Received: 23 March 2022 | Revised: 3 March 2023 | Accepted: 8 March 2023

DOI: 10.1002/alz.13069

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

RESEARCH ARTICLE

Alzheimer's disease heterogeneity explained by polygenic risk scores derived from brain transcriptomic profiles

Jaeyoon Chung[1] | Nathan Sahelijo[1] | Toru Maruyama[2] | Junming Hu[1] | Rebecca Panitch[1] | Weiming Xia[3,4] | Jesse Mez[5] | Thor D. Stein[4,6,7] | the Alzheimer's Disease Neuroimaging Initiative[1] | Andrew J. Saykin[8,9] | Haruko Takeyama[2,10,11,12] | Lindsay A. Farrer[1,5,13,14,15] | Paul K. Crane[16] | Kwangsik Nho[8,9] | Gyungah R. Jun[1,13,14]

[1]Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, Massachusetts, USA

[2]Department of Life Science and Medical Bioscience, Waseda University, Tokyo, Japan

[3]Department of Pharmacology & Experimental Therapeutics, Boston University School of Medicine, Boston, Massachusetts, USA

[4]Department of Veterans Affairs Medical Center, Bedford, Massachusetts, USA

[5]Department of Neurology, Boston University School of Medicine, Boston, Massachusetts, USA

[6]Department of Pathology & Laboratory Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

[7]Boston VA Healthcare Center, Boston, Massachusetts, USA

[8]Department of Radiology and Imaging Sciences and Indiana Alzheimer's Disease Research Center, Indiana University School of Medicine, Indianapolis, Indiana, USA

[9]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, Indiana, USA

[10]Computational Bio Big-Data Open Innovation Laboratory, AIST-Waseda University, Tokyo, Japan

[11]Research Organization for Nano and Life Innovations, Waseda University, Tokyo, Japan

[12]Institute for Advanced Research of Biosystem Dynamics, Waseda Research Institute for Science and Engineering, Graduate School of Advanced Science and Engineering, Waseda University, Tokyo, Japan

[13]Department of Ophthalmology, Boston University School of Medicine, Boston, Massachusetts, USA

[14]Departments of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA

[15]Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, USA

[16]Department of Medicine, University of Washington, Seattle, Washington, USA

Correspondence
Gyungah R. Jun, Department of Medicine (Biomedical Genetics), Boston University School of Medicine, 72 East Concord Street, Boston, MA 02118, USA.
E-mail: gyungah@bu.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

Abstract

Introduction: Alzheimer's disease (AD) is heterogeneous, both clinically and neuropathologically. We investigated whether polygenic risk scores (PRSs) integrated with transcriptome profiles from AD brains can explain AD clinical heterogeneity.

Methods: We conducted co-expression network analysis and identified gene sets (modules) that were preserved in three AD transcriptome datasets and associated with AD-related neuropathological traits including neuritic plaques (NPs) and neurofibrillary tangles (NFTs). We computed the module-based PRSs (mbPRSs) for each module

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
NIA, Grant/Award Numbers: U01-AG068057, U19-AG068753, P30-AG072978, RF1-AG057519, R56-AG069130

and tested associations with mbPRSs for cognitive test scores, cognitively defined AD subgroups, and brain imaging data.

Results: Of the modules significantly associated with NPs and/or NFTs, the mbPRSs from two modules (M6 and M9) showed distinct associations with language and visuospatial functioning, respectively. They matched clinical subtypes and brain atrophy at specific regions.

Discussion: Our findings demonstrate that polygenic profiling based on co-expressed gene sets can explain heterogeneity in AD patients, enabling genetically informed patient stratification and precision medicine in AD.

KEYWORDS
Alzheimer's disease, co-expression network, cognitive performance, genetic subtyping, module-based polygenic risk score, patient stratification, precision medicine

HIGHLIGHTS

- Co-expression gene-network analysis in Alzheimer's disease (AD) brains identified gene sets (modules) associated with AD heterogeneity.
- AD-associated modules were selected when genes in each module were enriched for neuritic plaques and neurofibrillary tangles.
- Polygenic risk scores from two selected modules were linked to the matching cognitively defined AD subgroups (language and visuospatial subgroups).
- Polygenic risk scores from the two modules were associated with cognitive performance in language and visuospatial domains and the associations were confirmed in regional-specific brain atrophy data.

1 | BACKGROUND

Late onset Alzheimer's disease (AD) is a complex disorder with clinical and neuropathological heterogeneity.[1,2] Types of clinical heterogeneity include progression rate, predominant cognitive symptoms, and whether psychotic symptoms manifest.[1] AD neuropathology can also be varied with complications of other neuropathological traits beyond plaques and tangles.[1,2] Clinical and neuropathological heterogeneity may have contributed to the repeated failure of AD clinical trials.[3] Classification of heterogeneous AD patients into biologically relevant subgroups may improve our understanding of biological mechanisms underlying the variability of cognitive symptoms and trajectories of decline, as well as lead to development of subgroup-specific treatment options.[4]

Different AD subtypes have been previously proposed based on neuropsychological and neuropathological characteristics,[5-7] domain-specific cognitive functions, magnetic resonance imaging (MRI) brain imaging data,[4] and metabolic profiling.[8] However, our understanding of molecular mechanisms underlying disease heterogeneity is still limited. A recent report illustrates that genetic variants with large effect sizes can distinguish six cognitively defined subgroups of AD compared to elderly controls.[9] A previous study showed that polygenic risk scores (PRSs) derived from clusters (i.e., gene sets) in genome-wide association studies (GWASs) of type 2 diabetes (T2D)–related phenotypes have successfully classified T2D patients into different subtypes.[10] These studies demonstrate that PRSs from biologically connected gene sets may explain disease heterogeneity and improve scientific understanding of biological mechanisms underlying disease subtypes. In addition, co-expression network analyses have been shown to be useful for identifying biologically connected and disease-relevant gene sets using transcriptome data.[11,12] Taken together, these findings led to the hypothesis that network analysis using transcriptome data of AD brains could capture biologically relevant gene sets responsible for distinct disease subtypes and PRSs derived from the gene sets could explain clinical heterogeneity of AD.

In this study, we identified modules (sets of biologically relevant genes) by co-expression analysis and thereby generated module-based PRSs of AD patients. Then, using domain-specific cognitive functions, previously defined AD cognitive subtypes, and brain imaging data, we evaluated whether the module-based PRSs can explain cognitive impairment heterogeneity among the AD patients.

2 | METHODS

2.1 | Sources of RNA sequencing data in autopsied AD brains for network analysis

Co-expression network analysis was performed using previously generated gene expression data from the dorsolateral prefrontal cortex (DLPFC) area of 65 autopsy-confirmed non-Hispanic White AD cases from the Framingham Heart Study and Boston University Alzheimer's Disease Research Center (FHS/BUADRC).[13] Details of procedures for quality control (QC) of RNA sequencing (RNA-Seq) data and neuropathological AD diagnosis are presented in supporting information and previously reported elsewhere.[13] Additional RNA-Seq datasets for validation were obtained from the CommonMind portal (http://www.synapse.org) including post-QC normalized gene expression data (version #1) from the DLPFC area of 363 neuropathologically confirmed AD cases in the Religious Orders Study and Rush Memory and Aging Project (ROSMAP)[14] and from the temporal cortex area of 82 autopsy-confirmed AD cases in the Mayo Clinic Study of Aging (MAYO).[15].

