



## ORIGINAL ARTICLE

## Brain structure and allelic associations in Alzheimer's disease

Seok Woo Moon<sup>1,2</sup>  | Lu Zhao<sup>2</sup> | William Matloff<sup>2</sup> | Sam Hobel<sup>2</sup> | Ryan Berger<sup>3</sup> | Daehong Kwon<sup>4</sup> | Jaebum Kim<sup>4</sup> | Arthur W. Toga<sup>2</sup> | Ivo D. Dinov<sup>2,5,6</sup>  | for the Alzheimer's Disease Neuroimaging Initiative

<sup>1</sup>Department of Neuropsychiatry, Research Institute of Medical Science, Konkuk University School of Medicine, Seoul, Korea

<sup>2</sup>USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, California, Los Angeles, USA

<sup>3</sup>Microbiology & Immunology, University of Michigan, Ann Arbor, Michigan, USA

<sup>4</sup>Department of Biomedical Science and Engineering, Konkuk University, Seoul, Korea

<sup>5</sup>Department of Health Behavior and Biological Sciences, Statistics Online Computational Resource (SOCR), Michigan Institute for Data Science (MIDAS), University of Michigan, Ann Arbor, Michigan, USA

<sup>6</sup>Department of Statistics, University of California, Los Angeles, California, USA

## Correspondence

Ivo D. Dinov, Department of Statistics, University of California, Los Angeles, CA 90095, USA.

Email: [dinov@ucla.edu](mailto:dinov@ucla.edu)

Seok Woo Moon, Dementia Center, Department of Neuropsychiatry, Konkuk University Hospital, 82 Gukwon-daero, Chungju-si, Chungbuk-do 380-704, Korea. Email: [hessem@kku.ac.kr](mailto:hessem@kku.ac.kr)

## Funding information

National Institutes of Health (NIH), Grant/Award Number: R01 CA233487, R01 MH121079, R01 MH126137, T32 GM141746 and UL1 TR002240; National Research Foundation of Korea (NRF), Grant/Award Number: NRF-2019R111A3A01058477; National Science Foundation (NSF), Grant/Award Number: 1734853, 1916425, 1023115, 1416953 and 1636840

## Abstract

**Background:** Alzheimer's disease (AD), the most prevalent form of dementia, affects 6.5 million Americans and over 50 million people globally. Clinical, genetic, and phenotypic studies of dementia provide some insights of the observed progressive neurodegenerative processes, however, the mechanisms underlying AD onset remain enigmatic. **Aims:** This paper examines late-onset dementia-related cognitive impairment utilizing neuroimaging-genetics biomarker associations.

**Materials and Methods:** The participants, ages 65–85, included 266 healthy controls (HC), 572 volunteers with mild cognitive impairment (MCI), and 188 Alzheimer's disease (AD) patients. Genotype dosage data for AD-associated single nucleotide polymorphisms (SNPs) were extracted from the imputed ADNI genetics archive using sample-major additive coding. Such 29 SNPs were selected, representing a subset of independent SNPs reported to be highly associated with AD in a recent AD meta-GWAS study by Jansen and colleagues.

**Results:** We identified the significant correlations between the 29 genomic markers (GMs) and the 200 neuroimaging markers (NIMs). The odds ratios and relative risks for AD and MCI (relative to HC) were predicted using multinomial linear models.

**Discussion:** In the HC and MCI cohorts, mainly cortical thickness measures were associated with GMs, whereas the AD cohort exhibited different GM-NIM relations. Network patterns within the HC and AD groups were distinct in cortical thickness, volume, and proportion of White to Gray Matter (pct), but not in the MCI cohort.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database ([adni.loni.usc.edu](http://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 analyzing or writing this report. A complete listing of ADNI investigators can be found at: [https://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf)

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *CNS Neuroscience & Therapeutics* published by John Wiley & Sons Ltd.

Multinomial linear models of clinical diagnosis showed precisely the specific NIMs and GMs that were most impactful in discriminating between AD and HC, and between MCI and HC.

**Conclusion:** This study suggests that advanced analytics provide mechanisms for exploring the interrelations between morphometric indicators and GMs. The findings may facilitate further clinical investigations of phenotypic associations that support deep systematic understanding of AD pathogenesis.

#### KEYWORDS

ADNI, Alzheimer's disease, genetics, mild cognitive impairment, networking, neuroimaging

## 1 | INTRODUCTION

Alzheimer's disease (AD) is by far the most common form of dementia among the elderly.<sup>1,2</sup> Late onset Alzheimer's disease (LOAD), defined by the onset of symptoms after age 65, is sporadic, non-familial AD.<sup>3,4</sup> Genetic studies have provided significant insights on the molecular basis of AD, but the mechanisms underlying AD onset and progression remain largely unexplained. While the underlying causes of LOAD are still unknown, there is evidence from familial aggregation, transmission pattern, and twin studies that AD has a substantial genetic component that has an estimated heritability of 58%–79%, and the lifetime risk of AD among first-degree relatives of patients may be twice that of the general population.<sup>5,6</sup> Recent genome-wide association studies (GWASs), which examine associations of AD diagnosis with genetic markers (single-nucleotide polymorphism [SNP]) across the genome, have discovered more than 20AD gene variants that confer genetic risk.<sup>7,8</sup> These findings improve the understanding of risks and causes for AD, and may guide diagnosis and therapy on a patient-specific basis.<sup>9</sup> However, case-control GWAS cannot completely characterize the exact roles of the identified genetic susceptibility loci in the pathophysiology of AD. Joint analysis of genetic and neuroimaging data could uncover the genetic mechanism in the disease's underlying biology.<sup>5</sup>

This study investigates holistically the significance of multi-gene patterns associated with neuroimaging markers (NIMs) of AD using imaging and genomic data of the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort.<sup>10</sup> Several single-nucleotide polymorphisms (SNPs) present in apolipoprotein E (APOE) gene have previously been associated with neuroimaging measures in both cognitively healthy control (HC) or impaired (such as mild cognitive impairment [MCI] and AD dementia) patients.<sup>11–13</sup> However, a single gene or a few imaging measures may be insufficient to understand the multiple mechanisms and imaging manifestations of the complex disease. Recent and ongoing advances in neuroimaging and genetics, including high-throughput genotyping techniques, have made it possible to scan populations with multimodality neuroimaging to collect genome-wide data and to study the influence of genetic variation on the brain structure and function.

In this work, we related high-throughput neuroimaging-derived phenotypes of brain structure to the clinical states of AD, and then associated the significant AD-related NIMs with GWAS-supported susceptibility genetic variants for AD to obtain true system-level gene-brain associations in dementia. Specifically, we used structural brain imaging to obtain biomarkers of a wide variety of brain morphological properties, allelic data to capture genotypic variation, and functional connectivity to evaluate imaging-genetic-phenotypic variation. We present a neuroimaging genetics framework that uses a whole-genome-and-whole-brain strategy to systematically evaluate genetic effects on neuroimaging phenotypes to discover quantitative trait (QT) loci. QT association studies have been shown to have increased statistical power and thus decreased sample size requirements.<sup>14,15</sup> In addition, neuroimaging phenotypes may be closer to the underlying biological etiology of the disease, making it easier to identify underlying genes. The methodology proposed in this paper is based on the identification of strong associations between regional neuroimaging phenotypes as QTs and SNP genotypes as QT loci.

The genetics of AD are complex because the practical effects may be weak, albeit statistical effects could still be strong, sample-sizes are often unbalanced (number of cases  $\ll$  genomic markers [GMs]), and considerable difficulties with result replication and validation.<sup>16,17</sup> Large-scale GWAS shows promise in untangling the genetic footprint of this neurodegenerative disease. Considering the limited sample size in the ADNI cohort ( $n = \sim 1200$ ), we used the AD-related genetic variants identified by the largest ( $n = \sim 450,000$ ) case-control GWAS in AD to date<sup>8</sup> instead of performing GWAS on the ADNI cohort.

We hypothesized that there exist significant relationships between the AD-related NIMs and the GWAS-supported susceptibility genetic variants for AD. Several prior studies have been conducted on the relationship between neuroimaging phenotypes and genetic variants.<sup>18–20</sup> However, few reports have previously performed functional analyses of neuroimaging genetics. This study expands the knowledge about dementia phenotypes using modern neuroimaging genetics and the network analysis to explore the relationships between genetic, phenotypic, and NIMs.

## 2 | METHODS

### 2.1 | Inclusion–Exclusion criteria

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<https://adni.loni.usc.edu>). ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see <https://www.adni-info.org>.

This study is based on participants from ADNI-1, ADNI-GO, and ADNI-2. Only data from the baseline visit of each participant was considered. Each baseline participant has multiple structural MRI scans available with varying scan parameters (e.g., accelerated vs. non-accelerated) and at various stages of preprocessing. Image selection aimed to maximize consistency across participants. First, the maximally preprocessed T1-weighted MRI scans were downloaded for each participant.

