Received: 2 May 2024 | Revised: 24 July 2024 | Accepted: 15 August 2024

DOI: 10.1002/alz.14244

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

RESEARCH ARTICLE

Integrating amyloid imaging and genetics for early risk stratification of Alzheimer's disease

Bing He[1] | **Ruiming Wu**[2] | **Neel Sangani**[3] | **Pradeep Varathan Pugalenthi**[1] | **Alice Patania**[4] | **Shannon L. Risacher**[3] | **Kwangsik Nho**[3] | **Liana G. Apostolova**[2] | **Li Shen**[2] | **Andrew J. Saykin**[3] | **Jingwen Yan**[1] | **for the Alzheimer's Disease Neuroimaging Initiative**

[1]Department of Biomedical Engineering and Informatics, Indiana University Luddy School of Informatics, Computing and Engineering, Indianapolis, Indiana, USA

[2]Department of Biomedical Engineering and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[3]Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, Indiana, USA

[4]Department of Mathematics Statistics, University of Vermont, Burlington, Vermont, USA

Correspondence
Jingwen Yan, Luddy School of Informatics, Computing and Engineering, Indiana University Indianapolis, Indianapolis, IN 46202, USA.
Email: jingyan@iu.edu

Funding information
NIH; U.S. National Library of Medicine, Grant/Award Number: R01 LM013463; National Institute on Aging, Grant/Award Numbers: U19 AG074879, R21AG072101, P30AG010133, P30AG072976, R01AG019771, R01AG057739, U19AG024904, R01AG068193, T32AG071444, U01AG068057, U01AG072177; National Science Foundation, Grant/Award Number: 1942394

Abstract

INTRODUCTION: Alzheimer's disease (AD) initiates years prior to symptoms, underscoring the importance of early detection. While amyloid accumulation starts early, individuals with substantial amyloid burden may remain cognitively normal, implying that amyloid alone is not sufficient for early risk assessment.

METHODS: Given the genetic susceptibility of AD, a multi-factorial pseudotime approach was proposed to integrate amyloid imaging and genotype data for estimating a risk score. Validation involved association with cognitive decline and survival analysis across risk-stratified groups, focusing on patients with mild cognitive impairment (MCI).

RESULTS: Our risk score outperformed amyloid composite standardized uptake value ratio in correlation with cognitive scores. MCI subjects with lower pseudotime risk score showed substantial delayed onset of AD and slower cognitive decline. Moreover, pseudotime risk score demonstrated strong capability in risk stratification within traditionally defined subgroups such as early MCI, apolipoprotein E (*APOE*) ε4+ MCI, *APOE* ε4− MCI, and amyloid+ MCI.

DISCUSSION: Our risk score holds great potential to improve the precision of early risk assessment.

KEYWORDS
Alzheimer's disease, imaging genetics, longitudinal association analysis, pseudotime analysis, survival analysis

Highlights

- Accurate early risk assessment is critical for the success of clinical trials.
- A new risk score was built from integrating amyloid imaging and genetic data.
- Our risk score demonstrated improved capability in early risk stratification.

1 | BACKGROUND

Alzheimer's disease (AD) is a neurodegenerative disorder that results in progressive and irreversible decline in cognitive performance. It is highly heritable[1] and large-scale genome-wide association studies (GWASs) have revealed a significant set of genetic variants associated with AD.[2,3] Despite the increasing number of AD patients and elderly population at risk,[4] there are so far no clinically validated cures for AD. Instead, a recent disease-modifying medication produced only a relatively small reduction in the rate of cognitive decline, while older medications alleviate symptoms.[5] Currently, numerous AD clinical trials are underway targeting a variety of mechanisms, but the overall success rate of these trials remains relatively low.[6] There is a growing consensus that early risk assessment is critical to enhance the chance of success in AD trials, as it will enable the initiation of timely intervention when the treatment is most likely to be effective.[7] This is, however, challenging due to the limited capability to capture accurate risk for future decline in the earliest stages of disease.

Amyloid plaques and neurofibrillary tangles are two pathological hallmarks of AD identified in *post mortem* AD brains. Substantial evidence suggests a general pattern of AD progression with amyloid pathology preceding tau pathology.[8] Furthermore, growing evidence indicates that amyloid plaques begin to accumulate up to two decades before the onset of clinical symptoms. Researchers have postulated the presence of an asymptomatic early stage characterized by "silent" amyloid beta (Aβ) accumulation.[9] This hypothesis laid the foundation for the current biological classification framework (A/T/N), which has been used to assist with early detection, staging, and patient screening for clinical trials.[10] This framework assigns individuals into different stages based on amyloid (A), tau pathology (T), and neurodegeneration (N) captured by neuroimaging scans and cerebrospinal fluid (CSF) measures. Particularly, amyloid plays a central role in this framework as one of the earliest pathological features and has been widely used in AD trials to provide biological evidence of the disease. Although Aβ accumulation takes place on a continuum, amyloid positron emission tomography (PET) scans are commonly dichotomized as pathological (positive) versus normal (negative) in the current A/T/N classification system.[11] Consequently, this approach is limited in capturing the risk of AD progression or future cognitive decline, particularly in the early stage when all participants are simply categorized as amyloid negative without differentiation. Recent findings from AD progression studies, despite having different focuses such as onset prediction,[12] staging,[13] and temporal ordering of biomarkers,[14,15] consistently showed that the rate of cognitive decline varies among individuals with different genetic predispositions and across disease stages, which is tightly linked with brain amyloid burden. Therefore, we hypothesize that integrating amyloid imaging and genotype data could enhance our ability to estimate and stratify the risk of future progression.

In this study, we propose a multi-factorial pseudotime approach to integrate amyloid imaging and AD GWAS findings for estimation of a pseudo-continuous risk score. The fundamental question underlying this approach is to estimate the risk of developing AD based on an individual's genetic profile and current distribution of amyloid in the brain. Given that amyloid is one of the earliest pathological features and carrying risk alleles could predispose an individual to an elevated risk of developing AD,[16] this new risk score is expected to offer significant potential improvement for early risk assessment and facilitate the recruitment of early-stage subjects who will be most likely to benefit from clinical trials. We first used a similarity network fusion technique to integrate the heterogeneous imaging and genotype data, and then performed pseudotime analysis to generate the final risk score for each individual. We validated our risk score by comparing longitudinal cognitive changes and clinical progressions across risk groups, with a particular focus on early-stage subjects with mild cognitive impairment (MCI). This longitudinally validated scheme could improve the precision of AD risk stratification and demonstrate the feasibility of predicting diverse disease progressions.

RESEARCH IN CONTEXT

1. **Systematic review**: PubMed was searched for studies investigating early risk stratification for Alzheimer's disease (AD). However, no existing studies were found that integrate amyloid imaging and genetic data for risk estimation.
2. **Interpretation**: Results revealed that individuals in the preclinical and prodromal stages, such as mild cognitive impairment, could possibly be further stratified using pseudotime risk score. Those with higher risk score will likely experience a faster decline in cognition along with a shorter transition to AD. Targeting this group is essential for clinical trials to maximize treatment outcomes.
3. **Future directions**: These findings underscore the importance of multi-factorial approaches for assessment of AD risk. Integration of genotype and other more accessible early biomarkers warrants further investigation. Additional studies in more diverse populations are necessary to improve the generalizability of risk score.

2 | METHODS

2.1 | Participants

Data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort (http://adni.loni.usc.edu/). ADNI is a longitudinal multi-center study launched in 2003, aiming to track the progression of AD using clinical and cognitive tests, magnetic resonance imaging (MRI), fluorodeoxyglucose PET, amyloid PET, CSF, and blood biomarkers. More details can be found in previous reports.[17,18] Informed consent was obtained from all participants or their authorized representatives.