2.2 | Identifying preserved and AD-associated modules

Co-expression gene sets (i.e., modules) were generated using the transcriptome data from the 65 AD brains in FHS/BUADRC using the weighted gene co-expression network analysis (WGCNA) approach, which computes pairwise correlations for all gene pairs and clusters genes by the correlated expression levels.[16] Transcriptome data of AD-free controls were not included in our co-expression study because our interest is to identify gene sets related to the disease heterogeneity, not the disease risk (e.g., AD cases versus controls). Details of co-expression module construction are presented in supporting information and previously described.[17] Preservation of the discovery modules was evaluated in the two independent validation datasets, including ROSMAP and MAYO datasets, using z summary statistics.[16] We considered a module to be preserved if z summary scores were > 5.0 in both validation datasets.[16] Among the preserved modules, we selected AD-associated modules by enrichment analyses using gene sets for AD-related neuropathological traits[18] including neuritic plaques (NP) and neurofibrillary tangles (NFT), as well as AD risk.[19] We used AD-associated genes for enrichment analyses that contained at least one single nucleotide polymorphism (SNP) with $P < 10^{-3}$ located within ±20 kb from the gene associated with one of the AD phenotypes (NP, NFT, or AD risk). We selected significant enrichment P-values < 0.05 using Fisher's exact test after false discovery rate (FDR) correction. Based on the result from the enrichment analysis for each module, we assigned the AD phenotypes (NP, NFT, or AD risk) for which the module was most significantly enriched and used it to calculate module-based PRSs. The selected AD-associated modules were considered to generate module-based PRSs.

RESEARCH IN CONTEXT

1. **Systematic Review**: We reviewed the literature using traditional (e.g., PubMed) as well as preprinted (e.g., medRxiv) sources on studies about Alzheimer's disease (AD) heterogeneity using genetic information.
2. **Interpretation**: Our co-expression network analysis among only AD brains without controls identified gene sets (modules) that are likely to be responsible for AD heterogeneity. The polygenic risk scores derived from the modules associated with cognitive performance for certain domains (language and visuospatial functioning) were also associated with cognitively defined AD subgroups for the matching domains and cortical thickness at the specific brain regions. These findings imply that genetics can be a useful source for dissecting the disease heterogeneity along with other resources including domain-specific cognitive measures, brain imaging scans, and neuropathological traits.
3. **Future Directions**: Follow-up analysis will repeat the analysis in large independent samples to validate our approach and findings.

We also examined expression coherence and cellular signatures of genes in each of the AD-associated modules using single cell RNA-seq data in five different cell types (astrocytes, microglia, oligodendrocytes, endothelia, and neurons) from the temporal lobe area (Gene Expression Omnibus ID: GSE67835)[20] and single nucleus RNA-seq data in seven cell types (astrocytes, microglia, oligodendrocytes, pericytes, endothelia, and excitatory/inhibitory neurons) from the prefrontal cortex in the ROSMAP.[21] Details of methods for deriving cell-type–specific gene sets and their expression profiling are presented in supporting information and reported elsewhere.[13] Enrichment of cell-type specificity for each AD-associated module was tested using Fisher's exact test.

2.3 | Genotypic and phenotypic data in ADNI

The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a longitudinal study assessing clinical, neuroimaging, genetic, and biomarker data from participants in various stages of cognitive impairment including cognitively normal (CN), mild cognitive impairment (MCI), and AD. Genetic and phenotypic data of ADNI participants were obtained from the Laboratory of Neuro Imaging (LONI) website (http://adni.loni.usc.edu). We used the ADNI genetic data for computing module-based PRSs (mbPRSs) and phenotype data for evaluating the relationships between mbPRSs and cognitive impairment heterogeneity. Genome-wide genotype data from two different arrays (ADNI-1, $n = 679$ and ADNI-GO/2, $n = 397$) were imputed using the

Haplotype Reference Consortium data. Details of QC, imputation, and population substructure procedures are described in the supporting information. Characteristics of the sample after QC are presented in Table S1 in supporting information.

Because the clinical spectrum of AD can be largely affected by impairment of specific cognitive functions,[22] we hypothesized that deficits in particular cognitive domains explain, at least in part, disease heterogeneity (i.e., cognitive impairment heterogeneity). To explore this heterogeneity, we used domain-specific cognitive tests at the last exam from the ADNI dataset (Table S1): logical memory immediate (LIMMTOTAL) and delayed (LDELTOTAL) recall tests for memory; Trail Making Test Parts A/B (TRAASCOR and TRABSCOR) for executive functioning; category fluency animal score (CATANIMSC) and Boston Naming Test total (BNTTOTAL) score for language; and clock test total score (COPYSCORE) for visuospatial functioning. Cognitive test scores were adjusted for age at last exam, sex, and education using linear regression, and the residuals derived from the regression models were ranked-transformed as previously described.[17]

2.4 | Computing and assessing PRSs for AD-associated modules in ADNI

We selected SNPs in each AD-associated module from the enrichment analysis for the assigned AD outcome and generated mbPRSs using effect estimates of the selected SNPs for NP, NFT, or AD risk from the enrichment analysis. For comparison, we also generated PRSs for NP, NFT, and AD risk in a conventional approach, which aggregates effect estimates of SNPs with $P < 0.001$ across the genome, defined as genome-wide PRS (gwPRS). Details about computing these two types of PRSs (gwPRS and mbPRS) are included in the supporting information. We excluded modules with low mean and standard error and/or skewed distributions.

After generating those PRSs in ADNI, we evaluated correlations among the mbPRSs and gwPRSs. To assess relevance of those PRSs to disease stages/progression, we stratified ADNI subjects by disease stages (CN, MCI, and AD) at the last exam and compared mean values of PRSs between different disease stages. We also tested associations between PRSs and conversion status for AD progression (e.g., CN to MCI or AD; MCI to AD) excluding AD at baseline using logistic regression models after adjusting for age, sex, the first four principal components (PCs) and the array information.

Next, we conducted association tests with mbPRSs or gwPRSs for specific cognitive domains using rank-transformed cognitive test scores as quantitative outcomes in linear regression models after adjusting the first four PCs and genotype platform as covariates. We followed up the nominally significant modules ($P < 0.05$) with domain-specific cognitive test scores as cognitive impairment heterogeneity (CIH) modules.

We also attempted to replicate the associations between mbPRSs of the selected CIH modules and domain-specific cognitive test scores among 134 AD cases in FHS (dbGaP Study Accession ID: phs000056.v5.p3). Details of sample characteristics, imputation, computation of mbPRSs, and association tests with cognitive test scores in Neuropsychological Test Battery in FHS are described in the supporting information.

2.5 | Validating CIH modules with cognitively defined AD subtypes in ADNI

Previously, 672 AD cases in ADNI have been classified into cognitively defined subtypes based on relative impairments at the time of AD diagnosis,[9] consisting of 196 as AD-Memory, 16 as AD-Executive functioning, 52 as AD-Language, 91 as AD-Visuospatial functioning, and 317 other domains (Table S2 in supporting infomation). Details about these cognitively defined subtypes are described in supporting information and reported elsewhere.[9] We evaluated whether mbPRSs of CIH modules are linked into one of the four cognitively defined subgroups (AD-Memory, AD-Executive, AD-Language, and AD-Visuospatial domains). Each subject was assigned into membership of one subgroup coded as 1, and otherwise coded as 0 excluding subjects with overlapping memberships. We tested the association between mbPRSs and a dichotomized membership of cognitively defined subtypes in a logistic regression model adjusting for age, sex, the first four PCs, and genotype platform as covariates.

