Received: 9 January 2024 | Revised: 19 March 2024 | Accepted: 17 April 2024

DOI: 10.1002/dad2.12597

Alzheimer's & Dementia
Diagnosis, Assessment & Disease Monitoring

RESEARCH ARTICLE

Detecting time-varying genetic effects in Alzheimer's disease using a longitudinal genome-wide association studies model

Xiaowei Zhuang[1,2,3] | **Gang Xu**[4,5] | **Amei Amei**[5] | **Dietmar Cordes**[3,6] | **Zuoheng Wang**[4] | **Edwin C. Oh**[1,2,7] | **For the Alzheimer's Disease Neuroimaging Initiative**

[1]Interdisciplinary Neuroscience PhD Program, University of Nevada Las Vegas, Las Vegas, Nevada, USA

[2]Laboratory of Neurogenetics and Precision Medicine, University of Nevada, Las Vegas, Las Vegas, Nevada, USA

[3]Lou Ruvo Center for Brain Health, Cleveland Clinic, Las Vegas, Nevada, USA

[4]Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut, USA

[5]Department of Mathematical Sciences, University of Nevada, Las Vegas, Las Vegas, Nevada, USA

[6]Institute of Cognitive Science, University of Colorado Boulder, Boulder, Colorado, USA

[7]School of Medicine, Department of Internal Medicine, University of Nevada, Las Vegas, Las Vegas, Nevada, USA

Correspondence
Edwin C. Oh, Interdisciplinary Neuroscience PhD Program, University of Nevada Las Vegas, Las Vegas, NV89154, USA.
Email: edwin.oh@unlv.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 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
NIH, Grant/Award Numbers: R01-MH109706, P20-GM103440, R01LM014087, RF1-AG071566

Abstract

INTRODUCTION: The development and progression of Alzheimer's disease (AD) is a complex process, during which genetic influences on phenotypes may also change. Incorporating longitudinal phenotypes in genome-wide association studies (GWAS) could unmask these genetic loci.

METHODS: We conducted a longitudinal GWAS using a varying coefficient test to identify age-dependent single nucleotide polymorphisms (SNPs) in AD. Data from 1877 Alzheimer's Neuroimaging Data Initiative participants, including impairment status and amyloid positron emission tomography (PET) scan standardized uptake value ratio (SUVR) scores, were analyzed using a retrospective varying coefficient mixed model association test (RVMMAT).

RESULTS: RVMMAT identified 244 SNPs with significant time-varying effects on AD impairment status, with 12 SNPs on chromosome 19 validated using National Alzheimer's Coordinating Center data. Age-stratified analyses showed these SNPs' effects peaked between 70 and 80 years. Additionally, 73 SNPs were linked to longitudinal amyloid accumulation changes. Pathway analyses implicated immune and neuroinflammation-related disruptions.

DISCUSSION: Our findings demonstrate that longitudinal GWAS models can uncover time-varying genetic signals in AD.

Xiaowei Zhuang and Gang Xu contributed equally to this study.

KEYWORDS
AD clinical impairment status, amyloid accumulation, longitudinal GWAS, RVMMAT, time-varying genetic effects

Highlights

- Identify time-varying genetic effects using a longitudinal GWAS model in AD.
- Illustrate age-dependent genetic effects on both diagnoses and amyloid accumulation.
- Replicate time-varying effect of APOE in a second dataset.

1 | INTRODUCTION

The biological constructs and hallmark pathologies in Alzheimer's disease (AD) are characterized by extracellular β-amyloid protein deposition, intraneuronal pathological tau protein accumulation, accompanied by neurodegeneration and neuroinflammation.[1,2] As a highly heritable disorder, genetic factors contribute significantly to the development of AD. The heritability of AD is estimated to be approximately 60% to 80%, initially derived from genetic twin studies,[3] and supported by large-scale genome-wide association studies (GWAS).[4,5] Delineating the strong genetic component in AD has become a major objective in AD research, as it provides an opportunity to (1) understand the disease etiology and risks, (2) characterize pathophysiological pathways, and (3) identify potential diagnostic and prognostic biomarkers.

To date, large-scale AD-GWAS have reported more than 80 putative associated loci and genes,[4–8] among which, the *apolipoprotein E (APOE)* E4 allele on chromosome 19 has been shown to have the largest genetic risk for AD. Several recent studies have focused on changes in longitudinal measures as phenotypes and identified genetic contributions toward cognitive decline or disease progression in AD.[9,10] In these studies, the reliance on single time-point measurements or the computation of a single measurement of changes can limit statistical power and hinder the identification of potential time-varying genetic contributions to dynamic phenotypes.

The genetic architecture of gene expression regulation has been shown to be unstable over time and linked with aging.[11] Genetic contributions toward complex traits such as body mass index (BMI)[12] and hypertension[13] also exhibit time- or age-dependent variability. Pertinent to AD, Studies on *APOE* have additionally shown that E4 allele counts demonstrate an inconstant hazard in developing AD, which declines with increasing age.[14] Therefore, as an aging and complex disorder, AD may have critical timelines for the onset and progression, during which genetic influences on phenotypes may also fluctuate. Thus, our main objective of this study was to identify time-varying genetic contributions to phenotypes in AD, with the expectation that the identified single nucleotide polymorphisms (SNPs) could further assist in delineating genetic mechanisms relevant to AD.

Current AD-GWAS has mostly been conducted using clinically diagnosed case-control subjects. Association studies with other amyloid, tau, or neurodegeneration (ATN) biomarkers, or within biologically defined AD participants, may be restricted by the limited sample size and thus suffer from reduced statistical power. Longitudinal models in this case could take advantage of repeated phenotypic measures from the same subject to potentially boost the statistical power in association analyses, which could in turn allow GWAS to be performed on ATN-specific biomarkers, or within biologically defined AD participants. Therefore, in the current study, we focused on both clinical and biological phenotypes using longitudinal GWAS models.

More recently, several large-scale data initiatives in AD have collected and measured longitudinal diagnoses and ATN biomarkers over time.[15–17] Among these databases, the Alzheimer's neuroimaging data initiative (ADNI, https://ida.loni.usc.edu/) is a multicenter, multiphase study dedicated to assessing clinical, imaging, and genetic biomarkers in AD. The availability of such comprehensive data empowers the research community to explore time-varying genetic effects on various clinical and biological phenotypes throughout the course of the disease in AD.

