Whole-exome rare-variant analysis of Alzheimer's disease and related biomarker traits

Küçükali, Fahri; Neumann, Alexander; Van Dongen, Jasper; De Pooter, Tim; Joris, Geert; De Rijk, Peter; Ohlei, Olena; Dobricic, Valerija; Bos, Isabelle; Vos, Stephanie J B; Engelborghs, Sebastiaan; De Roeck, Ellen; Vandenberghe, Rik; Gabel, Silvy [**and 26 more**]

How to cite

KÜÇÜKALI, Fahri et al. Whole-exome rare-variant analysis of Alzheimer's disease and related biomarker traits. In: Alzheimer's & dementia, 2023, vol. 19, n° 6, p. 2317–2331. doi: 10.1002/alz.12842

This publication URL: https://archive-ouverte.unige.ch/unige:177199

Publication DOI: 10.1002/alz.12842

Received: 6 April 2022 | Revised: 16 September 2022 | Accepted: 28 September 2022

DOI: 10.1002/alz.12842

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

RESEARCH ARTICLE

Whole-exome rare-variant analysis of Alzheimer's disease and related biomarker traits

Fahri Küçükali[1,2] | Alexander Neumann[1,2] | Jasper Van Dongen[1,2] | Tim De Pooter[2,3] | Geert Joris[2,3] | Peter De Rijk[2,3] | Olena Ohlei[4] | Valerija Dobricic[4] | Isabelle Bos[5] | Stephanie J. B. Vos[6] | Sebastiaan Engelborghs[2,7] | Ellen De Roeck[2,8] | Rik Vandenberghe[9,10] | Silvy Gabel[9,10] | Karen Meersmans[9,10] | Magda Tsolaki[11] | Frans Verhey[6,12,13] | Pablo Martinez-Lage[14] | Mikel Tainta[14] | Giovanni Frisoni[15,16] | Oliver Blin[17] | Jill C Richardson[18] | Régis Bordet[19] | Alzheimer's Disease Neuroimaging Initiative (ADNI) | Philip Scheltens[20] | Julius Popp[21,22] | Gwendoline Peyratout[23] | Peter Johannsen[24] | Lutz Frölich[25] | Yvonne Freund-Levi[26,27,28] | Johannes Streffer[2] | Simon Lovestone[29,30] | Cristina Legido-Quigley[31,32] | Mara ten Kate[20,33] | Frederik Barkhof[33,34] | Henrik Zetterberg[35,36,37,38,39] | Lars Bertram[4,40] | Mojca Strazisar[2,3] | Pieter Jelle Visser[6,20] | Christine Van Broeckhoven[2,41] | Kristel Sleegers[1,2] | the EMIF-AD Study Group

[1]Complex Genetics of Alzheimer's Disease Group, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium

[2]Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium

[3]Neuromics Support Facility, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium

[4]Lübeck Interdisciplinary Platform for Genome Analytics (LIGA), University of Lübeck, Lübeck, Germany

[5]Netherlands Institute for Health Services Research, Utrecht, Netherlands

[6]Alzheimer Centrum Limburg, Maastricht University, Maastricht, Netherlands

[7]Department of Neurology and Brussels Integrated Center for Brain and Memory (Bru-BRAIN), Universitair Ziekenhuis Brussel (UZ Brussel) and Center for Neurosciences (C4N), Vrije Universiteit Brussel (VUB), Brussels, Belgium

[8]Department of Neurology and Memory Clinic, Hospital Network Antwerp (ZNA) Middelheim and Hoge Beuken, Antwerp, Belgium

[9]University Hospital Leuven, Leuven, Belgium

[10]Department of Neurosciences, Laboratory for Cognitive Neurology, KU Leuven, Leuven, Belgium

[11]1st Department of Neurology, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, Makedonia, Thessaloniki, Greece

[12]Department of Psychiatry and Neuropsychology, Maastricht University, Maastricht, Netherlands

[13]School for Mental Health and Neuroscience, Maastricht University, Maastricht, Netherlands

[14]Center for Research and Advanced Therapies, CITA-Alzheimer Foundation, San Sebastian, Spain

[15]Department of Psychiatry, Faculty of Medicine, Geneva University Hospitals, Geneva, Switzerland

[16]RCCS Instituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy

[17]Clinical Pharmacology & Pharmacovigilance Department, Marseille University Hospital, Marseille, France

[18]Neurosciences Therapeutic Area, GlaxoSmithKline R&D, Stevanage, United Kingdom

[19]Neuroscience & Cognition, CHU de Lille, University of Lille, Inserm, France

[20]Alzheimer Center and Department of Neurology, VU University Medical Center, Amsterdam, Netherlands
[21]Department of Geriatric Psychiatry, University Hospital of Psychiatry Zürich, Zürich, Switzerland
[22]Department of Psychiatry, Old Age Psychiatry, University Hospital of Lausanne, Lausanne, Switzerland
[23]Department of Psychiatry, University Hospital of Lausanne, Lausanne, Switzerland
[24]Clinical Drug Development, Novo Nordisk, Copenhagen, Denmark
[25]Department of Geriatric Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany
[26]School of Medical Sciences, Örebro University, Örebro, Sweden
[27]Departments of Geriatrics, University Hospital Örebro, Örebro, Sweden
[28]Department of Clinical Science and Education, Södertälje Hospital, Södersjukhuset Karolinska Institutet, Stockholm, Sweden
[29]Department of Psychiatry, University of Oxford, Oxford, United Kingdom
[30]Janssen Medical Ltd, High Wycombe, United Kingdom
[31]Steno Diabetes Center, Copenhagen, Denmark
[32]Institute of Pharmaceutical Sciences, King's College London, London, United Kingdom
[33]Department of Radiology and Nuclear Medicine, VU University Medical Center, Amsterdam, Netherlands
[34]Queen Square Institute of Neurology and Centre for Medical Image Computing, University College London, London, United Kingdom
[35]Department of Psychiatry and Neurochemistry, University of Gothenburg, Gothenburg, Sweden
[36]Department of Neurodegenerative Disease, UCL Institute of Neurology, London, United Kingdom
[37]Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden
[38]UK Dementia Research Institute, University College London, London, United Kingdom
[39]Hong Kong Center for Neurodegenerative Diseases, Hong Kong, China
[40]Centre for Lifespan Changes in Brain and Cognition, University of Oslo, Oslo, Norway
[41]Neurodegenerative Brain Diseases Group, VIB Center for Molecular Neurology, VIB, Antwerp, Belgium

Correspondence
Kristel Sleegers, Complex Genetics of Alzheimer's Disease Group, VIB Center for Molecular Neurology, University of Antwerp – CDE, Universiteitsplein 1, B-2610, Antwerp, Belgium.
Email: kristel.sleegers@uantwerpen.vib.be

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

Abstract

Introduction: Despite increasing evidence of a role of rare genetic variation in the risk of Alzheimer's disease (AD), limited attention has been paid to its contribution to AD-related biomarker traits indicative of AD-relevant pathophysiological processes.

Methods: We performed whole-exome gene-based rare-variant association studies (RVASs) of 17 AD-related traits on whole-exome sequencing (WES) data generated in the European Medical Information Framework for Alzheimer's Disease Multimodal Biomarker Discovery (EMIF-AD MBD) study ($n = 450$) and whole-genome sequencing (WGS) data from ADNI ($n = 808$).

Results: Mutation screening revealed a novel probably pathogenic mutation (*PSEN1* p.Leu232Phe). Gene-based RVAS revealed the exome-wide significant contribution of rare coding variation in *RBKS* and *OR7A10* to cognitive performance and protection against left hippocampal atrophy, respectively.

Discussion: The identification of these novel gene–trait associations offers new perspectives into the role of rare coding variation in the distinct pathophysiological processes culminating in AD, which may lead to identification of novel therapeutic and diagnostic targets.