For ADNI-1 participants, quality assessment data available on the LONI IDA<sup>21</sup> (MRI MPAGE Ranking: MRIMPRANK) was used to select the higher-rated scan for each visit. Accelerated scans were removed and 3T scans were chosen over 1.5T scans where possible. For those with multiple scans after this filtering, the scan with the larger Image ID was kept. For the ADNI-GO and ADNI-2 participants, accelerated scans were removed and among the remaining participants with more than one scan, the scan with the larger Image ID was kept. After these filtering steps were completed, the scans were combined into a final set of scans consisting of 1242 participants with only one scan each. The ADNI database query yielded a pool of 1242 volunteers with both brain MRI and genetic data from the study phases of ADNI-1, ADNI-GO and ADNI-2. Three subjects were discarded due to failed MRI processing. We further removed 110 participants who did not have CEU ancestry (*ancestry from Northern and Western Europe*) and employed the remaining 1129 subjects for identifying the AD-related NIMs. In the final imaging-genetics association analysis another 103 subjects were excluded as they did not pass the genetic data quality control. We ended up with 1026 subjects who successfully completed all imaging and genetics processing protocols. For each participant, clinical severity of dementia was assessed using an annual semi-structured interview, which yielded an overall Clinical Dementia Rating (CDR) score and the CDR Sum of Boxes (CDR-SB). In addition, the Mini-Mental State Examination (MMSE) and a neuropsychological battery were also recorded.

All the ADNI participants included in this study were those with a baseline age of 65–85 years, a baseline MMSE score of 20–30, and available genetic and imaging data. The 1026 ADNI participants included: 266 HC's (CDR = 0, Male:138, Female:128), 572 MCI's (CDR = 0.5, Male:227, Female:245), and 188 AD dementia patients

(CDR = 0.5/1, Male:102, Female:86). We analyzed NIMs among participants with different APOE haplotype in the individual HC, MCI, and AD dementia cohorts. Subjects with AD dementia were probable AD dementia according to the NIA-AA diagnostic criteria for AD. The ADNI study-design specified that the time period between initial screening and clinical exams and the subsequent MRI imaging is about 3 months.

### 2.2 | MRI processing and analysis

The MRI acquisition protocols can be found on the ADNI website (<https://adni.loni.usc.edu>) and have been previously described elsewhere.<sup>22,23</sup> Baseline structural MRI scans of the ADNI subjects were processed for reconstructing cortical surfaces, brain parcellation and extracting morphological phenotypes using the FreeSurfer (v6.0) software package (<https://surfer.nmr.mgh.harvard.edu/>).<sup>24</sup> The FreeSurfer processing includes motion correction and averaging of volumetric T1-weighted images,<sup>25</sup> removal of non-brain tissue,<sup>26</sup> automated Talairach transformation, brain volume segmentation,<sup>27,28</sup> intensity normalization,<sup>29</sup> tessellation of the boundary between gray matter and white matter, automated topology correction<sup>30</sup> and surface deformation.<sup>31</sup>

Once the cortical models are completed, a number of deformable procedures were performed for further data processing and analysis, including surface inflation, registration to a spherical atlas using individual cortical folding patterns to match cortical geometry across subjects,<sup>32</sup> and finally creation of a variety of surface-based data including maps of surface area, cortical thickness, curvature features, etc.

For each subject, 1380 imaging-derived biomarkers were extracted using FreeSurfer, including measures of surface area, volume, thickness, standard deviation of thickness, mean curvature, Gaussian curvature, folding index, curvature index and/or gray matter/white matter contrast for different cortical, subcortical, and white matter regions.

All the imaging phenotypes were adjusted for age, gender, education, handedness, and intracranial volume (ICV) using linear mixed-effects regressions. ANOVA tests were then performed to find the NIMs associated with AD diagnosis at the GWAS significance level  $p < 5 \times 10^{-8}$  in the cohort of  $N = 1129$  subjects with a CEU ancestry.

### 2.3 | Genetics data processing

Genetic SNP data was downloaded from the ADNI database (<https://www.loni.usc.edu/ADNI>) through the LONI imaging data archive (IDA) interface (<https://ida.loni.usc.edu/pages/access/geneticData.jsp>) onto the LONI Cranium high-performance computing (HPC) cluster. The processing resulted in a single dataset containing the genetics information of all 1026 participants. The ADNI-1 genetics data was downloaded as PLINK bed/bim/bam files in the hg18 (build 36) format. The genome build was converted from hg18 to hg19

using liftOver, as described in (<https://www.nature.com/articles/nprot.2015.077>). The ADNI-GO and ADNI-2 genetics data, which is in the hg19 (build 37) format, was downloaded as PLINK bed/bim/bam files for sets 1–9 and as intensity data CSV files for sets 10–15. The intensity data CSV files were converted to PLINK files at a multiple GenCall Score (GC) threshold of 0.15 based on the procedure described in Ref. [13].

ADNI-1 and ADNI-GO/2 used different genotyping chips. Used genetic imputation, we harmonized the genetic data across the different ADNI studies. To prepare for imputation, population stratification analysis was first used to remove all of the non-CEU participants. We used PLINK for population stratification. PLINK relies on genome-wide average proportion of alleles shared between any two individuals to cluster subjects into homogeneous subsets and perform classical multidimensional scaling (MDS) to visualize substructure and provide quantitative indices of population genetic variation. Next, we used the “HRC or 1000G Imputation preparation and checking” tool (HRC-1000G-check-bim-v4.2.9) from the McCarthy Group to conduct common pre-imputation checks, such as strand, reference allele assignment, and frequency differences (<https://www.well.ox.ac.uk/~wrayner/tools>).

Imputation was completed using the Michigan Imputation Server v1.0.4 (Sept. 14, 2018) (<https://imputationserver.sph.umich.edu/index.html>). This imputation, based on Minimac3,<sup>33</sup> was completed with the accompanying quality control offered by the service. The reference panel used was the HRC r1.1 2016, phasing was competing using Eagle, and quality control was based on the European population. The ADNI-1 input consisted of 694 samples and 568,933 SNPs, 10,583 of which were excluded for imputation due to being monomorphic or having a SNP call rate of <90%. The ADNI-GO/2 input consisted of 723 samples and 696,245 SNPs, 38,605 of which were excluded due to being monomorphic or having a SNP call rate of <90%. The output of the Michigan Imputation Server was in the minimac3 output format, including both info and dosage files. The HRC-imputed data for the ADNI-1 and ADNI-GO/2 datasets were merged with *bcftools* (<https://samtools.github.io/bcftools/bcftools.html>). After this imputation and combination, the sites were filtered to only include those with  $R^2 > 0.6$  and with a minor allele frequency >0.5% using *bcftools* and *tabix* (<https://samtools.github.io/bcftools/bcftools.html>, <http://www.htslib.org>). Finally, the data was converted into the *pgen* format (PLINK 2 binary format) using PLINK 2.00 alpha (<https://www.cog-genomics.org/plink/2.0/>).

## 2.4 | The pipeline computational environment

To manage the large and complex raw and derived data, design and execute the end-to-end processing protocols, and to track provenance, we employed the LONI Pipeline.<sup>34–37</sup> The Pipeline is a graphical workflow environment facilitating the collaborative design, execution, validation, visualization, modification and sharing of complex heterogeneous computational protocols.

To promote reproducible open-science development and validation, we designed a Pipeline workflow that represents an end-to-end computational protocol for high-throughput data preprocessing. The pipeline workflow includes skull-stripping, volumetric registration, brain anatomical parcellation, extraction of volume and cortical thickness and between group statistical analyses of shape regional differences. The output of the pipeline workflow is a collection of 3D scenes illustrating the statistically significant regional anatomical differences between the study cohorts.

Rank-ordering the complete collection of NIMs, we chose the 200 most salient NIMs which provided the highest discrimination between the AD and HC groups. These 200 NIMs were derived from all structural imaging data using the workflow and are based on the automated ROI extractions generated by FreeSurfer. Finally, the pipeline workflow, computed the most significant genotypic discriminants among AD, MCI and HC subjects. The 200 NIMs were then associated with the top 29 SNPs, which were chosen by the PLINK.<sup>8</sup>

## 2.5 | Analytical protocol

The end-to-end data analysis protocol was implemented via the Pipeline graphical workflow environment and involved the following steps (1) Imputation on the Michigan Server using IMPUTE2, a genotype imputation and haplotype phasing program based on ideas from,<sup>38</sup> and (2) Beagle is a software package for phasing genotypes and for imputing ungenotyped markers.<sup>39</sup> The SOCR statistical computing infrastructure (<https://SOCR.umich.edu>)<sup>40–45</sup> was utilized to implement and execute the end-to-end computational statistics protocol, which included multivariate linear modeling and general parametric and non-parametric statistical analyses.