Several statistical models have been applied in longitudinal GWAS. Among which, linear mixed effects models and generalized estimating equations are commonly used to account for dependent structure in longitudinal observations.[18,19] Using these methods, studies have reported novel variants and genes associated with disease progression or cognitive resilience/decline in AD.[20–23] Varying coefficient models, as an extension of generalized mixed effects models, have been designed to specifically capture the time-varying genetic effects on dynamic traits in longitudinal GWAS.[13] These varying-coefficient models in longitudinal GWAS have led to the identification of time-dependent genetic effects in cocaine users,[24] subjects with hypertension,[13] and hippocampal volumes in AD subjects.[25]

In this study, we applied a varying coefficient model to perform longitudinal GWAS on both a binary phenotype of clinical impairment status and a continuous phenotype of brain amyloid accumulation in AD. We hypothesized that, with increased statistical power and modeling of varying coefficients, longitudinal GWAS models can support the detection of time-varying genetic effects in repeatedly measured phenotypes. We anticipate that our results will improve the identification of genetic variants associated with fluctuating pathological or clinical phenotypes in AD.

2 | METHODS

We performed longitudinal GWAS using a retrospective varying coefficient mixed model association test (RVMMAT), and Figure 1 depicts our overall pipeline.

Data from ADNI and the National Alzheimer's Coordinating Center (NACC) were utilized as the main and replication datasets, respectively. Both ADNI (https://ida.loni.usc.edu/) and NACC (https://naccdata.org/) are publicly available databases that can be accessed upon reasonable requests. No approval or participant consent was obtained locally from these participants.

2.1 | Primary dataset: ADNI participants

Descriptions of ADNI participants are detailed in Supplementary Method 1. Briefly, we included 1877 older participants with genome-wide genotyping data available from the ADNI database (Table 1). We downloaded participants' (1) genotyping data in PLINK format, (2) clinical diagnoses at each visit; and (3) composite standardized uptake value ratios (SUVRs) computed from florbetapir amyloid PET scan at each visit.

Genotyping data were preprocessed and imputed at the Michigan imputation server,[26] with the 1000 Genomes phase3 data (European) as a reference panel. SNPs meeting the following quality-control conditions were retained: (1) call rate > 99%, (2) Hardy-Weinberg χ^2 statistic p-value > 10^{-6}, and (3) minor allele frequencies > 1%. A total of 9,573,130 SNPs were examined in the following longitudinal GWAS analyses.

Clinical diagnoses from 10,825 clinical visits (as of Nov. 2022) of 1877 participants were considered as the binary phenotype (5.76 ± 2.50 visits/participant, detail in Supplementary Method 1), with 4010 clinical visits with a cognitively nonimpaired (i.e., normal) diagnosis and 6815 visits with a cognitively impaired diagnosis (including diagnoses of mild cognitive impairment [MCI] and dementia, Table 1). SUVRs of 2598 longitudinal PET scans from a subset of 1096 participants (2.34 ± 1.38 visits/participant, Table 1) were considered as the continuous phenotype in the following analyses.

RESEARCH IN CONTEXT

1. **Systematic review**: While longitudinal genome-wide association studies (GWAS) signals with time-varying effect in Alzhemer's disease (AD) have not been reported, traditional GWAS have been conducted at scale in AD. Longitudinal GWAS methods with time-varying coefficients have been applied to several other disease conditions. The relevant citations are appropriately cited.
2. **Interpretation**: We have identified single nucleotide polymorphisms (SNPs) with time-varying genetic effects in AD. This finding not only corroborates previous research on the age-dependent genetic impact of the *apolipoprotein E (APOE)* E4 allele, but also introduces new SNPs that may exert similar age-related effects.
3. **Future directions**: This manuscript establishes a framework for investigating time-varying and age-dependent genetic signals in AD using longitudinal GWAS methods. Our approach focuses on both clinical (binary) and biological (continuous) phenotypes. Looking ahead, with an expanded sample size, we aim to develop age-specific polygenic risk scores to more precisely assess an individual's genetic predisposition to AD. Additionally, future GWAS will target amyloid, tau, or neurodegeneration (ATN)-specific and dynamic phenotypes, enhancing our understanding of the disease's progression and variability.

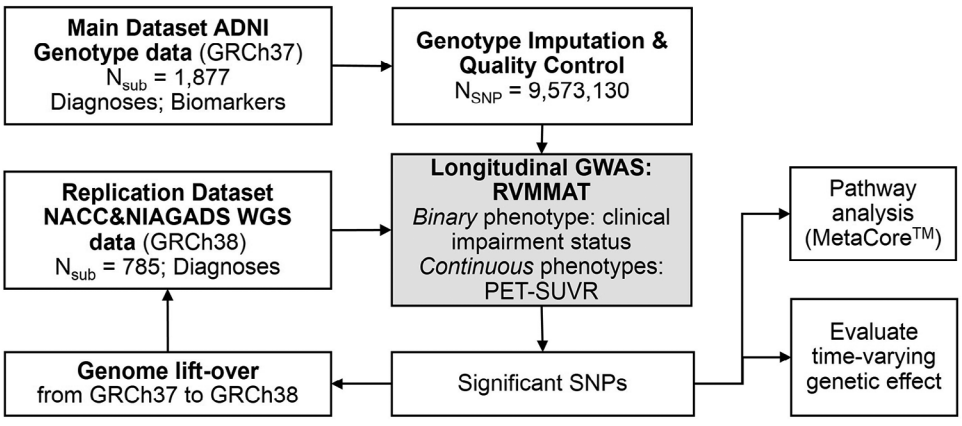


FIGURE 1 Analyses pipeline. RVMMAT was first performed to identify significant SNPs for both binary and continuous phenotypes using ADNI participants (center gray boxes). Posthoc analyses on significant SNPs were then performed, including identifying (1) enriched functional pathways and (2) time-varying genetic effect on phenotypes. Replication analyses were next performed on significant SNPs using NACC participants. AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; NACC, National Alzheimer's Coordinating Center; RVMMAT, retrospective varying coefficient mixed model association test; SNP, single nucleotide polymorphism.

TABLE 1 Demographics of ADNI (main dataset) and NACC (replication dataset) participants.

Parameter	Main: ADNI (*N* = 1877)	Replication: NACC (*N* = 785)
Sex		
Men	1020	310
Women	857	475
Race		
American Indian or Native Alaskan	4	2
Native Hawaiian	2	
Asian	29	5
African American	74	185
White	1745	593
More than one or unknown	23	0
Amyloid status (latest amyloid PET)		
No. of subjects positive	620	N/A
Negative	476	
Unknown	781	
Diagnostic visits		
Impaired	6815	1786
Nonimpaired	4010	4708
Age at diagnoses (yr)		
Impaired	75.75 ± 7.82	79.28 ± 10.06
Nonimpaired	75.78 ± 6.68	77.19 ± 7.43

Abbreviations: ADNI, Alzheimer's disease neuroimaging initiative; NACC, National Alzheimer's Coordinating Center.

