \mathbb{R}^{genes}\}_{i=1}^{n}$ and $\{y_j \in \mathbb{R}^{genes}\}_{j=1}^{m}$, respectively. The solution Γ_{ij} is the OT of moving the probability mass from x_i to y_j. After we obtained the optimal Γ_{ij}, we could co-embed the transcriptomic data in ROSMAP and ANDI by the barycentric projection. The same approach was also used to co-embed ROSMAP with ANMerge transcriptomics.

2.2.2 | The KNN-based graph to assign BBS labels

After the OT step, ROSMAP patients' and ADNI patients' gene expressions were co-embedded into the same space. We then built the KNN graph, in which a node represents a patient in either the ROSMAP or ADNI studies and an edge is drawn from each node to its k-nearest neighbors based on a distance computed from the expression of the nodes. In the KNN graph, patients in the ROSMAP study bear BBS labels, but patients in the ADNI study do not. We could then assign the BBS label to the ANDI patients based on their k-nearest neighbors' BBS consensus. In this paper, we set k to 1. We followed the same procedure to assign BBS to ANMerge patients from ROSMAP patients.

2.2.3 | AD subtyping in ADNI and ANMerge

Once we obtained the inferred BBSs for each patient in ADNI and ANMerge based on the competing methods, we could use the definition of AD subtypes introduced in Section 2.1 to classify patients in ADNI and ANMerge into four AD subtypes.

2.3 | Model selection and comparison metrics

2.3.1 | Compare label transfer methods to choose the best model

Many label transfer methods have been developed for single-cell data analysis.[20–23] We compared our OT-based label transfer method with the state-of-the-art methods to evaluate the performance against them. Here, we repurposed the methods originally developed for integrating and label transfer for single-cell RNA-seq data to transfer BBS from ROSMAP patients to ADNI and ANMerge patients. In our study,

patient-level instead of single-cell–level transcriptomic data from the three cohorts were used to do the BBS label transfer. The methods we chose to compare with OT are Ingest,[23] Monocle,[24] and maximum mean discrepancy manifold alignment (MMD-MA).[21] Both Ingest and Monocle use a principal component analysis (PCA)-based method to co-embed the reference and target data. The method assumes that the reference and target data share the same PCs, so it first computes the PCs of the reference data and then projects the target data to the same low-dimensional space using the PCs. MMD-MA co-embeds the reference and target data in the same kernel space and then aligns their manifolds to put them together.

2.3.2 | Model selection and comparison metrics

We compared the label transfer performance between our OT method[20] and Ingest,[23] Monocle,[24] and MMD-MA,[21] in terms of the AD subtyping accuracy in ADNI and ANMerge. The AD subtyping accuracy was defined using the diagnosis provided in ADNI and ANMerge. If a patient is classified as "Asym AD" and the patient's diagnosis is either "Control" or "MCI," we considered the patient to be accurately classified. If a patient is classified as "Typical AD" and the diagnosis of the patient is either "MCI" or "Dementia," we considered the patient to be accurately classified. If a patient is classified as "Low-NFT AD" and the diagnosis of the patient is either "MCI" or "Dementia," we considered the patient to be accurately classified. Lastly, if a patient is classified as "Control" and the diagnosis of the patient is also "Control," we considered the patient to be accurately classified. Further subtype-matched clinical measurements (cognitive and pTau levels) were performed after OT-based label transfer was performed to further evaluate the label transfer accuracy.

2.3.3 | Mini-Mental State Examination distribution

Once the AD subtypes were identified for ADNI and ANMerge using the OT method, we proceed to validate the preservation of cognitive decline across cohorts. We employed the Mini-Mental State Examination (MMSE) measure, a widely utilized assessment comprising 11 questions, to evaluate cognitive impairment encompassing thinking, communication, understanding, and memory. With scores ranging up to 30, an MMSE score below 23 typically indicates moderate to severe cognitive decline. Utilizing ROSMAP as a reference, we compared patient cognitive statuses and confirm the accuracy of label transfer for AD subtyping.

2.4 | Transcriptomic profiling and differential expression gene analysis

For ROSMAP blood transcriptomics, the publicly available RNA-seq data from monocytes were downloaded from the AMP-AD website (ID: syn22024496) and used for label transfer and transcriptomic analysis. The sample preparation and sequencing information are described in detail on the above website and in.[25] Three batches of RNA sequencing data were generated using Illumina HiSeq 2500 (reads length 2 × 100 bp and 2 × 76 bp) and NovaSeq (2 × 50 bp) with a target depth of about 50 million reads per sample,[25] which is equivalent to about 75X to 115X in sequencing depth regarding the human GRCh38 exome. One hundred and fifty-four RNAs-seq monocyte samples were available for our analysis, which is a part of the mix of the above three batches. The Cutadapt tool was used to trim the adapter contamination in the FASTQ sequences. These high sequencing quality sequences were then aligned to the GRCh38 human reference genome using STAR aligner. The software featureCounts from the Subread R package was used to generate a counts table of uniquely mapped reads to the genes.

For the ADNI cohort, peripheral blood samples were collected. RNA profiling was performed using Affymetrix Human Genome U219 Array with 530,467 probes targeting 49,293 transcripts. The detailed information is described in.[26] For the ANMerge cohort, transcriptional profiling was conducted on PBMCs using Illumina HumanHT-12 v3 and v4 BeadChip Kits. The data underwent thorough processing, including background correction, log transformation, normalization, and sophisticated statistical adjustments for batch effects. Outliers were identified and removed, and the dataset was filtered to retain only reliable probes. The details are provided in.[27]

We utilize the voom-limma package[28] to perform differential expression gene (DEG) analysis across patient subtypes, with the apolipoprotein E, ε4 allele (*APOE4*) genotype adjusted as a confounding factor (covariate), as illustrated in Figure 1B. For ADNI and ANMerge, the *APOE* genotype was coded as the number of ε4 alleles in each patient, that is, 0, 1, or 2. For ROSMAP RNA monocyte samples, since only one patient contained two ε4 alleles (genotype ε4/ε4), we grouped this patient with patients with one copy of the ε4 allele and coded both types of *APOE4* status as 1, then 0 for the other patients with no *APOE* ε4 allele. We also imputed the 5% of patients with missing genotype information as 0 for simplicity. A gene was considered differentially expressed (DE) if its *p*-value was less than or equal to 0.1 in ROSMAP and 0.05 in ADNI and ANMerge. Qiagen Ingenuity Pathway Analysis (IPA) was used for pathway enrichment analysis, as illustrated in Figure 1C.

2.5 | Trajectory inference and pseudotime calculation

We applied the partition-based graph abstraction (PAGA) algorithm[29] to infer a graph-like trajectory and calculate pseudotime for samples from the ROSMAP, ADNI, and ANMerge cohorts. Initially, the PAGA algorithm constructs a partition-based graph abstraction to represent the manifold of the input data. Pseudotime is then determined by establishing a root point on the PAGA graph and computing the geodesic distance between each sample and the root. In the case of ROSMAP data, trajectory inference incorporates clinical data, including BBS, COGDX, and CERAD. For ADNI and ANMerge data, trajectories are inferred using clinical data, including inferred BBS and CDR. To calculate pseudotime, a control patient is designated as the root point, with branches on the trajectory identified by finding the minimum spanning tree on the PAGA graph, as illustrated in Figure 1D.

2.6 | Statistical analysis

To identify pseudotime-dependent genes, which may indicate disease progress, we computed the correlation between gene expression and pseudotime. For each gene, a Pearson correlation coefficient (PCC) and Spearman correlation coefficient (SCC) were calculated along with corresponding *p*-values. A gene was considered pseudotime-dependent if the PCC *p*-value was less than 0.05. Note that when computing the PCC, we standardized the RNA-seq and microarray data for all three cohorts with zero mean and unit variance. To compare sex-specific differences at the pathway level, Student's *t*-test was applied to the normalized mean expression data. The measures for pTau, cerebrospinal fluid (CSF) tau, and AV45 between Asym AD and Typical AD were compared through the Wilcoxon rank sum test. Fisher's exact test was used to test the association between *APOE* $\varepsilon 2$ and $\varepsilon 3$ alleles and $\varepsilon 4$ alleles among Asym AD and Typical AD.

2.7 | Survival analysis

To assess the clinical relevance of the identified DE genes as potential biomarkers, we selected patients from the Asym AD to Typical AD branch based on a specific gene's quantile expression in that group, as illustrated in Figure 1E. Specifically, we focused on the lower quartile of the examined gene in the Asym AD group and the upper quartile of that in the Typical AD group, aiming to determine whether the expression levels of this gene are associated with disease progression from Asym AD to Typical AD over time. Additionally, as a comparison, we evaluated the same association with disease progress by dividing the lower and upper quartiles of gene expression levels across all patients. This was to determine whether the examined gene's expression level changes are exclusively correlated to the disease progression for the trajectory from Asym AD to Typical AD. The significance of the progression difference was analyzed with the log-rank test. We used the 13-item Alzheimer's Disease Assessment Scale–Cognitive Subscale (ADAS13) and ventricles to measure the disease progression and define the occurrence of an event if there is a greater than 10% increase in the metrics over the baseline during follow-up visits. ADAS13 cognitive measure and ventricular volume are both commonly used brain cognitive/atrophy metrics in AD clinical trials and are known to be associated with disease progression.[30–32]

3 | RESULTS

3.1 | OT distinguishes AD subtypes based on uncoupled tauopathy and cognitive decline

We screened a list of methods, in an attempt to find the one that best preserves intrinsic data structure across diverse cohorts with different data modalities (RNA-seq vs microarray), which included Ingest,[23] Monocle,[24] MMD-MA,[21] and OT.[20] We utilized transcriptomic data from the ROSMAP cohort and employed all the above methods to quantify the dissimilarity between tauopathy severity in terms of BBS across different subgroups in ADNI and ANMerge. When coupled with clinical diagnosis, our study demonstrated that OT surpasses alternative methods in discerning those tauopathy-cognition uncoupled AD subtypes, achieving an impressive average accuracy of 86%. Specific accuracies for each subtype are detailed in Supplementary Table S2.