TABLE 1 Demographic information of all participants subjects. CN EMCI LMCI AD

Subjects	CN	EMCI	LMCI	AD
Number	259	298	208	185
Sex (M/F)	126/133	172/126	123/85	111/74
APOE status (−/+)	190/69	173/125	103/105	64/121
Amyloid (−/+)	178/79	150/147	70/136	26/157
Age (mean ± std)	75.52 ± 6.98	72.04 ± 7.32	74.47 ± 8.41	75.51 ± 7.99
Educ (mean ± std)	16.55 ± 2.61	16.13 ± 2.63	16.25 ± 2.81	15.84 ± 2.69
PRS (mean ± std)	0.26 ± 0.77	0.60 ± 0.95	0.77 ± 1.01	1.14 ± 1.03

Abbreviations: AD, Alzheimer's disease; *APOE*, apolipoprotein E; CN, cognitively normal; EMCI, early mild cognitive impairment; LMCI, late mild cognitive impairment; PRS, polygenic risk score; std, standard deviation.

The study population was composed of participants from the ADNI-1, ADNI-2, and ADNI-GO phases.[19] In total, we have 950 subjects with both genotype and amyloid PET imaging data, including 259 cognitively normal (CN), 298 early MCI (EMCI), 208 late MCI (LMCI) and 185 AD patients. These subjects on average were followed 3.47 years (standard deviation [SD] 3.11 years), with up to 6 visits (median: 2, Figure S1 in supporting information). The average time gap between the imaging baseline and the cognitive baseline is 1.03 months (SD 1.85 months). Out of the 950 subjects, 773 were used to generate the disease trajectory and for estimation of pseudotime as risk score. The remaining 177 subjects were kept as a test set to validate the potential of pseudotime in early risk stratification through longitudinal association analysis and survival analysis. Detailed demographics of the subjects are summarized in Table 1. To evaluate the reliability of our results, we further randomly picked 80% of those 773 training subjects to repeat the analyses.

2.2 | Amyloid imaging data

Downloaded amyloid imaging data has been quality controlled and pre-processed.[20] Briefly, amyloid PET with florbetapir (18F) as a tracer was used to measure Aβ plaques inside the brain.[21] For each subject, brain regions of interest (ROIs) were defined from subject-specific structural MRI scans, which went through segmentation and parcellation using FreeSurfer (version 5.3). Then, each florbetapir scan was co-registered to the corresponding MRI scan and the mean florbetapir uptake within the predefined ROIs was calculated. In this study, standardized uptake value ratio (SUVR) of 68 cortical ROIs, indicating the level of amyloid deposition, were included and further normalized using COMPOSITE_REF_SUVR (a summary measure provided by the ADNI) as reference. Subcortical regions were excluded for amyloid analysis because their amyloid burden has been commonly considered non-specific and not related to AD risk.[22] More detailed image processing information can be found in Landau et al.[20] Effect of age, sex, and years of education on amyloid measures were regressed out with the weight derived from CN individuals.

2.3 | Genotype data

Quality controlled (QCed) and pre-processed genotype data were obtained from the ADNI cohort. We focused on 31 AD risk single nucleotide polymorphisms (SNPs) that were previously reported in polygenic risk studies.[23] Polygenic risk scores derived from these 31 SNPs has been proven to effectively identify individuals at risk for AD across age groups. Out of those, 27 SNPs were found to pass the QC process in the ADNI cohort, and their genotype data were included for the subsequent analysis.

2.4 | Cognitive performance

Cognitive performance has been commonly administered in clinical routines to aid in diagnosing and monitoring the progression of AD. As a form of validation, derived risk score was examined for association with longitudinal cognitive data, including scores from Alzheimer's Disease Assessment Scale (ADAS), Mini-Mental State Examination (MMSE), Rey Auditory Verbal Learning Test (RAVLT), and Trail Making Test. These tests evaluate various cognitive domains such as memory, attention, language, and executive function to assess the severity of cognitive impairment.[24] In addition, two composite scores from the ADNI cohort were also included, ADNI-Mem for memory[25] and ADNI-EF for executive function,[26] each summarized from multiple cognitive tests. These composite scores provide a more accurate representation of cognitive function with minimal impact of random fluctuations in performance.

2.5 | Fusion of amyloid imaging and GWAS risk SNPs

Amyloid imaging and genotype data are of heterogeneous types (e.g., categorical and continuous) and direct integration without proper consideration could significantly bias the results. In this project, we performed similarity network fusion to integrate information from imaging and genotype data.[27] First, two similarity networks across all

subjects were computed using amyloid imaging data and genotype of 27 SNPs. Suppose we have n subjects $\{x_1, x_2, \ldots, x_n\}$, two subject similarity matrices will be computed: W^A from amyloid imaging and W^G from genotype of 27 risk SNPs. Both similarity matrices have dimension of $n \times n$ and are symmetric. For amyloid imaging data, similarity between subjects was evaluated using Euclidean distance $\rho_A(x_i, x_j)$, followed with a scaled exponential similarity kernel (Eq. 1).

$$W_{ij}^A = \exp\left(-\frac{\rho_A^2(x_i, x_j)}{\mu\varepsilon_{i,j}}\right), \quad W_{ij}^G = \exp\left(-\frac{\rho_G^2(x_i, x_j)}{\mu\varepsilon_{i,j}}\right) \tag{1}$$

Here, μ is a hyper-parameter that was set to 0.5 as default and ε_{ij} helps eliminate the scaling problem (Eq. 2). $\rho(x_i, N_i)$ is the average distance between x_i and each of its neighbors N_i.

$$\varepsilon_{ij} = \frac{mean(\rho(x_i, N_i)) + mean(\rho(x_j, N_j)) + \rho(x_i, x_j)}{3} \tag{2}$$

The genetic similarity matrix W^G was computed similarly as shown in Eq. 1. The genetic distance between subjects $\rho_G(x_i, x_j)$ was calculated using two different approaches for comparison: (1) genetic relationship matrix (GRM), and (2) Euclidean distance of polygenic risk scores (PRS) derived similarly as Desikan et al.[23] GRM is the Euclidean distance calculated from the SNP information of the individuals, that is, from the minor allele counts. The PRS is based on the genotype of 27 risk SNPs and the weight from the European International Genomics of Alzheimer's Project (IGAP), a large-scale GWAS of AD with 11,480,632 SNPs from 21,982 AD cases and 41,944 controls.[28]

W^A and W^G in Eq. 1 were further normalized to P^A and P^G, respectively, to ensure that each subject's self-similarity is always higher than its similarity to other neighbors (Eq. 3). In addition, two local affinity matrices S^A and S^G were generated to encode the relative similarity of each subject to their nearest neighbors (Eq. 3). As such, we obtained four matrices P^A, S^A, P^G, and S^G for amyloid and genotype data. Here, P^A and P^G were defined as full kernel, which carries the information from one subject to all the other subjects for the amyloid and genotype data; S^A and S^G were defined as the local affinity, which encodes the information from one subject to its neighbors.

$$P_{ij}^A = \begin{cases} \frac{W_{ij}^A}{2\sum_{k\neq i} W_{ij}^A}, & j \neq i \\ \frac{1}{2}, & j = i \end{cases}, \quad P_{ij}^G = \begin{cases} \frac{W_{ij}^G}{2\sum_{k\neq i} W_{ij}^G}, & j \neq i \\ \frac{1}{2}, & j = i \end{cases}$$

$$S_{ij}^A = \begin{cases} \frac{W_{ij}^A}{2\sum_{k\in N_i} W_{ik}^A}, & j \in N_i \\ 0, & otherwise \end{cases}, \quad S_{ij}^G = \begin{cases} \frac{W_{ij}^G}{2\sum_{k\in N_i} W_{ik}^G}, & j \in N_i \\ 0, & otherwise \end{cases} \tag{3}$$