2.2 | Longitudinal GWAS model

We performed longitudinal GWAS using RVMMAT for both binary and continuous AD phenotypes. Details about RVMMAT could be found in Supplementary Method 2 and Xu et al., 2024.[13]

Briefly, RVMMAT can be viewed as an extension of the generalized linear mixed model (GLMM)[18] and retrospective GLMM-based association test (RGMMAT)[24] to a varying coefficient mixed effects model.[28] It models dynamic genetic effects using cubic smoothing splines. By embracing more flexible assumptions on the genetic effect function, unlike methods that assume constant genetic effects, RVMMAT can detect time-varying genetic variants linked with dynamic phenotypes, resulting in increased statistical power.[13]

2.3 | Post hoc analyses on significant SNPs

2.3.1 | Pathway analyses

Significant SNPs (with multiple-comparison-corrected *p*-values < 0.05 for the binary phenotype and *raw p*-values < 1E-04 for the continuous phenotype) were annotated to the genes that were located within, or to their nearest genes based on genome positions. Annotated genes were then included in Pathway analyses using MetaCore (Clarivate Analytics PLC, https://portal.genego.com/).

2.3.2 | Analysis of time-varying genetic effects

We next examined genetic effects at each time point (i.e., within each age interval) for significant SNPs identified by RVMMAT. All longitudinal visits were grouped into various age ranges with a 5-year interval. For each age interval, phenotype and covariates at each longitudinal visit, and corresponding participants' genotype were obtained. Note that only one visit was kept from the same subject with the same phenotype in each age interval.

We analyzed the time-varying genetic effects on phenotypes using both chi-square (χ^2) statistics and regression models. For the binary phenotype, within each age interval, we first performed a χ^2 test to examine the genotypic differences between phenotype groups for each significant SNP. A larger χ^2 value indicates a greater genetic difference between phenotype groups for the SNP within this age interval, as compared to other age intervals. For both binary and continuous phenotypes, within each age interval, we further performed a logistic regression and a linear regression with genotype as predictor and phenotype as outcome, respectively. The same set of covariates as in RVMMAT were included in the regression analyses. A larger regression coefficient (in amplitude) of genotype for a given age interval indicates a greater genetic effect for that SNP, as compared to other age intervals.

2.4 | Replication analyses

NACC participants were utilized as a replication dataset. The NACC has been established in collaboration with more than 42 previous and current Alzheimer's Disease Research Centers (ADRCs) throughout the United States over more than 20 years.[29] Details about NACC data and replication analysis is included in Supplementary Method 3.

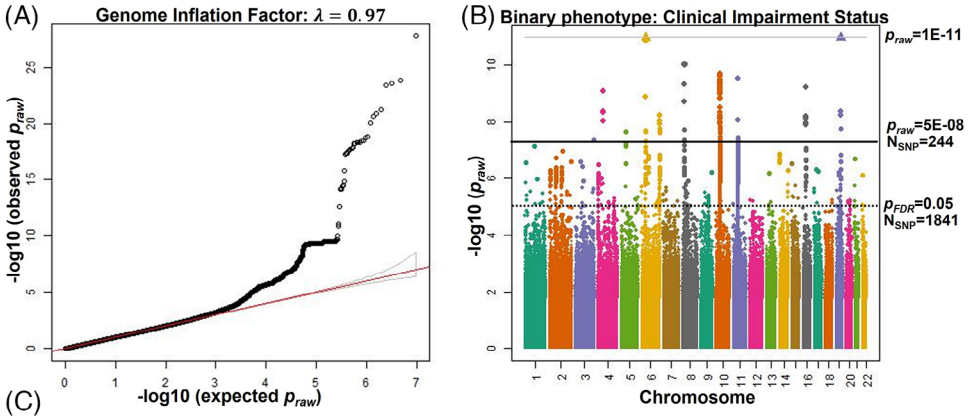
Briefly, for 785 NACC participants, we downloaded (1) whole-genome sequencing data in genome variant calling format (VCF) from the NIAGAD-data sharing service; and (2) clinical diagnoses at each visit from the NACC Uniform Data Set[30] (Table 1). We divided the total 6494 clinical visits into a nonimpaired (cognitively normal, N_{visit} = 4708) and an impaired (self-reported impairment, MCI, and AD, N_{visit} = 1786) group. We next performed RVMMAT on significant SNPs identified from ADNI data to examine whether the observed time-varying genetic effect on AD could be replicated using the NACC dataset.

3 | RESULTS

3.1 | Longitudinal GWAS with the binary phenotype

3.1.1 | Significant SNPs

In ADNI participants, applying RVMMAT with clinical impairment status as a phenotype showed no evidence of inflation in the



(C)

Functional Pathways	p-value	FDR	Network Objects from Active Data
1) Maturation and migration of dendritic cells in skin sensitization	1.50E-04	1.79E-02	MHC class II alpha chain, MHC class II, MHC class II beta chain
2) Immune response: Antigen presentation by MHC class II	2.17E-04	1.79E-02	MHC class II alpha chain, MARCH1, MHC class II, MHC class II beta chain
3) Breakdown of CD4+ T cell peripheral tolerance in type 1 diabetes mellitus	2.55E-04	1.79E-02	HLA-DQB1, MHC class II, HLA-DQA1
4) Immune response: Induction of the antigen presentation machinery by IFN-gamma	3.22E-04	1.79E-02	HLA-DQB1, MHC class II, HLA-DQA1
5) Immune response: T cell co-signaling receptors, schema	3.60E-04	1.79E-02	Nectin-2, Collagen V, MHC class II
6) Ca(2+)-dependent NF-AT signaling in cardiac hypertrophy	4.00E-04	1.79E-02	G-protein alpha-11, G-protein alpha-q/11, GATA-4
7) Transcription: Transcription factor Tubby signaling pathways	8.33E-04	3.20E-02	G-protein alpha-11, G-protein alpha-q/11
8) Role of ZNF202 in regulation of expression of genes involved in atherosclerosis	1.28E-03	3.82E-02	HDL proteins, APOE

FIGURE 2 RVMMAT results on binary phenotype of clinical impairment status in AD using ADNI participants. (A) Quantile-quantile plot scatters the observed *p*-values (*y*-axis) and the expected *p*-values (*x*-axis), indicating no evidence of inflation. (B) Manhattan plot shows that 244 SNPs reached genome-wide significance with $p_{raw} < 5E\text{-}08$ (solid black line), and 1841 SNPs reached genome-wide significance with $p_{FDR} < 0.05$ (dashed black line). For better display purposes, SNPs with a $p_{raw} < 1E\text{-}11$ (solid gray line) were not shown individually and were clustered and represented by one triangle at $p_{raw} = 1E\text{-}11$ (solid gray line). (C) Significantly disrupted functional pathways on 1841 genome-wide significant SNPs with $p_{FDR} < 0.05$ (and their annotated genes) using MetaCore. AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; RVMMAT, retrospective varying coefficient mixed model association test; SNP, single nucleotide polymorphism.