2.6 | Brain imaging (MRI) data analysis with mbPRSs of the CIH modules in ADNI

To understand the relationships between our CIH modules and brain atrophy at specific brain regions, we tested the association between mbPRSs and surface-based cortical thickness of AD patients using general linear models after adjusting age, sex, magnetic field strength, and intracranial volume as covariates.[23] Detailed information about brain imaging data processing for surface-based measure of cortical thickness in ADNI are described elsewhere.[23]

2.7 | Biological functions of genes in the CIH modules

Gene Ontology (GO) analyses were conducted to discern biological pathways of AD-associated genes in CIH modules using Ingenuity Pathway Analysis software (QIAGEN). We also looked up associations between the CIH module-genes and AD-related neuropathological traits including Consortium to Establish a Registry for Alzheimer's Disease (CERAD) score and Braak stage, and quantitative measures of proteins including amyloid beta ($A\beta)_{42}$, phosphorylated tau at 181 (p-tau181) and 231 (p-tau231), postsynaptic density protein 95 (PSD95), C4a, C4b, and PPP2CA/B from the prefrontal cortex area of autopsied brains (FHS/BUADRC).[13]

TABLE 1 AD associated preserved networks in AD brains

	Z summary		Enrichment *P*-value		
Module	ROSMAP	MAYO	NP	NFT	AD RISK
M1	40.56	20.71	$3.74X10^{-6}$	$3.26X10^{-9}$	0.04
M2	37.45	20.68	$3.50X10^{-8}$	$2.57X10^{-7}$	0.002
M3	35.48	18.59	$5.38X10^{-5}$	$8.36X10^{-7}$	0.01
M4	32.32	18	$3.42X10^{-2}$	$1.25X10^{-5}$	0.07
M5	29.83	15.91	$1.46X10^{-6}$	$2.92X10^{-4}$	0.40
M6	26.56	14.46	$4.70X10^{-6}$	0.09	0.09
M7	25.84	14.23	$2.65X10^{-3}$	$2.47X10^{-4}$	0.12
M8	20.52	14.07	0.47	0.02	0.20
M9	20.51	12.71	$4.34X10^{-3}$	$8.89X10^{-4}$	0.37
M10	18.25	12.64	0.02	0.16	0.16
M11	17.32	11.81	$8.70X10^{-4}$	$7.48X10^{-4}$	0.01
M12	17.06	10.55	1.00	0.05	0.95
M13	15.57	8.59	0.02	0.03	0.81
M14	14.79	7.81	0.06	0.04	0.48

Note: Module is a co-expressed gene network in the discovery from the Framingham Heart Study and Boston University Alzheimer's Disease Center (FHS/BUADRC) study. *Z* summary is a network preservation score of a module between the discovery and two validation datasets, the Religious Orders Study and Rush Memory and Aging Project (ROSMAP) and the Mayo Clinic Study of Aging (MAYO).
Enrichment *P*-values for a module were computed using genes in the given module containing a SNP with $P < 10^{-3}$ from Beecham et al. for neuritic plaque (NP) and neurofibrillary tangles (NFT)[18] and from Kunkle et al. for Alzheimer's disease (AD) risk.[19] Fourteen modules were selected when a module in the FHS/BUADRC study was preserved with *Z* summary > 5 in both validation datasets, ROSMAP and MAYO, and significant at *P*-value < 0.05 with at least one of the gene sets for NP, NFT, or AD risk.

3 | RESULTS

3.1 | AD-associated modules in AD brains were preserved in independent studies

The overall study design including module selection process is provided in Figure S1 in supporting information. Eighty-three modules were identified in the discovery dataset (FHS/BUADRC), and 29 of these modules were preserved in the two validation datasets (Figure 1A). Fourteen of the 29 preserved modules (M1–M14) contained genes that were significantly enriched in at least one of the AD gene sets (NP, NFT, or AD risk) with FDR < 0.05 (Figure 1B and Table 1). Interestingly, only four modules (M1–M3 and M11) were nominally enriched in the AD risk gene set, and all 14 modules were at least three orders of magnitude more significantly enriched in either NP or NFT gene sets (Table 1). These findings may imply that our modules derived from the transcriptome datasets of AD brains (without AD-free controls) would capture the gene sets for underlying changes in AD pathology, rather than the overall disease risk. Therefore, we selected one outcome, either NP or NFT, but not AD risk, to compute fourteen mbPRSs (NP-linked modules: M2, M5, M6, M10, and M13; NFT-linked modules: M1, M3, M4, M7–M9, M11, M12, and M14), according to its most significantly enriched gene set and the GWAS summary data of the selected outcome (NP or NFT).

All 14 AD-associated modules were significantly enriched in specific cell types, and these results were consistent between temporal lobe and prefrontal cortex regions (Figure 1B and Table S3 in supporting information). M1 to M4 modules were predominantly enriched in excitatory neurons (best *P* with M2 from prefrontal cortex = 4.4×10^{-188}), M6 and M7 in astrocytes (best *P* with M7 from prefrontal cortex = 1.3×10^{-97}), M10 in endothelia (*P* from temporal lobe = 9.9×10^{87}), and M11 in microglia (*P* from temporal cortex = 1.7×10^{-129}). The other five modules (M5, M8, M9, M12, and M13) were significantly enriched in more than one cell type, while M5 and M8 (astrocytes), M9 (endothelia), M12 (neurons), and M13 (microglia) were significantly enriched in at least one cell type in both brain regions.

3.2 | Module-based PRSs explained heterogeneity in cognitive functions among AD patients

Of 14 preserved and AD associated modules, we computed mbPRSs for nine NFT-linked modules (M1, M3, M4, M7–M9, M11, M12, and M14) using NFT, while computing mbPRSs of the remaining five NP-linked modules (M2, M5, M6, M10, and M13) using NP. Three modules, including M2, M13, and M14, were excluded due to their low standard error (<0.05) and/or their extremely skewed distributions for the following analyses (Table S4 in supporting information), leading to 11 modules for further evaluation.

In comparison, three gwPRSs were significantly correlated with mbPRSs of the M3, M11, and M12 modules (correlation $r^2 \geq 0.1$), while the rest of the eight mbPRSs were not correlated with those for gwPRSs ($r^2 < 0.01$; Figure S2 in supporting information). The mean values of all three gwPRSs were sequentially increased from CN to MCI and AD (Figure 1C). In contrast, the mean values of mbPRSs were varied across the disease stages, and the mean values of modules M3–M6 and M10 were smaller in MCI or AD stages than in the CN stage (Figure 1C). For the disease progression, two gwPRSs (NFT and AD risk) and three mbPRSs (M3, M11, and M12) were significantly associated with the progressions from CN to both MCI and AD (Figure 1D). None of the PRSs were associated with the progression from MCI to AD. Interestingly, NFT-gwPRS and M9-mbPRS were associated with the progression from CN to MCI, while M6-mbPRS was associated with the progression from CN to AD. All three gwPRSs were significantly associated with most cognitive test scores, except for the visuospatial domain (COPYSCORE), with consistent effect directions across cognitive tests. This indicates that the gwPRSs are not likely to differentiate cognitive impairment heterogeneity among AD cases (Figure 1E).