Here, N_i means the K nearest neighbors (KNN) of the i-th subject. We experimented with K from 5 to 50 but didn't observe much difference. Therefore, it was set to default value 20 as suggested in Pearl.[29] Finally, we performed the network fusion of two similarity matrices using a non-linear message-passing theory method.[29] This is an iterative process in which both matrices continue to be updated until convergence. The final fused matrix is expected to represent the subject relationships supported by both amyloid imaging and genotype data. Let $P_{t=0}^A = P^A$ and $P_{t=0}^G = P^G$ be the initial matrices when $t = 0$. The fusion process will iteratively update two similarity matrices as follows (Eq. 4). Here, the alternating multiplication of the squared KNN similarity in two modalities essentially combines the local information across modalities. It helps reinforce shared information and thereby achieves a balanced fusion of two modalities. The final fused similarity network is simply the average of P_{t+1}^A and P_{t+1}^G.

$$P_{t+1}^A = S_A \times P_t^G \times (S_A)^T$$

$$P_{t+1}^G = S_G \times P_t^A \times (S_G)^T \tag{4}$$

2.6 | Pseudotime analysis

From the fused similarity network, pseudotime analysis tool PHATE was applied to learn a low-dimensional trajectory embedding in which subjects are ordered along a two-dimensional curve (or path). The relative position of each subject on the trajectory curve was computed as pseudotime, ranging between 0 and 1.[30] This approach has been recently applied to tau imaging data and pseudotime of each subject was interpreted as an estimation of disease progression, but not the risk of progression.[31] In addition, this pseudotime was built on autoencoder neural networks that are unable to handle the integration of heterogeneous imaging and genetic data. In this paper, we leveraged another pseudotime analysis tool (PHATE) and repurposed it to estimate the risk of progression from the fused similarity network. Briefly, PHATE involves several key steps to learn a two-dimensional trajectory from similarity network input. It starts with a diffusion process to learn global relationships within the similarity network. Next, it encodes these learned relationships using potential distance, which will finally go through eigen decomposition to generate a two-dimensional trajectory. The principal curve along the trajectory was captured using SlingShot[32,33] and all subjects were projected onto the curve for estimation of pseudotime, which will be leveraged as risk score. Subjects with similar amyloid and genetic profile are well connected in the similarity network and therefore are expected to stay close in the trajectory curve, leading to similar pseudotime (or risk score). AD patients were also included to set the reference end points in the pseudotime analysis. Pseudotime values between 0 and 1 reflect an individual's similarity to AD patients in terms of both amyloid deposition and genetic profiles. For those with EMCI, higher pseudotime values suggest a greater similarity to AD patients—whether in brain amyloid deposition, genetic profiles, or both—implying a higher risk of progression.

2.7 | Association with cross-sectional cognitive performance

We first assessed the pseudotime derived from integrated imaging genetic data regarding its correlation with clinical cognitive scores measured at baseline using a Spearman correlation. For comparison,

we also investigated the correlation with cognitive performance for pseudotime derived from GRM alone, pseudotime derived from amyloid imaging data alone, PRS from 27 risk SNPs, chronological age, apolipoprotein E (*APOE*) $\varepsilon 4$ status, and amyloid composite SUVR. For the amyloid imaging only pseudotime and GRM only pseudotime, we used similarity matrix P^A and P^G calculated in Eq. 4 before the fusion process, which further went through PHATE and SlingShot for pseudotime estimation.

2.8 | Longitudinal association analysis

Significant changes in cognitive test performance typically do not appear until MCI stages, at which time amyloid deposition may have been present for years. Further, the rate of cognitive decline in MCI is quite variable even in those at similar levels of impairment. Thus, it would be ideal if we could estimate the expected cognitive change in an earlier stage and more accurately so that effective interventions could potentially be applied. Toward this end, we investigated whether imaging genetics–based pseudotime time (i.e., risk score) is associated with rate of cognitive decline in MCI groups, including EMCIs and LMCIs. Leveraging the cognitive test scores across follow-up visits, we examined the association of baseline pseudotime with longitudinal cognitive performance using a linear mixed model (LLM),[34] with formula shown as

$$\begin{aligned}\text{Memory\_scores} \sim \text{years} * \text{group} + \text{years\_from\_baseline} \\ + \text{sex} + \text{education} + (\text{years} \mid \text{RID})\end{aligned} \tag{5}$$

In the random slopes model, it is assumed that individuals vary in terms of their changes (slope) in the mean response over time. Here, years from baseline was modeled as a random effect, and subgroups (pseudotime $\geq 0.5/ < 0.5$, $\varepsilon 4$ carriers/non-carriers, amyloid $\pm$, or EMCI/LMCI) were treated as fixed effects. We used lme4 to fit the LMM and lmerTest package[35] to obtain the *P* value, which helps determine whether there is a significant interaction between subgroups and time in predicting cognitive performance. This analysis allows us to assess whether the rate of change in cognitive performance differs significantly across different subgroups, providing insights into potential subgroup-specific trajectories of cognitive decline.

2.9 | Survival analysis

We conducted a Cox proportional hazard model[36] to explore whether MCI subgroups stratified by fused pseudotime exhibit different risks of developing AD. The event of interest is defined as the conversion from MCI to AD, and censoring occurs for MCI subjects if they did not convert to AD by the last visit. We used the survival package in R[37] to perform the Cox regression. This package provides functions to estimate survival curves and conduct statistical tests, allowing us to obtain a *P* value to quantify the difference in risk of developing AD between subgroups of subjects. We assessed the ability of pseudotime ($\geq 0.5/ < 0.5$) in early risk stratification and compared it to traditional subgroup classification including *APOE* $\varepsilon 4$ carriers/non-carriers, amyloid $\pm$, and diagnostic groups.

3 | RESULTS

3.1 | Pseudotime risk score

The fused similarity matrix, which captures the subject relationship supported by both amyloid imaging data and genotype data of AD risk SNPs, was projected into a two-dimensional trajectory embedding using PHATE, in which subjects were ordered along a principal curve (Figure 1A). Slingshot was applied to capture the principal curve of the low-dimensional embedding and to project each subject onto the principal curve. The relative position of each subject on the principal curve was computed as pseudotime, ranging between 0 and 1 (Figure 1B). A significant difference of pseudotime was observed across diagnosis groups (Figure 1C for training subjects and Figure 1D for test subjects). Estimated pseudotime was low in CN, and higher in EMCI and LMCI groups. AD patients were found with significantly higher pseudotime compared to all other diagnosis groups, indicating high genetic risk and severity in amyloid deposition. The pseudotime derived from amyloid+PRS showed inconsistent performance across randomly sampled training subjects and was therefore excluded from further analysis.