KEYWORDS
Alzheimer's disease, biomarkers, endophenotypes, rare coding variants, whole-exome sequencing

1 | BACKGROUND

Alzheimer's disease (AD) is the most common cause of dementia, which affects millions of individuals worldwide, an estimate that could be doubled by 2060 in the absence of effective medical breakthroughs.[1] AD is a progressive neurodegenerative disease whose pathological hallmarks in the brain are extracellular amyloid plaques and intracellular neurofibrillary tangles formed by aggregates of amyloid beta (Aβ) and tau proteins, respectively.[2] As these changes typically occur years before the onset of first dementia symptoms,[3] disease progression is classified into three phases, a preclinical, a mild cognitive impairment (MCI), and an Alzheimer's dementia phase.[2] Numerous biomarkers for AD have been developed and characterized to better understand the disease process, for early detection of the disease, and for developing new disease-modifying treatments and monitoring. These include biochemical (cerebrospinal fluid [CSF] and plasma) and imaging biomarkers for Aβ pathology, tau pathology, neurodegeneration, synaptic dysfunction, glial activation, and neuroinflammation.[4] From these biomarker studies, a temporal sequence of biomarker changes has become apparent, as reviewed in detail in Zetterberg et al.[4]. Briefly, in the cognitively normal preclinical phase, the first biomarker changes toward an abnormal state are typically related to Aβ pathology, followed by abnormal increases in total tau (t-tau) and phosphorylated tau at residue 181 (p-tau$_{181}$), changes that are indicative of tau pathology and neurodegeneration in response to Aβ pathology; and an increase in CSF neurogranin (Ng), a relatively early marker of synaptic dysfunction. In relatively late stages of the preclinical phase and the MCI phase, neurodegeneration biomarker abnormalities become more apparent, such as neurofilament light (NfL) and brain atrophy. Elevated CSF chitinase-3-like protein 1 (YKL-40) levels, indicative of astrocytic activation, are observed in the relatively later phases of AD.[4] Finally, during MCI and AD phases, cognitive impairment can be detected using neuropsychological screening instruments such as the Mini-Mental State Examination (MMSE).[5]

Several studies have analyzed the contribution of genome-wide common genetic variation to these AD-relevant biomarkers and endophenotypes. Examples of genome-wide significant loci associated with CSF and imaging traits include apolipoprotein E (*APOE*),[6–9] *SUCLG2*,[10] *GLIS1*,[8] and *SERPINB1*,[8] for CSF Aβ_{42}; *APOE*,[7,8,11] *GMNC*,[7–9] *SRRM4*,[11] and *CEP170B/PLD4*,[11] for CSF t-tau; *APOE*,[7,8,11] *GLIS3*,[7,8] *PCDH8*,[8] *CTDP1*,[8] *GMNC*,[8] *NCR2*,[7] and *C16orf95*,[12] for CSF p-tau; *TMEM106B*[13] and *ADAMTS1*[14] for CSF NfL; *CHI3L1*[13] and *CPOX*[13] for CSF YKL-40; *HRK*,[15,16] *MSRB3*,[15,16] *APOE*,[11] and others[15] for hippocampal volume; *C15orf54*,[17] *C16orf95*,[18] and others[17,18] for cortical thickness; and *TRIM65*[19,20] and others including *APOE*[20] for white matter lesions. However, despite the presence of common variant associations of AD biomarker traits, the contribution of rare variants to these processes remains largely unexplored. Except for the studies on plasma Aβ[21] and white matter hyperintensities,[22] and our recent analysis of principal components (PCs) of CSF biomarkers,[23] no systematic analyses were conducted previously to study the relationship between exome-wide coding variation in each gene and AD biomarker traits and endophenotype outcomes.

RESEARCH IN CONTEXT

1. **Systematic review**: We reviewed the literature using sources such as PubMed and Google Scholar. Few studies investigated the effect of rare variants on single biomarker traits, but systematic analyses examining the role of rare coding variation in a large collection of Alzheimer's disease (AD)–relevant biomarker traits are lacking.
2. **Interpretation**: Our analyses revealed novel exome-wide significant contributions of rare coding variation in *RBKS* and *OR7A10* to cognitive performance and protection against left hippocampal atrophy, respectively. Moreover, subthreshold hits included numerous plausible gene–trait associations.
3. **Future directions**: This study shows a new landscape of rare coding variation associated with various AD-relevant pathophysiological processes. Future studies in larger cohorts/biobanks will allow further elucidation of these genetically associated molecular processes, which may aid the development of better therapeutic and preventive strategies for AD.

In this study, we conducted a systematic exome-wide, gene-based, rare-variant association study (RVAS) of 17 Alzheimer-relevant traits (as described in Table 1 and Tables S1–S3), including clinical, cognitive, CSF, and volumetric magnetic resonance imaging (MRI) phenotypes. These analyses were performed on a European multicenter whole-exome sequencing (WES) data set generated for $n = 450$ participants of the European Medical Information Framework for Alzheimer's Disease Multimodal Biomarker Discovery (EMIF-AD MBD) study.[24] Meta-analysis was performed including the Alzheimer's Disease Neuroimaging Initiative (ADNI)[25] whole-genome sequencing (WGS) data set on $n = 808$ participants.

2 | METHODS

2.1 | EMIF-AD MBD WES cohort and whole-exome sequencing

Participants were derived from the EMIF-AD MBD study, a European multicenter cohort of individuals with AD, MCI, and normal cognition (NC), for whom extensive molecular and phenotypic information is available.[24] From this cohort, we received DNA samples meeting the requirements for WES (see Supporting Information) for $n = 450$ participants from 10 European countries (Belgium [Flanders population], Denmark, France, Germany, Greece, Italy, The Netherlands, Spain [Basque population], Sweden, and Switzerland). The local medical ethical committee of each participant recruitment center approved the

TABLE 1 Phenotypic characteristics of EMIF-AD WES and ADNI WGS cohorts (analysis subsets)

			EMIF-AD WES		ADNI WGS	
			Analysis subset ($n = 442$)		Analysis subset ($n = 747$)	
	Characteristic/trait	Unit	*n*	Percentage or mean ± SD	*n*	Percentage or mean ± SD
Characteristics	Sex	Female %	442	50.7%	747	43.2%
	Age at participation	Years	442	70.12 ± 8.53	747	73.44 ± 7.02
	Age at last follow-up	Years	253	72.52 ± 8.73	741	80.01 ± 7.68
	Baseline diagnosis	AD %	442	23.6%	747	6.0%
		MCI %	442	42.1%	747	60.0%
		NC %	442	34.4%	747	34.0%
	Last available diagnosis	AD % (EMIF-AD) and dementia (ADNI)	442	31.4%	747	32.1%
		MCI %	442	30.6%	747	39.0%
		Other dementia %	442	3.2%	747	–
		NC %	442	34.8%	747	29.0%
	MCI - AD converter	Converted %	137	30.6%	407	41.8%
	APOE status	ε4 frequency %	442	27.9%	747	24.2%
Traits	AD vs NC	AD %	293	47.4%	–	–
	MMSE score	Score (0–30 range)	440	25.79 ± 4.34	747	28 ± 2.09
	CSF $A\beta_{42}$	pg/ml	352	295.51 ± 181.7	570	1053.54 ± 461.1
	CSF p-tau_{181}	Z-score (EMIF-AD) and pg/ml (ADNI)	356	0.62 ± 1.45	570	275.09 ± 114.13
	CSF t-tau	Z-score (EMIF-AD) and pg/ml (ADNI)	356	0.78 ± 1.46	570	26.13 ± 12.59
	CSF NfL	pg/ml	352	1315.39 ± 2394.16	125	1332.58 ± 1188.65
	CSF Neurogranin	pg/ml	345	127.44 ± 193.41	125	440.68 ± 291.32
	CSF YKL-40	pg/ml (EMIF-AD) and Z-score (ADNI)	353	176946.43 ± 67909.97	157	−0.11 ± 0.94
	CSF $A\beta_{42}$ status	Abnormal %	356	54.8%	570	44.7%
	CSF p-tau_{181} status	Abnormal %	356	48.6%	570	37.5%
	CSF t-tau status	Abnormal %	356	55.6%	570	34.7%
	Total hippocampal volume	mm/cm3	233	7132.44 ± 1157.97	240	6704.88 ± 1067.11
	Left hippocampal volume	mm/cm3	233	3527.18 ± 604.14	240	3271.38 ± 515.11
	Right hippocampal volume	mm/cm3	233	3605.25 ± 608.3	240	3433.5 ± 591.69
	Average cortical thickness (all regions)	mm	205	2.25 ± 0.12	–	–
	Average cortical thickness (AD signature regions)	mm	205	2.58 ± 0.16	–	–
	Fazekas scale	Score (0–3 range)	234	0.97 ± 0.73	–	–

Note: For each cohort, available clinical and biomarker information for *n* subjects is provided as either percentage of the indicated category or mean and standard deviation (SD) of the continuous measures. The analysis subset represents the subset that was used in the gene-based, rare-variant association analyses in this study. Hippocampal volumes in EMIF-AD were adjusted for intracranial volumes. For full cohort characteristics, see Table S1; for measurement details, see Table S2. The distributions of continuous measures are provided in Figure S4.