Our findings reveal that following integration, transcriptomic data from ROSMAP monocytes and ADNI PBMC samples are relatively homogeneously embedded and evenly distributed in a unified space, even though they were generated from different platforms (RNA-seq vs microarrays), as shown in Figure 2A-B and Figure S1. To further check the label transfer accuracy, the cognitive test result MMSE scores were compared among all inferred subtypes. As demonstrated in Figure 2C-E, the distribution of MMSE scores across the four subtypes remains well-preserved in the inferred patients from the ADNI and ANMerge datasets. The cognitive differences between Asym AD and Typical AD are very significant, as shown by the extremely small adjusted *p*-values (3.84e-18, 1.22e-19, and 2.25e-37, for ROSMAP, ADNI, and ANMerge, respectively). This result suggests that the cognitive statuses are aligned with transcriptomics across datasets and platforms, and the label transfer is accurate despite the slightly different cell types and data types used for co-embedding. Further validations of the label transfer are discussed, in the Results Section 3.3 with the well-clustered subgroup samples on the pseudotime trajectories, the fact of no significant difference for pTau and CSF Tau between the Asym AD and Typical AD groups, as well as with the difference of potential clinical biomarkers in Section 3.4.

3.2 | Identify the neuroprotective pathways for AD resilience from the inferred Asym AD subtype

To reveal the molecular distinctions among AD subtypes, particularly to identify the neuroprotective factors that contribute to the Asym subtype's AD resilience, we focused on the comparison between the Asym AD and Typical AD subgroups across three cohorts. The comparison of Low NFT to Typical AD was not conducted due to the relatively small sample size and, therefore, lack of detection power. We conducted a differential gene expression analysis between the Asym AD and Typical AD subgroups. Then, the significant DE genes were input into the pathway enrichment analysis tool IPA to understand the common underlying pathways across all three cohorts. As shown in Figure 3A, our findings reveal several pathways involved in neuroprotection, including the repression of several immune pathways, such as neutrophil degranulation-like immune process (activation z-score −6.6), acute response signaling (activation z-score −2.5), IL-6 signaling (activation z-score −1.3), and the cachexia signaling pathway (activation z-score −3.7). There are also activated pathways that render neuron protection, which include the processing of capped intron-containing pre-mRNA (activation z-score 3.0) and enhanced mitochondrial function (activation z-score 2.0). For details of the enriched pathways, please see Table S3. The complete list of DE genes can be found in Tables S3-S9.

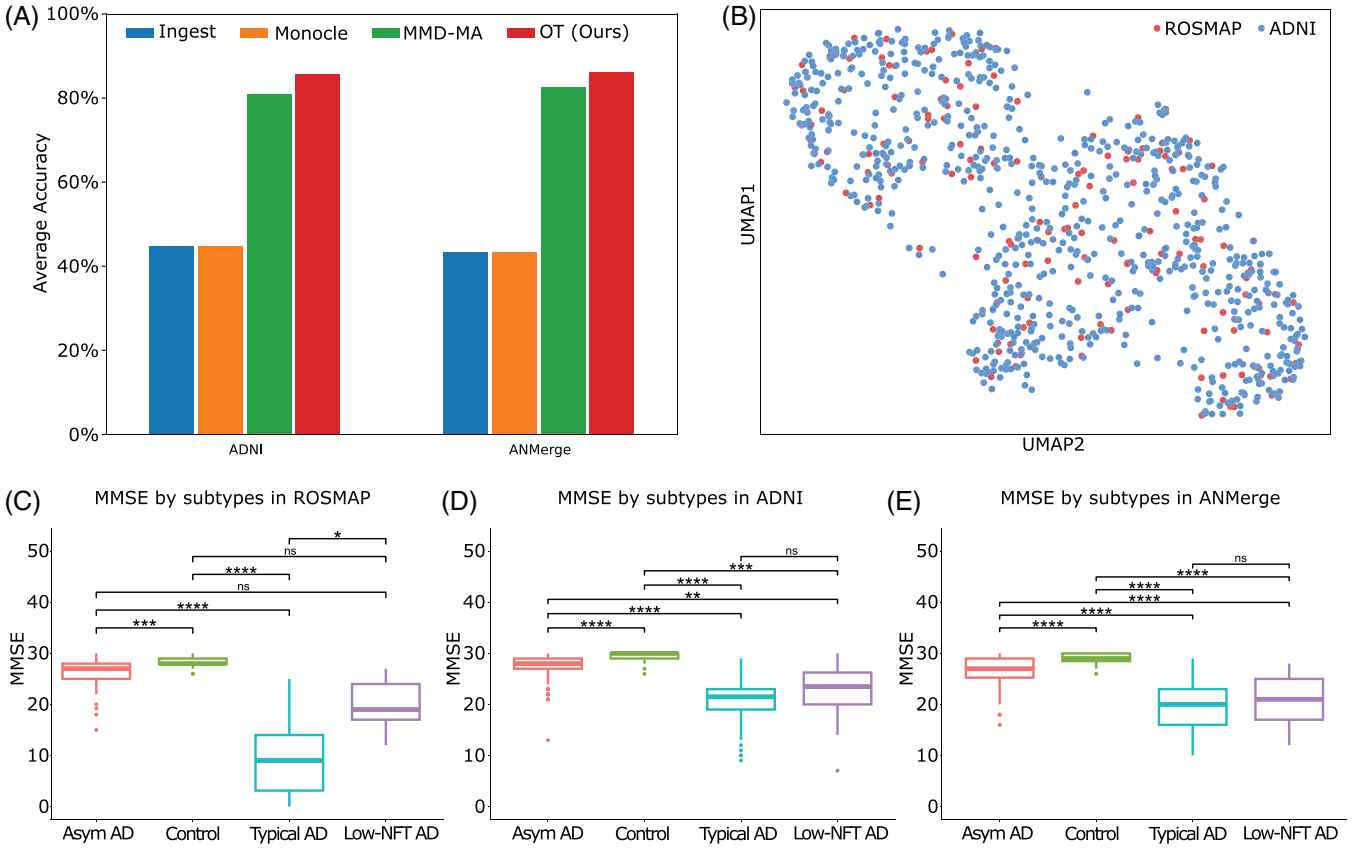

FIGURE 2 Optimal transport accurately infers the disease subtypes from ROSMAP to ANDI and ANMerge. (A) Bar plots of AD subtypes accuracy among four approaches. (B) Embedding visualization of four patient groups in the same space. (C-E) MMSE distribution across ROSMAP, ADNI, and ANMerge. AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; ANMerge, AddNeuroMed consortium; Asym, asymptomatic; MMD-MA, maximum mean discrepancy manifold alignment; MMSE, Mini-Mental State Examination; NFT, neural fibrillary tangle; ns, nonsignificant; OT, optimal transport; ROSMAP, Religious Orders Study and Rush Memory and Aging Project. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$.

As a representative among those genes from the enriched pathways, *CKAP4* (Figure 3B-C), demonstrates significantly lower expression levels in Asym AD compared to Typical AD across all our independent cohorts (*p*-value 0.05 in ROSMAP; 0.0002 in ADNI; 0.04 in ANMerge; Table S4). Notably, *CKAP4*'s role as a mediator in the neurochemical crosstalk with inflammatory signals such as IL-6 and IL-1 has been extensively studied in the context of AD and other diseases.[33–35] Furthermore, our study reveals sex-specific differences in *CKAP4* expression, with females exhibiting notably lower levels in Asym AD compared to males (*p*-value 0.0005 in ADNI).

The cachexia signaling pathway is known for its association with advanced dementia.[36] A previous study has shown that the dysregulation in energy balance tantalizes researchers with potential insights into its enigmatic workings.[37] One key player in this complex interplay is the CCAAT/enhancer binding protein delta (CEBPD), a transcription factor found prominently in activated astrocytes enveloping Aβ plaques. Elevated CEBPD expression stands as a hallmark feature observed predominantly within the astrocytes of AD patients and AppTg mice.[38,39] In our study, we found a significant elevation in *CEBPD* expression among Typical AD patients (*p*-value 0.008 in ADNI; 0.002 in ANMerge; Table S5), consistent with prior research findings[38,39] (Figure 3D). Furthermore, a sex-specific divergence for this gene is observed, with males exhibiting higher expression levels compared to females in both Asym AD and Typical AD groups (*p*-value 0.05 in Asym AD; 0.02 in Typical AD, Figure 3E).

Research has shown that amyloid genesis stems from an IL-1-/IL-6-mediated acute phase reaction within the brain.[40] Additionally, studies have detected activation of the IL-6 pathway within the hypothalamus and hippocampus of AD mice,[41,42] which are consistent with our observations in living human cohorts in Figure 3A, showing the repressed IL-6 signaling is associated with slower cognitive decline in Asym AD patients despite high BBS. Specifically, our analysis reveals a notable elevation in the expression of a novel component within this pathway, *RBP7*, elevated in Typical AD patients (*p*-value 0.005 in ADNI; *p*-value 0.002 in ANMerge; Table S6), as depicted in Figure 3F. Again, there is a sex difference in *RBP7* expression, with males exhibiting higher levels than females in both Asym AD and Typical AD groups (*p*-value 0.02 in Asym AD; *p*-value 0.02 in Typical AD), as shown in Figure 3G.