Five mbPRSs, for M3, M6, M9, M11, and M12, were robustly associated (all tests in domains with *P*-value < 0.05), while two mbPRSs for M7 and M10 were nominally associated with only one cognitive test in the domain (Figure 1E and Table S5 in supporting information). Four mbPRSs, for M1, M4, M5, and M8, showed no association

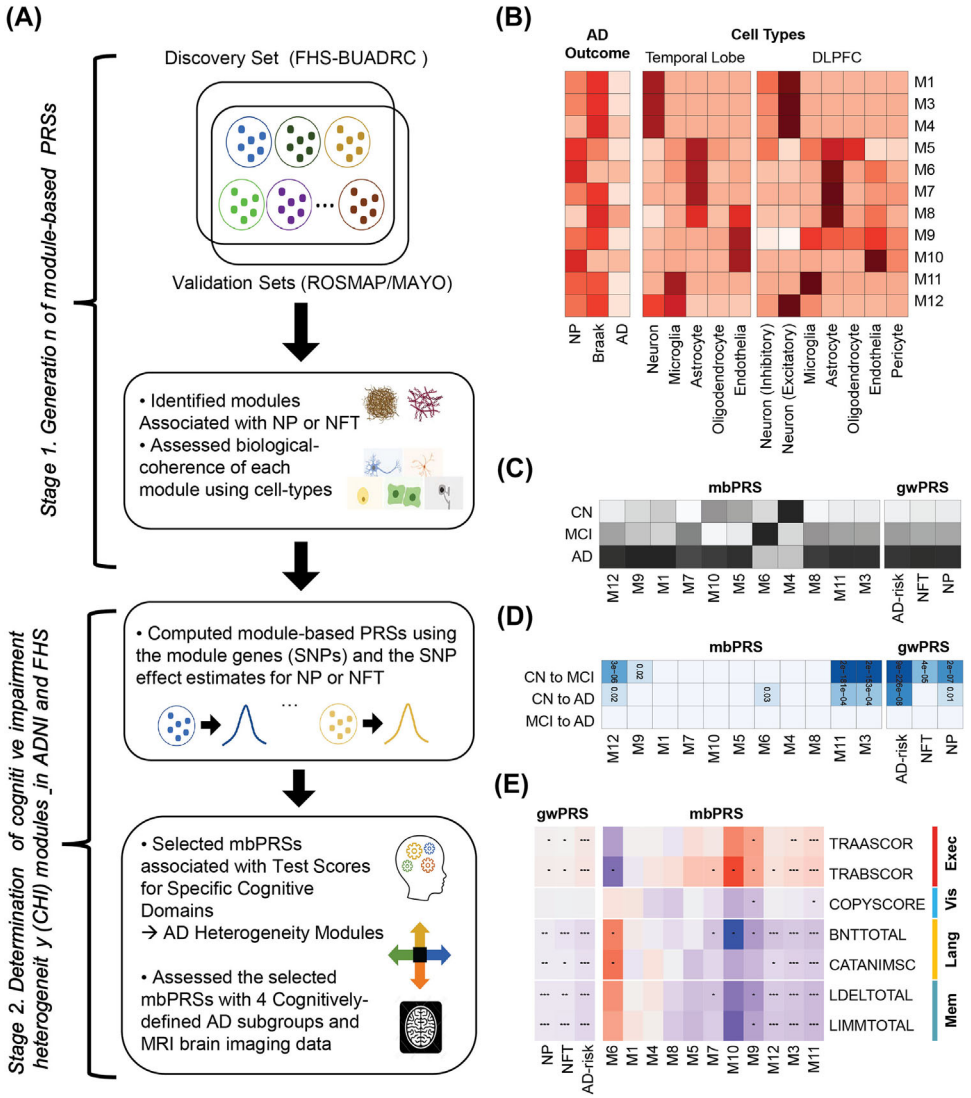

FIGURE 1 **A**, Schematic of our study design. We constructed co-expression modules (sets of genes), selected AD-associated modules, and generated module-based polygenic risk scores for explaining the AD heterogeneity, which were tested and evaluated with gene sets for AD-related neuropathological traits (NP and NFT), human brain cell-types, cognitive test scores, cognitively defined AD subgroups, and brain MRI imaging data. **B**, Enrichment analysis. The strength of enrichment results with the 11 AD-associated modules for AD phenotypes (NP, NFT, and AD risk; left) and cell type–specific gene sets in temporal lobe (middle) and DLPFC (right). The darker color indicates the more significant enrichment *P*-value. M2, M13, and M14 modules were excluded since these modules were not followed up by further analyses. **C**, Heatmap of mean values for mbPRSs and gwPRSs across the disease stages including CN, MCI, and AD (the darker color indicates the larger mean value of the PRS). **D**, Associations between PRSs (mbPRSs and gwPRSs) and disease progression (CN to MCI, CN to AD, MCI to AD). The darker color indicates the more significant *P*-value. **E**, Associations between PRSs (mbPRSs and gwPRSs) and seven test scores for four cognitive domains (executive functioning, visuospatial functioning, language, and memory). Red and blue color represents positive and negative effect direction, respectively. The number of asterisks (*) in the cells indicate the strength of associations. AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; CIH, cognitive impairment heterogeneity; CN, cognitively normal; DLPFC, dorsolateral prefrontal cortex; FHS, Framingham Heart Study; gwPRS, genome-wide polygenic risk score; mbPRS, module-based polygenic risk score; MCI, mild cognitive impairment; MRI, magnetic resonance imaging; NFT, neurofibrillary tangle; NP, neuritic plaque; PRS, polygenic risk score; SNP, single nucleotide polymorphism

(*P*-value > 0.05) with any cognitive test scores (Figure 1E and Table S5). Of the five mbPRSs with robust associations for cognitive domains, mbPRSs for M3 and M11 were strongly associated with the three cognitive domains except for the visuospatial functioning, indicating that mbPRSs from M3 and M11 did not differentiate cognitive impairment heterogeneity. The M6-mbPRS was nominally associated with all two language-domain test scores (BNTTOTAL *P*-value = 0.03 and CATANIMSC *P*-value = 0.01). The M9-mbPRS was associated with all tests in memory and executive function domains ($P < 0.05$), as well as with visuospatial functioning domain (COPYSCORE *P*-value = 0.05). The M12-mbPRS was strongly associated with language (best *P* with BNTTOTAL = 2.2×10^{-6}) and memory (best *P* with LIMMTOTAL = 4.4×10^{-6}) domains (Table 2). Therefore, we prioritized M6, M9, and M12 as CIH modules and attempted to validate the associations between mbPRSs