Abbreviations: AD, Alzheimer's disease; ADNI WGS, Alzheimer's Disease Neuroimaging Initiative whole-genome sequencing cohort; $A\beta_{42}$, amyloid beta 1-42 peptide; CSF, cerebrospinal fluid; EMIF-AD WES, European Medical Information Framework for Alzheimer's Disease whole-exome sequencing cohort; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination; NC, normal cognition; NfL, neurofilament light chain; p-tau_{181}, phosphorylated tau at amino acid 181; SD, standard deviation; t-tau, total tau; YKL-40, chitinase-3-like protein 1.

study. Subjects had provided written informed consent for use of data, samples, and scans.[24]

The EMIF-AD MBD WES cohort phenotypic characteristics are described in detail in Table 1 and Table S1. For the analysis sample ($n = 442$), 50.7% of the participants were female, mean age ± standard deviation (SD) at participation was 70.12 ± 8.53 years, and *APOE* ε4 allele prevalence was 27.9%. At baseline $n = 104$ individuals were diagnosed with AD, $n = 186$ individuals with MCI, and $n = 152$ individuals had NC. For 80% of the participants, CSF measurements were available for the following AD biomarkers (methods described previously[26]): amyloid beta 1-42 peptide ($A\beta_{42}$), phosphorylated tau at amino acid 181 (p-tau_{181}), total tau (t-tau), neurogranin (Ng), neurofilament light chain (NfL), and chitinase-3-like protein 1 (YKL-40). Furthermore, for $n = 233$ samples brain MRI scans were available, which include hippocampal volumes (total, left, and right), average cortical thickness (total and AD-signature region specific as defined in Jack et al.[27]) for $n = 205$; and Fazekas scale for grading the white matter lesion intensities for $n = 234$. Finally, baseline MMSE scores were available for $n = 440$ participants. The measurement details of these biomarkers, specific cutoffs to separate abnormal and normal groups for a given CSF biomarker, and primary references for these phenotypic details are provided in Table S2.

WES was performed at the Neuromics Support Facility of VIB-UAntwerp Center for Molecular Neurology, Belgium. DNA samples were hybridized with SeqCap EZ Human Exome Kit v3.0 (Roche). We sequenced a maximum number of 12 indexed sample libraries per run on a NextSeq500 (Illumina), generating 90.8 ± 11.2 million reads per sample on average and spanning 93.8% ± 1.63% of the targeted sites with at least 20 reads (> 20x coverage) per sample on average.

2.2 | ADNI WGS cohort

Whenever possible, data from the EMIF-AD MBD WES cohort were meta-analyzed with comparable traits from the ADNI (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of the ADNI has been to test whether serial MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. WGS of 808 ADNI participants was performed on a HiSeq200 (Illumina) at about 30–40x coverage.

The ADNI WGS cohort phenotypic characteristics are described in Table 1 and Table S1. For the analysis sample ($n = 747$), 43% of the participants were female, the mean age ± SD at participation was 73.44 ± 7.02 years, and *APOE* ε4 allele prevalence was 24.2%. At baseline, $n = 45$ individuals were diagnosed with AD, $n = 448$ with MCI, and $n = 254$ individuals had NC. For 570 subjects, baseline CSF $A\beta_{42}$, p-tau_{181}, and t-tau levels were available; CSF Ng and NfL measurements were available for 125 participants and CSF YKL-40 measurements for 157 participants. Furthermore, on 240 participants MRI-derived baseline left, right, and total hippocampal volumes were measured. Finally, for all participants baseline MMSE scores were measured. The details of these traits are provided in Table S2.

2.3 | Bioinformatic processing and quality control

For EMIF-AD WES data, the sequencing reads were aligned to hg19 human reference genome with Burrows-Wheeler Aligner (BWA) 0.7.15.[28] Variant calling was performed using the Genome Analysis Toolkit (GATK) 4.0.3,[29] followed by Variant Quality Score Recalibration (VQSR). For ADNI WGS data (accessed in April 2020 through LONI portal, https://ida.loni.usc.edu/), we accessed the multi-sample VCFs created with GATK Best Practices. Both data sets were processed with the same bioinformatic processing and sample and variant quality control (QC) pipeline (see Supporting Information). Three EMIF-AD participants and five ADNI participants were excluded from the genetic association analyses due to relatedness (PI-HAT > 0.1). Another 53 ADNI participants with estimated European ancestry proportion less than 80% were excluded from the genetic association analyses to avoid confounding due to population stratification. After sample QC and selection, we included 442 EMIF and 747 ADNI participants for downstream genetic association analyses (Table 1).

2.4 | Mutation screening and Sanger validations

We screened for known pathogenic neurodegenerative disease mutations as described in Alzforum Mutation Database and ClinVar (accessed December 2020; Table S4) and predicted loss-of-function (pLoF) deleterious rare variants (study level minor allele frequency [MAF] <1%, Combined Annotation Dependent Depletion [CADD] score ≥20) in *ABCA7* and *SORL1*.[30–36] These mutations in the EMIF-AD MBD cohort were validated using Sanger sequencing (see Supporting Information).

2.5 | Statistical analyses

Gene-based optimal sequence kernel association test (SKAT-O)[37] in the R package SKAT v2.0.1 was used to test the association of the combined effect of rare variants (MAF <1% and genotype missingness <15%) in each gene across the exome on the tested phenotypes. Details of the statistical analyses are provided in Supporting Information. Briefly, two different models were assessed: a protein-altering (missense, nonsense, frameshift, and splice-site disrupting) model and a predicted LoF model (excluding missense variants in the protein-altering model). The carriers of known neurodegenerative disease pathogenic mutations and the genes with <2 rare-variant carriers per cohort (<4 in meta-analysis) were excluded from the genetic association analyses. We normalized continuous outcomes using rank-based inverse normal transformation (INT) in R; and showed the untransformed and transformed distributions in Figure S4 and Shapiro-Wilk test results for normal distributions in Table S2. Because INT did

not perform well for MMSE, and count models are not implicated in SKAT-O, for significant gene associations with MMSE we performed an additional Quasi-Poisson regression model, which is limited to a burden-like model, to verify that findings were not driven by skewness. Covariates used in the statistical models included sex, age (age at measurement for biomarkers, and age at first AD diagnosis for patients or age at last clinical visit for controls), diagnosis at the time of measurement, first four genetic PCs (calculated separately with respect to the subsets of individuals included in each analysis), age squared, and number of *APOE ε4* alleles; a full covariate list for all tested traits can be found in Table S3. For meta-analysis of outcomes and genes that could be tested in both cohorts, we used a multimarker extension of the random-effects meta-analysis[38] allowing for heterogeneous genetic effects as implemented in MetaSKAT (v0.81)[39] in R. For parameters and settings of both SKAT-O and MetaSKAT-O, method was set as "SKATO," and default beta (1,25) weights were used. Estimates of size and direction of effect were obtained by fitting general linear model.