Mitochondrial dysfunction has been known as an early event in AD pathogenesis, manifesting prior to clinical symptoms, and has

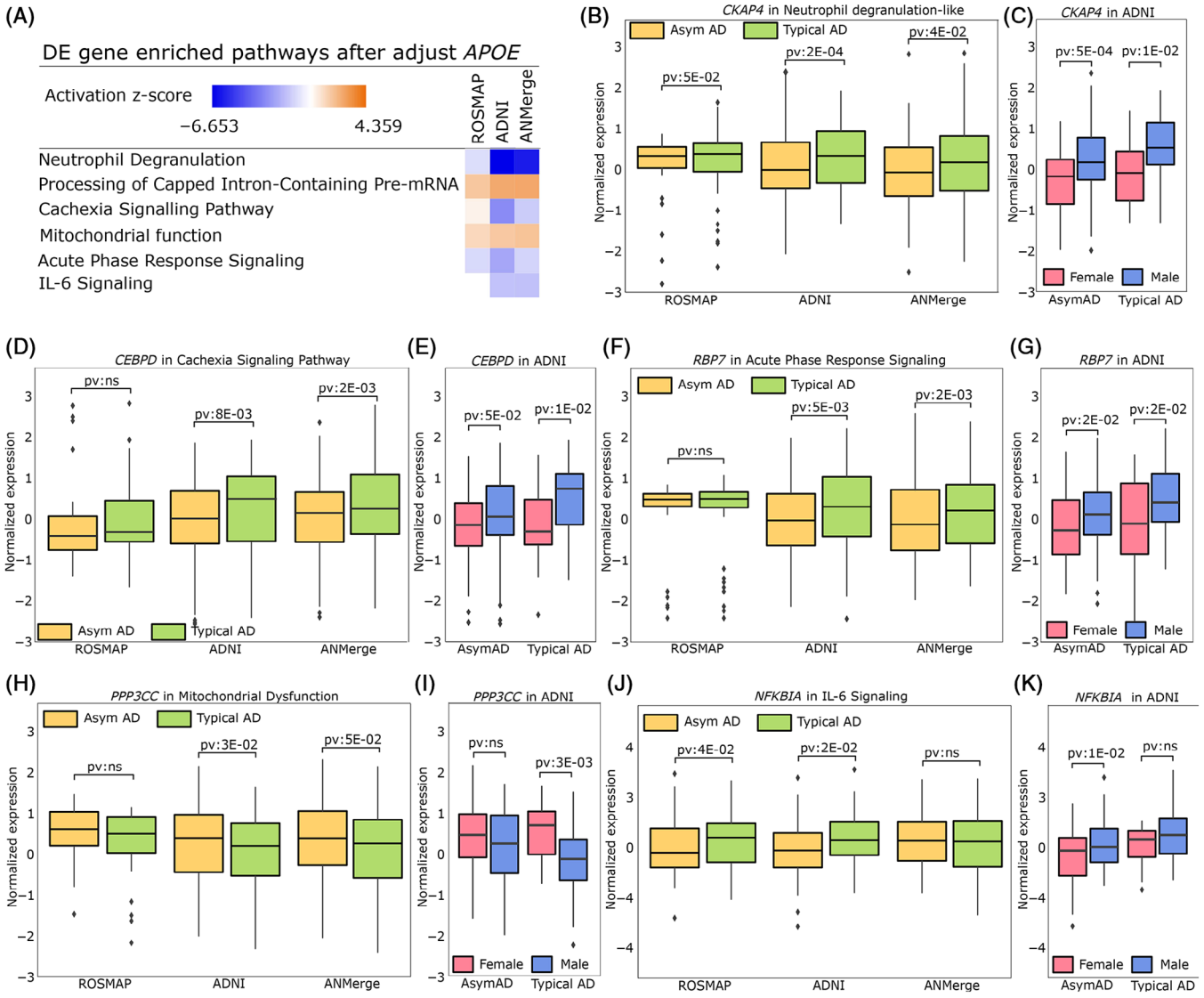

FIGURE 3 Pathway enrichment comparison (Asym AD vs Typical) across ROSMAP, ADNI, and ANMerge. (A) IPA DE comparison analysis at the pathway level. Representative DE gene and sex-specific difference in the neutrophil degranulation-like immune process (B,C), cachexia signaling (D,E), acute phase response (F,G), mitochondria function (H,I), and IL-6 signaling (J,K). AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; ANMerge, AddNeuroMed consortium; Asym, asymptomatic; *APOE4*, apolipoprotein E, ε4 allele; DE, differentially expressed; IPA, Ingenuity Pathway Analysis; pv, *p*-value; ROSMAP, Religious Orders Study and Rush Memory and Aging Project.

been widely accepted in the AD field.[43–46] One study on postmortem analysis of patient brains also revealed the reduced expression and activity of mitochondrial respiratory chain complexes.[45] We also found repressed mitochondria function in Typical AD patients, as shown by the representative gene *PPP3CC* (Figure 3H), which encodes calcineurin, a calcium-dependent, calmodulin-stimulated protein phosphatase. Besides its role in neuronal signaling transfer, it also participates in energy metabolism and is found in the mitochondria.[47] As shown in Figure 3H, we observed *PPP3CC* elevated in Asym AD (*p*-value 0.03 in ADNI; 0.05 in ANMerge; Table S7), especially in the female group (*p*-value 0.003 in Typical AD), as shown in Figure 3I.

One more AD-associated gene, *NFKBIA*, a nominated AD target gene,[48] is upregulated in Typical AD compared to Asym AD. Studies have reported elevated *NFKBIA* inhibits the expression of NF-κB signaling pathways, leading to cognitive decline and AD. Hence, inhibiting the expression of *NFKBIA* could be a target for the treatment of AD.[49,50] We found a significantly increased *NFKBIA* expression in the Typical AD group (*p*-value 0.04 in ANMerge; 0.02 in ADNI, Table S8), especially in males (*p*-value 0.01 in Asym AD), as shown in Figure 3J-K.

The consistency of these findings across three independent cohorts and their consistency with previous work demonstrated the success of subtype label transferring using OT and the subsequent understanding of the different disease progression on the molecular scale. Furthermore, our research sheds light on potential neuroprotective mechanisms within the complex landscape of AD pathology.

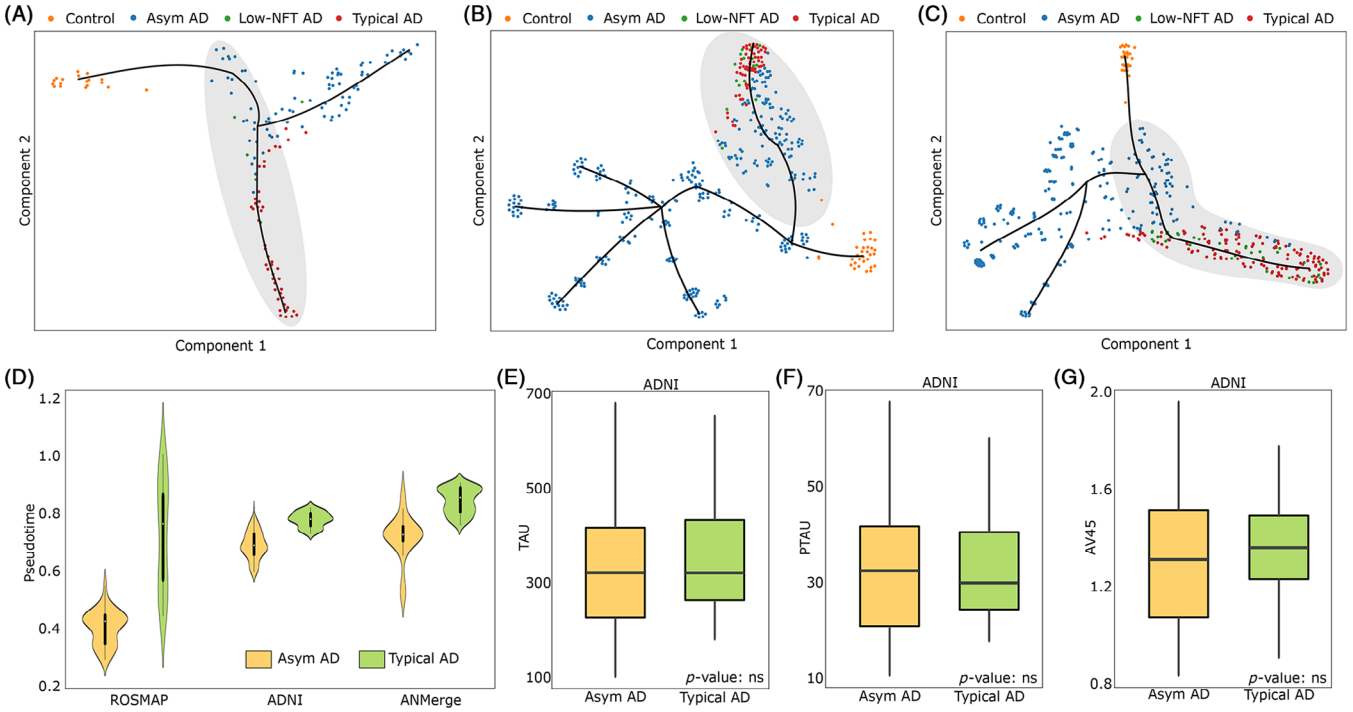

FIGURE 4 Trajectory analysis, pseudotime, and biomarkers from ROSMAP, ADNI, and ANMerge. (A-C) Disease progression trajectory in ROSMAP, ADNI, and ANMerge. (D) Pseudotime distribution for Asym AD and Typical AD in ROSMAP, ADNI, and ANMerge. Branch-specific CSF tau (E), CSF pTau (F), and PET marker AV45 (G) distributions in ROSMAP, ADNI, and ANMerge. AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; ANMerge, AddNeuroMed consortium; Asym, asymptomatic; NFT, neural fibrillary tangle; ns, nonsignificant; PET, positron emission tomography; ROSMAP, Religious Orders Study and Rush Memory and Aging Project.