TABLE 2 Associations between cognitive test scores and three module-based PRSs

		M6			M9			M12		
Cognitive domain	**Cognitive test**	**BETA**	**SE**	***P*-value**	**BETA**	**SE**	***P*-value**	**BETA**	**SE**	***P*-value**
Executive functioning	TRAASCOR	−0.16	0.12	0.17	0.18	0.06	1.08×10^{-3}	0.02	0.02	0.15
	TRABSCOR	−0.26	0.12	0.03	0.18	0.06	1.58×10^{-3}	0.03	0.02	0.05
Visuospatial functioning	COPYSCORE	0.03	0.12	0.79	−0.12	0.06	0.049	−0.01	0.02	0.77
Language	BNTTOTAL	0.26	0.12	0.03	−0.17	0.06	2.57×10^{-3}	−0.08	0.02	2.23×10^{-6}
	CATAANIMSC	0.29	0.12	0.01	−0.11	0.06	0.05	−0.04	0.02	0.01
Memory	LDELTOTAL	0.22	0.12	0.06	−0.18	0.06	1.14×10^{-3}	−0.07	0.02	4.55×10^{-5}
	LIMMTOTAL	0.18	0.12	0.12	−0.15	0.06	7.18×10^{-3}	−0.08	0.02	4.42×10^{-6}

Note: LIMMTOTAL and LDELTOTAL: logical memory immediate and delayed recall tests; TRAASCOR and TRABSCOR: Trail Making Test Parts A/B. CATANIMSC: category fluency animal score; BNTTOTAL: Boston Naming Test total; COPYSCORE: clock test total score. Beta estimates (BETA), standard error (SE), and *P*-values were calculated with module-based polygenic risk scores of the M2, M6, and M9 modules for domain-specific cognitive functions.

TABLE 3 Membership assignment of three module-based PRSs in previously defined cognitive subtypes of AD in ADNI

Source of PRS	Executive functioning ST		Language ST		Memory ST		Visuospatial functioning ST	
	OR (95% CI)	***P***	**OR (95% CI)**	***P***	**OR (95% CI)**	***P***	**OR (95% CI)**	***P***
Genome-wide PRS (gwPRS)								
NP	0.99 (0.93–1.0)	0.69	0.99 (0.95–1.0)	0.43	1.00 (0.99–1.00)	0.57	1.00 (0.97–1.00)	0.75
NFT	1.00 (0.91–1.10)	0.98	0.99 (0.93–1.01)	0.74	**1.01 (1.00–1.11)**	**0.04**	0.96 (0.91–1.01)	0.09
AD risk	1.01 (0.93–1.10)	0.65	0.97 (0.91–1.00)	0.24	1.00 (1.00–1.10)	0.08	0.96 (0.91–1.00)	0.07
Module-based PRS (mbPRSs)								
M6	0.37 (0.04–3.60)	0.39	**5.50 (1.50–20.0)**	**0.009**	0.97 (0.44–2.20)	0.95	0.88 (0.28–2.80)	0.83
M9	0.95 (0.34–2.60)	0.92	0.81 (0.42–1.50)	0.51	1.20 (0.84–1.80)	0.29	**1.91 (1.11–3.31)**	**0.04**
M12	1.10 (0.78–1.50)	0.66	0.92 (0.75–1.10)	0.40	1.00 (0.92–1.20)	0.56	1.00 (0.87–1.20)	0.76

Note: Cognitively defined AD subtypes (ST) in ADNI have been previously defined.[9] Bold values with significance at $P < 0.01$ and OR > 1.0 indicate these modules are likely members of the corresponding cognitively defined subtypes.
Abbreviations: AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; CI, confidence interval; NP, neuritic plaques; NFT, neurofibrillary tangles; OR, odds ratio; PRS, polygenic risk score.

of the CIH modules (M6, M9, and M12) and the cognitive test scores among the AD cases in FHS (Table S6 in supporting information). We replicated nominally significant associations ($P < 0.05$) between the M6-mbPRS and two language-domain cognitive test scores in AD cases from the FHS (BNT30 *P*-value = 0.03 and BNT30cue *P*-value = 0.03; Table S7 in supporting information). Although we did not find associations of two modules (M9 and M12) with the cognitive test scores in FHS, three CIH modules (M6, M9, and M12) were further tested with cognitively defined subgroups[9] and brain atrophy among AD patients.

3.3 | Module-based PRS associations with cognitively defined AD subtypes and brain atrophy

Of the three CIH mbPRSs, the mbPRSs of M6 and M9 showed nominal associations at $P < 0.05$ with odds ratio (OR) > 1.0 with previously defined cognitive subtypes for AD-Language (OR = 5.5; $P = 0.01$) and AD-Visuospatial functioning (OR = 1.9; $P = 0.04$), respectively (Table 3). The M12-mbPRS was associated with none of the cognitive subtypes (Table 3). In contrast, the gwPRSs for NP and AD risk failed to differentiate any of the subgroups, while only NFT-gwPRS was nominally associated with the AD-Memory subtype (OR = 1.01; $P = 0.04$).

The mbPRSs of M6 and M9 were significantly associated with cortical thickness at specific brain locations (M6: bilateral frontal, parietal, and temporal lobes; M9: bilateral frontal lobes; Figure 2A). Particularly, the brain atrophy for the M6-mbPRS was localized at the Wernicke area where lesions have been associated with severe impairments of word comprehension.[24]

3.4 | Functional profiling of M6 and M9

Among the genes in M6 and M9, we focused on the GWAS genes (i.e., seed genes) containing a SNP with $P < 0.001$ for NP or NFT (# of GWAS