We conducted meta-analyses on all traits of interest, with the exception of three MRI-derived imaging traits (cortical thickness of all regions, of AD-signature regions, and Fazekas scale) and the AD case-control diagnosis comparison because of data availability in $n < 50$ cases in ADNI and/or differences in phenotype definition. The exome-wide and suggestive significance thresholds for each tested phenotype were determined with Bonferroni correction (Table S3), $\alpha = 0.05$/number of genes tested per trait analyzed, as recommended by Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA).[40] Because fewer genes harbor pLoF mutations, these analyses do not cover the full exome; therefore, the significance threshold for these models is referred to as "multiple testing-adjusted significance." In addition, we provide an alternative conservative threshold, which adjusts for testing seven non-derivative outcomes as determined according to Li and Ji's methodology.[41] We caution that this approach is likely too conservative, as outcomes are correlated and the need for multiple outcome adjustment is debated (see also Discussion).[42,43]

3 | RESULTS

3.1 | Mutation screening for dementia genes, SORL1, and ABCA7

We identified pathogenic mutations in known dementia genes (Table S5), including a novel *PSEN1* mutation (p.Leu232Phe) identified in an early-onset Alzheimer's disease (EOAD; ≤65 years old at the time AD diagnosis) patient from The Netherlands. Moreover, for well-established AD risk genes *SORL1* and *ABCA7*, for which pLoF mutations of intermediate-to-high penetrance were reported previously, we identified previously reported and novel deleterious mutations (Table S6).

Figure S7 shows normalized CSF $A\beta_{42}$, p-tau_{181}, and t-tau profiles of the carriers of these screened mutations, where *ABCA7* pLoF mutation carriers were at increased odds of having abnormal $A\beta_{42}$ (odds ratio $[OR]_{sum} = 4.38$, 95% CI 1.09–17.7), abnormal p-tau_{181} ($OR_{sum} = 6.27$, 95% CI 1.7–23.1), and abnormal t-tau ($OR_{sum} = 3.81$, 95% CI 1.22–11.9) at nominally significant levels compared to non-carriers.

3.2 | Exome-wide, gene-based, rare-variant association analyses

We performed exome-wide, gene-based RVAS on 17 AD-related phenotypes using both a protein-altering and an LoF model in the EMIF-AD WES cohort. For 13 phenotypes, a meta-analysis could be performed using the ADNI WGS cohort. Cohort-specific and meta-analysis results are presented in a tabular format in Table 2 and Tables S8–S14; and as Manhattan plots in Figures 1–2 and Figures S8–S15. The quantile-quantile (QQ) plots of all gene-based, rare-variant association tests are shown in Figures S5 and S6.

Two phenotypes (MMSE and left hippocampal volume) showed an exome-wide significant rare-variant association in meta-analysis for *RBKS* and *OR7A10*, respectively, as described in subsequent text; and one phenotype (Fazekas scale) that was available only in EMIF-AD MBD WES cohort showed an multiple testing-adjusted significant signal for *ZBTB4* (see Supporting Information). Details on other genes reaching significance in a specific cohort or suggestive association in meta-analysis are provided in Supporting Information.

3.2.1 | RBKS pLoF rare variants and MMSE

For rare variant meta-analysis of MMSE scores across EMIF-AD and ADNI cohorts, a total of 1187 individuals were included. We identified a multiple testing-adjusted signific association signal for ribokinase gene (*RBKS*) pLoF rare variants (MetaSKAT-O $p = 1.58 \times 10^{-5}$; Figure 1A). Quasi-Poisson regression analysis in EMIF-AD and ADNI was in line with SKAT-O analysis on INT MMSE ($Quasi\text{-}Poisson_{untransformed\text{-}burden}$ $p = 2.37 \times 10^{-5}$ and 5.77×10^{-3} for EMIF-AD and ADNI; $Gaussian_{INT\text{-}burden}$ $p = 1.14 \times 10^{-4}$ and 9.43×10^{-3} for EMIF-AD and ADNI; $SKAT\text{-}O_{INT}$ $p = 1.18 \times 10^{-4}$ and 9.44×10^{-3}, respectively). The gene harbors two pLoF mutations, that is, rs140948699, a splice acceptor site mutation (CADD = 33), and rs142879777, a frameshift deletion mutation (CADD = 34) (Figure 1B). Together, they were identified in 21 individuals across both cohorts, and were associated with relatively lower MMSE scores ($\beta_{sum} = -0.72$, 95% CI −1.19 to −0.24) (Figure 1C and Table S7). Furthermore, *RBKS* pLoF mutations were also nominally associated with decreased CSF $A\beta_{42}$ levels in EMIF-AD (SKAT-O $p = 0.028$, $\beta = -0.83$, 95% CI −1.54 to −0.11).

3.2.2 | *OR7A10* protein-altering rare variants and left hippocampal volume

We observed an exome-wide significant association between the olfactory receptor family 7 subfamily A member 10 gene (*OR7A10*) and

TABLE 2 Exome-wide significant, multiple testing-adjusted significant, and suggestive gene-based coding rare variant meta-analysis results

Domain	Trait	Gene	Model	*n*	No. variants	No. variants shared	cMAC	No. carriers	MetaSKAT-O *P*	Summary β or *OR*	Lower 95% CI	Upper 95% CI
Cognitive	MMSE score	HID1	Protein-alt.	1187	11	1	19	19	4.22×10^{-5}	0.85	0.00	1.71
		RBKS	LoF	1187	2	2	21	21	$\mathbf{1.58 \times 10^{-5}}$	−0.72	−1.19	−0.24
		FSIP1	LoF	1187	4	1	5	5	7.90×10^{-4}	1.04	0.43	1.65
CSF	CSF $A\beta_{42}$	CYR61	Protein-alt.	922	4	2	21	21	2.65×10^{-5}	−0.54	−1.41	0.32
		ZNF90	LoF	922	2	1	7	7	1.87×10^{-4}	−0.85	−1.97	0.28
		HAP1	LoF	922	1	1	4	4	5.63×10^{-4}	−1.52	−2.35	−0.70
	CSF p-tau_{181}	COLGALT2	Protein-alt.	926	13	1	20	18	5.33×10^{-5}	−0.76	−1.13	−0.40
		PLA2R1	LoF	926	4	1	10	10	2.24×10^{-4}	−1.07	−1.63	−0.51
	CSF t-tau	ZNF365	Protein-alt.	926	9	1	12	11	4.81×10^{-5}	−1.01	−1.47	−0.54
		PLA2R1	LoF	926	4	1	10	10	6.57×10^{-5}	−1.17	−1.81	−0.54
	CSF NfL	NLRC3	Protein-alt.	477	26	2	36	35	1.11×10^{-5}	0.31	0.03	0.59
	CSF YKL-40	CHI3L1	Protein-alt.	510	7	2	12	12	8.05×10^{-5}	−0.82	−1.33	−0.31
	CSF p-tau_{181} status	GDPD4	Protein-alt.	926	9	1	23	22	4.98×10^{-5}	*NA*	*NA*	*NA*
	CSF t-tau status	STK3	Protein-alt.	926	4	2	22	21	7.03×10^{-5}	*7.67*	*2.28*	*25.77*
MRI	Left hippo. vol.	**OR7A10**	Protein-alt.	473	9	–	15	7	$\mathbf{1.94 \times 10^{-6}}$	0.44	0.24	0.63
		VCPKMT	Protein-alt.	473	4	–	4	4	1.88×10^{-5}	1.50	0.74	2.25
		C1QTNF9B	Protein-alt.	473	8	1	12	12	3.20×10^{-5}	0.67	−1.37	2.70
		SLC17A8	Protein-alt.	473	5	1	6	6	4.46×10^{-5}	0.97	0.25	1.69
		PAPD4	Protein-alt.	473	6	1	7	7	7.54×10^{-5}	0.87	−0.13	1.87
		APCDD1	Protein-alt.	473	10	2	14	14	8.26×10^{-5}	0.48	0.07	0.89
		FAM173B	LoF	473	2	1	8	8	2.55×10^{-3}	−0.89	−1.44	−0.34
	Right hippo. vol.	FAM173B	LoF	473	2	1	8	8	1.45×10^{-3}	−0.91	−1.44	−0.37
	Total hippo. vol.	RD3	Protein-alt.	473	2	1	8	8	7.91×10^{-5}	1.11	0.58	1.64
		FAM173B	LoF	473	2	1	8	8	4.67×10^{-4}	−0.98	−1.50	−0.46