quantile-quantile plot. Specifically, 99.93% SNPs were close to the diagonal line, indicating an absence of significant inflation (genome inflation factor = 0.97, Figure 2A). RVMMAT significance levels (raw *p*-values [p_{raw}]) for all SNPs were shown in a Manhattan plot (Figure 2B). Although we performed genome-wide tests across 9,573,130 imputed SNPs, many SNPs were highly correlated; and therefore, we considered a widely used threshold $p_{raw} < 5E\text{-}08$ as our genome-wide significance level (solid black line in Figure 2B). We further performed a false discovery rate (FDR) correction on p_{raw} across all SNPs at a lower genome-wide significance level ($p_{FDR} < 0.05$, dashed black line in Figure 2B) using the method developed by Benjamini and Hochberg in 1995.[31]

At $p_{raw} < 5E\text{-}08$, 244 SNPs reached genome-wide significance (Figure 2B). Among the 244 SNPs, the most significant signals were derived from 36 SNPs on chromosome 19, and were clustered at the *APOE*, *TOMM40*, *APOC1*, and *NECTIN2* (also known as *PVRL2*) genes. Among which, the most significant SNP was rs429358 at position 45411941 on chromosome 19 ($p_{raw} = 1.66E\text{-}28$), which was the *APOE* E4 determinant SNP. Besides chromosome 19, we identified significant SNPs on chromosomes 4, 5, 6, 8, 10, 11, and 16. Detailed genome positions, RVMMAT statistics (p_{raw} and p_{FDR}), minor allele frequencies (MAFs), annotated genes, and distances to annotated genes (in base-pairs) are listed in Table S1 and described in Supplementary Result 1.

We next examined functional pathways associated with the 1841 genome-wide significant SNPs at $p_{FDR} < 0.05$ (Figure 2C and Tables S2 and S3). We identified eight significantly enriched functional pathways (FDR-$p < 0.05$ in Fisher's exact test) associated with immune response (four pathways), G-protein signaling (two pathways), lipid-associated gene expression (one pathway) and dendritic cell maturation (one pathway).

3.1.2 | Time-varying genetic effect

The genotypic differences and the estimated genetic effect over time are shown in Figure 3 for 244 genome-wide significant SNPs at $p_{raw} < 5E\text{-}08$. The χ^2 statistics (Figure 3A) and estimated logistic regression coefficients (Figure 3B) at each time point (i.e., age interval) were obtained by using the observed phenotypic values within that age range. A straight line was used to connect the estimated values at two adjacent age intervals, and a zero-line was added to indicate no genotypic effect in Figure 3B.

As shown in Figure 3A, for SNPs on chromosome 19, we observed an overall greater genotypic difference between clinically impaired and nonimpaired participants in a wider age interval of 65–80 years old (maximum $\chi^2 = 77.86$, bottom plot in Figure 3A), as compared to other