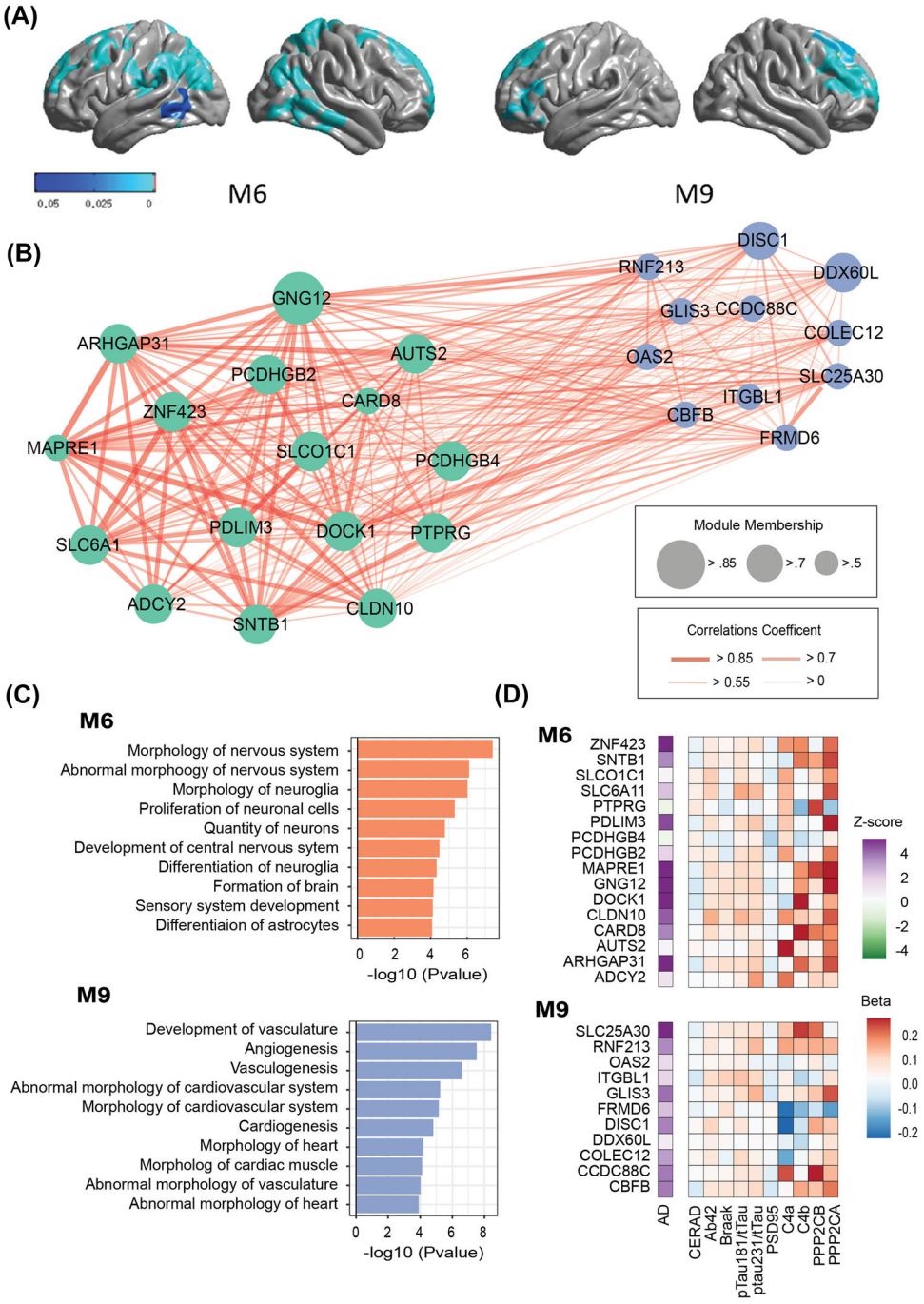

FIGURE 2 M6 and M9 were selected as cognitive impairment heterogeneity (CIH) modules. **A**, Association between cerebral cortical thickness and module-based polygenic risk scores for M6 and M9. *P*-value map with threshold at $P < 0.05$ indicated that the darker blue color showed more significant *P*-value. **B**, Co-expression network of genes in M6 and M9. **C**, Pathways enriched for M6 and M9. **D**, Associations of the expression levels of genes in M6 and M9 with AD and AD-related neuropathological traits (previously published[13]). AD. Alzheimer's disease

genes, M6 = 16 and M9 = 11; Figure 2B and Table S8 in supporting information). The seed genes in M6 were significantly enriched in pathways (Figure 2C and Table S9 in supporting information) including morphology of nervous system ($P = 4.0 \times 10^{-8}$), abnormal morphology of nervous system ($P = 7.8 \times 10^{-7}$) and differentiation of astrocytes ($P = 8.4 \times 10^{-5}$), while the M9 seed genes were enriched in pathways including vascular system including development of vasculature ($P = 3.8 \times 10^{-9}$), angiogenesis ($P = 3.0 \times 10^{-8}$), and vasculogenesis ($P = 2.5 \times 10^{-7}$; Figure 2C and Table S9).

According to our previous report,[13] the majority of the seed genes in M6 (best: *DOCK1*, $P = 3.0 \times 10^{-7}$) and M9 (best: *SLC25A30*, $P = 6.3 \times 10^{-6}$) were upregulated in AD compared to control brains (Table S10 in supporting information). In addition, we observed significant associations between expression levels of the seed genes in

M6 and M9 and AD-related protein levels (Figure 2D and Table S11 in supporting information). The seed genes in M6 were significantly associated ($P < 0.05$) with CERAD scores, Braak stages, Aβ_{42}, p-tau$_{181}$/total tau (t-tau) ratio, p-tau$_{231}$/t-tau ratio, C4a, C4b, and PSD95 (Table S11), with the most significant association observed with expression of *ADCY2* with p-tau$_{181}$/t-tau ratio (P-value = 1.1×10^{-3}). The expressions of the seed genes in M9 were nominally associated with Braak stages, Aβ_{42}, p-tau$_{231}$/t-tau ratio, and C4a levels with the best P-value between expression of *DISC1* and C4a (P-value = 1.0×10^{-3}).

4 | DISCUSSION

4.1 | Key findings

The goal of this study was to identify gene sets responsible for the biological mechanisms underlying AD heterogeneity. We generated modules (gene sets) that were commonly observed in multiple transcriptome datasets of AD brains. We closely evaluated biological coherence and disease relevance of networks of genes (modules) using profiling of human brain cell types and genetics of AD neuropathology. Then, we selected the CIH modules (M6, M9, and M12) that are likely to explain the disease heterogeneity in cognitive impairment of the AD patients, by testing with domain-specific cognitive test scores in ADNI (clinic-based study) and in FHS (population-based study). We identified and validated two CIH modules (M6 and M9) that showed significant associations for language and visuospatial domains with matching cognitive AD subtypes (AD-Language and AD-Visuospatial), respectively. These results were further linked to atrophy in specific brain areas (M6: Wernicke's area in temporoparietal cortex; M9: frontal cortex), which were previously reported to underpin language comprehension[24,25] and visuospatial deficit.[26] This study demonstrated the novel concept that can be generalizable and applicable to diverse populations, although not all the modules are available in all populations. The process and approach used in this study indicate that polygenic risk profiling in co-regulated and biologically connected genes provide unique and distinct frameworks to explain AD heterogeneity.

4.2 | Advantage of mbPRSs for AD subgrouping

The three gwPRSs for NP, NFT, and AD risk showed high correlations with each other and largely similar patterns from associations with disease conversion and cognitive test scores. In contrast, the mbPRSs showed almost no correlations with each other and were associated with the performance of specific cognitive domains. These findings, after comparing our novel mbPRSs to conventional gwPRSs, demonstrate that mbPRSs can be more useful for explaining the clinical heterogeneity in AD patients, while gwPRSs (i.e., traditional PRS) can be more relevant to predict overall disease risk. Our mbPRSs successfully distinguished differences in clinical (cognitive domains) and structural brain imaging patterns, indicating representation of different disease mechanisms and thereby would be effective tools for dissecting disease heterogeneity. The gwPRSs for NP and AD risk failed to recognize the AD subgroups. Only the gwPRS for NFT discerned the most typical cognitive subgroup, AD-Memory domain. In contrast, newly identified mbPRSs for M6 and M9 modules recognized different types of AD subgroups. This indicates that the conventional gwPRS approach is less likely to recognize differences among AD subtypes. Further, these results support our hypothesis that subgrouping genetic markers from gene sets responsible for a distinct disease mechanism leading to an AD subtype is important for precision medicine and genome-guided clinical trials.

There have been huge efforts to improve prediction and distinguish disease subtypes using polygenic profiling for early detection of subjects at risk.[27–29] Polygenic risk scores can be useful to predict disease development or treatment responses in particular patient subgroups.[30] Our module-based polygenic profiling has innovative features compared to those previously conducted co-expression studies[11,12] and conventional PRS approaches for AD.[29–32] First, our co-expression modules were developed from only AD brains excluding CN and MCI brains, while previous co-expression studies used transcriptome data of AD cases together with controls.[11,12] Biological processes underlying disease heterogeneity in AD brains may be different from CN or MCI brains.[33,34] Inclusion of non-AD transcriptome data would well differentiate gene sets relevant to the disease risk but not explain disease heterogeneity. Second, previous polygenic profiling studies have generated PRSs by aggregating genetic estimates of genome-wide or most significant SNPs, which may have improved prediction rates[30] but cannot explain specific biological functions. In contrast, our mbPRSs derived from biologically coherent gene sets enable us to interpret biological functions of the modules and thereby provide insights on functional/mechanistic pathways for the AD subtypes. A previous study demonstrated that genomic annotations at the single tissue level can improve our understanding on the etiology of complex human diseases.[35] A recent simulation study with failed AD trials confirms that the main reason for failure reason is that variability between individuals in trials masks efficacy.[3] Therefore, our mbPRSs relevant to cell/tissue-level transcriptome profiles, brain imaging data, and cognitively defined subgroups can be used for studying disease subtypes, prognosis, and response of treatment.