3.3 | Trajectory analysis unveils the progression and branching of the disease state

As shown in Figure 4A-C, after the label transfer, the disease state transitions of all three datasets exhibit distinct clustering into four patient groups (ie, Asym AD, Typical AD, Low-NFT AD, and Control). We randomly designated a sample in the Control group as the initial time point and applied PAGA[29] to infer diffusion pseudotime for all other samples based on their distance in the manifold, ranging from 0 to 1. Consequently, each sample is assigned a pseudotime point, serving as a snapshot of the disease state for the corresponding AD subgroup, as detailed in Table S10. It is worth noting that due to the well-clustered nature of control samples, selecting a different initial sample within this group does not alter our pseudotime significantly, indicating those subgroup samples are more homogeneous within each cluster and distinct among clusters, and the generated trajectories are robust. To further test the robustness of these trajectories, we assess the resilience of the backbone map by randomly removing 10% of our patients. Remarkably, even with this perturbation, the shape of disease branching and trajectory remains consistently robust, as illustrated in Figure S2.

It becomes evident that starting from Control samples, diverse trajectories exist, leading to different subtypes and progression, ranging from remaining in a stable Control state to progressing towards Asym AD, Low-NFT AD, or Typical AD, as shown in Figure 4A-C. Importantly, variability exists even within the same AD subtype, mirroring the heterogeneous disease trajectories observed in clinical practice. For example, we observe a star-like structure of Asym AD patients (blue dots) with the multiple branches or states generated in ADNI (Figure 4B) and ANMerge (Figure 4C). Some individuals experience rapid progression, while others remain stable for extended periods with consistent clinical outcomes. However, it needs to be pointed out that stability at the clinical level does not necessarily imply the absence of molecular or pathological changes that are not reflected by the clinical measurements.

Specifically, the disease transitioning from early stage (Asym AD) to late stage (Typical AD), highlighted in the shaded branch in Figure 4A-C, is well reflected in the pseudotime shown in Figure 4D (Wilcoxon test, p-value: 6.2e-14 in ROSMAP; p-value 7.8e-24 in ADNI; p-value 2.8E-30 in ANMerge). Furthermore, to validate the label transfer accuracy, we conducted further analyses using CSF tau (Figure 4E) and pTau (Figure 4F) biomarkers, as well as the PET imaging biomarker AV45 (Figure 4G) in living patients. Interestingly, we found no significant differences between Asym AD and Typical AD on this branch (p-value $>$ 0.05 in ADNI). This reinforces our findings regarding the mechanisms underlying the independence of tauopathy and cognitive decline over time in living patients of AD subgroups.

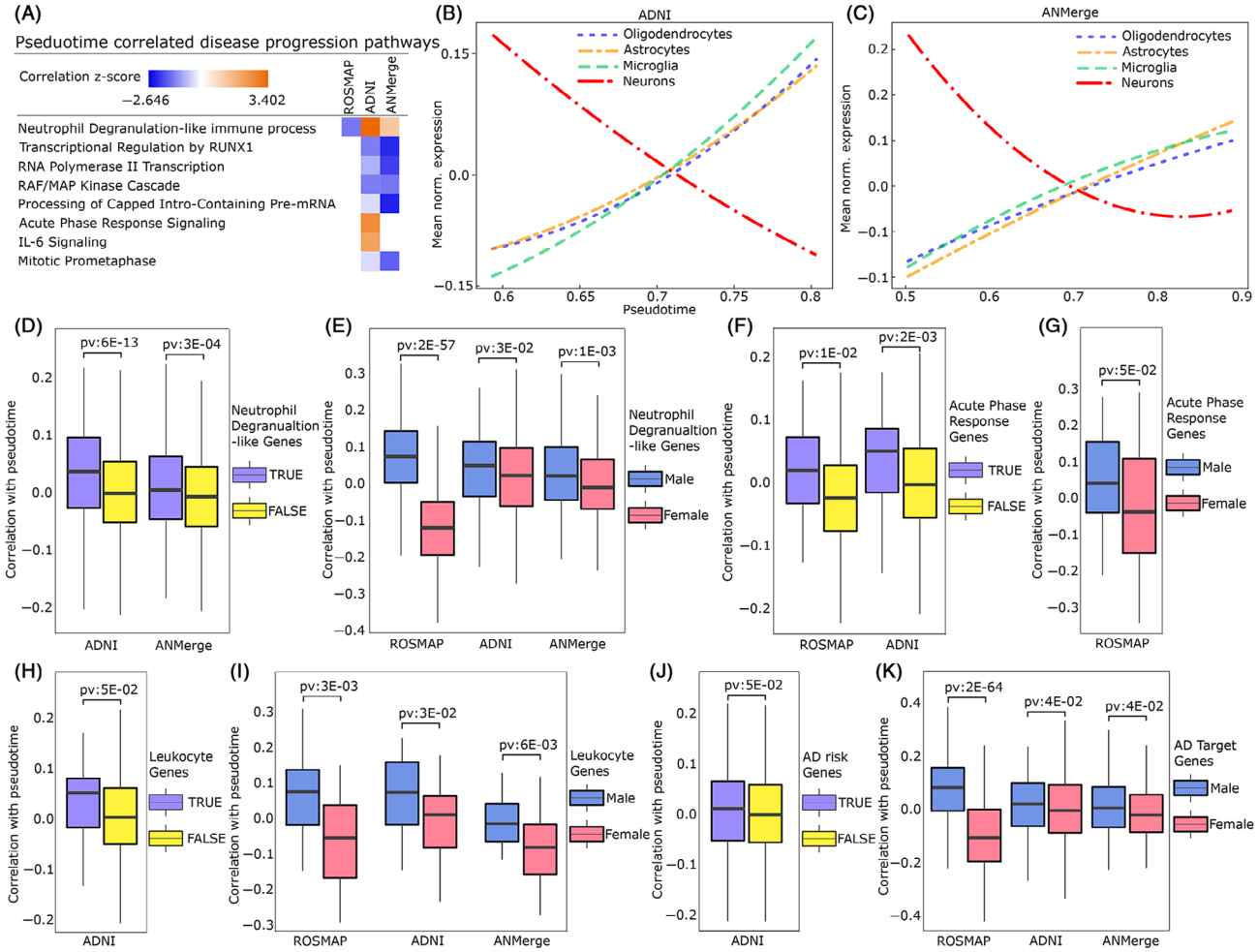

FIGURE 5 Branch-specific molecular characterization with pseudotime (from Asym AD to Typical AD). (A) An IPA comparison of enriched pathways was significantly correlated with pseudotime. (B,C) Cell type disease progression trajectory in ADNI and ANMerge. Neutrophil degranulation-like immune process (D,E), acute phase response (F,G), leukocyte (H,I), and AD risk (J,K) gene- and sex-specific-pseudotime correlations in ROSMAP, ADNI, and ANMerge. AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; ANMerge, AddNeuroMed consortium; Asym, asymptomatic; IPA, Ingenuity Pathway Analysis; pv, *p*-value; ROSMAP, Religious Orders Study and Rush Memory and Aging Project.

3.4 | Branch-specific pseudotime-dependent molecules and genetic factors explain the disease progression

To investigate the genes and pathways contributing to the changes that happen along the branch from the early stage (Asym AD) to the late stage (Typical AD) (shaded in Figure 4A-C), we conducted a branch-specific correlation analysis between the gene expression and pseudotime. This analysis yields PCCs and their respective significance with pseudotime, which is documented in Table S10-S11. Furthermore, we identify enriched pathways from the genes significantly correlated with pseudo time (p-value < 0.05) across all three cohorts, as shown in Figure 5A, Table 1, and Table S12. Furthermore, we dissected the significant temporal genes into functional groups, including pathways, AD-nominated target genes, and cell types, to delineate their temporal effects on disease progression across pseudotime. Notably, we identified genes associated with microglia, astrocytes, and oligodendrocytes that exhibit overexpression in the late stage of AD but remain lowly expressed in the early stage of the disease across three datasets on the Asym AD to Typical AD branch, as shown in Figure 5B-C and Figure S3. For details, see Table S13. In particular, we identified disease-associated astrocyte (DAA) genes overexpressed in late AD, such as *ABCA1*, *S100A6*, and *SLC22A4*, supported by recent studies using transgenic mice models of AD (5xFAD)[51] and brain injury lipopolysaccaride astrocyte-induced models, suggesting the involvement of those detrimental cell types in the AD.[48] Similarly, we identified disease-associated microglia (DAM) genes overexpressed in late AD, such as *RPS11*, *CTSB*, *SLC11A1*, *TYROBP*, *RPS2*, *GPX1*, SRGN, and *VSIG4*. These genes were also identified by single-nucleus transcriptome studies, where they were found to be unregulated at late stages of the disease.[12] Moreover, in a mouse study, genes including *NFKB2*, *IL1B*, *SOCS3*, *IL1RN*, *EPSTI1*, *CTSB*, *PLK3*, *NCF4*, *ITGAX*, *SLC11A1*, *CLTA*, *CD48*, *TLR2*, *CSF3R*, *B2M*, *TYROBP*, *PTPRC*, *RPL18A*, *ICAM1*, *PLAUR*, *P2RX4*, *RENBP*, *IFI30*, *PFKFB3*, *NPC2*, and *NFE2L2* were found to be

TABLE 1 Pseudotime-dependent representative genes among ROSMAP, ADNI, and ANMerge.