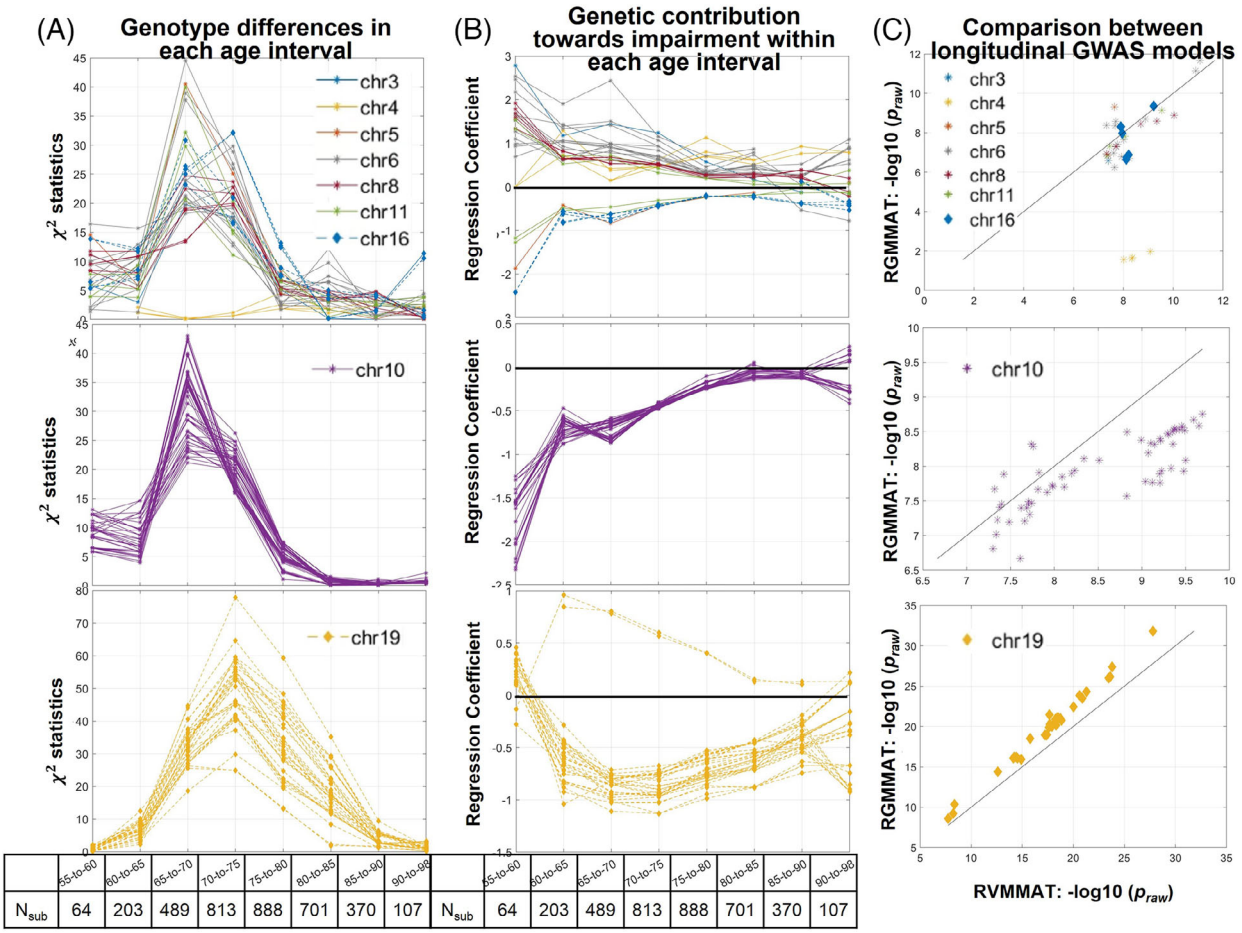


FIGURE 3 Estimated genotypic difference and genetic effect of 244 genome-wide significant SNPs at $p_{RAW} < 5E\text{-}08$ on binary clinical impairment status in AD at each time point and age interval. SNPs on chromosomes 10 and 19 are plotted separately in the middle (chromosome 10 SNPs) and bottom panels (chromosome 19 SNPs) due to the relatively larger number of significant SNPs. SNPs on other chromosomes are plotted in the top panel. (A) Estimated genotypic differences (χ^2 statistics) between phenotype groups within each age interval. The number of subjects in each age group are listed at the bottom. (B) Estimated genotypic effect in predicting impairment status (regression coefficient) in logistic regression model for subjects within each age interval. Deviations from the zero-line (i.e., no genotypic effect (solid black line)) indicate greater genotypic effects on this phenotype. (C) Comparisons of significance levels (-log10(p_{raw})) between longitudinal GWAS models with time-varying coefficients (RVMMAT, *x*-axis) and assuming time-constant genetic effect (RGMMAT, *y*-axis). AD, Alzheimer's disease; GWAS, genome-wide association studies; RVMMAT, retrospective varying coefficient mixed model association test; SNP, single nucleotide polymorphism.

SNPs (top and middle plots in Figure 3A). An overall greater genotypic effect on this phenotype was also observed in the same age interval for these SNPs, as reflected by larger (in amplitude) logistic regression coefficients that deviate from the zero-line (Figure 3B, bottom plot). Since in the logistic regression model, we coded clinically impaired status as zero and clinically nonimpaired status as one, a negative regression coefficient here indicated the possession of alleles contributing negatively toward clinically normal status, (i.e., increased effect toward clinical impairment status). Furthermore, for these chromosome 19 SNPs, both longitudinal GWAS models with and without time varying coefficient (RVMMAT and RGMMAT) demonstrated high statistical power, as most SNPs reached significant *p*-values ($p_{raw} < 1E\text{-}10$) in both models (Figure 3C, bottom plots).

We additionally observed a larger genotypic difference on AD clinical impairment status in a smaller age interval of 70–75 years old for SNPs on chromosomes 6, 8, 10, and 11 (top and middle plots in Figure 3A). These genotypic differences were not observed before 60 years old, and diminished after 80 years old, as shown by the reduced χ^2 values. Deviations of regression coefficients from the zero-line in Figure 3B also indicated that the genetic effect of these SNPs on AD clinical impairment status decreased during aging (top and middle plots in Figure 3B). For most of these SNPs, longitudinal GWAS models that assume a time-constant genetic effect (RGMMAT) generated less significant *p*-values compared to our model, which incorporates a time-varying coefficient (RVMMAT, Figure 3C upper and middle plots).

3.1.3 | Replication

We repeated the RVMMAT method on 244 genome-wide significant SNPs ($p_{raw} < 5E\text{-}08$) using the NACC dataset. In this replication, 29 SNPs reached $p_{raw} < 0.05$ (Table S1). Twenty-four SNPs clustered

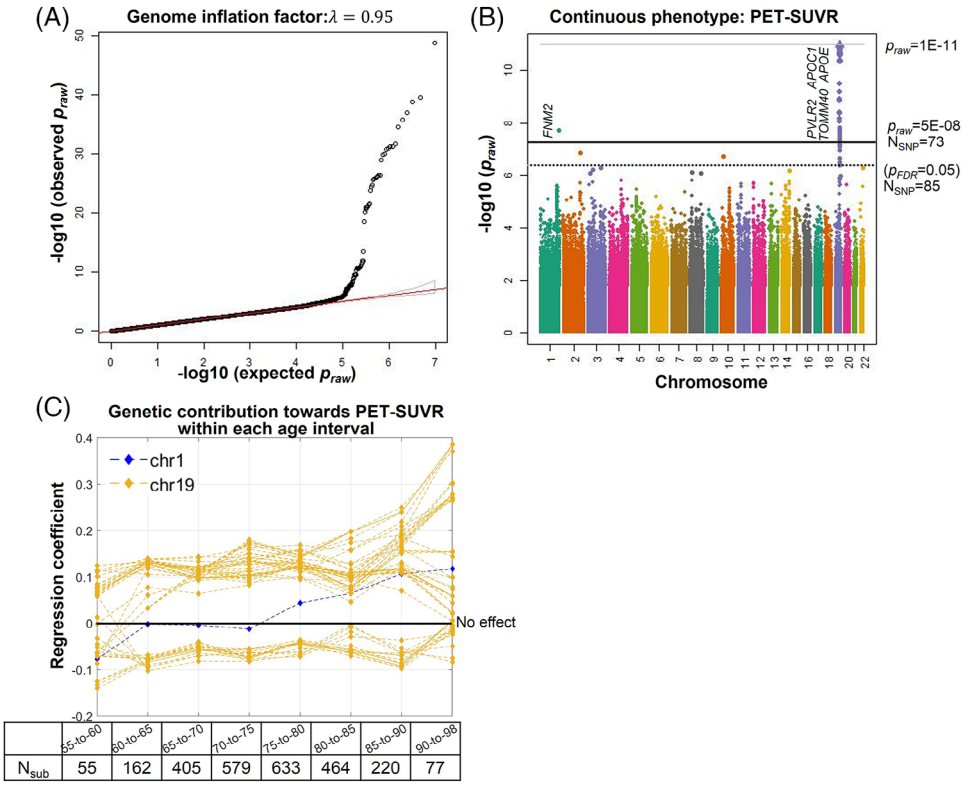


FIGURE 4 RVMMAT results on continuous phenotype of brain amyloid accumulation (PET-SUVR) in AD using ADNI participants. (A) Quantile-quantile plot indicates no evidence of inflation with a genome inflation factor $\lambda = 0.95$. (B) Manhattan plot shows that 73 SNPs reached genome-wide significance with $p_{raw} < 5E\text{-}08$ (solid black line), and 85 SNPs reached genome-wide significance with $p_{FDR} < 0.05$ (dashed black line). For better display purposes, SNPs with a $p_{raw} < 1E\text{-}11$ (solid gray line) were not shown individually and were clustered and represented by one triangle at $p_{raw} = 1E\text{-}11$ (solid gray line). (C) Estimated genotypic effect (regression coefficient) in predicting PET-SUVRs in linear regression model for subjects within each age interval. Deviations from the zero-line (solid black line) indicate greater genotypic effects on this phenotype. ADNI, Alzheimer's disease neuroimaging initiative; PET-SUVR, positron emission tomography-standardized uptake value ratio; RVMMAT, retrospective varying coefficient mixed model association test; SNP, single nucleotide polymorphism.

on chromosome 19, with 12 SNPs survived this FDR-correction (described in Supplementary Result 2 and Figure S1). Five SNPs clustered at *KSR1P1* pseudogene on chromosome 10. Two additional SNPs clustered to *HLA-DQB1* gene on chromosome 6 reached a trend level significance ($p_{raw} < 0.10$).