The imputation protocol relied on the default setting. Following quality control (QC) and imputation, there were 1026 subjects with a European ancestry remaining with minimac3 outputs, including both info and dosage files. Next, we combined the ADNI-1 and ADNI-GO/2 imputed data into a single PLINK file. ADNI1 and ADNIGO/2 are genotyped using different chips, so we imputed the arrays prior to their integration and ran GWAS on the combined array. The resulting data were filtered to include only the subjects with a CEU ancestry and contained the following four tensors:

1. genetic.markers.associated.with.ADdx.CEU: included the SNPs, chromosomes, and positions corresponding to each SNP reference sequence ID, see Table 2. We replaced the APOE SNP (X19.45351516) with the separate APOE genotype test conducted by ADNI giving the exact APOE allele.
2. ADNI\_baseline\_CEU\_metadata: contains the APOE genotype meta-data.
3. imaging.markers.associated.with.ADdx.ANOVA: includes 200 NIMs and their corresponding association with clinical phenotypes (*p*-values).
4. FS.stats.imaging.markers.associated.with.ADdx.CEU: contains the FreeSurfer-extracted neuroimaging volume, surface and thickness measures for cortical and subcortical regions of interest (ROIs).

Category	HC	MCI	AD	p-Value
No. of subjects (1026)	266	572	188	
Gender (M/F)	138/128	227/245	102/86	0.580 (*)
Age	73.51 ± 3.89	70.74 ± 6.07	71.72 ± 6.34	<0.001 (*)
MMSE	29.07 ± 1.94	27.05 ± 1.79	27.75 ± 1.78	<0.0001 (**)
ADAS-Cog	6.15 ± 2.86	11.43 ± 4.40	18.46 ± 6.28	<0.0001 (*)
Education (years, mean ± SD)	16.18 ± 2.68	15.98 ± 2.78	15.40 ± 2.87	0.010 (*)
Handedness (R/L)	207/18	348/35	188/12	0.418 (*)
APOE (ε2 / ε3 / ε4)	34/421/77	39/709/396	9/184/183	<0.0001 (**)

**TABLE 1** Participants demographic data

Note: p-Values correspond to the appropriate chi-square (\*) or ANOVA (\*\*) test statistics.

Abbreviations: AD, Alzheimer's disease; HC, healthy controls; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination.

We extracted the AD-related GMs and NIMs for the network analysis. Genotype dosage data for AD-associated SNPs were extracted from the imputed ADNI genetics dosage data using sample-major additive (0/1/2) coding. A set of 29 SNPs were selected, representing a subset of independent SNPs found to be highly associated with AD in a recent AD meta-GWAS study by Jansen et al.,<sup>8</sup> which met the MAF threshold determined by the imputation process. The genotyping data of these SNPs are included in the genetic markers.associated.with.ADdx.CEU data object along with SNP IDs,<sup>8</sup> Table 1. We replaced the APOE SNP (X19.45351516) with the separate APOE genotype test conducted by ADNI giving the exact APOE allele.

Age and gender were included as confounding factors in the analysis. The top 200 NIMs were identified to be associated with AD diagnosis at the GWAS significance level ( $p < 5 \times 10^{-8}$ ). The 200 NIMs and their p-values are stored in imaging.markers.associated.with.ADdx.ANOVA tensor. These NIMs were extracted and used to generate association heatmaps (SNPs by diagnosis and Imaging biomarker by diagnosis).

## 2.6 | Neuroimaging-genetics association analytics protocol

More details are provided in the Appendix S1 (Methods).

## 2.7 | Network analysis between 29 genomic and 200 neuroimaging markers

The WGCNA (Weighted graph correlation network analysis) R package (version 1.68)<sup>8</sup> was used to perform network analysis using 29 genomic and 200 neuroimaging markers. The aim of this analysis was to connect the NIMs having similar patterns observed from the GMs. WGCNA was originally developed to find the

network of co-expressed genes based on their expression patterns in multiple conditions. Specifically, for each of three cohorts, such as AD, MCI, and HC, the values of the NIMs and GMs were measured, and correlation coefficient values between the two kinds of markers were calculated. The correlation coefficient values constitute the correlation coefficient matrix (CCM) between the two kinds of markers. The CCM was then converted to matrices representing an unsigned adjacency matrix (using soft thresholding of 7) and a topology overlap matrix (using a score threshold of 0.05). The NIMs with similar genetic patterns were predicted by WGCNA using the converted matrices. The predicted networks for each of the AD, MCI, and HC cohort were visualized using Cytoscape (version 3.7.1).<sup>46</sup>

## 2.8 | Statistically significant MCI/HC and AD/HC odds ratios

Using multinomial linear modeling of diagnosis, we studied the associations between the three individual cohorts (HC, MCI and AD dementia). The differences of the 200 NIMs and 29 SNPs between HC, MCI, and AD dementia cohorts. The results of a 3-way ANOVA (ROI, Dx, SNP) may be less interpretable compared to a multinomial linear modeling (Outcome = Dx). We computed the odds ratios (ORs) and relative risks (RRs) for AD and MCI, relative to HC.

The MCI and AD effects quantified the metrics "relative to HC." These represent extensions of the binary outcome in logistic regression, but reflect 3 categorical outcomes (HC, MCI, AD), which may also be analyzed via more general multi-nominal linear modeling. In general, to assess statistical significance a customary false-positive rate of  $\alpha = 0.05$  may be used for many different tests. However, in many GWAS studies, it's common that a correction for multiple comparison (e.g., false discovery rate, family wise error rate), or other strategies are used to control the false positive rate of significance.<sup>47,48</sup>



**TABLE 2** Twenty nine genomics markers

[Top 29 genomics markers <sup>8</sup> ]			
Index	GM (ADNI)	SNPs	Genes
1	X1.161155392_G	rs4575098	[ADAMTS4]
2	X1.207786828_A	rs2093760	[CR1]
3	X2.127891427_A	rs4663105	[BIN1]
4	X2.233981912_G	rs10933431	[INPP5D]
5	X4.11026028_A	rs6448453	[CLNK]
6	X4.11723235_A	rs7657553	[HS3ST1]
7	X6.32583357_A	rs6931277	[HLA-DRB1]
8	X6.47432637_C	rs9381563	[CD2AP]
9	X7.99971834_A	rs1859788	[ZCWPW1]
10	X7.143108158_T	rs11763230	[EPHA1]
11	X8.27464929_A	rs4236673	[CLU/PTK2B]
12	X10.11717397_T	rs11257238	[ECHDC3]
13	X11.59958380_C	rs2081545	[MS4A6A]
14	X11.85776544_G	rs867611	[PICALM]
15	X11.121435587_T	rs11218343	[SORL1]
16	X14.92938855_G	rs12590654	[SLC24A4]
17	X15.59022615_T	rs442495	[ADAM10]
18	X15.63569902_C	rs117618017	[APH1B]
19	X16.31133100_G	rs59735493	[KAT8]
20	X17.5138980_G	rs113260531	[SCIMP]
21	X17.47450775_G	rs28394864	[ABI3]
22	X17.56409089_G	rs2632516	[BZRAP1-AS1]
23	X18.29088958_C	rs8093731	[SUZ12P1]
24	X18.56189459_T	rs76726049	[ALPK2]
25	X19.1039323_C	rs111278892	[ABCA7]
26	X19.45351516_C	rs41289512	[APOE]
27	X19.46241841_C	rs76320948	[AC074212.3]
28	X19.51727962_C	rs3865444	[CD33]
29	X20.54998544_A	rs6014724	[CASS4]

Abbreviation: SNP, single-nucleotide polymorphism.

### 3 | RESULTS

#### 3.1 | Demographic characteristics

The demographics and clinical data of the ADNI participants at the baseline are summarized in Table 1 and include *p*-values computed using Chi-square or ANOVA, as appropriate. The 1,026 subjects (aged 65~85 years) were chosen from the ADNI datasets. The AD, MCI and HC subjects had statistically significant differences in age, MMSE, ADAS-Cog, Education and APOE genotype.

#### 3.2 | AD genetic and imaging markers for the network analysis

We extracted the AD-related genotypes and NIMs for the network analysis. Genotype dosage data for AD-associated SNPs

were extracted from the imputed ADNI genetics dosage data using sample-major additive (0/1/2) coding. The genotypes, SNPs, chromosomes, and positions can be found in<sup>8</sup> Table 2.

To streamline the analyses, we chose the top 200 NIMs corresponding to the lowest *p*-values for the discrimination between HC and AD subjects to correlate with 29 GMs, Table S1.

#### 3.3 | Association analysis among genetic and neuroimaging biomarkers

In-house R-scripts were developed to generate three heatmaps for each of the three cohorts. These plots represent the association analyses (SNPs\*diagnosis, NIMs\*diagnosis), see Figure 1 and Table S2.

HC cohort, 46 cortical thickness measures, 4 volumes, 3 pct (the proportion of White to Gray Matter – a proxy measure of cortical thickness), and 1 Gaussian-curvature (gauscurve) measure among the 200 NIMs were significantly associated with 9 markers among the 29 GMs at the level of *p* < 0.01. There were 9 ROIs (7 thicknesses, 1 volume, and 1 gauscurve) which were associated with 6 GMs at the level of *p* < 0.001, Table 3A.

Additional details are provided in Figure 1A and Table S2A.

MCI cohort, 116 thickness measures, 37 volumes, 22 pct, 1 area, and 1 folding index (foldind) among 200 NIMs were significantly associated with the 6 markers among the 29 GMs at the level of '*p* < 0.01'. There were 43 ROI measures (31 thicknesses, 11 volume, and 1 pct) which were associated with 5 GMs at the level of '*p* < 0.001'. There were 10 ROIs (9 thicknesses and 1 volume) which were associated with 3 GMs at the level of '*p* < 0.0001', Table 3B.