Note: Meta-analysis results are shown for each exome-wide and suggestive significant association identified per respective domain, trait, and model type. In order, *n* is total number of subjects included in the meta-analysis, No. variants is the number of total unique variants tested and No. variants shared is the number of variants that were observed in both cohorts. cMAC is cumulative minor allele count and No. carrier shows total number of carriers. Exome-wide significant and multiple testing-adjusted significant hits are in bold, and random-effects meta-analysis of effect sizes are shown in the last three columns as summary odds ratio (OR, in italics) or beta coefficient (β) effect sizes (estimated from generalized linear models) with 95% Wald's confidence interval. Analyses are adjusted for sex, age, and diagnosis at the time of measurement, first four genetic principal components, age squared, number of *APOE* e4 alleles, and intracranial volume for hippocampal measurements. OR_{sum} of *GDPD4* is not available because the model did not converge. Details of cohort-specific associations for these highlighted meta-analysis associations are available in Table S6.

Abbreviations: $A\beta_{42}$, amyloid beta 1-42 peptide; cMAC, cumulative minor allele count; CSF, cerebrospinal fluid; hippo. vol., hippocampus volume; LoF, loss-of-function only model; MMSE, Mini-Mental State Examination; MRI, magnetic resonance imaging; NfL, neurofilament light chain; p-tau_{181}, phosphorylated tau at amino acid 181; Protein-alt, protein-altering model; t-tau, total tau.

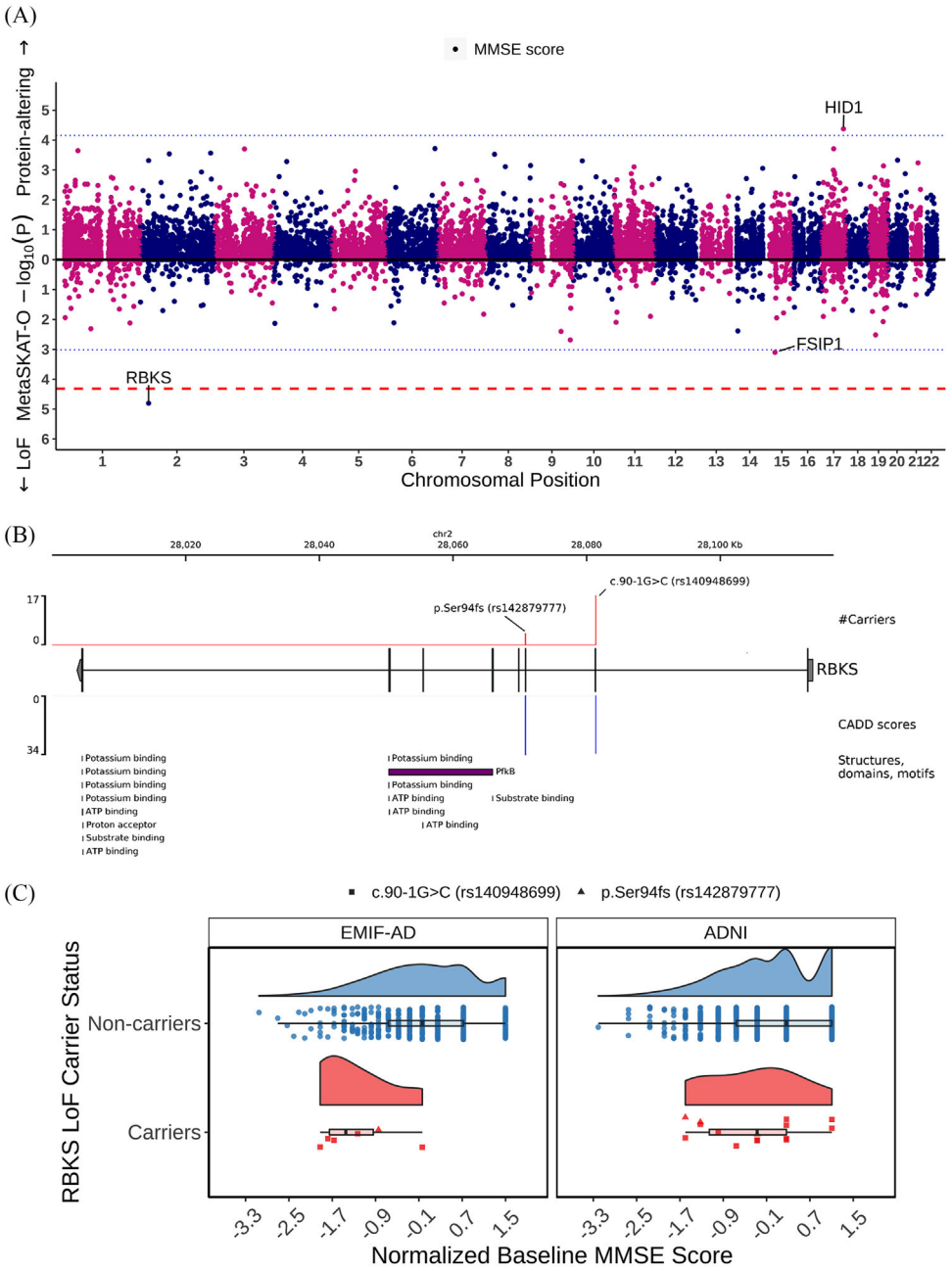

FIGURE 1 (A) Manhattan plot of MMSE score MetaSKAT-O results. Protein-altering (positive y-axis) and LoF-only (negative y-axis) gene-based, rare-variant association results on MMSE scores are plotted as two mirrored Manhattan plots on the x-axis. Exome-wide significance threshold is indicated with a red dashed line and suggestive significance threshold with a blue dotted line (as described in Table S3), and all the genes passing these thresholds are labeled on the plot. (B) Schematic representation of the identified pLoF mutations in *RBKS* associated with MMSE score. The canonical transcript of *RBKS* (ENST00000302188.3) was plotted, where the light blue color represents the protein-coding sequences and the gray color represents non-coding (UTR) sequences of the transcript. From top to bottom, the track descriptions are: hg19-based chromosomal position, the number of mutation carriers in both cohorts in red, CADD PHRED scores (v1.6) for predicted deleterious effects of these variants shown in blue, and known structures, motifs, post-translational modification sites, topological domains, and functional domains of canonical protein isoform retrieved from UniProt shown in purple. (C) Raincloud plot of MMSE scores of study cohorts stratified by *RBKS* pLoF mutation carrier status. The distribution of the normalized baseline MMSE scores are shown in both cohorts based on *RBKS* pLoF mutation carrier status (blue: non-carriers, red: carriers). For the carrier group, specific non-circular shapes were additionally used to represent the distinct *RBKS* pLoF mutations they have. Abbreviations: ADNI, Alzheimer's Disease Neuroimaging Initiative; CADD, Combined Annotation Dependent Depletion; EMIF-AD, European Medical Information Framework for Alzheimer's Disease; LoF, loss-of-function; MMSE, Mini-Mental State Examination; pLoF, predicted loss-of-function.