4.3 | Role of omics and genetic profiling in AD subgrouping

Profiling using omics data including transcriptome data at tissue- or cell-level helped identify clinically and neuropathologically heterogeneous modules but also aided understanding of the biological functions of the modules. For example, the identified M6 module genes were enriched for astrocytes, neuritic plaque scores, and language domain of cognitive function. This confirms the previous report that astrocytes are involved in amyloid clearance[36] and damaged astrocytes impact language domain among AD patients.[37] Our discovery showed that M9 module genes are linked to endothelial cells, Braak stages, and visuospatial functioning in this study. Increased vascular inflammation

in endothelial cells has been observed among AD patients with poor short-term visuospatial functioning.[38]

Genes in the M6 and M9 modules have been previously reported for association with neurodegenerative diseases. Most of the genes in the two modules have biological functions relevant to the nervous system or have been previously reported in genetic or experimental studies for neurodegenerative diseases. For example, *SLC6A11* in M6 has been targeted for drug development of different neurodegenerative diseases including epilepsy.[39] *GLIS3* in M9 has been associated with T2D[40,41] and a longer life expectancy.[42] SNPs from *GLIS3* in M9 showed genome-wide significant associations from GWASs for Aβ and p-tau proteins in cerebrospinal fluid (CSF).[43]

4.4 | Limitations

Our study has several limitations. First, the sample size of discovery AD brains was modest. Therefore, we did not have statistical power for explaining the subtle phenotypic variations among AD patients, which might lead to detection of modules associated with a few specific cognitive domains. In addition, our current study exclusively relied on cognitive test scores for prioritizing CIH modules, which may not be useful for detecting unknown or brain imaging–based subtypes of the disease. Second, our findings in ADNI may not represent AD heterogeneity in other populations. However, because one of modules was replicated in an independent study (FHS), there are shared mechanisms across diverse populations. Third, because we focused on AD patients, our sample size of AD subgroups remained underpowered, so we could not apply multiple testing correction in the current study. This limitation was mitigated by replicating one of the mbPRSs in FHS. Fourth, we limited our mbPRSs calculations using GWAS summary statistics for AD risk and neuropathological outcomes regardless of available GWAS studies for CSF biomarker[44] or brain imaging data.[1] We decided to focus on neuropathological outcomes instead of biomarkers, because our goal is to explain AD heterogeneity by linking clinical subtypes to neuropathological outcomes. Fifth, we did not observe significant associations between PRSs and uncommon subgroups (e.g., AD-Executive). This may be because most of the previously defined cognitive subtypes in ADNI were predominantly classified as subgroups of memory (31.8%), and the executive functioning subgroup (2%) was relatively limited, especially in small datasets. In addition, datasets with GWAS for enough AD patients with carefully classified clinical phenotypes and clinically and/or pathologically defined subtypes are extremely limited. Finally, we recognize that the GWAS summary statistics for AD neuropathological traits (NP and NFT) in this study were generated based on genotype imputation using a previous reference panel (1000 Genomes),[18] which may affect quality and accuracy of our gene sets and PRSs. However, we used common SNPs (minor allele frquency > 5%) for constructing gene sets and PRSs, and the imputation qualities of common SNPs are still relatively acceptable even in the previous reference panel.[45] Therefore, potential problems caused by low imputation quality would be relatively limited in our study.

Future work in other independent GWAS samples with cognitively defined subgroups (or relevant subgroups based on cognitive tests) will be required to validate our module-based subgrouping of AD patients. Furthermore, linking genetics of various AD-related phenotypes including endophenotypes would enhance our ability to dissect further disease heterogeneity.[10] Other AD-related GWAS summary data including cerebral amyloid angiopathy, hypertension, cholesterol, and insulin resistance can be added for extending AD phenotype gene sets, which will lead us to detect novel gene sets and to recognize other subgroups beyond AD-Language/Visuospatial domains.

4.5 | Conclusion

In conclusion, PRSs developed using biologically coherent gene sets and disease-related phenotypes can successfully differentiate cognitively defined subgroups and brain region–specific atrophy, which likely represent specific mechanistic pathways responsible for the corresponding disease subtypes. Classification of patients using genetic information will allow patient subgrouping and target prioritization for the subgroups, which may eventually lead to precision medicine in AD. However, AD heterogeneity explained by the specific polygenic risk profiles in this study does not mean that mbPRSs can predict subjects in different disease stages of AD or at risk to AD progression in the future, as our mbPRSs can only differentiate AD patients into different cognitive subgroups. By comparing high- and low-risk groups of each mbPRSs using cognitively normal and MCI subjects, this aspect may be tested in the future. Our study warrants further validations in large datasets.

ACKNOWLEDGMENTS

This study was supported by the National Institute on Aging (NIA) grants, U01-AG068057, U19-AG068753, P30-AG072978, RF1-AG057519, and R56-AG069130. GWAS summary statistics used in this study were distributed by the National Institute on Aging Alzheimer's Disease Data Storage Site (NIAGADS) at the University of Pennsylvania (U24-AG041689-01). Collection of study data provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, was supported through funding by NIA grants P30-AG10161, R01-AG15819, R01-AG17917, R01-AG30146, R01-AG36836, U01-AG32984, U01-AG46152, and U01-AG61358, and funding from the Illinois Department of Public Health and the Translational Genomics Research Institute. Study data were also provided by Dr. Nilüfer Ertekin-Taner and Dr. Steven G. Younkin, Mayo Clinic, Jacksonville, FL, using samples from the Mayo Clinic Study of Aging, the Mayo Clinic Alzheimer's Disease Research Center, and the Mayo Clinic Brain Bank. Collection of these data was supported through funding by NIH grants P50-AG016574, R01-AG032990, U01-AG046139, R01-AG018023, U01-AG006576, U01-AG006786, R01-AG025711, R01-AG017216, R01-AG003949, and R01-NS080820, and by funding from the CurePSP Foundation and the Mayo Foundation. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI; National Institutes of Health

Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie; Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

CONFLICT OF INTEREST STATEMENT

All authors report no conflicts of interest. Author disclosures are available in the supporting information.

DATA AVAILABILITY STATEMENT

Genotypes and clinical/neuropathological phenotype data are accessible by directly applying to the LONI portal for the ADNI at http://adni.loni.usc.edu. Summary statistics for GWAS studies can be accessed by applying directly to the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS), a NIA/NIH-sanctioned qualified-access data repository, under accession NG00075. Data supporting the findings of this study are available from the NIAGADS website (https://www.niagads.org/).

CONSENT STATEMENT

The study protocol, design, and performance of the current study were approved by the Boston University Institutional Review Board.