3.2 | Association with cross-sectional cognitive performance

As shown in Figure 2, we found that pseudotime derived from amyloid imaging data alone was more significantly associated with baseline MMSE, RAVLT, and composite memory scores than the amyloid composite SUVR. Associations with ADAS scores were comparable between amyloid pseudotime and amyloid composite SUVR. While AD PRS and *APOE* $\varepsilon 4$ status showed reasonable correlation with cognitive performance, we did not observe any significant correlation with cognition for pseudotime derived from GRM alone. Pseudotime derived from the fused similarity network demonstrated the highest correlation with all cognitive scores, particularly the ADAS and memory composite scores, suggesting the complementary contribution of amyloid imaging and AD risk SNPs on cognitive function. It is worth noting that amyloid composite SUVR is the driving factor of AD diagnosis in the current AT framework. Therefore, our result suggests the great potential of fused pseudotime score as a supplement of the current AT framework for more precise risk stratification and prediction of disease progression. A detailed scatter plot of these correlations can be found in Figure S2 in supporting information. For fused pseudotime and amyloid pseudotime, we further tested the significance of difference between the correlations with cognitive scores using the cocor package,[38] which provides a comprehensive statistical

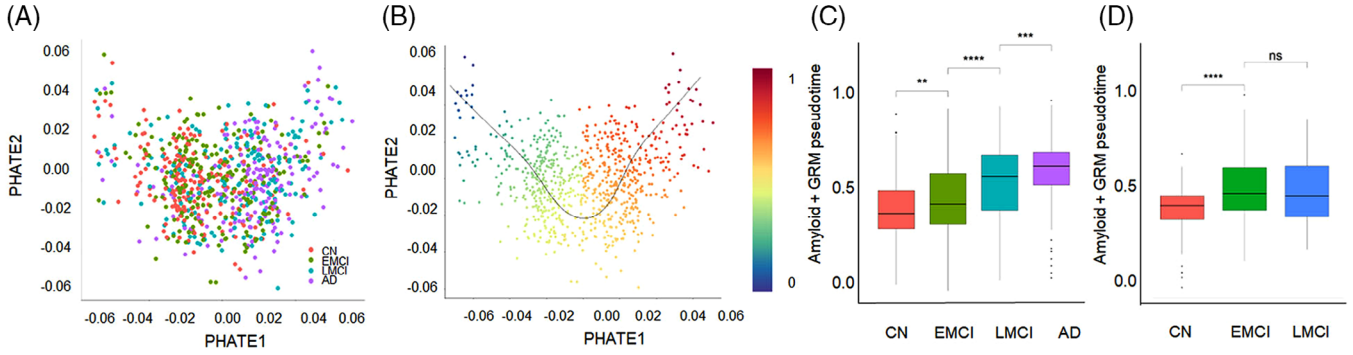

FIGURE 1 A, Subjects were ordered along a two-dimensional progression trajectory learned from PHATE, which took the input as the fused similarity network. B, Pseudotime of each subject is estimated as the relative position on the trajectory, ranging from 0 to 1. C, Pseudotime distribution across diagnosis groups for training data. D, Pseudotime distribution across diagnosis groups for testing data. ***: $P \leq 0.001$, ****: $P \leq 0.0001$. AD, Alzheimer's disease; CN, cognitively normal; EMCI, early mild cognitive impairment; GRM, genetic relationship matrix; LMCI, late mild cognitive impairment

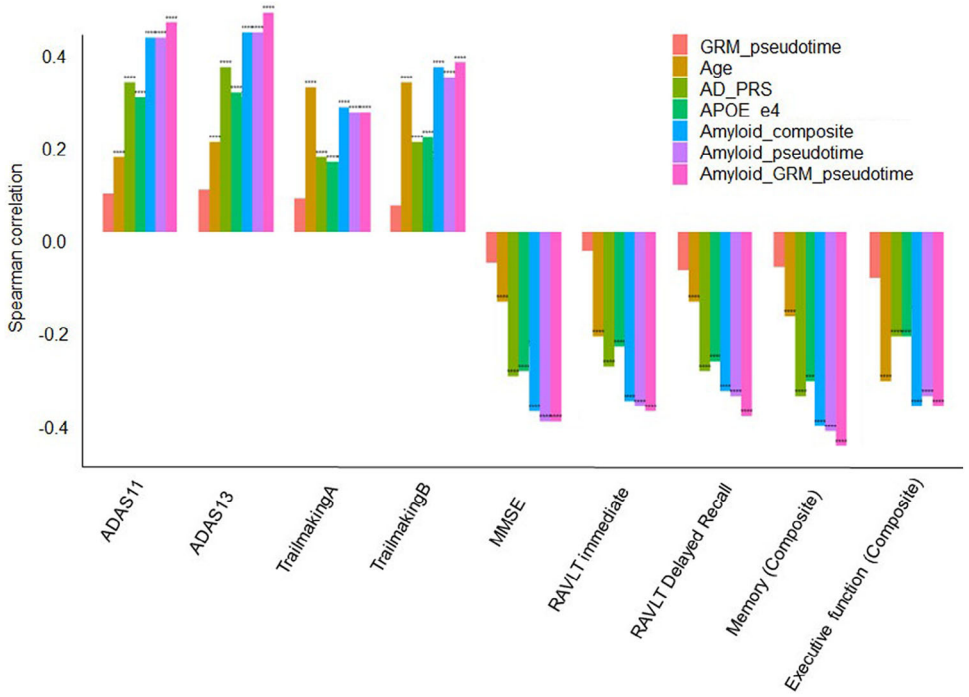

FIGURE 2 Spearman correlation with clinical cognitive performance. Memory and executive function on the right end are two composite scores. AD, Alzheimer's disease; ADAS, Alzheimer's Disease Assessment Scale; *APOE*, apolipoprotein E; GRM, genetic relationship matrix; MMSE, Mini-Mental State Examination; PRS, polygenic risk score; RAVLT, Rey Auditory Verbal Learning Test

comparison of correlations accounting for the overlap between samples and correlation between variables. For all six cognitive scores that showed differential correlation with fused pseudotime and amyloid pseudotime, we tested whether correlation from fused pseudotime is greater than that from amyloid pseudotime. As shown in Table S1 in supporting information, we found that the improved correlation observed with fused pseudotime, although modest, are all statistically significant, particularly for ADAS-13 (P = 2.65e-14), RAVLT delayed recall (P = 1.57e-13), and composite memory score (P = 9.09e-09).

3.3 | Longitudinal association with cognitive changes

LMMs were applied to investigate the association of the fused pseudotime risk score with rate of cognitive decline (Figure 3). First, we examined the association of fused pseudotime with longitudinal cognitive changes in MCI training subjects. Among all cognitive scores, ADAS-13 and composite memory score exhibited significantly different progression rate in the two pseudotime subgroups ($\geq$ 0.5 or $\leq$ 0.5). MCI subjects with pseudotime $\geq$ 0.5 showed a significantly