3.2 | Longitudinal GWAS with a continuous phenotype

Figure 4 shows RVMAMT results on a continuous phenotype of brain amyloid accumulation using 2598 longitudinal PET-SUVR measures from ADNI participant. Using this phenotype, RVMMAT produced a genomic inflation factor of 0.95 (Figure 4A). At $p_{raw} < 5E\text{-}08$, 73 SNPs reached genome-wide significance (Figure 4B). Most of these SNPs were again clustered on chromosome 19 and annotated to *APOE*, *APOC1*, *TOMM40*, and *NECTIN2* genes. In addition to chromosome 19, one SNP on chromosome 1 reached genome-wide significance and was located at the *FMN2* gene. In addition, as shown in Figure 4C, an increasing genotypic effect on PET-SUVR was observed when participants' age increased, as indicated by the larger (in amplitude) regression coefficient that deviates from the zero-line. We next performed a functional pathway analysis using top SNPs with $p_{raw} < 1E\text{-}04$ ($N_{SNP} = 1039$, Table S4) and identified 17 significantly disrupted biological processes that were involved in immune response, signal transduction, development, and neurophysiological process (Table S5).

4 | DISCUSSION

The development and progression of AD is a complex process that could change over time, during which the impact of genetic variation on phenotypes may also fluctuate. Incorporating longitudinal phenotypes with time-varying coefficients in GWAS provides the opportunity to identify this changing genetic effect on phenotypic variations over time along the disease continuum and, therefore, may shed new light on understanding AD patho-mechanisms. In this study, we utilized a varying coefficient model, RVMMAT, to perform longitudinal GWAS on repeated measurements of clinical and biological phenotypes in AD. Benefiting from the improved statistical power with longitudinal measures in RVMMAT, a relatively limited number of subjects from the publicly available ADNI database were used. Our study led to the

identification of genome-wide significant SNPs that could convey time-varying and age-dependent genetic effects on phenotypes for both binary clinical impairment status and continuous amyloid accumulation in AD.

Modeling varying AD genetic risk over time has been shown to result in a 5%–10% statistical power gain in GWAS and has assisted in discovering novel AD-associated variants.[20,25,32] The method we adopted here, RVMMAT, (1) utilized time-varying coefficients to model fluctuating genetic effects on dynamic phenotypes, and (2) applied retrospective tests that provide robustness against model misspecification. With the adaptability and robustness, Xu et al., 2024[13] have shown RVMMAT can lead to improved statistical power in GWAS when tested on simulated data with known ground truth. In our results, the RVMMAT method also produced more significant results for SNPs with moderate effects than models assuming a constant genetic effect over time (i.e., RGMMAT, Figure 3C, upper and middle plots). Note that for SNPs that convey relatively large genetic effects, both models with and without time-varying genetic effect tend to produce significant results (Figure 3C bottom plots).

With increased statistical power using RVMMAT, we identified 244 genome-wide significant SNPs relevant to clinical impairment status in AD (Figure 2). The most predominant signals were clustered on chromosome 19, with the most significant SNP to be the *APOE* E4 determinant SNP (Table S1). Our post hoc age-stratified analyses further demonstrated that the maximum genotypic effect of these SNPs on AD impairment status existed at the age interval of 70–75 years old, and then declined with age (Figure 3). These observations were consistent with several previous reports. For instance, *APOE* E4 allele has been shown to convey an age-dependent hazard in developing AD that declines with increasing age.[14] Logistic regression analyses (similar to what we performed in this study) have further demonstrated that the *APOE* genotype shared a greater regression coefficient and a larger area under the ROC curve in predicting clinical AD cases in participants younger than 80 years old, as compared to more senior participants.[33] In our study, these observed time-varying genetic effects on chromosome 19 SNPs, especially the *APOE* E4 determinant SNP, were also replicated using the NACC dataset (Table S1). Our replication study combined with previous reports have strengthened the potential time-varying effect of these chromosome 19 SNPs on AD clinical status.

Different from clinical impairment status, our study identified 72 SNPs on chromosome 19 that may convey an increasing genotypic effect on brain amyloid accumulation with increased age (Figure 4). More specifically, our results showed that, before the age of 85, the genotypic effect of these SNPs on brain amyloid accumulation remained constant, with a significant increase in more senior participants. These results indicated that distinct time-varying genetic effects might occur for these chromosome 19 SNPs toward AD clinical and biological phenotypes, respectively. These results will benefit from further validation, as our study contained a relatively small number of senior participants for the amyloid phenotype.

In addition to chromosome 19, our longitudinal GWAS models identified significant SNPs on other chromosomes. Among them, SNPs on chromosome 6 were clustered to *HLA-DQB1* gene and reached a trend level significance in our replication analysis using the NACC cohort (Figure 2 and Table S1). *HLA-DQB1* has been reported to convey a significant age-of-onset risk in developing AD,[34] which might explain our observation of the decreased genotypic effect of these SNPs on AD clinical impairment status along aging. Our study further identified significant SNPs clustered to lincRNA or pseudogenes on chromosome 10, with a time-varying genetic effect on AD clinical impairment status. More recently, there has been an increasing number of studies focused on the role of long noncoding RNAs in regulating expression and modulating protein levels in AD.[35,36] Therefore, these SNPs could be potential novel signals in AD genetic mechanisms, but require further replication and validation studies.

Our pathway analyses of top SNPs highlighted the involvement of immune responses, lipid metabolism, G-protein signaling and neurophysiological processes in clinical impairment status and brain amyloid accumulations related to AD (Figure 2 and Table S3). Given the established role for these pathways in neurodegeneration,[7] our findings suggest that the longitudinal GWAS model can provide enhanced statistical power in detecting biologically relevant genetic loci that are associated with phenotypic dynamics, and highlight the role of neuroinflammation in AD.

Following the ATN framework[1] and AD clinical impairment status, we have further limited our longitudinal GWAS analyses to participants that were classified as amyloid positive on their latest PET scans in ADNI (detailed in Supplementary Result 3). In this analysis, chromosome 19 SNPs clustered to *APOE*, *APOC1*, or *TOMM40* did not reach genome-wide significance (Figure S2). Given the reduced sample size ($N_{sub} = 620$), we may lack the statistical power of the model to detect relevant signals. The lack of significance of these chromosome 19 SNPs might also provide additional evidence of previous findings that SNPs clustered to *APOE* loci were associated with amyloid accumulation in AD.[37] Furthermore, pathway analyses on these top SNPs have highlighted oxidative stress, particularly reactive oxygen species-induced cellular signaling, to be the most disrupted functional pathway (Figure S2C).