REFERENCES

1. Dong A, Toledo JB, Honnorat N, et al. Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression and biomarkers. *Brain*. 2017;140:735-47.
2. Qiu Y, Jacobs DM, Messer K, Salmon DP, Feldman HH. Cognitive heterogeneity in probable Alzheimer disease: clinical and neuropathologic features. *Neurology*. 2019;93:e778-e90.
3. Anderson RM, Hadjichrysanthou C, Evans S, Wong MM. Why do so many clinical trials of therapies for Alzheimer's disease fail? *Lancet*. 2017;390:2327-9.
4. Ten Kate M, Dicks E, Visser PJ, et al. Atrophy subtypes in prodromal Alzheimer's disease are associated with cognitive decline. *Brain*. 2018;141:3443-56.
5. Murray ME, Graff-Radford NR, Ross OA, Petersen RC, Duara R, Di DW. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. *Lancet Neurol*. 2011;10:785-96.
6. Crane PK, Trittschuh E, Mukherjee S, et al. Incidence of cognitively defined late-onset Alzheimer's dementia subgroups from a prospective cohort study. *Alzheimers Dement*. 2017;13:1307-16.
7. Groot C, Grothe MJ, Mukherjee S, et al. Differential patterns of gray matter volumes and associated gene expression profiles in cognitively-defined Alzheimer's disease subgroups. *Neuroimage Clin*. 2021;30:102660.
8. Mendez MF. Early-onset Alzheimer disease and its variants. *Continuum (Minneap Minn)*. 2019;25:34-51.
9. Mukherjee S, Mez J, Trittschuh EH, et al. Genetic data and cognitively defined late-onset Alzheimer's disease subgroups. *Mol Psychiatry*. 2020;25:2942-51.
10. Udler MS, Kim J, von Grotthuss M, et al. Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: a soft clustering analysis. *PLoS Med*. 2018;15:e1002654.
11. Zhang B, Gaiteri C, Bodea LG, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*. 2013;153:707-20.
12. Mostafavi S, Gaiteri C, Sullivan SE, et al. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat Neurosci*. 2018;21:811-9.
13. Panitch R, Hu J, Chung J, et al. Integrative brain transcriptome analysis links complement component 4 and HSPA2 to the APOE ε2 protective effect in Alzheimer disease. *Mol Psychiatry*. 2021;26:6054-64.
14. De Jager PL, Ma Y, McCabe C, et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci Data*. 2018;5:180142.
15. Allen M, Carrasquillo MM, Funk C, et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci Data*. 2016;3:160089.
16. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.
17. Chung J, Wang X, Maruyama T, et al. Genome-wide association study of Alzheimer's disease endophenotypes at prediagnosis stages. *Alzheimers Dement*. 2018;14:623-33.
18. Beecham GW, Hamilton K, Naj AC, et al. Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. *PLoS Genet*. 2014;10:e1004606.
19. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet*. 2019;51:414-30.
20. Darmanis S, Sloan SA, Zhang Y, et al. A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci U S A*. 2015;112:7285-90.
21. Mathys H, Davila-Velderrain J, Peng Z, et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*. 2019;570:332-7.
22. Scheltens NM, Galindo-Garre F, Pijnenburg YA, et al. The identification of cognitive subtypes in Alzheimer's disease dementia using latent class analysis. *J Neurol Neurosurg Psychiatry*. 2016;87:235-43.
23. Nho K, Risacher SL, Crane PK, et al. Voxel and surface-based topography of memory and executive deficits in mild cognitive impairment and Alzheimer's disease. *Brain Imaging Behav*. 2012;6:551-67.
24. Mesulam MM, Thompson CK, Weintraub S, Rogalski EJ. The Wernicke conundrum and the anatomy of language comprehension in primary progressive aphasia. *Brain*. 2015;138:2423-37.
25. Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T. Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci*. 1997;17:353-62.
26. Nobre AC, Sebestyen GN, Gitelman DR, Mesulam MM, Frackowiak RS, Frith CD. Functional localization of the system for visuospatial

attention using positron emission tomography. *Brain*. 1997;120(Pt 3):515-33.
27. Escott-Price V, Myers AJ, Huentelman M, Hardy J. Polygenic risk score analysis of pathologically confirmed Alzheimer disease. *Ann Neurol*. 2017;82:311-4.
28. 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:e1002258.
29. Chaudhury S, Brookes KJ, Patel T, et al. Correction: Alzheimer's disease polygenic risk score as a predictor of conversion from mild-cognitive impairment. *Transl Psychiatry*. 2019;9:167.
30. Wand H, Lambert SA, Tamburro C, et al. Improving reporting standards for polygenic scores in risk prediction studies. *Nature*. 2021;591:211-9.
31. Wang T, Han Z, Yang Y, et al. Polygenic risk score for Alzheimer's disease is associated with Ch4 volume in normal subjects. *Front Genet*. 2019;10:519.
32. Leonenko G, Sims R, Shoai M, et al. Polygenic risk and hazard scores for Alzheimer's disease prediction. *Ann Clin Transl Neurol*. 2019;6:456-65.
33. Sanchez-Valle J, Tejero H, Fernandez JM, et al. Interpreting molecular similarity between patients as a determinant of disease comorbidity relationships. *Nat Commun*. 2020;11:2854.
34. Dahl A, Zaitlen N. Genetic influences on disease subtypes. *Annu Rev Genomics Hum Genet*. 2020;21:413-35.
35. Lu Q, Powles RL, Abdallah S, et al. Systematic tissue-specific functional annotation of the human genome highlights immune-related DNA elements for late-onset Alzheimer's disease. *PLoS Genet*. 2017;13:e1006933.
36. Thal DR. The role of astrocytes in amyloid beta-protein toxicity and clearance. *Exp Neurol*. 2012;236:1-5.
37. Resende EPF, Nolan AL, Petersen C, et al. Language and spatial dysfunction in Alzheimer disease with white matter thorn-shaped astrocytes. *Neurology*. 2020;94:e1353-e64.
38. Liu X, Ma Y, Ouyang R, et al. The relationship between inflammation and neurocognitive dysfunction in obstructive sleep apnea syndrome. *J Neuroinflammation*. 2020;17:229.
39. Ricciarelli R, Argellati F, Pronzato MA, Domenicotti C. Vitamin E and neurodegenerative diseases. *Mol Aspects Med*. 2007;28:591-606.
40. Dimitri P, Warner JT, Minton JA, et al. Novel GLIS3 mutations demonstrate an extended multisystem phenotype. *Eur J Endocrinol*. 2011;164:437-43.
41. Barker A, Sharp SJ, Timpson NJ, et al. Association of genetic Loci with glucose levels in childhood and adolescence: a meta-analysis of over 6,000 children. *Diabetes*. 2011;60:1805-12.
42. Dimitri P, Habeb AM, Gurbuz F, et al. Expanding the clinical spectrum associated with GLIS3 mutations. *J Clin Endocrinol Metab*. 2015;100:E1362-9.
43. Deming Y, Li Z, Kapoor M, et al. Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. *Acta Neuropathol*. 2017;133:839-56.
44. Tijms BM, Gobom J, Reus L, et al. Pathophysiological subtypes of Alzheimer's disease based on cerebrospinal fluid proteomics. *Brain*. 2020;143:3776-92.
45. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*. 2021;590:290-9.

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: Chung J, Sahelijo N, Maruyama T, et al. Alzheimer's disease heterogeneity explained by polygenic risk scores derived from brain transcriptomic profiles. *Alzheimer's Dement*. 2023;19:5173–5184.
https://doi.org/10.1002/alz.13069