		Differential Expressed			Pseudotime Correlated			
Key genes	Direction[c]	ROSMAP	ADNI	ANMerge	ROSMAP	ADNI	ANMerge	Enriched pathway
NFKBIA[a]	Positive	X	X			X		IL-6 Signaling, Acute Phase Response
IL6ST[a]	Positive			X			X	
SHC1	Negative			X			X	
CRABP2, MAPK14, MAPK3, TNFRSF1A[a]	Positive		X			X		Acute Phase Response
NUP214	Positive		X	X			X	Processing of Capped Intron-containing Pre-mRNA
NUP205	Negative		X	X				
PRPF4	Negative		X				X	
CTNNBL1, HNRNPA2B1[a], *PNN, SF3A3, ZNF830*	Negative			X			X	
MAPK1[a]	Positive					X		RAF/MAP Kinase Cascade
FGR, NBEAL2, PYGL, QPCT	Positive		X	X		X		Neutrophil Degranulation
PGM1[b]	Positive		X		X	X		
CD300A[a], *CD93*[b]	Positive			X		X		
CYBB[a]	Positive			X			X	
CTSH, LAMP2, RAB24, RAB31	Positive		X	X			X	
CSTF3, EAF2, MNAT1, TAF4B	Negative					X		RNA Polymerase II Transcription
DHK38, POLR2C	Positive					X		
CCNK, IWS1, PAF1, SRS4	Negative						X	
SMARC4	Negative						X	Transcriptional Regulation by RUNX1
TUBB	Negative						X	Mitotic Prometaphase
DCTN1[a]	Negative				X			

Abbreviations: AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; ANMerge, AddNeuroMed consortium; *APOE4*, apolipoprotein E, ε4 allele; DE, differentially expressed; ROSMAP, Religious Orders Study and Rush Memory and Aging Project.

[a]AD target gene.

[b]Gene becoming marginal significant after adjusting for *APOE4* status in DE analyses.

[c]A direction of "Negative" means underexpressed as pseudotime progresses, while "Positive" means overexpressed as pseudotime progresses.

unregulated in late-response microglia.[52] They were also observed to be overexpressed in Typical AD patients in the Asym AD to Typical AD branch.

In addition, we compared gene groups that are enriched in identified pathways, shown in Figure 5A, to the rest of the genes and checked the sex difference in terms of pseudotime correlation. Notably, we have identified differences in the correlation between pseudotime and expression levels for the neutrophil degranulation-like immune process pathway (correlation z-score: 3.402 in ADNI), with males exhibiting significantly higher correlations than females (*p*-value: 2e-57), particularly evident in ROSMAP, as in Figure 5D-E. This discrepancy can also explain the observed opposite direction of the neutrophil degranulation-like immune process in ROSMAP, shown in Figure 5A. Furthermore, we conducted similar comparisons for acute phase response genes (Figure 5A, correlation z-score: 2.646 in ADNI), leukocyte extravasation signaling (Table S14), and AD target genes sourced from the Agora portal (Table S15). We observed that the genes enriched in those pathways all have a higher correlation with pseudotime than the rest of the genes on the genome (Figure 5F, *p*-value 0.01 in ADNI, *p*-value 0.002 in ANMerge; Figure 5H, *p*-value 0.05 in ADNI; Figure 5J, *p*-value 0.05 in ADNI). Additionally, we report sex-specific differences for acute phase response genes (Figure 5G, *p*-value 0.05 in ADNI), leukocyte extravasation signaling (Figure 5I, *p*-value 0.0003 in ROSMAP, *p*-value 0.03 in ADNI, *p*-value 0.006 in ANMerge), and AD target genes from the Agora portal (Figure 5K, *p*-value 2e-64 in ROSMAP, *p*-value 0.04 in ADNI, *p*-value 0.04 in ANMerge).

3.5 | Gene expression trajectory on a global scale provides insights into the molecular landscape of AD progression

Plotting the gene expression trajectory on a global scale offers a comprehensive view of the molecular changes occurring throughout AD

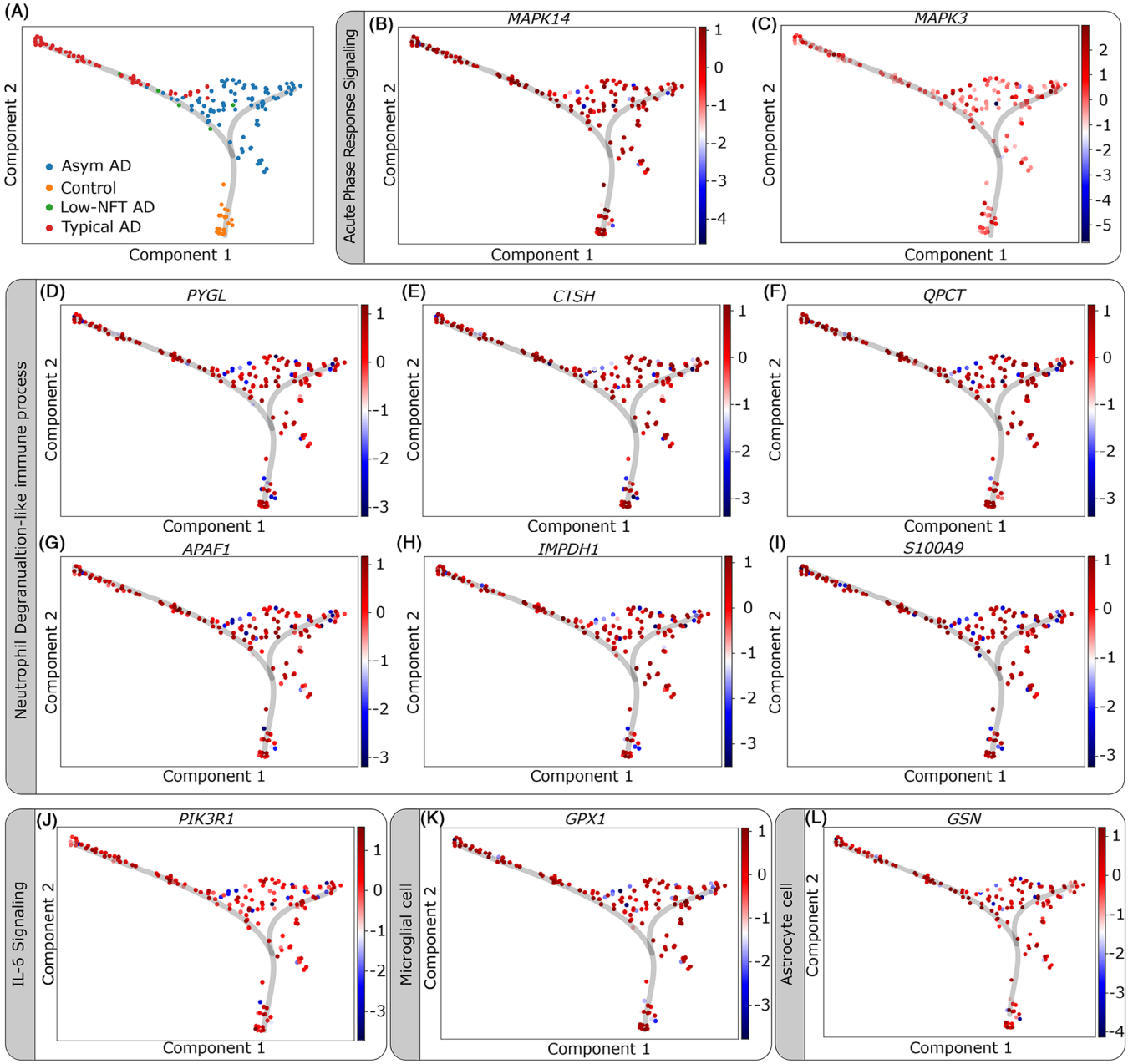

FIGURE 6 Selected gene expression correlation to predicted AD progression on a global scale. (A) ROSMAP trajectory backbone map. (B,C) Representative genes for acute phase response signaling. (D-I) Representative genes for neutrophil degranulation-like immune process. (J) Representative gene for IL-6 signaling. (K) Disease-associated microglial marker gene. (L) Disease-associated astrocyte marker gene. AD, Alzheimer's disease; ROSMAP, Religious Orders Study and Rush Memory and Aging Project.

progression. Shown in Figure 6A, the ROSMAP trajectory backbone map clusters the four subgroups within the manifold. By dissecting this trajectory, we may identify crucial time points or stages in AD progression marked by significant gene alterations. Here, we showed several example genes: *MAPK14* (Figure 6B) and *MAPK3* (Figure 6C) in acute phase response signaling; as well as *PYGL* (Figure 6D), *CTSH* (Figure 6E), *QPCT* (Figure 6F), *APAF1* (Figure 6G), *IMPDH1* (Figure 6H), and *S100A9* (Figure 6I) in neutrophil degranulation-like immune process; along with *PIK3R1* in IL-6 signaling (Figure 6J). Notably, the blue dots (representing patients with lower normalized expression) are enriched in Control and Asym AD, while red dots (indicating patients with higher normalized expression) are predominant in Typical AD patients. Similar expression patterns are observed for the identified DAM gene (Figure 6K) and DAA gene *GSN* (Figure 6L), which are consistent with recent reports.[12,51–54] These findings support the significant clinical and molecular heterogeneity inherent in AD, wherein distinct molecular pathways may be activated through various gene sets, leading to phenotypic variability among individuals; refer to Figure S4 for more

detailed time trend information, and Figure S5 shows the global trend at the pathway level.