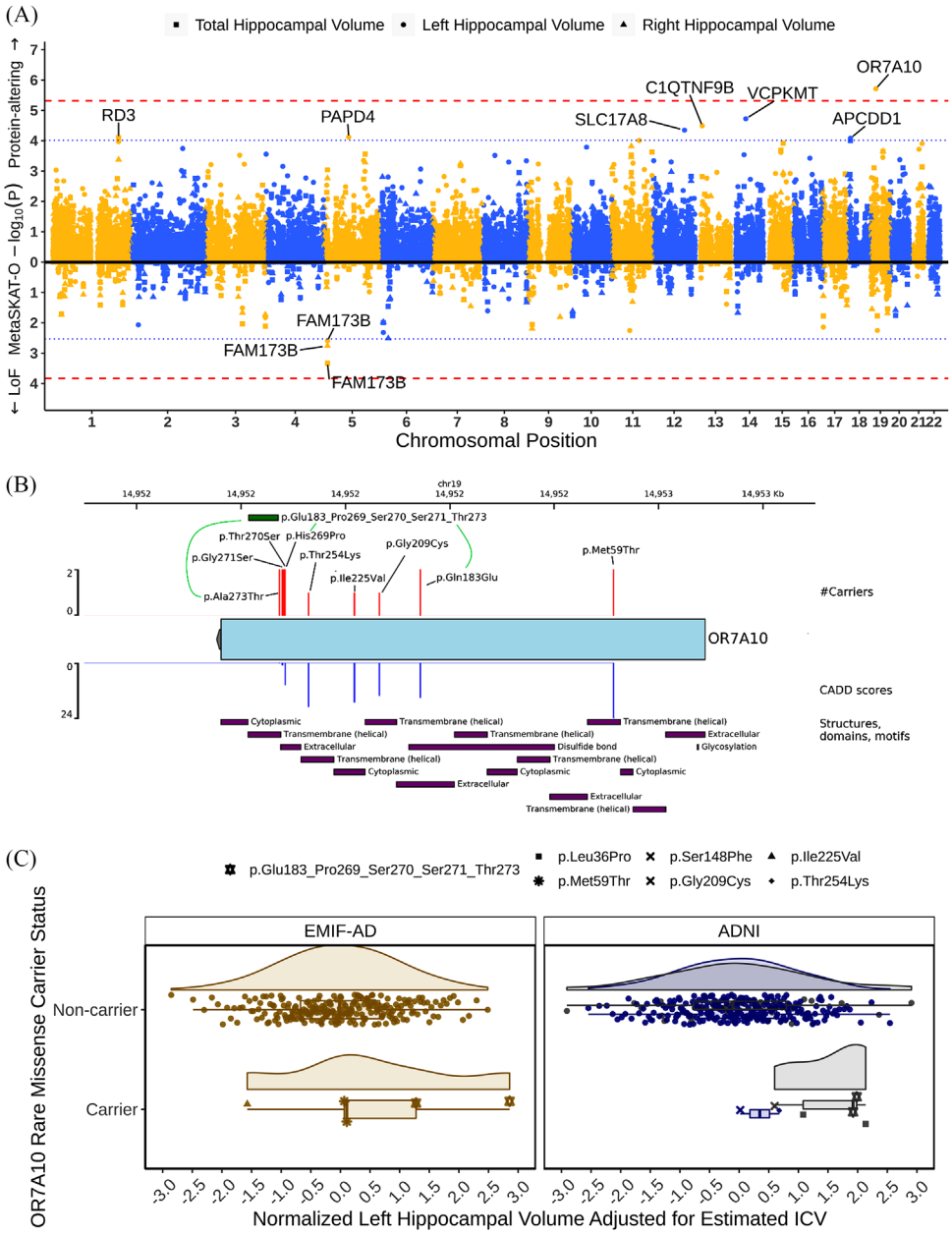

FIGURE 2 (A) Manhattan plot of total, left, and right hippocampal volume MetaSKAT-O results. Protein-altering (positive y-axis) and LoF-only (negative y-axis) gene-based, rare variant association results on the tested phenotypes are plotted as two, mirrored Manhattan plots on the x-axis, separated by different shapes for associations that represents the tested trait according to the legend. Exome-wide significance threshold is indicated with a red dashed line and suggestive significance threshold with a blue dotted line (as described in Table S3), and all the genes passing these thresholds are labeled on the plot. (B) Schematic representation of the identified protein-altering mutations in *OR7A10* associated with left hippocampal volume. The canonical transcript of *OR7A10* (ENST00000248058.1) was plotted, where the light blue color represents the protein-coding sequences and the gray color represents non-coding (UTR) sequences of the transcript. From top to bottom, the track descriptions are: hg19-based chromosomal position, the number of mutation carriers in both cohorts in red, CADD PHRED scores (v1.6) for predicted deleterious effects of these variants shown in blue, and known structures, motifs, posttranslational modification sites, topological domains, and functional domains of canonical protein isoform retrieved from UniProt shown in purple. Furthermore, the p.Glu183_Pro269_Ser270_Ser271_Thr273 haplotype in *OR7A10* consisting of five, rare missense variants is indicated in green. (C) Raincloud plot of left hippocampal volumes of study cohorts stratified by *OR7A10* protein-altering mutation carrier status. The distribution of the normalized left hippocampal volumes (adjusted for EICV) were shown in yellow for EMIF-AD MBD participants; meanwhile for ADNI participants, it was indicated in two colors, with the analysis subset being shown in blue and outliers (total of 35 samples, 16 excluded due to non-EUR genetic ancestry, and 19 excluded for lacking a baseline measurement) shown in gray. For the carrier group, specific non-circular shapes were additionally used to represent distinct *OR7A10* protein-altering mutations or the p.Glu183_Pro269_Ser270_Ser271_Thr273 haplotype they have. Abbreviations: ADNI, Alzheimer's Disease Neuroimaging Initiative; CADD, Combined Annotation Dependent Depletion; EUR, European; EMIF-AD, European Medical Information Framework for Alzheimer's Disease; ICV: intracranial volume; EICV: estimated intracranial volume; LoF, loss-of-function.

left hippocampal volume (MetaSKAT-O $p = 1.94 \times 10^{-6}$; Figure 2A and Table S7). The gene harbored nine rare variants found in seven carriers (cMAC = 15), which were associated with increased hippocampal volume ($\beta_{sum} = 0.44$, 95% CI 0.24–0.63). Of note, two individuals carried a haplotype (p.Glu183_Pro269_Ser270_Ser271_Thr273) consisting of five rare missense variants (four located in the extracellular domain), greatly contributing to the association signal (Figure 2B,C the 1st and the 24th highest measures in $n = 233$ EMIF subjects). This haplotype was not detected in the European (EUR) ancestry subset of 1000 Genomes (1KG)[44]; however, its frequency was between 1.7% and 10.3% in non-EUR ancestry participants of the 1KG data set (Figure S17). In fact, we also observed this haplotype in the admixed American (AMR) and African (AFR) ancestry participants from the ADNI cohort, which were excluded from the association analyses due to genetic ancestry differences. These two non-EUR ancestry carriers also had relatively large left hippocampal volumes (see Figure 2C; the seventh and the eighth highest measures in ADNI). Of note, two of these haplotype carriers had MCI at the time of hippocampal volume measurement (at the ages of 68 and 76), and the other two were cognitively normal (at the age of 73). Moreover, the association between rare variants in *OR7A10* and right hippocampal volume was in the same positive direction ($\beta_{sum} = 0.08$, 95% CI −0.11 to 0.27) for right hippocampal volume, but of lower magnitude and overall gene effects were not statistically significant (MetaSKAT-O $p = 0.31$). However, considering both sides, these variants were nominally associated with total hippocampal volumes in the expected direction as well (MetaSKAT-O $p = 1.9 \times 10^{-3}$, $\beta = 0.25$, 95% CI 0.06–0.44). Furthermore, rare variation in *OR7A10* was also nominally associated with decreased CSF t-tau (SKAT-O $p = 0.024$, $\beta = -0.39$, 95% CI −0.72 to −0.06) and CSF p-tau$_{181}$ levels (SKAT-O $p = 0.048$, $\beta = -0.34$, 95% CI −0.66 to −0.02), but only in ADNI.