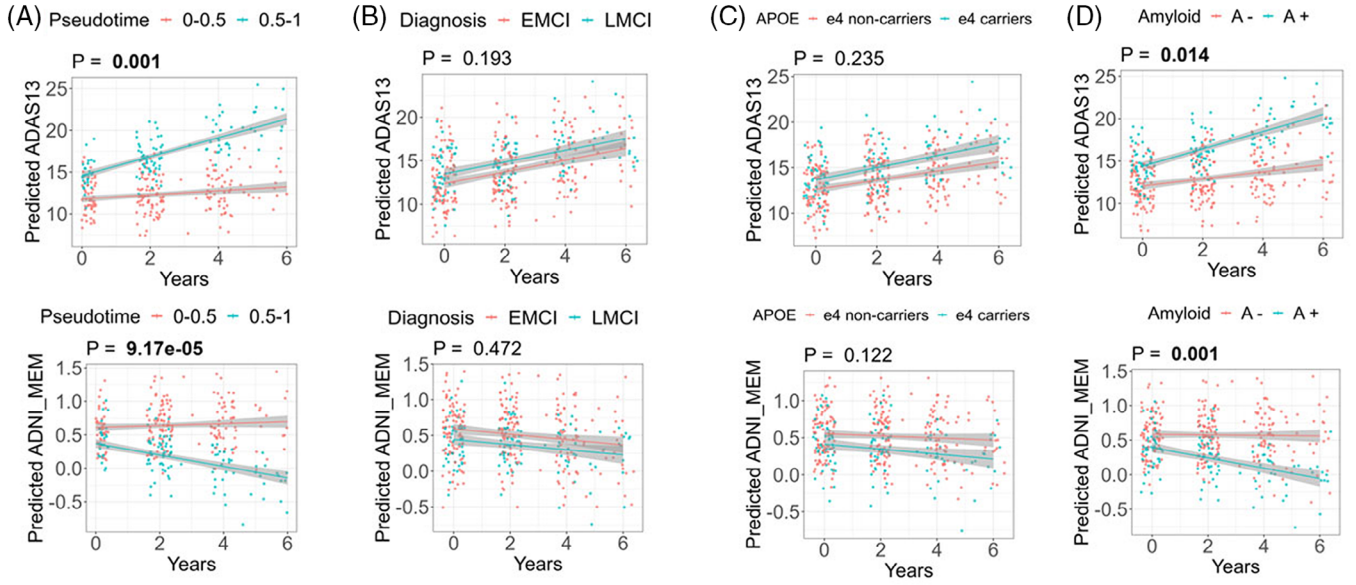

FIGURE 3 Differential rate of cognitive decline in MCI training subjects stratified by fused pseudotime (A), diagnostic groups (B), *APOE* ε4 status (C), and amyloid positivity (D). Significant longitudinal association was only observed for ADAS13 score (top) and composite memory score (bottom). AD, Alzheimer's disease; ADAS, Alzheimer's Disease Assessment Scale; ADNI, Alzheimer's Disease Neuroimaging Initiative; *APOE*, apolipoprotein E; CN, cognitively normal; EMCI, early mild cognitive impairment; LMCI, late mild cognitive impairment

faster decline in ADAS-13 ($P = 0.001$) and composite memory scores ($P = 9.17e-05$), compared to those with pseudotime ≤ 0.5 (Figure 3A). Similarly, a significant difference in rate of cognitive decline was also observed in subgroups from dichotomous classification of amyloid (i.e., amyloid ±), but with much less significant results (Figure 3D). Diagnostic subgroups EMCI/LMCI and *APOE* ε4 status are not associated with any longitudinal cognitive change (Figure 3B-C). The detailed summary statistics of the linear mixed model are in Table S2 in supporting information. Next, we performed a longitudinal association analysis in CN/EMCI subjects, which didn't yield any significant findings. We also repeated the longitudinal association analysis twice for both MCIs and CN/EMCIs by randomly sampling 80% of training subjects and obtained similar results (Figure S3 in supporting information). Additionally, we conducted a longitudinal analysis using amyloid pseudotime. This approach generally showed similar performance to fused pseudotime, but with less statistical significance (Figure S4 in supporting information).

3.4 | Survival analysis

We further evaluated the performance of fused pseudotime in risk stratification of MCI training subjects, combining EMCIs and LMCIs, and compared to conventional stratification criteria such as diagnosis of *APOE* ε4 status and amyloid positivity (Figure 4). As expected, EMCI and LMCI subjects showed distinct risk of developing AD ($P = 1.26e-05$; Figure 4B). *APOE* ε4+ MCIs exhibited a faster progression to AD compared to *APOE* ε4− MCIs ($P = 5.73e-05$; Figure 4C). Amyloid-positive MCIs showed a remarkably accelerated disease progression trajectory ($P = 0.002$; Figure 4D). Compared to these well-established subgroups, fused pseudotime unveiled distinct subgroups within the MCI population with further improved differentiation of survival risk with a significantly lower *P* value ($P = 7.03e-07$; Figure 4A).

More importantly, fused pseudotime also demonstrated strong capability in further stratifying MCI subgroups. EMCIs with pseudotime ≥ 0.5 demonstrated significantly poorer survival outcomes than those with pseudotime ≤ 0.5 ($P = 0.011$; Figure 4E). In addition, we found that *APOE* ε4+ MCIs with lower levels of fused pseudotime experienced a significantly delayed onset of AD ($P = 0.006$; Figure 4F). For MCIs that are *APOE* ε4−, a lower level of pseudotime is also associated with significantly lower risk of developing AD ($P = 0.003$; Figure 4G). Last, a similar level of risk stratification by fused pseudotime was also observed within amyloid-positive MCIs ($P = 0.012$; Figure 4H). Survival analysis was not conducted within the amyloid-negative subgroup due to insufficient subjects who progressed to AD.

Similarly, we also evaluated the performance of fused pseudotime in risk stratification of CN and EMCI subjects and compared to conventional stratification criteria such as diagnosis of *APOE* ε4 status and amyloid positivity (Figure 4I-P). The event for the survival analysis was defined as the conversion from CN/EMCI to LMCI/AD, and censorship was applied to CN/MCI subjects who did not convert up to their last visit. While neither diagnosis subgroups nor amyloid positivity subgroups showed a significant difference (Figure 4J, L), *APOE* ε4+ CN/EMCIs exhibited a faster progression to AD compared to *APOE* ε4− CN/EMCIs ($P = 0.017$; Figure 4K). High- and low-fused pseudotime again successfully stratified CNs/EMCIs into subgroups with differential risk of progression ($P = 0.003$; Figure 4I). Moreover, fused pseudotime further identified high-risk CN/EMCI subjects from pre-defined subgroups like *APOE* ε4+ CN/EMCIs ($P = 0.026$; Figure 4N), and amyloid-positive CN/EMCIs ($P = 0.037$; Figure 4P),