3.6 | Evaluate potential gene biomarkers for AD prognosis with survival analysis

To validate our finding on the trajectory branch highlighted in Figure 4A-C and to look for potential gene biomarkers for early-stage AD prognosis, we used longitudinal cognitive measurement data from the ADNI dataset to investigate the relationship between identified pseudotime-dependent markers and their potential to predict the patients' later cognitive decline. To ensure accuracy, we selected patients from the Asym AD to Typical AD branch based on their gene expression levels, focusing on the lower quartile of Asym AD patients and upper quartile of Typical AD patients for the specific gene marker identified. The spaghetti plots with the event (10% increase) in red at the individual patient level are shown in Figure S6. The survival curves illustrating the mean change for the two selected groups are shown in Figure 7. To gain insights into the distinct disease progression mechanisms of the Asym AD subgroup, we compared counterpart spaghetti plots and survival curves for patients not specific to the defined shaded branch. This involves patients falling within the lower quartile and upper quartile expression levels for the selected gene marker, regardless of AD subtype labels, as shown in Figure S7. This approach allowed us to comprehensively assess the association of the identified gene marker and disease progression from the Asym AD state to Typical AD.

The *IL6ST* gene, a part of the cytokine receptor complex, exhibits significant expression level association with cognitive decline/brain atrophy in the Asym AD and Typical AD groups. We found ADAS13 and ventricle show a faster decline in Typical AD patients (in Figure S6A,C) expressing high *IL6ST* (in Table 1) versus Asym AD patients expressing low *IL6ST* (in Figure S6B,D), with log-rank *p*-values 4.5e-5 (in Figure 7A) and 0.018 (in Figure 7C), respectively, indicating significant differences in disease progression trajectories between the two subtypes. On the other side, when analyzing all patients irrespective of AD subtype, no significant variations (in Figure S7I,K) in event frequencies were observed when solely based on *IL6ST* expression levels (in Figure S7A-D). A similar trend was noted for the *CTSH* gene, which encodes a lysosomal cysteine proteinase (Table 1). For Typical AD patients (in Figure S6E,G) with high expression of *CTSH*, those patients exhibit a faster decline compared to Asym AD patients with low expression (in Figure S6F,H), with log-rank *p*-values 2.2e-3 (in Figure 7B,D). Again, when considering all patients, no significant differences (in Figure S7J,L) were detected based solely on *CTSH* expression levels (in Figure S7E-H). More genes from Table 1 with similar trends are shown in Figure S8. These findings indicate the clinical significance of the above-mentioned genes as prognostic biomarkers and highlight the importance of branch-specific trajectory analysis in comprehending the diverse states and underlying mechanisms of AD. The results from our survival analysis further validate the heterogeneity of AD and emphasize the need for tailored approaches to disease management.

4 | DISCUSSION

In this study, a novel approach is developed for predicting AD subgroups in living patients. Through OT learning, the intrinsic structure of the transcriptomic data is preserved, which allows us to map molecular characteristics specific to individual subgroups, facilitating the subtype predictions to living patients. Consequently, we can further understand the molecular mechanisms underlying AD heterogeneity. To achieve this, we conducted a DEG analysis between Asym AD and Typical AD to locate genes and pathways potentially associated with neuroprotection. We observed that several pathways are involved in the transition from the early to the late stage of AD progression, such as neutrophil degranulation-like immune process, acute phase response signaling, IL-6 signaling, processing of capped intron-containing pre-mRNA, cachexia signaling, mitochondrial function, and others. Consistent with previous findings,[55–57] the repression of the immune response-related pathways and the enhanced mitochondrial functions are associated with Asym AD. Several of the above immune pathway genes also significantly correlate with pseudotime as disease advances. This may suggest that through long-lasting activation, these immune pathways play a role as chronic drivers of disease progression. Targeting these pathways could be a potential new direction for AD therapy.

In the predicted subtypes of ADNI and ANMerge cohorts, tauopathy is uncoupled with cognitive decline, which is different from the general belief that tauopathy is highly correlated to cognitive decline in AD.[58–61] Moreover, using longitudinal clinical data, potential genetic biomarkers (*IL6ST* and *CTSH*) were found, with expressions linked to significantly different disease progression trajectories among subtypes (Asym AD vs Typical AD). Tracking these markers longitudinally may allow us to further assess their potential as biomarkers for disease progression, which in turn may help clinicians intervene early for certain individual patients.

Neutrophil degranulation-like immune process emerged as a top-ranked pathway in our IPA comparative analysis on DEG (Asym AD vs Typical AD) comparison across AD cohorts. Although neutrophils are usually absent in the PBMC preparation, neutrophil-like functions may appear in the pathway enrichment analysis in several scenarios. For example, the presence could be simply an increasing presence of low-density neutrophils (LDNs), a particular type of neutrophil found in PBMCs.[62,63] This could explain why neutrophil degranulation pathway enrichment is greater in ADNI and ANMerge than in ROSMAP monocyte data, as LDNs are more likely to present in PBMCs. Moreover, studies have shown enriched neutrophil degranulation-like functions in mononuclear cells[62] and crosstalk between immune cells.[64,65] Under specific conditions, such as inflammation or infection, monocytes and other PBMCs can express genes typically associated with neutrophil functions.[63,66,67] However, as IPA only outputs one annotation label for a set of genes' diverse functions, such immune response is annotated by IPA as neutrophil degranulation. In our gene ontology analysis, the 26 genes from Table S3A, enriched in neutrophil degranulation, are also enriched significantly in the innate immune system, as shown in

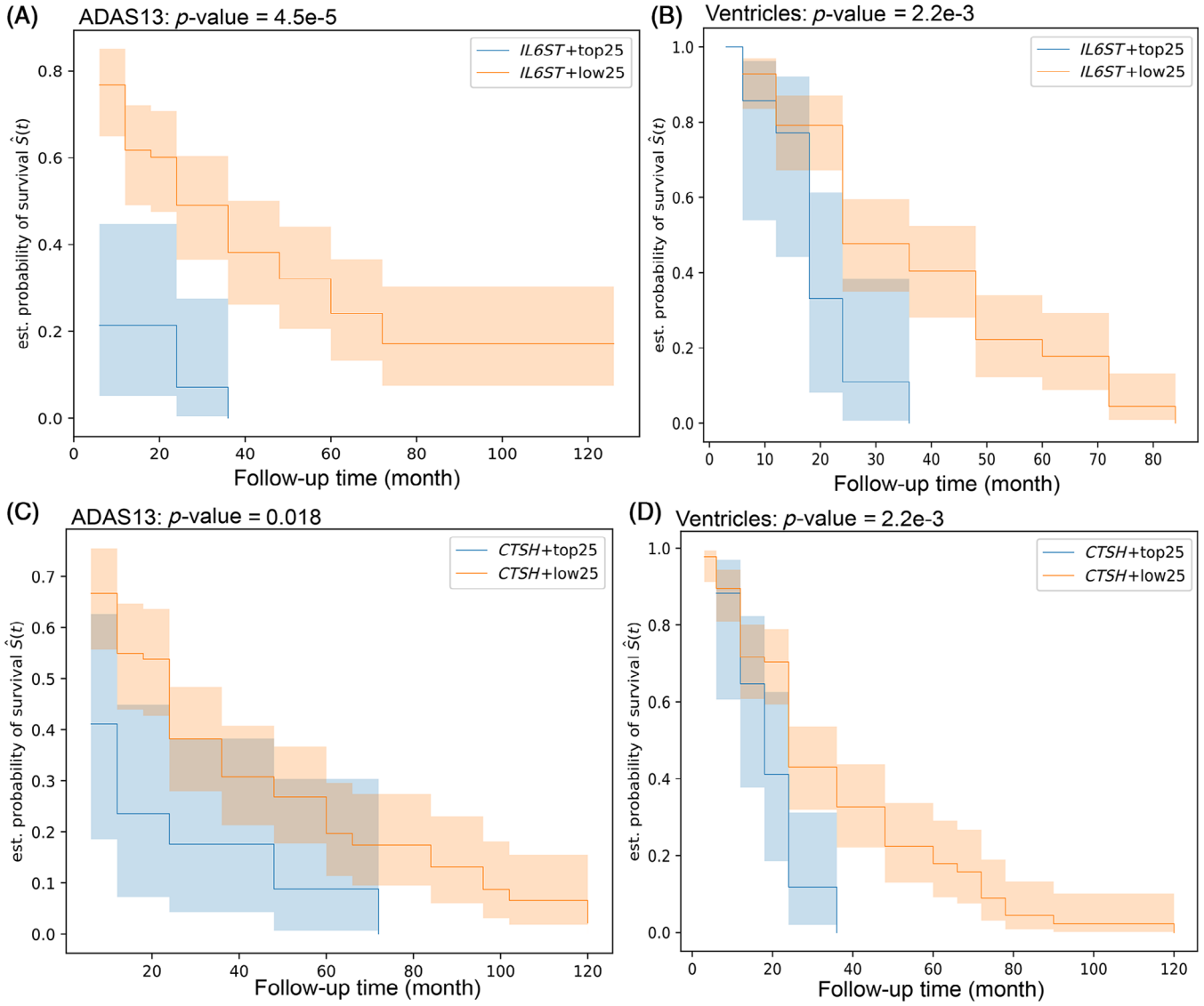

FIGURE 7 Survival analysis for potential AD prognostic biomarkers. The Kaplan-Meier plot of survival curves for the two groups of subjects (*IL6ST* low and high) in terms of 10% increase in ADAS13 (A) and ventricles (B). The Kaplan-Meier plot of survival curves for Asym AD and Typical AD patients (*CTSH* low and high) in terms of 10% increase in ADAS13 (C) and ventricles (D). AD, Alzheimer's disease; ADAS13, 13-item Alzheimer's Disease Assessment Scale–Cognitive Subscale; Asym, asymptomatic.