Additional details are provided in Figure 1B and Table S2B.

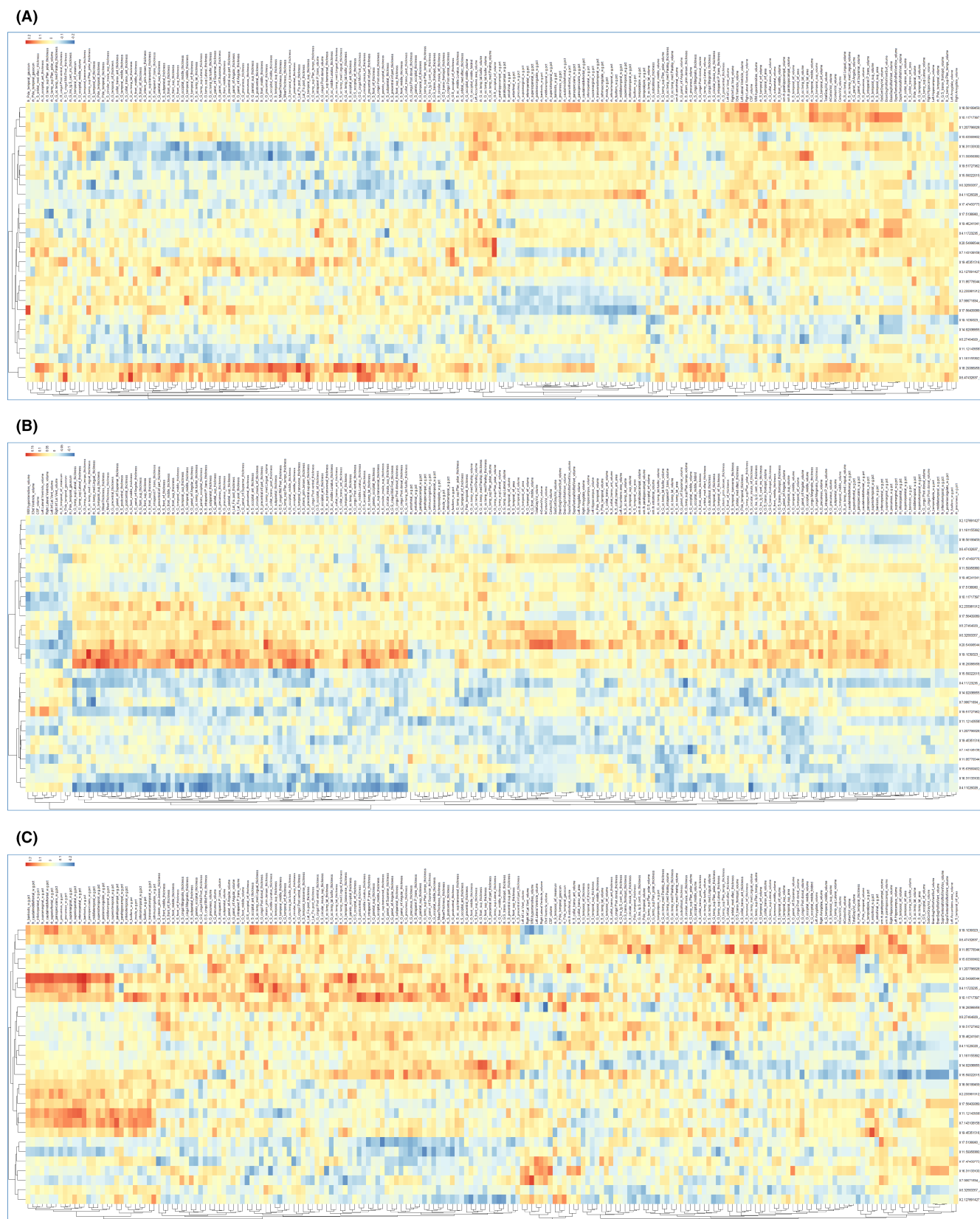
AD cohort, 23 thickness measures, 14 volumes, 15 pct, 1 area, 1 gauscurve, and white matter hyperintensity (WMHI) among 200 NIMs were significantly associated with the 11 markers among the 29 GMs at the level of '*p* < 0.01'. There were three ROIs (1 thickness, 1 volume, and 1 pct) which were associated with three GMs at the level of '*p* < 0.001', Table 3C.

Additional details are provided in Figure 1C and Table S2C.

#### 3.4 | Network analysis

The interactions between the 29 genes and the 200 NIMs were also explored using network analysis for each of the three cohorts, HC, MCI and AD. We did not find strong network patterns solely within the 29 genes themselves, see Figure 2. However, we found strong networking patterns within the HC group. Three types of NIMs were divided into thickness, volume, and proportion of white to gray matter (pct). These three measurement groups were networking separately under the control of the 29 genes in the HC group, regardless of the *p*-values in association analysis which is described earlier in the association analysis section.

In thickness measure group, these NIMs (mainly lh\_G\_precentral, rh\_S\_precentral.inf.part, rh\_S\_precentral.sup.part,



**FIGURE 1** Dendrograms illustrating Neuroimaging-Genetics Associations. (A) Cognitively asymptomatic control (HC) cohort; (B) Mild Cognitive Impairment (MCI) cohort; (C) Alzheimer's Disease (AD) cohort.

**TABLE 3** GM-NIM associations for the HC, MCI, and AD cohorts.

SNP	Reference	Assoc. direction	Associated covariates
<b>(A) GM-NIM associations for the HC cohort</b>			
X6.47432637	rs9381563, <i>CD2AP</i>	+	rh_S_oc.temp_lat_thickness ( $p = 0.0004$ ) rh_S_temporal_sup_thickness ( $p = 0.0009$ ) lh_G_S_cingul.Mid.Post_thickness ( $p = 0.0010$ )
X7.143108158	rs11763230, <i>EPHA1</i>	+	rh_G_S_occipital_inf_volume ( $p = 0.0010$ )
X11.59958380	rs2081545, <i>MS4A5A</i>	-	rh_G_occipital_middle_thickness ( $p < 0.0004$ )
X17.56409089	rs2632516, <i>BZRAP1-AS1</i>	+	lh_Pole_temporal_gauscurve ( $p = 0.0010$ )
X18.29088958	rs8093731, <i>SUZ12P1</i>	+	rh_G_S_cingul.Mid.Post_thickness ( $p = 0.0002$ ) rh_Lat_Fis.post_thickness ( $p = 0.0003$ ) rh_G_occipital_middle_thickness ( $p = 0.0010$ )
X20.54998544	rs6014724, <i>CASS4</i>	+	rh_G_S_occipital_inf_volume ( $p = 0.0010$ )
<b>(B) GM-NIM associations for the MCI cohort</b>			
X18.29088958	rs8093731, <i>SUZ12P1</i>	+	lh_MeanThickness ( $p = 0.0001$ ) rh_MeanThickness ( $p = 0.0001$ ) rh_S_temporal_sup_thickenss ( $p = 0.0001$ ) lh_G_pariet_inf.Supramar_thickness ( $p = 0.0001$ ) lh_G_temp_sup.Plan tempo_thickness ( $p = 0.0001$ ) rh_G_occipital_middle_thickness ( $p = 0.0001$ ) lh_G_occipital_middle_thickness ( $p = 0.0001$ ) rh_S_precentral.sup. part_thickness ( $p = 0.0001$ )
X19.1039323	rs111278892, <i>ABCA7</i>	+	lh_S_oc.temp_med.Lingual_thickness ( $p = 0.0001$ )
X20.54998544	rs6014724, <i>CASS4</i>	+	rh_S_oc.temp_med.Lingual_volume ( $p = 0.0001$ )
<b>(C) GM-NIM associations for the AD cohort</b>			
X10.11717397	rs11257238, <i>ECHDC3</i>	+	rh_S_precentral.inf. part_thickness ( $p = 0.0009$ )
X18.29088958	rs8093731, <i>SUZ12P1</i>	-	3rd.Ventricle_volume ( $p = 0.0007$ )
X20.54998544	rs6014724, <i>CASS4</i>	+	rh_bankssts_pct ( $p = 0.0009$ )

Abbreviations: AD, Alzheimer's disease; HC, healthy controls; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; SNP, single nucleotide polymorphism.

rh\_S\_temporal\_sup, both\_G\_precuneus, rh\_G\_cingul.Post.dorsal, rh\_S\_cingul.Marginalis, and rh\_G\_S\_cingul.Mid.Post) were networking according to the involvement of the 29 genes.

In the volume measure group, these NIMs (mainly TotalGrayVol, CortexVol, lhCortexVol\_volume, rhCortexVol, BrainSegVolNotVent, BrainSegVolNotVentSurf, rh\_S\_subparietal, SupraTentorialVolNotVentVox, SupraTentorialVolNotVent, SupraTentorialVolNotVentVox, rh\_S\_temporal\_sup, and rh\_G\_temporal\_middle) were networking according to the involvement of the 29 genes.