Interestingly, we were only able to replicate time-varying genetic signals on chromosome 19 using the NACC dataset for the phenotype of AD clinical impairment status. This limited replication might be explained partially by the different racial and phenotypic group distributions between ADNI and NACC participants (Table 1). Future analyses using additional cohorts from diverse backgrounds could then take full advantage of the RVMMAT method in longitudinal GWAS to produce novel and stable results. With the larger sample-size, it would also be interesting to restrict longitudinal GWAS, or even traditional GWAS, to biologically defined AD cases, that is, to amyloid-positive and tau-positive participants (Supplementary Discussion). Another possible future direction that might not rely on additional participants would be to compute an age-specific polygenic risk score (PRS) for each age-group with the genetic risk estimated particularly for that age interval (Supplementary Discussion). Age-specific PRS might be a less biased and more accurate measure to estimate participants' genetic risk, as it accounts for the potential time-varying genotypic effects on dynamic phenotypes.

AUTHOR CONTRIBUTIONS

Xiaowei Zhuang: Conceptualization; methodology; formal analysis; investigation; visualization; writing—original draft; review and editing. *Gang Xu*: Conceptualization; methodology; formal analysis; investigation; visualization; writing—original draft; review and editing. *Zuoheng Wang*: Methodology; visualization; review and editing. *Dietmar Cordes*: Visualization; review and editing. *Amei Amei*: Visualization; review and editing. *Edwin C. Oh*: Conceptualization; methodology; formal analysis; investigation; visualization; infrastructure provision; review and editing; supervision.

ACKNOWLEDEMENTS

The authors additionally thank all NACC and ADNI participants and their families for their support in AD research. Dr. Edwin C. Oh is supported by NIH grants R01-MH109706 and P20-GM103440. Dr. Gang Xu and Dr. Zuoheng Wang are supported by the NIH grant R01LM014087. Ms. Xiaowei Zhuang and Dr. Dietmar Cordes are supported by the NIH grant RF1-AG071566.

CONFLICT OF INTEREST STATEMENT

All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. Author disclosures are available in the supporting information.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

ADNI participants. Main data used in this study are collected by ADNI. Therefore, the study was approved by each participating ADNI site's local Institutional Review Boards, as documented on the ADNI website. All participants gave written, informed consent. No ethics approval or participant consent was obtained locally for ADNI participants. *NACC participants.* Replication data used in this study are collected by each ADRC and uploaded to the NACC database. Therefore, the study was approved by each ADRC site's local Institutional Review Boards. No ethics approval or participant consent was obtained locally for NACC participants.

AVAILABILITY AND IMPLEMENTATION

The implementation of RVMMAT is freely available on the web at: https://github.com/ZWang-Lab/RVMMAT. Details of RVMMAT are presented in Xu et al., 2024.

The data underlying this article are accessed from ADNI (https://ida.loni.usc.edu/) and NACC (https://naccdata.org) databases. ADNI data used in this study are from 1,877 individuals with whole-genome genotyping data available on https://ida.loni.usc.edu/pages/access/geneticData.jsp through ADNI data request process. De-identified data from NACC participants used in this study are available through NACC data request process (https://naccdata.org/requesting-data/data-request-process) with qualified researchers.

ORCID

Edwin C. Oh https://orcid.org/0000-0002-6042-1478

REFERENCES

1. Jack CR, Bennett DA, Blennow K, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement*. 2018;14:535-562. doi:10.1016/j.jalz.2018.02.018
2. Heneka MT, Carson MJ, Khoury JEl, et al. Neuroinflammation in Alzheimer's disease. *Lancet Neurol*. 2015;14:388-405. doi:10.1016/S1474-4422(15)70016-5
3. Gatz M, Reynolds CA, Fratiglioni L, et al. Role of genes and environments for explaining Alzheimer disease. *Arch Gen Psychiatry*. 2006;63:168-174. doi:10.1001/archpsyc.63.2.168
4. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013;45:1452-1458. doi:10.1038/ng.2802
5. Wightman DP, Jansen IE, Savage JE, et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet*. 2021;53:1276-1282. doi:10.1038/s41588-021-00921-z
6. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat Genet*. 2019;51:414-430. doi:10.1038/s41588-019-0358-2
7. Bellenguez C, Küçükali F, Jansen IE, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet*. 2022;54:412-436. doi:10.1038/s41588-022-01024-z
8. 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;51:404-413. doi:10.1038/s41588-018-0311-9
9. Euesden J, Gowrisankar S, Qu AX, St Jean P, Hughes AR, Pulford DJ. Cognitive decline in Alzheimer's disease: limited clinical utility for GWAS or polygenic risk scores in a clinical trial setting. *Genes*. 2020;11:501. doi:10.3390/genes11050501
10. Kumar A, Shoai M, Palmqvist S, et al. Genetic effects on longitudinal cognitive decline during the early stages of Alzheimer's disease. *Sci Rep*. 2021;11:19853. doi:10.1038/s41598-021-99310-z
11. Bryois J, Buil A, Ferreira PG, et al. Time-dependent genetic effects on gene expression implicate aging processes. *Genome Res*. 2017;27:545-552. doi:10.1101/gr.207688.116
12. Chu W, Li R, Liu J, Reimherr M. Feature selection for generalized varying coefficient mixed-effect models with application to obesity GWAS. *Ann Appl Stat*. 2020;14:276-298. doi:10.1214/19-AOAS1310
13. Xu G, Amei A, Wu W, et al. Retrospective varying coefficient association analysis of longitudinal binary traits: application to the identification of genetic loci associated with hypertension. *Ann Appl Stat*. 2024;18:487-505. doi:10.1214/23-AOAS1798
14. Liu L, Caselli RJ. Age stratification corrects bias in estimated hazard of APOE genotype for Alzheimer's disease. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*. 2018;4:602-608. doi:10.1016/j.trci.2018.09.006
15. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology*. 2010;74:201. doi:10.1212/WNL.0B013E3181CB3E25
16. Beekly DL, Ramos EM, Van Belle G, et al. The National Alzheimer's Coordinating Center (NACC) database an Alzheimer disease database purpose of the database. *Alzheimer Dis Assoc Disord*. 2004;18:270.
17. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12:1001779. doi:10.1371/journal.pmed.1001779
18. Bolker BM. Linear and generalized linear mixed models. In: Fox GA, Negrete-Yankelevich S, Sosa VJ, eds. *Ecological Statistics: Contemporary theory and Application*. Oxford University Press; 2015: 309-333.