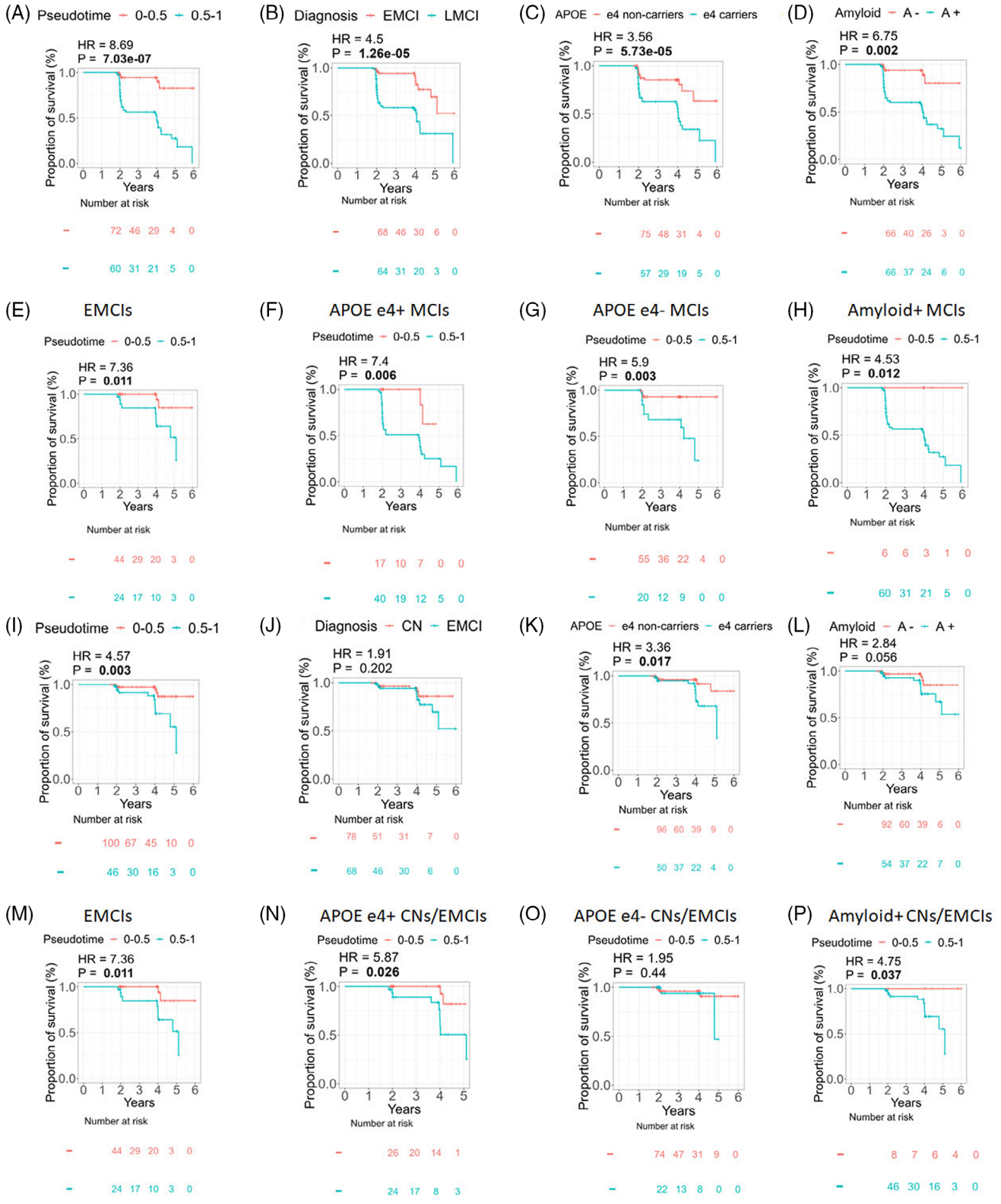

FIGURE 4 Differential progression risk of MCI subjects (Row 1-2) and CN/EMCI subjects (Row 3-4). Row 1: MCIs stratified by (A) pseudotime derived from integration of amyloid imaging and candidate SNPs, (B) diagnosis, (C) *APOE* ε4 status, and (D) amyloid positivity. Row 2: pseudotime can further stratify MCI subjects with differential survival risk within groups like (E) EMCIs, (F) *APOE* ε4+ MCIs, (G) *APOE* ε4− MCIs, and (H) amyloid positive MCIs. Row 3: CN/EMCIs stratified by (I) pseudotime derived from integration of amyloid imaging and candidate SNPs, (J) diagnosis, (K) APOE ε4 status, and (L) amyloid positivity. Row 4: differential survival risk across pseudotime groups within (M) EMCIs, (N) APOE ε4+ CN/EMCIs, (O) APOE ε4− CN/EMCIs, and (P) amyloid positive CN/EMCIs. AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; *APOE*, apolipoprotein E; CN, cognitively normal; EMCI, early mild cognitive impairment; HR, hazard ratio; LMCI, late mild cognitive impairment; MCI, mild cognitive impairment

implying its potential in preclinical risk stratification. Additionally, we performed survival analysis using amyloid pseudotime. Like the longitudinal analysis, amyloid pseudotime yielded results comparable to fused pseudotime. However, fused pseudotime showed a much higher hazard ratio and demonstrated improved effectiveness in capturing amyloid-positive MCIs and CN/EMCIs with higher progression risk (Figure S4). For MCIs and CN/EMCIs, we repeated the survival analysis with random subsampling of training subjects and observed similar performance as shown in Figures S5 and S6 in supporting information, which confirmed the reliability of our results.

3.5 | Validation on test subjects

We validated the above observed association and risk stratification using test subjects not involved in the trajectory learning (Figure 5). For those 177 testing subjects, one at a time, each test subject was included in the training dataset to calculate the pseudotime risk. We compared the pseudotime of training subjects before and after retraining with each test subject, and only observed minimal differences (average correlation > 0.999, mean absolute error: 0.000359 ± 0.000292). Subsequently, longitudinal association analysis and survival analysis were performed on the test subjects. All test subjects were divided into two subgroups: those with pseudotime ≥ 0.5 and those with pseudotime ≤ 0.5. Starting from MCIs, we stratified them as EMCI/LMCI, *APOE* $\varepsilon 4$ carriers/non-carriers, and amyloid $\pm$. Consistent with earlier observation in the training data, individuals with elevated fused pseudotime exhibited a notably faster decline in composite memory score compared to those with lower pseudotime (Figure 5A; $P = 0.036$). However, we didn't observe any notable difference in rate of cognitive decline for subgroups delineated by diagnosis, $\varepsilon 4$ status, and amyloid positivity (Figure 5B-D). The detailed summary statistics about the LMM is in Table S3 in supporting information. In the survival analysis (Figure 5E-H), subgroups defined by fused pseudotime exhibited a notable difference of progression risk ($P = 0.002$; Figure 5E), in which MCI individuals with higher fused pseudotime were observed with elevated risk of developing AD throughout the follow-up period.

For the CN/EMCI test subjects, no significant association was observed between fused pseudotime and longitudinal cognitive changes (Figure 5I-L). Subgroups defined by fused pseudotime, but not the other factors (Figure 5N-P), exhibited a notable difference of progression risk ($P = 0.002$; Figure 5M), in which individuals with higher pseudotime were observed with elevated risk of developing AD throughout the follow-up period.

4 | DISCUSSION

In this study, we introduced a novel multi-factorial pseudotime approach to derive an AD risk score through integration of heterogeneous amyloid imaging and genotype data. Pseudotime analysis translates a fused similarity network into a pseudo-continuous risk score for each subject, reflecting not only its genetic predisposition but also severity in amyloid imaging. Compared to an amyloid composite SUVR, PRS, or pseudotime derived from amyloid or genotype alone, fused pseudotime exhibited a modest but statistically significant improvement in association with both individual clinical cognitive scores and two composite scores for memory and executive function. This result suggests the complementary contribution of amyloid imaging and genetics data in cognitive function. Compared to amyloid composite SUVR, pseudotime derived from brain-wide amyloid showed either stronger or comparable correlation with cognitive performance in most clinical tests, which validated the effectiveness of pseudotime in capturing the AD progression.

In addition, fused pseudotime outperformed traditional stratification strategies like *APOE* $\varepsilon 4$ status, amyloid positivity, and diagnosis and exhibited significantly improved differentiation power in stratifying MCI subjects with accelerated cognitive decline. Given that cognitive performance is part of current clinical routines to evaluate AD progression, yet substantial changes often do not manifest until late stages,[39] strong association of baseline pseudotime risk with longitudinal cognitive changes suggests its great clinical promise for early detection and subtyping of progression. Moreover, subjects with higher fused pseudotime at baseline exhibit substantially shorter transition time to the onset of AD, which is observed not only within MCI subjects but also EMCIs, *APOE* $\varepsilon 4$ carriers, *APOE* $\varepsilon 4$ non-carriers, and amyloid-positive subgroups. In other words, fused pseudotime can effectively identify individuals at higher risk within previously established subgroups, which could enable more precise targeting of participants during clinical trial recruitment. When applied to test subjects, it was observed that comparable levels of differentiation can also be extended to cognitively normal individuals, suggesting the significant promise of fused pseudotime for early risk assessment. In both longitudinal and survival analyses, amyloid pseudotime demonstrated performance close to fused pseudotime, highlighting the benefits of pseudotime approaches. However, amyloid pseudotime by itself is less effective in stratifying progression among amyloid-positive MCIs or CN/EMCIs and showed a lower hazard ratio, which underscores the importance and value of incorporating genetic information. Taken together, these findings suggest that studies of preclinical and prodromal AD could benefit from using the fused pseudotime as risk score instead of separating groups simply by *APOE* $\varepsilon 4$ status or amyloid positivity. This score could enable clinical trials to better pinpoint individuals at elevated risk, further facilitating early intervention and enhancing treatment outcomes.