In the pct measure group, these NIMs (mainly rh\_inferiortemporal\_w.g., rh\_transversetemp\_w.g., rh\_isthmuscingulate\_w.g., rh\_inferiorparietal\_w.g., and rh\_fusiform\_w.g.) were networking according to the involvement of the 29 genes, see [Figure 2A](#).

In the MCI cohort, the 200 NIMs were not divided as in the HC subjects. All the NIMs intermingled without a clear subgroup clustering. These three measurement groups were networking diffusely without any grouping under the control of the 29 genes, regardless of the  $p$ -values in association analysis, which was described earlier in the association analysis section, see [Figure 2B](#).

In the AD cohort, we found some patterns between the 29 genes and the 200 NIMs. Three types of NIMs were divided into thickness, volume, and pct.

These three measurement groups independently tracked neuroimaging-genetic associations for the 29 genes in the AD subjects. The protocol for estimating the corresponding association  $p$ -values was described earlier in the association analysis section.

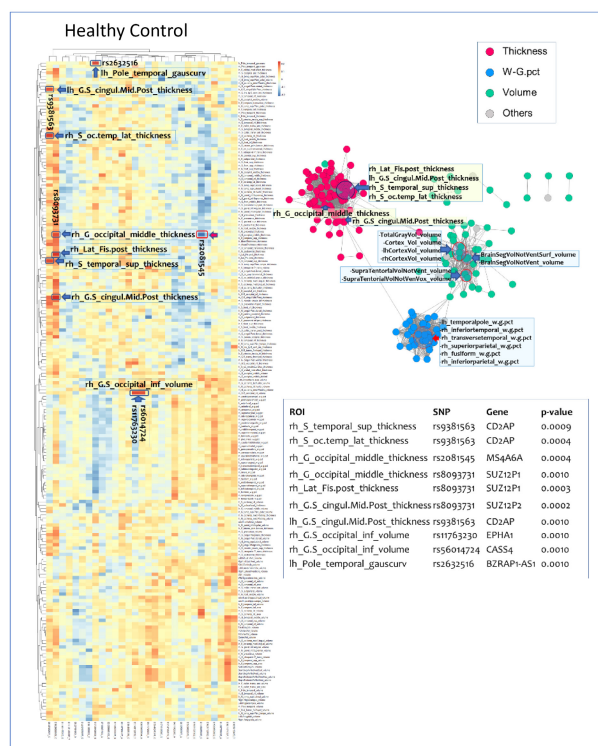
In the thickness measure group, these NIMs (mainly both\_MeanThickness, rh\_S\_parieto\_occipital, lh\_G\_S\_ocipital\_inf, rh\_G\_ocipital\_middle, rh\_G\_parietal\_sup, rh\_G\_pariet\_inf.Supramar, lh\_G\_postcentral, and lh\_S\_oc.middle.Lunatus) were involved in networks associated with the 29 SNPs.

In the volume measure group, several NIMs (mainly TotalGrayVol, CortexVol, lh/rhCortexVol, Left.Hipocampus, rh\_G\_temporal\_middle, lh\_G\_temp\_sup.Plan\_polar, BrainSegVolNotVentSurf, and SupraTentorialVolNotVentVox) were associated with the 29 genes.

In the pct measure group, a network of NIMs (mainly rh\_transversetemporal\_w.g., rh\_inferiortemporal\_w.g., rh\_superiortemporal\_w.g., rh\_middletemporal\_w.g., lh\_bankssts\_w.g., and rh\_insula\_w.g.) were associated with the 29 genes, see [Figure 2C](#).



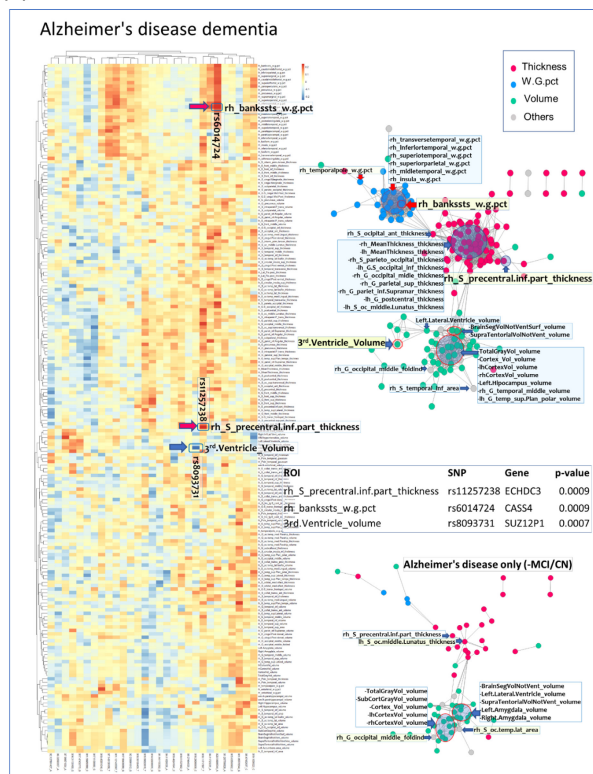
(A)



(B)



(C)



**FIGURE 2** Network analysis for (A) healthy controls (HC), (B) Mild Cognitive Impairment (MCI), and (C) Alzheimer's disease (AD).

### 3.5 | Statistically significant ORs: MCI/HC vs. AD/HC

Figures S1 and S2 show the NIMs and GMs which discriminated between AD and HC and between MCI and HC. The values of the "OR" and "RR",  $>1$  or  $<1$ , identify specific biomarkers (NIM or GM) whose increase or decrease affect the risk of MCI or AD dementia, relative to the asymptomatic HC volunteers.

The AD RR of Rh\_G\_occipital\_middle\_foldind is 0.9924 in AD RR depiction, but it is a rather higher point of 1.0136 in MCI RR depiction, see Figures S1 and S2. The AD RRs of both left and right (lh/rh) parahippocampal\_wg.pct, as well as (lh/rh)\_entorhinal\_wg.pct, lh\_midtemporal\_wg.pct, lh\_inferiotemporal\_wg.pct, and rh\_G\_occipital\_middle\_foldind were 0.9850/0.9882, 0.9887/0.9916, 0.9924, 0.9919, and 0.9924 in AD\_RR depiction, see Figure S1. In addition, the MCI RRs of (lh/rh) temporal pole\_wg.pct and (lh/rh)\_entorhinal\_wg.pct, and rh\_G\_occipital\_middle\_foldind was 0.9902/0.9871, 0.9923/0.9880, and 1.014 in MCI RR depiction. The RRs for MCI of both (lh/rh)\_parahippocampal\_wg.pct were 1.0036/0.9957. Thickness and volume measures were not associated with MCI, relative to HC, see Figure S2.

In this model, the 29 SNPs did not appear to be significantly associated with AD, relative to HC, see Figure S1. The RRs for MCI of X7.143108158\_T (EPHA1) and X19.1039323\_C (ABCA7) were 0.9964 and 0.9973, see Figure S2.

The observed accuracy of the multinomial model (AD/HC OR; multinomial linear model of diagnosis) was 0.62 (the expectation for a null model would be 0.33) with a 95% CI of (0.6025, 0.6371), corresponding to a  $p$ -value  $[\text{Acc} > \text{NIR}] < 2.2 \times 10^{-16}$  and Kappa = 0.3978. Figure S3 shows us predicted diagnosis depiction.

## 4 | DISCUSSION

### 4.1 | SNP selection

There are two options for identifying GMs associated with AD. The first one involves GWAS analysis on the ADNI data, which includes a relatively small sample compared to other GWASs with sample sizes of tens of thousands. Another approach is to narrow the scope by utilizing the top ranked SNPs reported in various online databases, e.g., IGAP ([http://web.pasteur-lille.fr/en/recherche/u744/igap/igap\\_download.php](http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php)), which includes summary results data from the largest (70K samples) GWAS in AD to date. This study utilized the second approach to identify the SNPs from existing databases like GWAS in IGAP, which is based on a much larger sample compared to ADNI (IGAP ~70,000 vs. ADNI ~1200). This enormous variation between sample sizes leads to differences in statistical power to detect effects of interest, i.e., identify SNPs highly associated in NIMs and/or specific clinical outcomes. In addition, IGAP results have been extensively validated and are widely assumed to be highly reliable.

Furthermore, Jansen et al. identified the 29 SNPs extracted by meta-analysis using PCG-ALZ, ADSP, UKB, and deCODE as well as IGAP. We used already identified 29 SNPs by Jansen et al.<sup>8</sup> Genotype

dosage data for AD-associated SNPs were extracted from the imputed ADNI genetics dosage data using sample-major additive coding. Such 29 SNPs were selected, representing a subset of independent SNPs found to be associated with AD in the recent AD meta-GWAS study.

### 4.2 | Neuroimaging-genetics association (prior to network analysis)

Several prior neuroimaging-genetics GWAS studies analyzed and reported specific NIMs as QTs.<sup>5,11,13</sup> There are some similarities with our study, however previous reports do not employ network analysis.