Figure S9A from the Toppgene pathway enrichment analysis.[68] Additionally, upon reaching the site of inflammation, monocytes and other PBMCs can take on roles similar to those of neutrophils, and engage in processes such as degranulation, typically associated with neutrophils, to further stimulate and sustain the innate immune response.[69,70] Based on all the above reasons, in order to make the annotation more precise and unbiased, we changed the annotation to "neutrophil degranulation-like immune process" and used it throughout this work.

Our sex-specific DE results on several genes (*CKAP4*, *CEBPD*, *RBP7*, *PPP3CC*, *NFKBIA*) suggest their roles in neuroprotection in females may be different from those in males over time, adding another layer of complexity to our understanding of AD pathogenesis. This also indicates the intricate interplay between sexual hormone-induced physiology and the disease molecular mechanisms in AD progression for those genes and more. Further investigation is needed to understand how these genes interplay with the AD risk in each subtype.

Additionally, we examined specific *APOE* genotype enrichment in each subtype for all cohorts, and found the enrichment is not consistent among the three. In ROSMAP, Asym AD patients exhibit enriched *APOE* ε2 and ε3 alleles (odds ratio 2.47, Fisher one-sided *p*-value 0.008), while Low-NFT AD subtype has a pronounced prevalence of *APOE* ε4 in the ANMerge cohort (odds ratio 6.08, Fisher *p*-value 1.95E-14). Except for these, no significant *APOE* genotype enrichment in a particular subtype was observed in other cohorts (see Table S16 for details).

Moreover, when *APOE4* status was adjusted for DE analysis, as shown in Figure 3A and Figure S9, in the ROSMAP data, a slight change in the total number of significant genes was observed with the *APOE4* adjusted (Figure S9C). Compared to the *APOE4* unadjusted results, the

enriched pathways based on DE genes do not change; only the activation z-score changed slightly, as shown in Figure 3A versus Figure S9B (*APOE4* not adjusted). It is important to note that our DE significance thresholds are intentionally broad to capture a wide range of potential changes for exploration purposes. As ROSMAP DE results show more change with *APOE4* status adjustment, we postulate that gene expression in single-cell types, such as monocytes, is more sensitive to genetic factors like *APOE4*. In contrast, in mixed-cell populations, such as PBMCs, gene expression may be moderated due to interactions among various cell types, potentially offsetting some of the effects seen in individual cell types.

Please note that the blood samples from ROSMAP do not contain exactly the same cell type as those from ADNI and ANMerge. The microarray data from ADNI and ANMerge were generated from PMBC samples, while ROSMAP data were from monocytes. However, since PBMCs contain monocytes, we think the transcriptome-based label transfer is still feasible between the three blood transcriptomic datasets, and the results have demonstrated that the label transfer is accurate by the facts that (1) a highly homogeneous co-embedding into the same low-dimensional space is achieved by OT (Figure 1,2, Figure S1; Table S2); (2) the inferred subtypes are well separated and clustered on the pseudotime trajectory plot (Figure 4A-C); (3) inferred samples' cognitive difference agrees with the subtype cognition levels during the MMSE score comparison (Figure 2C-E); and (4) blood pathological tau biomarkers such as pTau agree with the inferred subtype labels (Figure 4E-G). This cross-cohort approach enhances our ability to identify robust biomarkers and deepen our understanding of AD's systemic impacts.

Examining the DE results of Asym AD versus Typical AD regarding the number of overlapping DE genes among three cohorts, we observed limited gene overlaps across different sample types or platforms (monocytes vs PBMCs, RNA-seq vs microarray), as shown in Figure S9D. The observed consistency across the three cohorts is not at the gene level but at the pathway level, as shown in Figure 3A. Additionally, we also performed cell type enrichment analysis with the tool ToppGene[68] on the three DE lists separately to check how the different composition of cell types affects the DE results in different studies. We found that the DE genes from each cohort are not enriched with the cell type marker genes limited only to that specific cell type; genes involved in other white blood cell functions are also observed. For example, we observed monocyte co-expression genes in DE lists from PBMC samples (ADNI and ANMerge), as well as dendritic cell co-expression genes in monocyte samples (ROSMAP) (Table S17). Please note that we focused mostly on the pathway level, which is consistently enriched among the three cohorts, and only discussed the DE results from overlapping DEs in at least two cohorts.

Please note that our subtyping characterization has been primarily focused on understanding the resilience in the Asym AD subtype. Due to an insufficient sample size and statistical power, the analysis of Low-NFT AD has been limited.

Our study harnesses the vast datasets from three cohorts, ROSMAP, ADNI, and ANMerge, demonstrating the generalizability and robustness of our label-transferring approach. Leveraging multiple cohorts not only validates our subtyping results across diverse populations but also provides a deeper comprehension of AD heterogeneity. While our pathway results are generally consistent across these three datasets, it is important to acknowledge that potential biases or disparities may exist. Besides cell type and platform differences, variations in cohort demographics, clinical information, and data collection protocols may introduce other confounding variables. However, despite the vast variations mentioned above, low-dimensional co-embedding with OT can still effectively mitigate the noises and capture the transcriptomic similarity with an accurate label transfer.

In conclusion, this study introduces a novel workflow for predicting AD subtypes and molecular characteristics in living patients, leveraging OT and disease progression pseudotime inference. By applying these approaches across multiple studies, we discovered that consistently repressed immune response pathways and enhanced mitochondrial functions are associated with the Asym AD subtype among three studies, and two genes can be potential prognostic biomarkers. This graph-based OT label transferring on multi-AD transcriptomics enhanced our understanding of AD progression in different subtypes, offering insights into potential biomarkers and personalized interventions for improved patient care.

ACKNOWLEDGMENTS

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. We gratefully acknowledge the funding received for this work from the Indiana University Precision Health Initiative NIH TREAT-AD U54 grant 5U54AG065181, as well as the 5R01NS119280 grant. This support has been essential in facilitating our research endeavors.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest. This research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest. Additionally, no external funding was received for this study. Author disclosures are available in the supporting information.

CONSENT STATEMENT

This research did not involve direct interactions with human subjects or the collection of personal data. The study exclusively utilized publicly available datasets, which were obtained in accordance with all applicable laws and regulations concerning the ethical use of secondary data. Therefore, informed consent was not required for this study.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database atadni.loni.usc.edu, the ROSMAP data (syn23446022), and ANMerge data (syn4988768). The code is available upon request.

ORCID

Xiaoqing Huang https://orcid.org/0009-0003-2183-8398

REFERENCES

1. Park JC, Han SH, Yi D, et al. Multi-omic analyses of Alzheimer's disease uncover early disease-specific molecular signatures. *Alzheimer's & Dementia*. 2022;18(12):345-345.
2. Swarup V, Hinz FI, Rexach JE, et al. Identification of evolutionarily conserved gene networks mediating neurodegenerative dementia. *Nat Med*. 2019;25(1):152-164.
3. 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;5:180185.
4. Zhang B, Gaiteri C, Bodea L-G, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*. 2013;153(3):707-720.
5. Mattsson N, Zetterberg H, Janelidze S, et al. Plasma tau in Alzheimer disease. *Neurology*. 2016;87(17):1827-1835.
6. Ossenkoppele R, Rabinovici GD, Smith R, et al. Discriminative Accuracy of [18F]flortaucipir positron emission tomography for Alzheimer disease vs other neurodegenerative disorders. *JAMA*. 2018;320(11):1151-1162.
7. Palmqvist S, Janelidze S, Stomrud E, et al. Performance of fully automated plasma assays as screening tests for Alzheimer disease-related β592 amyloid status. *JAMA Neurol*. 2019;76(6):1060-1069.
8. Ritchie C, Smailagic N, Noel-Storr AH, Ukoumunne O, Ladds EC, Martin S. CSF tau and the CSF tau/ABeta ratio for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev*. 2017;3(3):CD010803.
9. Jr J, Bennett CR, Blennow DA, et al. Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*. 2018;14:535-562.
10. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):280-292.
11. Klein HU, Mccabe C, Gjoneska E, et al. Epigenome-wide study uncovers large-scale changes in histone acetylation driven by tau pathology in aging and Alzheimer's human brains. *Nat Neurosci*. 2019;22(1):37-46.
12. Mathys H, Davila-Velderrain J, Peng Z, et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*. 2019;570(7761):332-337.
13. 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.
14. Paterson RW, Heywood WE, Heslegrave AJ, et al. A targeted proteomic multiplex CSF assay identifies increased malate dehydrogenase and other neurodegenerative biomarkers in individuals with Alzheimer's disease pathology. *Transl Psychiatry*. 2016;6(11):e952.
15. Haghverdi L, Büttner M, Wolf FA, et al. Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods*. 2016;13(10):845-848.
16. Mukherjee S, Heath L, Preuss C, et al. Molecular estimation of neurodegeneration pseudotime in older brains. *Nat Commun*. 2020;11(1):5781.
17. Tong TaZB. Optimal transport analysis of transcriptomic profiles reveals Alzheimer's disease subtypes with distinct molecular characteristics. *Front Genet*. 2023;14(768):953.
18. Young AL, Marinescu RV, Oxtoby NP, et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. *Nat Commun*. 2018;9(1):4273.
19. Young AL, Oxtoby NP, Huang J, et al. Multiple orderings of events in disease progression. *Inf Process Med Imaging*. 2015;24:711-722.
20. Flamary R, Courty N, Gramfort A, et al. POT: python optimal transport. *J Mach Learn Res*. 2021;22(78):1-8.
21. Liu J, Huang Y, Singh R, Vert JP, Noble WS. Jointly embedding multiple single-cell omics measurements. *Algorithms Bioinform*. 2019;143:10.
22. Qiu X, Mao Qi, Tang Y, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*. 2017;14(10):979-982.
23. Weinreb C, Rodriguez-Fraticelli A, Camargo FD, Klein AM. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science*. 2020;367(6479):eaaw3381.
24. Cao J, Spielmann M, Qiu X, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*. 2019;566(7745):496-502.
25. Panitch R, Hu J, Xia W, et al. Blood and brain transcriptome analysis reveals APOE genotype-mediated and immune-related pathways involved in Alzheimer disease. *Alzheimers Res Ther*. 2022;14(1):30.
26. Nho K, Kim S, Soares H, et al. Microarray gene expression of ADNI participants. *ADNI Microarry Gene Expression Methods*. 2015.
27. Birkenbihl C, Westwood S, Shi L, et al. ANMerge: a comprehensive and accessible Alzheimer's disease patient-level dataset. *J Alzheimers Dis*. 2021;79(1):423-431.
28. Ritchie ME, Phipson B, Wu Di, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
29. Wolf FA, Hamey FK, Plass M, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol*. 2019;20(1):59.
30. Ihl R, Ferris S, Robert P, Winblad B, Gauthier S, Tennigkeit F. Detecting treatment effects with combinations of the ADAS-cog items in patients with mild and moderate Alzheimer's disease. *Int J Geriatr Psychiatry*. 2012;27(1):15-21.
31. Cho SH, Woo S, Kim C, et al. Disease progression modelling from preclinical Alzheimer's disease (AD) to AD dementia. *Sci Rep*. 2021;11(1):4168.
32. Ferrarini L, Palm WM, Olofsen H, Van Buchem MA, Reiber JHC, Admiraal-Behloul F. Shape differences of the brain ventricles in Alzheimer's disease. *Neuroimage*. 2006;32(3):1060-1069.
33. Beydoun MA, Beydoun HA, Noren Hooten N, et al. Hospital-treated prevalent infections, the plasma proteome and incident dementia among UK older adults. *iScience*. 2023;26(12):108526.