The study has several limitations that merit future consideration. First, using similarity networks as input requires re-training a new model to predict the risk of each new patient. While our findings indicate that retraining with one additional test subject at a time does not impact the training outcomes, new supervised training methods that can be directly applied to new patients are critical to facilitate clinical use. Second, due to limited data sets with both amyloid imaging and genotype data, we were only able to validate the findings using test subjects excluded from trajectory learning. Participants included in this analysis are White only, therefore limiting the generalizability of these data to the population at large. This study could be further enhanced with larger and more diverse samples when they are available. In

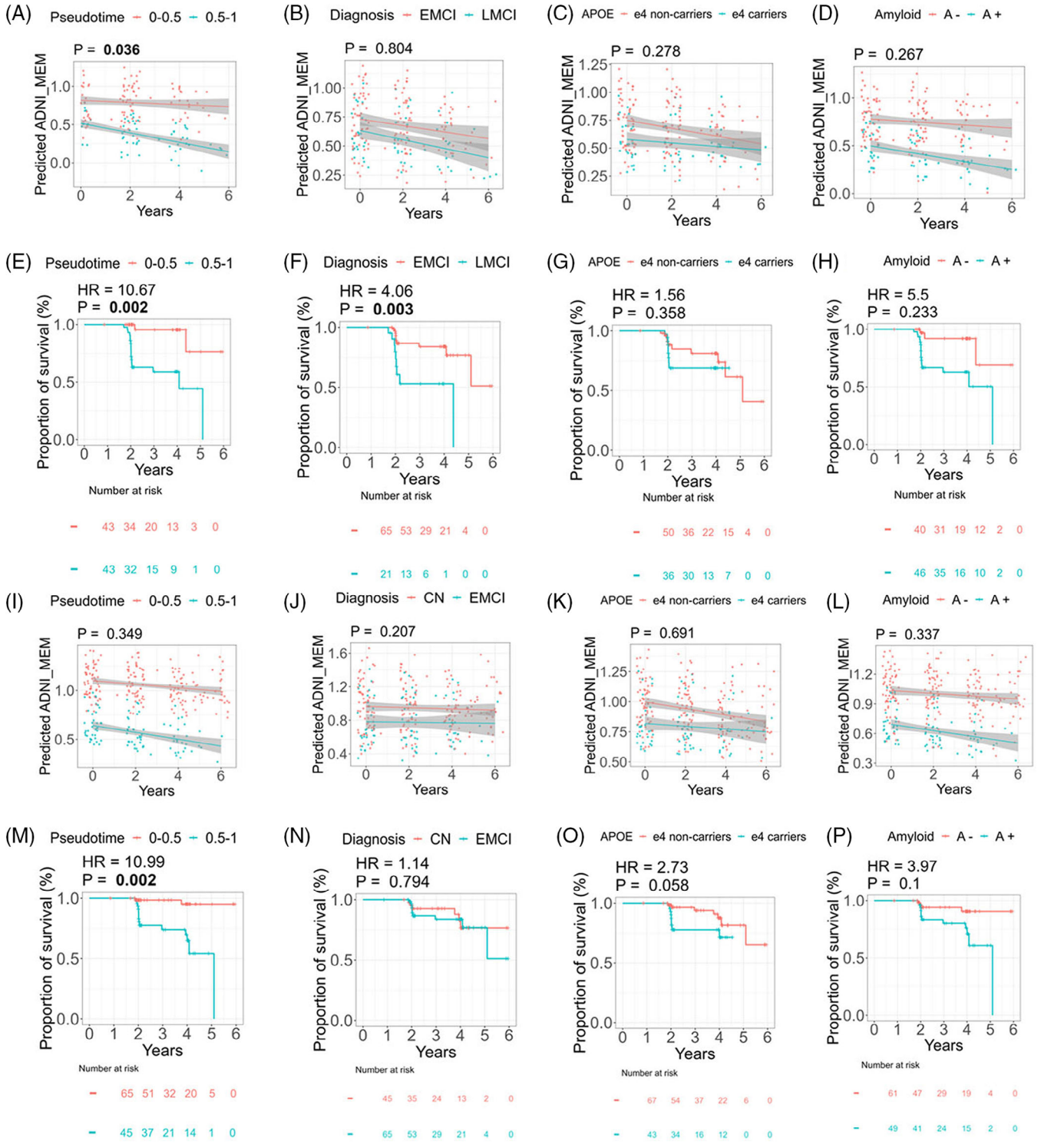

FIGURE 5 Validation results of longitudinal association and survival analyses on test subjects. A-D, Rate of decline in composite memory score across MCI subgroups delineated by fused pseudotime, diagnosis, *APOE* ε4 status, and amyloid positivity. E-H, Progression risk of MCI subgroups delineated by pseudotime, diagnosis, *APOE* ε4 status, and amyloid positivity. I-L, Rate of decline in composite memory score across CN/EMCI subgroups delineated by pseudotime, diagnosis, *APOE* ε4 status, and amyloid positivity. M-P, progression risk of CN/EMCI subgroups delineated by pseudotime, diagnosis, *APOE* ε4 status, and amyloid positivity. AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; *APOE*, apolipoprotein E; CN, cognitively normal; EMCI, early mild cognitive impairment; HR, hazard ratio; LMCI, late mild cognitive impairment; MCI, mild cognitive impairment

addition, pseudotime analysis methods themselves are so far limited in handling longitudinal follow-up data, which could provide additional support to the learned progression trajectory and lead to improved precision of pseudotime risk score. Finally, despite some debate, substantial evidence suggests that changes in CSF amyloid levels may precede detectable amyloid accumulation on PET scans.[40] Therefore, pseudotime derived from integration of amyloid PET and genotype may be still not optimal for early risk stratification. While CSF amyloid is invasive and hard to administer in clinical trials, there is potential for exploring the integration of genotype and other more accessible blood biomarkers, especially with growing evidence of early changes in blood metabolites and proteins.[41,42]

ACKNOWLEDGMENTS

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 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. This research was supported by NIH/National Institute on Aging grants R21 AG072101, P30 AG010133, P30 AG072976, R01 AG019771, R01 AG057739, U19 AG024904, R01 AG068193, T32 AG071444, U01 AG068057, U01 AG072177, U19 AG074879, U.S. National Library of Medicine grant R01 LM013463, and National Science Foundation grant 1942394.

CONFLICT OF INTEREST STATEMENT

Dr. Saykin has received support from Avid Radiopharmaceuticals, a subsidiary of Eli Lilly (in kind contribution of PET tracer precursor) and holds advisory roles with Siemens Medical Solutions USA, Inc., NIH NHLBI, and Eisai. His editorial commitments include serving as editor-in-chief for the journal *Brain Imaging and Behavior*, and he participates in various NIH/NIA advisory committees. Liana G. Apostolova receives funding from NIA U01 AG057195, NIA R01 AG057739, NIA P30 AG010133, Alzheimer Association LEADS GENETICS 19-639372, and Roche Diagnostics RD005665; research support from AVID Pharmaceuticals and Life Molecular Imaging; does consulting for Biogen, Two Labs, Eli Lilly, NIH, Florida Dept. Health, NIH Biobank; participates on data safety monitoring boards for IQVIA, NIA R01 AG061111, UAB Nathan Schick Center; owns stock in Semiring, Inc. and Cassava Neurosciences; receives honoraria from AAN, MillerMed, ASiM, Health and Hospitality Corporation, Mayo Clinic; and is editor-in-chief for *Disease Assessment and Disease Monitoring*. Bing He, Ruiming Wu, Neel Sangani, Pradeep Varathan Pugalenthi, Alice Patania, Shannon L. Risacher, Kwangsik Nho, Li Shen, Jingwen Yan have nothing to disclose. Author disclosures are available in the supporting information.