In our HC cohort, CD2AP (rs9381563), MS4A6A (rs2081545) and SUZ12P1 (rs8093731) were associated with thickness measures, and EPHA1 (rs11763230) and CASS4 (rs6014724) were associated with volume measures, especially in the occipital lobe. In general, however, cortical thickness measures were more significantly associated with the GMs compared to volume measures.

CD2AP (CD2 [cluster of differentiation 2] associated protein) loss of function is linked to enhanced A $\beta$  metabolism, tau-induced neurotoxicity, abnormal neurite structure modulation and reduced blood-brain barrier integrity.<sup>49</sup> CD2AP is expressed in both neuron and microglia in the brain and postulated to be involved in immune system regulation.<sup>46</sup> To the best of our knowledge, this is the first report of neuroimaging association of CD2AP gene to date. MS4A6A (membrane-spanning 4-domains, subfamily A, member 6A) has been identified as susceptibility loci of AD by several recent GWAS, whereas little is known about the potential roles of these variants in the brain structure of AD. While MS4A6A genotype was strongly positively related to atrophy rate in the left middle temporal, precuneus and entorhinal on MRI in the MCI subgroup in a study,<sup>50</sup> in our study, this GM was mainly negatively influencing cortical thickness in the right middle occipital gyrus in the HC. Although the exact mechanisms underlying the effects of MS4A6A on AD pathogenesis are still largely unknown, the role of MS4A6A in AD progression might be mediated by modifying neuroimaging changes, possibly by affecting immune system function.<sup>51,52</sup> SUZ12P1 (SUZ12 [suppressor of zeste 12] pseudogene 1) is a pseudogene, which is a long noncoding RNA promoting the proliferation of and inhibiting apoptosis of prostate cancer.<sup>53</sup> To our knowledge, the SUZ12P1 gene is not known to be associated with NIMs and this is the first report of such SUZ12P1-neuroimaging association. EPHA1 (ephrin type-A receptor 1) plays a role in immunity and endocytosis and regulates cell morphology and motility, including permeability of the blood-brain barrier to leucocytes.<sup>54</sup> EPHA1 is known to be highly expressed in cerebral cortex and hippocampus and plays a crucial role in cortical and axonal development. Wang et al. reported that AD patients with an allele of EPHA1 (A allele) having enhanced rate of cerebral glucose metabolism in the right lateral occipitotemporal gyrus may not have hippocampal atrophy. These authors suggested that in AD, EPHA1 expression can play a protective role,<sup>55</sup> even though in our study, we could not confirm this protective role. CASS4 (Cas scaffolding protein family

member 4) regulates focal adhesion integrity and cell spreading and plays a role in cytoskeleton and axon development and tau metabolism, affecting cellular adhesion, migration and motility.<sup>54</sup>

In the MCI cohort, *SUZ12P1* (rs8093731) was very significantly associated with a number of cortical thickness measures, and *CASS4* (rs6014724) was associated with volume measures. Cortical thickness and volume were significantly associated with various GMs, however, the statistical significance of thickness metrics was more profound. *SUZ12P1* was positively associated with some thickness measures in the HC cohort. There was some evidence that *SUZ12P1* may be associated with some thickness measures, which may facilitate tracking the transition from HC to the MCI stage. Similarly to the findings in the HC cohort, *CASS4* was positively associated with a volume measure in the occipital region. It is possible that *CASS4* may also be associated with volume measures in the occipital area and may be useful in forecasting transition from HC to the MCI stage. On the other hand, *ABCA7* (rs11278892) was positively associated with a thickness measure of the occipital area. Associations of *ABCA7* (ATP-binding cassette sub-family A member 7) with both brain atrophy<sup>56</sup> and amyloidosis, especially in the MCI stage, have been previously reported.<sup>57,58</sup> *ABCA7* may influence AD risk via cholesterol transfer to *APOE* or by clearing A $\beta$  aggregates.<sup>59,60</sup> The *ABCA7* gene is strongly expressed in the hippocampus subfield CA1<sup>61</sup> and showed a significant association with hippocampal atrophy<sup>56</sup> and gray matter density.<sup>62</sup> Like this, *ABCA7* is quite well known gene that is associated with some NIMs in AD.

In the AD cohort, some metrics of thickness, pct and volume (3rd ventricle) were significantly associated with the GMs. *ECHDC3* (rs11257238) was significantly associated with a thickness measure. The *ECHDC3* (enoyl CoA hydratase domain containing 3) is known to be involved in fatty acid biosynthesis in mitochondria.<sup>63</sup> Desikan et al. reported that the gene expression of *ECHDC3* was changed in opposite directions in the AD cohort,<sup>64</sup> and in our study, the gene was significantly associated with a thickness measure in the AD cohort. This gene may need to be further investigated in the future neuroimaging genetics studies. The *CASS4* was significantly associated with a pct measure in the AD cohort, which was different from the HC and MCI cohorts. The *SUZ12P1* was significantly associated with the 3rd ventricle, however, ventricular volume may also inversely reflect cortical volume changes. Finally, *SUZ12P1* and *CASS4* were significantly and differentially associated with some ROIs in the HC, MCI, and AD cohorts.

Cortical thickness is a signature marker for memory functioning across the adult lifespan. Among asymptomatic healthy individuals, the degree of cortical thinning predicts progression to clinical AD.<sup>65,66</sup> Moreover, our study suggests that cortical thickness may be an important measure of early detection of cognitive impairment progressing from HC to MCI and eventually leading to AD pathogenesis.<sup>67-69</sup> Previous findings suggest that (1) cortical thickness and cortical surface area are independent, both globally and regionally; and (2) gray matter volume is tracked by both metrics, even though cortical thickness is less influential than surface area.<sup>70,71</sup> We found meaningful associations between 200 NIMs and 29 GMs for cortical

thickness and regional ROI volume measures. However, cortical surface area appears as a less sensitive measure in individualized analysis within each cohort. Further studies are necessary to determine the intricate relationships between regional morphometry metrics, such as cortical volume, surface area, and thickness, specific genotypic markers, such as SNPs, and different clinical phenotypes.

On the other hand, our research finding suggests that cortical thickness may represent an important factor for tracking and discriminating subtle differences between HC and MCI cohorts. Yet, the importance of this association may not yet be extended to the AD cohort.<sup>5</sup> Our current research provides evidence that cortical thickness measures are important early on, prior to dementia onset, but their importance may taper off after dementia diagnosis.

All 29 GMs reported by Jansen, et al.<sup>8</sup> were found to play a crucial role in the immune system. Biological implications potentiate the hypothesis that AD pathogenesis involves an interplay between inflammation and lipids, as lipid changes might harm immune responses of microglia and astrocytes, and vascular health of the brain.<sup>72</sup> Whereas, our findings indicate that during the very early phases of the HC-MCI-AD progression, the disease pathogenesis may be detected by accurate measurements of cortical thickness change. Based on this assumption, cortical thickness changes might be detected initially as subtle brain changes before other measurements, such as volume, pct, surface area, folding index, gausscurve, etc. In the AD pathogenesis, most of these 29 GMs seem to be functioning in immune and lipid systems. However, we found that the associations between the genetic and the 200 NIMs are rather subtle and we suspect that the 29 GMs can affect specific neuroimaging changes that may be mediated by the broader functioning of immune and lipid systems. Delineating the precise molecular mechanisms linking 'genomic traits' with 'cognitive deficits' via 'immune system dysfunction' may reveal underlying mechanistic effects manifesting as brain anatomical changes (NIMs), genetic phenotypes (SNPs), and specific clinical outcomes (pathological states). Some evidence for this immune and lipid system hypothesis was recently reported.<sup>72,73</sup>

### 4.3 | Network analysis

To obtain the most reliable networks of 200 NIMs, threshold values for the network prediction were carefully chosen to meet the scale-free topology criterion as used in other recent studies based on biological network analysis.<sup>74-76</sup> No specific network patterns were identified jointly within all the 29 genes, which was contrary to our initial expectations. Hence, we performed network analyses using individual GMs.

We found specific network patterns within the respective HC and AD groups. The NIMs were divided into three thickness, volume and pct metrics under the control of 29 GMs. In the MCI cohort, transient intermingled stages without the patterning were traced and compared with the HC and AD cohorts. We found some specific AD-proper networking patterns (HC and MCI patterns were subtracted) as well between the GMs and the NIMs. In general, the two types of

NIMs were divided into thickness and volume measures. Pct measures were not highly impactful in this AD-proper networking, Figure 2D. This may imply that in the demented AD cohort, pct measures are not homogeneously tracking the observed disease pathogenesis.

#### 4.4 | Statistically significant ORs and RRs in MCI/HC and AD/HC comparisons

We also calculated the ORs to study the relationships of AD and MCI to HC in terms of AD pathogenesis. In AD RR, *rh\_G\_occipital\_middle\_foldind* (cortical folding index) is protective, but in MCI RR, it appears detrimental rather than protective. The folding index is a specific morphometric measure of cortical integrity. It complements other measures like surface area, volume, and cortical thickness. The folding index captures the regional cortical curvature patterns. We hypothesize that the folding index might play an inhibiting role (detrimental factor) for changing from HC to MCI, but its impact reverses (as a protective factor) for changing from MCI to AD dementia. Higher folding index values correspond to highly convolved surface folding, and lower index values correspond to flatter cortical surface patches.