34. Rahman MA, Islam K, Rahman S, Alamin M. Neurobiochemical cross-talk between COVID-19 and Alzheimer's disease. *Mol Neurobiol*. 2020;58(3):7571527.
35. Shajahan SR, Kumar S, Ramli MDC. Unravelling the connection between COVID-19 and Alzheimer's disease: a comprehensive review. *Front Aging Neurosci*. 2024;15:1274452.
36. 2022 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2022;18(4):700-789. doi:10.1002/alz.12638
37. Minaglia C, Giannotti C, Boccardi V, et al. Cachexia and advanced dementia. *J Cachexia Sarcopenia Muscle*. 2019;10(2):263-277.
38. Ko CY, Wang W-L, Wang S-M, Chu Yu-Yi, Chang W-C, Wang Ju-M. Glycogen synthase kinase-3beta-mediated CCAAT/enhancer-binding protein delta phosphorylation in astrocytes promotes migration and activation of microglia/macrophages. *Neurobiol Aging*. 2014;35(1):24-34.
39. Li R, Strohmeyer R, Liang Z, Lue L-F, Rogers J. CCAAT/enhancer binding protein delta (C/EBPdelta) expression and elevation in Alzheimer's disease. *Neurobiol Aging*. 2004;25(8):991-999.
40. Ayton S, Janelidze S, Roberts B, et al. Acute phase markers in CSF reveal inflammatory changes in Alzheimer's disease that intersect with pathology, APOE epsilon4, sex and age. *Prog Neurobiol*. 2021;198:101904.
41. Barreto Chang OL, Maze M. Defining the role of Interleukin-6 for the development of perioperative neurocognitive disorders: evidence from clinical and preclinical studies. *Front Aging Neurosci*. 2022;14:1097606.
42. Lyra ESNM, Gonçalves RA, Pascoal TA, et al. Pro-inflammatory interleukin-6 signaling links cognitive impairments and peripheral metabolic alterations in Alzheimer's disease. *Transl Psychiatry*. 2021;11(1):251.
43. Varma VR, Varma S, An Y, et al. Alpha-2 macroglobulin in Alzheimer's disease: a marker of neuronal injury through the RCAN1 pathway. *Mol Psychiatry*. 2017;22(1):13-23.
44. Bhatia S, Rawal R, Sharma P, Singh T, Singh M, Singh V. Mitochondrial dysfunction in Alzheimer's disease: opportunities for drug development. *Curr Neuropharmacol*. 2022;20(4):675-692.
45. Pessoa J, Duarte AI. Overcoming mitochondrial dysfunction in neurodegenerative diseases. *Neural Regen Res*. 2023;18(7):1486-1488.
46. Atlante A, Amadoro G, Latina V, Valenti D. Therapeutic potential of targeting mitochondria for Alzheimer's disease treatment. *J Clin Med*. 2022;11(22):6742.
47. Pfluger PT, Kabra DG, Aichler M, et al. Calcineurin links mitochondrial elongation with energy metabolism. *Cell Metab*. 2015;22(5):838-850.
48. *Agora*. Available from: https://agora.adknowledgeportal.org/genes/nominated-targets
49. Ryu S, Han J, Norden-Krichmar TM, et al. Genetic signature of human longevity in PKC and NF-kappaB signaling. *Aging Cell*. 2021;20(7):e13362.
50. Zhang Q, Li J, Weng L. Identification and validation of aging-related genes in Alzheimer's disease. *Front Neurosci*. 2022;16:905722.
51. Habib N, Mccabe C, Medina S, et al. Disease-associated astrocytes in Alzheimer's disease and aging. *Nat Neurosci*. 2020;23(6):701-706.
52. Mathys H, Adaikkan C, Gao F, et al. Temporal tracking of microglia activation in neurodegeneration at single-cell resolution. *Cell Rep*. 2017;21(2):366-380.
53. McKenzie AT, Wang M, Hauberg ME, et al. Brain cell type specific gene expression and co-expression network architectures. *Sci Rep*. 2018;8(1):8868.
54. Zamanian JL, Xu L, Foo LC, et al. Genomic analysis of reactive astrogliosis. *J Neurosci*. 2012;32(18):6391-6410.
55. Balsa E, Perry EA, Bennett CF, et al. Defective NADPH production in mitochondrial disease complex I causes inflammation and cell death. *Nat Commun*. 2020;11(1):2714.
56. Yang Y, Ouyang Y, Yang L, et al. Pink1 regulates mitochondrial dynamics through interaction with the fission/fusion machinery. *Proc Natl Acad Sci*. 2008;105(19):7070-7075.
57. Tracy TE, Madero-Pérez J, Swaney DL, et al. Tau interactome maps synaptic and mitochondrial processes associated with neurodegeneration. *Cell*. 2022;185(4):712-728. e14.
58. Oka M, Fujisaki N, Maruko-Otake A, et al. Ca2+/calmodulin-dependent protein kinase II promotes neurodegeneration caused by tau phosphorylated at Ser262/356 in a transgenic Drosophila model of tauopathy. *J Biochem*. 2017;162(5):335-342.
59. Rossi D, Gruart A, Contreras-Murillo G, et al. Reelin reverts biochemical, physiological and cognitive alterations in mouse models of Tauopathy. *Prog Neurobiol*. 2020;186:101743.
60. Huang X, Jury-Garfe N, Liu J, Lasagna Reeves CA, Zhang J. Understand the uncoupling of tauopathy and dementia through comparative analysis of subgroups of atypical Alzheimer's disease patients. *Alzheimer's & Dementia*. 2021;17:e052163.
61. Paquet D, Bhat R, Sydow A, et al. A zebrafish model of tauopathy allows in vivo imaging of neuronal cell death and drug evaluation. *J Clin Invest*. 2009;119(5):1382-1395.
62. Rosales C. Neutrophil: a cell with many roles in inflammation or several cell types? *Front Physiol*. 2018;19:113.
63. Sun R, Huang J, Yang Y, et al. Dysfunction of low-density neutrophils in peripheral circulation in patients with sepsis. *Sci Rep*. 2022;12(1):685.
64. Wang X, Fan D, Yang Y, Gimple RC, Zhou S. Integrative multi-omics approaches to explore immune cell functions: challenges and opportunities. *iScience*. 2023;26(4):106359.
65. Riaz , Sohn S. Neutrophils in inflammatory diseases: unraveling the impact of their derived molecules and heterogeneity. *Cells*. 2023;12(22):2621.
66. Malech HL, Deleo FR, Quinn MT. The role of neutrophils in the immune system: an overview. *Methods Mol Biol*. 2014;1124:3-10.
67. Hong Y, Chen L, Sun J, et al. Single-cell transcriptome profiling reveals heterogeneous neutrophils with prognostic values in sepsis. *iScience*. 2022;25(11):105301.
68. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res*. 2009;37:W305-W311. Web Server issue.
69. Villar J, Ouaknin L, Cros A, Segura E. Monocytes differentiate along two alternative pathways during sterile inflammation. *EMBO Rep*. 2023;24(7):e56308.
70. Bouchery T, Harris N. Neutrophil-macrophage cooperation and its impact on tissue repair. *Immunol Cell Biol*. 2019;97(3):289-298.

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: 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 Dement*. 2025;21:e14241. https://doi.org/10.1002/alz.14241