4 | DISCUSSION

Herein we describe the first comprehensive WES analysis of multiple biomarker modalities relevant to AD. Specifically, we performed a mutation screening and a systematic exome-wide gene-based RVAS in two multi-center case-control studies. We report two novel gene–endophenotype associations, which may shed new light on pathophysiological processes in the AD continuum. First, we found that rare pLoF variants in *RBKS* are associated with lower cognitive performance as measured by the MMSE score. Second, rare missense variants in *OR7A10* were found to be associated with left hippocampal volume.

For *RBKS*, two rare pLoF variants were observed in both cohorts, and both were associated with lower MMSE score. *RBKS* encodes ribokinase, which catalyzes phosphorylation of D-ribose to D-ribose-5-phosphate.[45] *RBKS* pLoF mutations could, therefore, possibly affect cognitive performance through a decrease in catalysis of D-ribose. Of interest, two recent studies reported that urine D-ribose levels were correlated negatively with MMSE scores in an AD case-control cohort[46] and in a larger sample of community-dwelling older individuals,[47] which would be in line with this hypothesis. Another study reported a potential rescue of D-ribose dysmetabolism in rats with benfotiamine (BTMP) treatment, leading to decreased aging, tau hyperphosphorylation, and neurodegeneration.[48] BTMP was previously shown to improve the cognitive performance of patients with mild-to-moderate AD, independent of brain amyloidosis.[49] In fact, a phase 2 clinical trial for BTMP in AD is ongoing (ClinicalTrials.gov ID: NCT02292238). Furthermore, *RBKS* is significantly downregulated in the frontal and temporal lobes of AD patients (Agora platform). Our new observation that *RBKS* rare pLoF variant carriers have lower MMSE scores complement these observations, thereby warranting further exploration for potential implications. Of note, the association between pLoF in *RBKS* only reached multiple testing-adjusted significance for MMSE, and not for more precise biomarkers such as CSF tau or Aβ_{42}. This could be due to smaller sample sizes for the latter traits (we did observe nominal association for CSF Aβ_{42}), but it could also suggest that loss of RBKS has an effect on cognitive function upstream or independent of amyloidosis, tauopathy, or neuronal loss, for example, disruption of cellular energy production via mitochondrial dysfunction[50] or formation of advanced glycation end products via ribosylation.[51] Moreover, it should be noted that we did not adjust for education years in MMSE score analyses: (1) because it was not available for 16% of subjects in EMIF-AD cohort and (2) because of limited informativeness due to different educational systems and cultural differences among the 10 countries participating in EMIF-AD.

OR7A10 is a member of olfactory receptor genes, positioned within an olfactory G protein-coupled receptor (*GPCR*) gene cluster locus on chromosome 19. Its function is not yet known; however, a recent ADNI imaging study based on common genetic markers revealed that two protein–protein interactions (PPIs) containing *OR7A10* were suggestively associated with cortical thickness.[52] We observed that *OR7A10* missense variants strongly affect left hippocampal volume, especially a five-variant haplotype that modifies the extracellular residues of the protein, which could potentially affect receptor-ligand interactions. The possible protective effect against left hippocampal atrophy of the five-variant haplotype could be studied further in populations of non-European ancestry with increased haplotype frequency.

By meta-analyzing whole-exome genetic and biomarker data of near 1200 EMIF-AD and ADNI participants, we detected exome-wide significant association for several gene-trait pairs. Compared to genome-wide association studies (GWASs), where association signals are often found in non-coding regions of the genome and determination of the causal gene typically requires post–GWAS analyses,[53] one of the main advantages of exome-wide analyses of rare coding variation is a more direct determination of the potential causal links between the gene and the trait. Differences in cohort characteristics or inter-site variability in biomarker measurements should be taken into account to avoid bias. Here, we used rank-based INT to normalize and standardize raw phenotype values, which allows better comparison of phenotypes between cohorts. In our study we opted for SKAT-O to run gene-based, rare-variant association tests, but we acknowledge that other tests such as pure burden, pure SKAT, and ACAT do exist and are being used in similar studies. Within the context and aims of our study, which used numerous traits and two distinct cohorts in association testing,

we aimed to limit extra multiple testing burden by using the SKAT-O framework, which tests for optimal association under a range of models ranging from pure burden to pure SKAT models, while correcting for the number of models tested. However, specific tests relevant for different research questions could be considered to detect additional signals in future studies.

Our study has several limitations as well. Despite being the largest study of its kind, combining a rich array of endophenotype data from two independent data sets, the study still lacked power for several traits. We did observe some plausible candidate genes among the subthreshold associations, described in full in Supporting Information, including *NLRC3* for CSF NfL, *FAM173B* for hippocampal volumes, and *WNK2* for cortical thickness. Of note, among these was also a suggestive association between CSF YKL-40 and rare variants in *CHI3L1*, which encodes the YKL-40 protein. This "proof-of-concept" observation strongly suggests that these subthreshold associations may harbor additional true signals, warranting further replication. Indeed, our initial subthreshold findings in this study can be a new starting point for larger-scale studies that may use our study for increasing sample size and boosting statistical power, in combination with emerging AD cohorts and biobanks with similar exome and phenotype data in near future. Second, because of power considerations, we performed only cross-sectional analyses. Future longitudinal analyses on sufficiently large data sets will be of clear interest to investigate how rare variants affect biomarker changes over time in relation to AD. In this light, it is noteworthy that the association between *RBKS* and MMSE is driven by two variants that were observed in both cohorts, one of which has an MAF close to 1%. This opens up opportunities for imputation in large-scale GWAS of longitudinal measures of cognitive decline. Third, to avoid confounding, this study was performed in individuals of European ancestry only, but efforts to generate similar data sets in populations of different ancestries is recommended to reveal novel insights due to population-specific variants or enrichment of alleles—as might be the case for the five-variant haplotype in *OR7A10*.

A strength of the study is the comprehensive assessment of many AD biomarkers; however, this may increase the chance of false-positive findings. Under the most stringent multiple testing adjustment, *RBKS* would not be considered significantly associated. However, this adjustment is likely too conservative, as it does not take into account the dependence between outcomes, and the need to adjust for multiple outcomes is debated. Several researchers[42,43] argue that the number of outcomes pertaining to a family of tests is arbitrarily defined and that adjustment for multiple outcomes increases type II error, and encourages paper splitting and the use of smaller studies.[42,43]

In a concurrent study,[23] we performed a joint multivariate analysis of multiple CSF biomarkers in $n = 480$ EMIF-AD and ADNI participants, which resulted in the identification of six novel exome-wide significant associations. *IFFO1*, *DTNB*, *NLRC3* and *SLC22A10* associated with a neuronal injury and inflammation PC, loading on NfL and YKL-40. In this study, these genes also associated with relevant biomarkers with at least nominal significance in univariate models. Similarly, *GABBR2* and *CASZ1* associated with a synaptic functioning component, loading on Ng, and also showed nominally significant associations with Ng in univariate analyses. Multivariate approaches may thus offer a power advantage in rare variant analyses, as reported previously in GWAS studies of common variants,[54] and could therefore be explored further for other biomarkers at the potential cost of interpretability.

In addition to reporting two new gene-trait associations, we identified and validated a novel *PSEN1* mutation (p.Leu232Phe) in a patient with EOAD. We propose the pathogenicity of this mutation as probable based on the Guerreiro classification,[55] as it is in a conserved site between *PSEN1* and *PSEN2*, and all mutations reported to date in the same TM5 domain were pathogenic (Alzforum Mutation Database), including a pathogenic mutation (p.Leu232Pro) at the same residue in a Korean patient with EOAD.[56] However, further investigation of the familial history is required to determine if *PSEN1* p.Leu232Phe is definitely pathogenic. We further identified and validated novel pLoF mutations in *SORL1* and *ABCA7*. In line with the literature,[36] *SORL1* mutations were detected only in patients, whereas relatively more frequent *ABCA7* mutations were also detected in cognitively normal individuals. However, cognitively normal *ABCA7* mutation carriers showed preclinical CSF biomarker abnormalities. In fact, although not reaching multiple testing-adjusted significance, the *ABCA7* LoF model was the top-associated hit for p-tau$_{181}$, increasing the likelihood of an abnormal p-tau$_{181}$ status ~6 times compared to non-carriers with a similar clinical diagnosis. This suggests that *ABCA7* pLoF mutations might be contributing to early AD pathology. In line with this, an ADNI PET imaging study[57] showed that the risk allele of the GWAS common lead variant in *ABCA7* is significantly associated with increased amyloidosis, an effect that is more pronounced in asymptomatic and early stages of AD.