In MCI RR, left and right parahippocampal\_wg.pct are involved as a risk factor and as a protective factor individually at the same time. Interestingly, *rh\_G\_occipital\_middle\_thickness* was negatively associated with rs2082545 (MS4A6A) in the HC cohort of our association analysis. This suggests that we need to pay attention to the occipital division as a NIM in terms of AD pathogenesis. Compared to the pct measures, thickness and volume measures did not appear as protective or detrimental factors, none of the 29 SNPs were highly associated with HC to AD transition. The genetic factors were less impactful, compared to the NIMs. This is not necessarily a problem as it's clear that single genes, or single SNPs, may not explain complex neurodegenerative phenomena such as memory loss, cognitive impairment, or clinical dementia. We aimed to identify statistically significant associations between regional, diagnostic, and genetic effects, using multinomial linear modeling and network analysis.

Basically, we suggest that similar genetic and epigenetic mechanisms continue to impact the structure and function of the brain throughout life. Early on, both genetics and experience guide neocortical and brain patterning, and these mechanisms continue to impact the maintenance of cortical areas and their boundaries as well as physiological area function throughout adulthood. Late in life, similar genetic mechanisms may be involved in the breakdown of brain microstructure as in early development, can either advance or ameliorate the deleterious effects of aging.<sup>5</sup> These ideas can be reflected in the pathogenesis of AD as well.

#### 4.5 | Limitations and future directions

Potential limitations of this study reflect the relatively small sample size to analyze genetic influence on NIMs. The ADNI sample was not

collected under a perfect epidemiological ascertainment strategy and the sample size was relatively small for a GWAS study, which may affect the generalizability of the findings. Because of our restricted statistical power, we were forced to constrain our analysis to SNPs that have been previously reported in Janssen et al. We used 29 SNPs from the Janssen et al. that are derived from the other diagnosis system from ADNI research subjects, which may also affect the generalizability of the findings.

For the neuroimaging genetics study, we used imputation tools to unify several separated ADNI data and to increase as much as possible the sample size of the genetic and neuroimaging ADNI data. This allowed us to aggregate the ADNI data and generate computable multimodal data objects including homogenous NIMs and GMs.

We did not manually inspect the brain scans of all participants (this is done by ADNI QC), to avoid potential subjective rater bias for location, size, or etiology of MRI-evident infarcts in the QC protocol. So, there is a potential that minor WMH effects may play a role in our analyses.

The sample only contained mild AD patients (CDR = 1), a relatively narrow range of illness, and is thus not fully representative of the HC-MCI-AD spectrum. At this point in time, ADNI does not collect gene expression/RNAseq data, and we could not complete a full network analysis in terms of neuroimaging genetics due to lack of available data.

Despite these challenges, the results are encouraging, and the proposed analytic framework appears to have a potential for enabling the discovery and localization of phenotypic imaging-genetics associations. We believe that imaging-genetics techniques offer important clues for the formulation of advanced methods of early detection, monitoring, and treatment of dementia.

## 5 | CONCLUSIONS

Structural brain changes are important indicators of progressive memory decline, cognitive impairment, and dementia progression. Network analysis pairing morphometric NIMs with genetic indicators allows investigation of clinical and phenotypic associations that facilitate deep systematic understanding of dementia pathogenesis. This neuroimaging-genetics study provides valuable clues to dementia onset and the prospective pathogenic trajectory. Our results are promising for untangling the intricate interrelations between brain anatomy and genetic phenotypes. Network analysis using neuroimaging measures and genotypic biomarkers provides cues to the structure of various deep brain-networks and assists with interpreting structural imaging-genomics association with disease. Further studies are necessary to reveal any specific mechanistic associations between GMs and NIMs and discover triggers or buffers of complex AD pathogenic traits.

#### ACKNOWLEDGMENTS

This research was partially supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF)



funded by the Ministry of Education (NRF-2019R11A3A01058477). Additional support came from the National Science Foundation (NSF) grants 1916425, 1734853, 1636840, 1416953, and 1023115, and the National Institutes of Health (NIH) grants UL1 TR002240, R01 CA233487, R01 MH121079, R01 MH126137, T32 GM141746. 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](http://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

The authors declare that this study has no conflicts of interest.

## DATA AVAILABILITY STATEMENT

The heterogeneous data types used in this study, which supported the reported research findings, are openly available from the archive of the Alzheimer's disease Neuroimaging Initiative (ADNI) ([adni.loni.usc.edu](http://adni.loni.usc.edu)).

## ORCID

Seok Woo Moon  <https://orcid.org/0000-0002-1074-8122>

Ivo D. Dinov  <https://orcid.org/0000-0003-3825-4375>

## REFERENCES

- Liu CC, Liu CC, Kanekiyo T, Xu H, Bu G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol*. 2013;9(2):106-118.
- Krstic D, Knuesel I. Deciphering the mechanism underlying late-onset Alzheimer disease. *Nat Rev Neurol*. 2013;9(1):25-34.
- Mayeux R. Alzheimer's disease: epidemiology. *Handb Clin Neurol*. 2008;89:195-205.
- Moon S, Dinov ID, Hobel S, et al. Structural brain changes in early-onset Alzheimer's disease subjects using the LONI pipeline environment. *J Neuroimaging*. 2015;25(5):728-737.
- Hofer E, Roshchupkin GV, Adams HHH, et al. Genetic correlations and genome-wide associations of cortical structure in general population samples of 22,824 adults. *Nat Commun*. 2020;11(1):4796.
- Ertekin-Taner N. Genetics of Alzheimer's disease: a centennial review. *Neurol Clin*. 2007;25(3):611-667. v.
- 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(12):1452-1458.
- Jansen IE, Savage JE, Watanabe K, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet*. 2019;51(3):404-413.
- Van Cauwenberghe C, Van Broeckhoven C, Sleegers K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genet Med*. 2016;18(5):421-430.
- Jack CR, Bernstein MA, Fox NC, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J Magn Reson Imaging*. 2008;27(4):685-691.
- Moon SW, Dinov ID, Kim J, et al. Structural neuroimaging genetics interactions in Alzheimer's disease. *J Alzheimers Dis*. 2015;48(4):1051-1063.
- Shen L, Kim S, Risacher SL, et al. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: a study of the ADNI cohort. *Neuroimage*. 2010;53(3):1051-1063.
- Hibar DP, Adams HHH, Jahanshad N, et al. Novel genetic loci associated with hippocampal volume. *Nat Commun*. 2017;8:13624.
- Potkin SG, Turner JA, Guffanti G, et al. Genome-wide strategies for discovering genetic influences on cognition and cognitive disorders: methodological considerations. *Cogn Neuropsychiatry*. 2009;14(4-5):391-418.
- Adams HH, Hibar DP, Chouraki V, et al. Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat Neurosci*. 2016;19(12):1569-1582.
- Sherva R, Tripodis Y, Bennett DA, et al. Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. *Alzheimers Dement*. 2014;10(1):45-52.
- Leduc V, De Beaumont L, Theroux L, et al. HMGCR is a genetic modifier for risk, age of onset and MCI conversion to Alzheimer's disease in a three cohorts study. *Mol Psychiatry*. 2015;20(7):867-873.
- Nathoo FS, Kong L, Zhu H, Initiative AsDN. A review of statistical methods in imaging genetics. *Can J Stat*. 2019;47(1):108-131.
- Harrison JR, Mistry S, Muskett N, Escott-Price V. From polygenic scores to precision medicine in Alzheimer's disease: a systematic review. *J Alzheimers Dis*. 2020;74:1271-1283.
- Gao W, Grewen K, Knickmeyer RC, et al. A review on neuroimaging studies of genetic and environmental influences on early brain development. *Neuroimage*. 2019;185:802-812.
- Kim H, Irimia A, Hobel SM, et al. LONI QC system: a semi-automated, web-based and freely-available environment for the comprehensive quality control of neuroimaging data. *Front Neuroinform*. 2019;13:60.
- Roussotte FF, Gutman BA, Madsen SK, Colby JB, Thompson PM. Alzheimer's disease neuroimaging I. combined effects of Alzheimer risk variants in the CLU and ApoE genes on ventricular expansion patterns in the elderly. *J Neurosci*. 2014;34(19):6537-6545.
- Wyman BT, Harvey DJ, Crawford K, et al. Standardization of analysis sets for reporting results from ADNI MRI data. *Alzheimers Dement*. 2013;9(3):332-337.
- Fischl B. FreeSurfer. *Neuroimage*. 2012;62(2):774-781.
- Reuter M, Rosas HD, Fischl B. Highly accurate inverse consistent registration: a robust approach. *Neuroimage*. 2010;53(4):1181-1196.