19. Wang M. Generalized estimating equations in longitudinal data analysis: a review and recent developments. *Adv Stat*. 2014;2014:1-11. doi:10.1155/2014/303728
20. Phillips J, Dumitrescu L, Archer DB, et al. Longitudinal GWAS identifies novel genetic variants and complex traits associated with resilience to Alzheimer's disease. *Alzheimers Dement*. 2022;18:e067816. doi:10.1002/alz.067816
21. Yuan M, Xu XS, Yang Y, et al. SCEBE: an efficient and scalable algorithm for genome-wide association studies on longitudinal outcomes with mixed-effects modeling. *Brief Bioinform*. 2021;22:bbaa130. doi:10.1093/bib/bbaa130
22. Lee E, Giovanello KS, Saykin AJ, et al. Single-nucleotide polymorphisms are associated with cognitive decline at Alzheimer's disease conversion within mild cognitive impairment patients. *Alzheimers Dement*. 2017;8:86. doi:10.1016/J.DADM.2017.04.004
23. Sherva R, Gross A, Mukherjee S, et al. Genome-wide association study of rate of cognitive decline in Alzheimer's disease patients identifies novel genes and pathways. *Alzheimers Dement*. 2020;16:1134-1145. doi:10.1002/alz.12106
24. Wu W, Wang Z, Xu K, et al. Retrospective association analysis of longitudinal binary traits identifies important loci and pathways in cocaine use. *Genetics*. 2019;213:1225-1236. doi:10.1534/genetics.119.302598
25. Li Y, Xu I, Liu C. Functional data modeling and hypothesis testing for longitudinal Alzheimer genome-wide association studies. In: Zhao Y, Chen DG, eds. *Modern Statistical Methods for Health Research. Emerging Topics in Statistics and Biostatistics*. Springer; 2021: 353-379. doi:10.1007/978-3-030-72437-5_16
26. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48:1284-1287. doi:10.1038/NG.3656
27. Chen H, Wang C, Conomos MP, et al. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *Am J Hum Genet*. 2016;98:653-666. doi:10.1016/j.ajhg.2016.02.012
28. Zhang D. Generalized linear mixed models with varying coefficients for longitudinal data. *Biometrics*. 2004;60:8-15. doi:10.1111/J.0006-341X.2004.00165.X
29. Beekly D, Schwabe-Fry K, Bollenbeck M, et al. [P1–405]: The National Alzheimer's Coordinating Center: development of the MRI, PET and CSF biomarker databases. *Alzheimer's & Dementia*. 2017;13:P429-P429. doi:10.1016/j.jalz.2017.06.421
30. Besser L, Kukull W, Knopman DS, et al. Version 3 of the National Alzheimer's Coordinating Center's uniform data set. *Alzheimer Dis Assoc Disord*. 2018;32(4):351-358.
31. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Methodol*. 1995;57:289-300.
32. Le Guen Y, Belloy ME, Napolioni V, et al. A novel age-informed approach for genetic association analysis in Alzheimer's disease. *Alzheimers Res Ther*. 2021;13:72. doi:10.1186/s13195-021-00808-5
33. Bellou E, Baker E, Leonenko G, et al. Age-dependent effect of APOE and polygenic component on Alzheimer's disease. *Neurobiol Aging*. 2020;93:69-77. doi:10.1016/j.neurobiolaging.2020.04.024
34. 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;14:1-17. doi:10.1371/journal.pmed.1002258
35. Shukla S, Srividya K, Nazir A. Not a piece of junk anymore: pseudogene T04B2.1 performs non-conventional regulatory role and modulates aggregation of α- synuclein and β-amyloid proteins in C. elegans. *Biochem Biophys Res Commun*. 2021;539:8-14. doi:10.1016/J.BBRC.2020.12.029
36. Li D, Zhang J, Li X, Chen Y, Yu F, Liu Q. Insights into lncRNAs in Alzheimer's disease mechanisms. *RNA Biol*. 2021;18:1-11. doi:10.1080/15476286.2020.1788848
37. Morris JC, Roe CM, Xiong C, et al. APOE predicts amyloid-beta but not tau Alzheimer pathology in cognitively normal aging. *Ann Neurol*. 2010;67:122-131. doi:10.1002/ANA.21843

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: Zhuang X, Xu G, Amei A, et al. Detecting time-varying genetic effects in Alzheimer's disease using a longitudinal genome-wide association studies model. *Alzheimer's Dement*. 2024;16:e12597. https://doi.org/10.1002/dad2.12597

APPENDIX

Collaborators

Data collection and sharing for this project were 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.

The NACC database is funded by NIA/NIH Grant U24 AG072122. NACC data are contributed by the NIA-funded ADRCs: P30 AG062429 (PI James Brewer, MD, PhD), P30 AG066468 (PI Oscar Lopez, MD), P30 AG062421 (PI Bradley Hyman, MD, PhD), P30 AG066509 (PI Thomas Grabowski, MD), P30 AG066514 (PI Mary Sano, PhD), P30

AG066530 (PI Helena Chui, MD), P30 AG066507 (PI Marilyn Albert, PhD), P30 AG066444 (PI John Morris, MD), P30 AG066518 (PI Jeffrey Kaye, MD), P30 AG066512 (PI Thomas Wisniewski, MD), P30 AG066462 (PI Scott Small, MD), P30 AG072979 (PI David Wolk, MD), P30 AG072972 (PI Charles DeCarli, MD), P30 AG072976 (PI Andrew Saykin, PsyD), P30 AG072975 (PI David Bennett, MD), P30 AG072978 (PI Neil Kowall, MD), P30 AG072977 (PI Robert Vassar, PhD), P30 AG066519 (PI Frank LaFerla, PhD), P30 AG062677 (PI Ronald Petersen, MD, PhD), P30 AG079280 (PI Eric Reiman, MD), P30 AG062422 (PI Gil Rabinovici, MD), P30 AG066511 (PI Allan Levey, MD, PhD), P30 AG072946 (PI Linda Van Eldik, PhD), P30 AG062715 (PI Sanjay Asthana, MD, FRCP), P30 AG072973 (PI Russell Swerdlow, MD), P30 AG066506 (PI Todd Golde, MD, PhD), P30 AG066508 (PI Stephen Strittmatter, MD, PhD), P30 AG066515 (PI Victor Henderson, MD, MS), P30 AG072947 (PI Suzanne Craft, PhD), P30 AG072931 (PI Henry Paulson, MD, PhD), P30 AG066546 (PI Sudha Seshadri, MD), P20 AG068024 (PI Erik Roberson, MD, PhD), P20 AG068053 (PI Justin Miller, PhD), P20 AG068077 (PI Gary Rosenberg, MD), P20 AG068082 (PI Angela Jefferson, PhD), P30 AG072958 (PI Heather Whitson, MD), P30 AG072959 (PI James Leverenz, MD).