CONSENT STATEMENT

All human subjects provided informed consent and signed consent forms.

ORCID

Jingwen Yan https://orcid.org/0000-0001-9066-4908

REFERENCES

1. Gatz M, Pedersen NL, Berg S, et al. Heritability for Alzheimer's disease: the study of dementia in Swedish twins. *J Gerontol A Biol Sci Med Sci*. 1997;52(2):M117-M125.
2. Sims R, Hill M, Williams J. The multiplex model of the genetics of Alzheimer's disease. *Nat Neurosci*. 2020;23(3):311-322.
3. Jansen IE, Savage JE, Watanabe K, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet*. 2019;51(3):404-413.
4. Organization WH, et al. Risk reduction of cognitive decline and dementia: WHO guidelines 2019.
5. Van Dyck CH, Swanson CJ, Aisen P, et al. Lecanemab in early Alzheimer's disease. *N Engl J Med*. 2023;388(1):9-21.
6. Cummings J. Lessons learned from Alzheimer disease: clinical trials with negative outcomes. *Clin Transl Sci*. 2018;11(2):147.
7. Rasmussen J, Langerman H. Alzheimer's disease–why we need early diagnosis. *Degener Neurol Neuromuscul Dis*. 2019:123-130.
8. Bloom GS. Amyloid-β and tau: the trigger and bullet in Alzheimer disease pathogenesis. *JAMA Neurol*. 2014;71(4):505-508.
9. Dubois B, Hampel H, Feldman HH, et al. Preclinical Alzheimer's disease: definition, natural history, and diagnostic criteria. *Alzheimers Dement*. 2016;12(3):292-323.
10. Counts SE, Ikonomovic MD, Mercado N, Vega IE, Mufson EJ. Biomarkers for the early detection and progression of Alzheimer's disease. *Neurotherapeutics*. 2017;14:35-53.
11. Jack Jr CR, Bennett DA, Blennow K, et al. A/T/N: an unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology*. 2016;87(5):539-547.
12. Betthauser TJ, Bilgel M, Koscik RL, et al. Multi-method investigation of factors influencing amyloid onset and impairment in three cohorts. *Brain*. 2022;145(11):4065-4079.
13. Bilgel M, Jedynak BM, Initiative ADN, et al. Predicting time to dementia using a quantitative template of disease progression. *Alzheimer's Dement: Diagn Assess Dis Monit*. 2019;11:205-215.
14. Li Y, Yen D, Hendrix RD, Gordon BA, et al. Timing of biomarker changes in sporadic Alzheimer's disease in estimated years from symptom onset. *Ann Neurol*. 2024;95(5):951-965.
15. Bilgel M, Koscik RL, An Y, et al. Temporal order of Alzheimer's disease-related cognitive marker changes in BLSA and WRAP longitudinal studies. *J Alzheimers Dis*. 2017;59(4):1335.
16. Selkoe DJ, Hardy J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol Med*. 2016;8(6):595-608.
17. Jack Jr CR, Bernstein MA, Borowski BJ, et al. Update on the magnetic resonance imaging core of the Alzheimer's disease neuroimaging initiative. *Alzheimers Dement*. 2010;6(3):212-220.
18. Saykin AJ, Shen L, Foroud TM, et al. Alzheimer's disease neuroimaging initiative biomarkers as quantitative phenotypes: genetics core aims, progress, and plans. *Alzheimers Dement*. 2010;6(3):265-273.
19. Weiner MW, Veitch DP, Aisen PS, et al. The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimers Dement*. 2013;9(5):e111-e194.
20. Landau SM, Breault C, Joshi AD, et al. Amyloid-β imaging with Pittsburgh compound B and florbetapir: comparing radiotracers and quantification methods. *J Nucl Med*. 2013;54(1):70-77.
21. Okamura N, Yanai K. Florbetapir (18F), a PET imaging agent that binds to amyloid plaques for the potential detection of Alzheimer's disease. *IDrugs*. 2010;13(12):890-899.
22. Edmonds EC, Bangen KJ, Delano-Wood L, et al. Patterns of cortical and subcortical amyloid burden across stages of preclinical Alzheimer's disease. *J Int Neuropsychol Soc*. 2016;22(10):978-990.

23. Desikan RS, Fan CC, Wang Y, et al. Genetic assessment of age-associated Alzheimer disease risk: development and validation of a polygenic hazard score. *PLoS Med*. 2017;14(3):e1002258.
24. Kueper JK, Speechley M, Montero-Odasso M. The Alzheimer's disease assessment scale–cognitive subscale (ADASCog): modifications and responsiveness in pre-dementia populations. a narrative review. *J Alzheimers Dis*. 2018;63(2):423-444.
25. Crane PK, Carle A, Gibbons LE, et al. Development and assessment of a composite score for memory in the Alzheimer's disease neuroimaging initiative (ADNI). *Brain Imaging Behav*. 2012;6:502-516.
26. Gibbons LE, Carle AC, Mackin RS, et al. A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain Imaging Behav*. 2012;6:517-527.
27. Wang B, Mezlini AM, Demir F, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods*. 2014;11(3):333-337.
28. Schellenberg GD, et al. International Genomics of Alzheimer's Disease Project (IGAP) genome-wide association study. *Alzheimers Dement*. 2012;4(8):P101.
29. Pearl J, Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan kaufmann; 1988.
30. Campbell KR, Yau C. Uncovering pseudotemporal trajectories with covariates from single cell and bulk expression data. *Nat Commun*. 2018;9(1):2442.
31. Mukherjee S, Heath L, Preuss C, Jayadev S, Garden GA, Greenwood AK, et al. Molecular estimation of neurodegeneration pseudotime in older brains. *Nat Commun*. 2020;11(1):5781.
32. Moon KR, van Dijk D, Wang Z, et al. Visualizing structure and transitions in highdimensional biological data. *Nat Biotechnol*. 2019;37(12):1482-1492.
33. Street K, Risso D, Fletcher RB, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *Bmc Genomics [Electronic Resource]*. 2018;19:1-16.
34. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv*. 2014:14065823.
35. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: tests in Linear Mixed Effects Models. *J Stat Softw*. 2017;82(13):1-26.
36. Van Dijk PC, Jager KJ, Zwinderman AH, Zoccali C, Dekker FW. The analysis of survival data in nephrology: basic concepts and methods of Cox regression. *Kidney Int*. 2008;74(6):705-709.
37. Therneau TM, Lumley T. Package 'survival'. *R Top Doc*. 2015;128(10):28-33.
38. Diedenhofen B, Musch J. cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS One*. 2015;10(4):e0121945.
39. van Loenhoud AC, van der Flier WM, Wink AM, et al. Cognitive reserve and clinical progression in Alzheimer disease: a paradoxical relationship. *Neurology*. 2019;93(4):e334-e346.
40. Hampel H, Hardy J, Blennow K, et al. The amyloid-β pathway in Alzheimer's disease. *Mol Psychiatry*. 2021;26(10):5481-5503.
41. Guo T, Zhang D, Zeng Y, Huang TY, Xu H, Zhao Y. Molecular and cellular mechanisms underlying the pathogenesis of Alzheimer's disease. *Mol Neurodegener*. 2020;15:1-37.
42. Lleó A, Núñez-Llaves R, Alcolea D, et al. Changes in synaptic proteins precede neurodegeneration markers in preclinical Alzheimer's disease cerebrospinal fluid. *Mol Cell Proteomics*. 2019;18(3):546-560.

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: He B, Wu R, Sangani N, et al. Integrating amyloid imaging and genetics for early risk stratification of Alzheimer's disease. *Alzheimer's Dement*. 2024;20:7819–7830. https://doi.org/10.1002/alz.14244