26. Ségonne F, Dale AM, Busa E, et al. A hybrid approach to the skull stripping problem in MRI. *Neuroimage*. 2004;22(3):1060-1075.
27. Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. 2002;33(3):341-355.
28. Fischl B, Salat DH, van der Kouwe AJ, et al. Sequence-independent segmentation of magnetic resonance images. *Neuroimage*. 2004;23(Suppl 1):S69-S84.
29. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging*. 1998;17(1):87-97.
30. Segonne F, Pacheco J, Fischl B. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. *IEEE Trans Med Imaging*. 2007;26(4):518-529.
31. Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci U S A*. 2000;97(20):11050-11055.
32. Fischl B, Sereno MI, Tootell RB, Dale AM. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp*. 1999;8(4):272-284.
33. Browning BL, Browning SR. Genotype imputation with millions of reference samples. *Am J Hum Genet*. 2016;98(1):116-126.
34. Moon S, Dinov ID, Zamanian A, et al. Alzheimer's disease neuroimaging initiative. Gene interactions and structural brain change in early-onset Alzheimer's disease subjects using the pipeline environment. *Psychiatry Invest*. 2015;12(1):125-135.
35. Dinov I, Torri F, Maciardi F, et al. Applications of the pipeline environment for visual informatics and genomics computations. *BMC Bioinformatics*. 2011;12(1):304.
36. Dinov I, Lozev K, Petrosyan P, et al. Neuroimaging study designs, computational analyses and data provenance using the LONI pipeline. *PLoS One*. 2010;5(9):e13070. doi:10.1371/journal.pone.0013070
37. Dinov I, Van Horn JD, Lozev KM, et al. Efficient, distributed and interactive neuroimaging data analysis using the LONI pipeline. *Front Neuroinform*. 2010;3(22):1-10.
38. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012;44(8):955-959.
39. Batenkov D. Open BEAGLE: a generic framework for evolutionary computations. *Genet Program Evolvable Mach*. 2011;12(3):329-331.
40. Dinov I, Velev M. *Data Science: Time Complexity, Inferential Uncertainty, and Space-time Analytics*. 1st ed. De Gruyter; 2021.
41. Dinov I. *Data Science and Predictive Analytics: Biomedical and Health Applications Using R*. Springer International Publishing; 2018.
42. Dinov ID. *Data Science and Predictive Analytics: Biomedical and Health Applications using R*. 2nd ed. Springer Series in Applied Machine Learning; 2023.
43. Husain S, Kalinin A, Truong A, Dinov ID. SOCR data dashboard: an integrated big data archive mashing medicare, labor, census and econometric information. *J Big Data*. 2015;2(13):1-18.
44. Christou N, Dinov ID. Confidence interval based parameter estimation—a new SOCR applet and activity. *PLoS One*. 2011;6(5):e19178.
45. Al-Aziz J, Christou N, Dinov I. SOCR motion charts: an efficient, open-source, interactive and dynamic applet for visualizing longitudinal multivariate data. *J Stat Educ*. 2010;18(3):1-29.
46. Srivatsan S, Swiecki M, Otero K, Cella M, Shaw AS. CD2-associated protein regulates plasmacytoid dendritic cell migration, but is dispensable for their development and cytokine production. *J Immunol*. 2013;191(12):5933-5940.
47. Yang Q, Cui J, Chazaro I, Cupples LA, Demissie S. Power and type I error rate of false discovery rate approaches in genome-wide association studies. *BMC Genet*. 2005;6(1):S134.
48. Brinster R, Köttgen A, Tayo BO, Schumacher M, Sekula P, CKDGen Consortium. Control procedures and estimators of the false discovery rate and their application in low-dimensional settings: an empirical investigation. *BMC Bioinformatics*. 2018;19(1):78.
49. Tao QQ, Chen YC, Wu ZY. The role of CD2AP in the pathogenesis of Alzheimer's disease. *Aging Dis*. 2019;10(4):901-907.
50. Ma J, Zhang W, Tan L, et al. MS4A6A genotypes are associated with the atrophy rates of Alzheimer's disease related brain structures. *Oncotarget*. 2016;7(37):58779-58788.
51. Ma J, Yu JT, Tan L. MS4A cluster in Alzheimer's disease. *Mol Neurobiol*. 2015;51(3):1240-1248.
52. Zuccolo J, Deng L, Unruh TL, et al. Expression of MS4A and TMEM176 genes in human B lymphocytes. *Front Immunol*. 2013;4:195.
53. Wan X, Huang W, Yang S, et al. Identification of androgen-responsive lncRNAs as diagnostic and prognostic markers for prostate cancer. *Oncotarget*. 2016;7(37):60503-60518.
54. Sims R, Hill M, Williams J. The multiplex model of the genetics of Alzheimer's disease. *Nat Neurosci*. 2020;23(3):311-322.
55. Wang HF, Tan L, Hao XK, et al. Effect of EPHA1 genetic variation on cerebrospinal fluid and neuroimaging biomarkers in healthy, mild cognitive impairment and Alzheimer's disease cohorts. *J Alzheimers Dis*. 2015;44(1):115-123.
56. Ramirez LM, Goukasian N, Porat S, et al. Common variants in ABCA7 and MS4A6A are associated with cortical and hippocampal atrophy. *Neurobiol Aging*. 2016;39:82-89.
57. Thambisetty M, An Y, Nalls M, et al. Effect of complement CR1 on brain amyloid burden during aging and its modification by APOE genotype. *Biol Psychiatry*. 2013;73(5):422-428.
58. Hughes TM, Lopez OL, Evans RW, et al. Markers of cholesterol transport are associated with amyloid deposition in the brain. *Neurobiol Aging*. 2014;35(4):802-807.
59. Kim WS, Li H, Ruberu K, et al. Deletion of Abca7 increases cerebral amyloid-beta accumulation in the J20 mouse model of Alzheimer's disease. *J Neurosci*. 2013;33(10):4387-4394.
60. Wildsmith KR, Holley M, Savage JC, Skerrett R, Landreth GE. Evidence for impaired amyloid beta clearance in Alzheimer's disease. *Alzheimers Res Ther*. 2013;5(4):33.
61. Rosenthal SL, Kamboh MI. Late-onset Alzheimer's disease genes and the potentially implicated pathways. *Curr Genet Med Rep*. 2014;2:85-101.
62. Stage E, Duran T, Risacher SL, et al. The effect of the top 20 Alzheimer disease risk genes on gray-matter density and FDG PET brain metabolism. *Alzheimers Dement*. 2016;5:53-66.
63. Eicher JD, Wakabayashi Y, Vitseva O, et al. Characterization of the platelet transcriptome by RNA sequencing in patients with acute myocardial infarction. *Platelets*. 2016;27(3):230-239.
64. Desikan RS, Schork AJ, Wang Y, et al. Polygenic overlap between C-reactive protein, plasma lipids, and Alzheimer disease. *Circulation*. 2015;131(23):2061-2069.
65. Dickerson BC, Stoub TR, Shah RC, et al. Alzheimer-signature MRI biomarker predicts AD dementia in cognitively normal adults. *Neurology*. 2011;76(16):1395-1402.
66. Busovaca E, Zimmerman ME, Meier IB, et al. Is the Alzheimer's disease cortical thickness signature a biological marker for memory? *Brain Imaging Behav*. 2016;10(2):517-523.
67. Batzu L, Westman E, Pereira JB. Cerebrospinal fluid progranulin is associated with increased cortical thickness in early stages of Alzheimer's disease. *Neurobiol Aging*. 2020;88:61-70.
68. Racine AM, Brickhouse M, Wolk DA, Dickerson BC. The personalized Alzheimer's disease cortical thickness index predicts likely pathology and clinical progression in mild cognitive impairment. *Alzheimers Dement*. 2018;10:301-310.
69. Yao Z, Hu B, Liang C, Zhao L, Jackson M, The Alzheimer's Disease Neuroimaging I. A longitudinal study of atrophy in amnesic mild cognitive impairment and normal aging revealed by cortical thickness. *PLoS One*. 2012;7(11):e48973.

70. Winkler AM, Kochunov P, Blangero J, et al. Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *Neuroimage*. 2010;53(3):1135-1146.
71. Panizzon MS, Fennema-Notestine C, Eyler LT, et al. Distinct genetic influences on cortical surface area and cortical thickness. *Cereb Cortex*. 2009;19(11):2728-2735.
72. Ai L, Me C. Inflammation in Alzheimer's Disease: Cross-Talk between Lipids and Innate Immune Cells of the Brain. 2015.
73. Misra A, Chakrabarti SS, Gambhir IS. New genetic players in late-onset Alzheimer's disease: findings of genome-wide association studies. *Indian J Med Res*. 2018;148(2):135-144.
74. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 2005;4:Article17.
75. Shi G, Shen Z, Liu Y, Yin W. Identifying biomarkers to predict the progression and prognosis of breast cancer by weighted gene co-expression network analysis. *Front Genet*. 2020;11:597888.
76. Pascut D, Pratama MY, Gilardi F, Giuffre M, Croce LS, Tiribelli C. Weighted miRNA co-expression networks analysis identifies

circulating miRNA predicting overall survival in hepatocellular carcinoma patients. *Sci Rep*. 2020;10(1):18967.

## 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:** Moon SW, Zhao L, Matloff W, et al. Brain structure and allelic associations in Alzheimer's disease. *CNS Neurosci Ther*. 2023;29:1034-1048. doi:[10.1111/cns.14073](https://doi.org/10.1111/cns.14073)