In summary, the systematic exome-wide gene-based RVAS of 17 AD-related traits in two independent cohorts of individuals along the AD continuum revealed the exome-wide significant contribution of rare coding variation in *RBKS* and *OR7A10* to cognitive performance and protection against left hippocampal atrophy, respectively. In addition, subthreshold hits included numerous plausible candidate genes as well, warranting further replication. The mutation screening revealed several new mutations in known causal or risk-increasing genes. Taken together, our results collectively revealed new perspectives into the contribution of rare coding variation to AD and its relevant biomarker traits that are indicative of distinct AD pathophysiological processes. Future work will be needed to better understand and resolve the underlying molecular processes that could be impacted by these newly identified rare variations, which may ultimately lead to the potential identification of novel therapeutic and diagnostic targets for AD.

ACKNOWLEDGMENTS

The authors thank the participants and families who took part in this research. The authors would also like to thank all people involved in data and sample collection and/or logistics across the different centers.

EMIF: The present study was conducted as part of the European Medical Information Framework for Alzheimer's Disease (EMIF-AD) project, which has received support from the Innovative Medicines Initiative Joint Undertaking under the European Medical Information Framework (EMIF) grant agreement no. 115372, the European

Prevention of Alzheimer's Dementia (EPAD) grant no. 115736, and from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 666992, resources of which are composed of financial contributions from the European Union's Seventh Framework Program (FP7/2007-2013) and the European Federation of Pharmaceutical Industries and Association (EFPIA) companies' in-kind contribution. The Development of Screening Guidelines and Criteria for Predementia Alzheimer's Disease (DESCRIPA) study was funded by the European Commission within the fifth framework program (QLRT-2001-2455). The Beta Amyloid Oligomers in Early Diagnosis of AD and as Marker for Treatment Response (EDAR) study was funded by the European Commission within the fifth framework program (contract no. 37670). The San Sebastian Gipuzkoa Alzheimer Project (GAP) study is partially funded by the Department of Health of the Basque Government (allocation 17.0.1.08.12.0000.2.454.01.41142.001.H). The Leuven cohort was funded by the Stichting voor Alzheimer Onderzoek (grant nos. 11020, 13007, and 15005). The Lausanne cohort was supported by grants from the Swiss National Research Foundation (SNF 320030_141179), Synapsis Foundation - Alzheimer Research Switzerland (grant no. 2017-PI01). F.K. is recipient of a PhD fellowship of the University of Antwerp Research Fund. H.Z. is a Wallenberg Scholar supported by grants from the Swedish Research Council (no. 2018-02532), the European Research Council (no. 681712), Swedish State Support for Clinical Research (no. ALFGBG-720931), the Alzheimer Drug Discovery Foundation (ADDF), USA (no. 201809-2016862), the AD Strategic Fund and the Alzheimer's Association (nos. ADSF-21-831376-C, ADSF-21-831381-C, and ADSF-21-831377-C), the Olav Thon Foundation, the Erling-Persson Family Foundation, Stiftelsen för Gamla Tjänarinnor, Hjärnfonden, Sweden (#FO2019-0228), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860197 (MIRIADE), and the UK Dementia Research Institute at University College London. F.B. is supported by the National Institute for Health and Care Research (NIHR) biomedical research centre at University College London Hospitals (UCLH). This work was supported for Y.F-L. by grants from the Petrus and Augusta Hedlunds Foundation, the Gun och Bertil Stohnes Foundation, the Loo and Hans Osterman Foundation, the Demensförbundet, Brain Foundation "Hjärnfonden" (grant FO2018-0315), "Särfond 31 Forskning Senil demens," Region Örebro län, "Stiftelsen for Gamla Tjänarinnor," and Demensfonden, Stockholm and Nyckelfonden Region Örebro Län.

ADNI: Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. 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 no. 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.

CONFLICTS OF INTEREST

H.Z. has served at scientific advisory boards and/or as a consultant for Abbvie, Alector, Annexon, AZTherapies, CogRx, Denali, Eisai, Nervgen, Pinteon Therapeutics, Red Abbey Labs, Passage Bio, Roche, Samumed, Siemens Healthineers, Triplet Therapeutics, and Wave; has given lectures in symposia sponsored by Cellectricon, Fujirebio, Alzecure, and Biogen; and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program. J.P. received consultation honoraria from Nestle Institute of Health Sciences, Ono Pharma, OM Pharma, and Fujirebio, unrelated to the submitted work. F.B. is on the editorial board of *Neurology*, *Radiology*, *MSJ*, and *Neuroradiology*, for the latter receiving compensation; receives personal fees from Springer, personal fees from Biogen, grants from Roche, grants from Merck, grants from Biogen, personal fees from IXICO Ltd, grants from the Innovative Medicines Initiative - European Union (IMI-EU), grants from GE Healthcare, grants from UK MS Society, grants from Dutch Foundation MS Research, grants from Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), grants from National Institute for Health and Care Research (NIHR), and personal fees from Combinostics, outside the submitted work. S.L. is currently an employee of Janssen Medical Ltd (UK), a cofounder of Akrivia Health Ltd (UK), and within the past 5 years has filed patents related to biomarkers unrelated to the current work and advised or given lectures for Merck, Optum Labs, and Eisai as well as having received grant funding from multiple companies as part of European Union (EU) Innovative Medicines Initiative (IMI) programmes and from Astra Zeneca. S.E. has served on scientific advisory boards for Biogen, Danone, icometrix, Novartis, Nutricia, and Roche, and has received unrestricted research grants from Janssen Pharmaceutica and ADx Neurosciences (paid to institution). The other authors declare that

there is no conflict of interest. Author disclosures are available in the Supporting Information.

DATA AVAILABILITY AND WEB SOURCES

To comply with EU law and participant privacy, individual-level clinical data from EMIF-AD cannot be shared publicly, however can be requested via EMIF-AD website (see https://emifcatalogue.eu and https://www.emif.eu/about/emif-ad). ADNI data can be accessed via ADNI portal (see https://adni.loni.usc.edu/) after registration and approval. Up to top ten associated genes for each trait and model are provided in Tables S11–S14 in supporting information. The full summary statistics results and analysis scripts will be made publicly available upon publication via https://github.com/SleegersLab-VIBCMN/AD_Biomarkers_RareVariantAnalyses repository.

The rest of the public online sources used in this study are listed below:

FastQC, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
BCFtools, https://samtools.github.io/bcftools/bcftools.html
PLINK, https://www.cog-genomics.org/plink/1.9/
gnomAD, https://gnomad.broadinstitute.org/
Healthy Exomes (HEX), https://www.alzforum.org/exomes/hex
Alzforum Mutations Database, https://www.alzforum.org/mutations
ClinVar, https://www.ncbi.nlm.nih.gov/clinvar/
Integrative Genomics Viewer (IGV), https://software.broadinstitute.org/software/igv/
Phase 3 VCFs of 1KG samples, https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/
UniProt, https://www.uniprot.org/
R, https://www.r-project.org/
pyGenomeTracks, https://github.com/deeptools/pyGenomeTracks
ggplot2, https://ggplot2.tidyverse.org/
rmeta, https://cran.r-project.org/web/packages/rmeta/index.html
UniProt, https://www.uniprot.org/
LDlink, https://ldlink.nci.nih.gov/
Agora Platform, https://agora.ampadportal.org/genes
ClinicalTrials.gov, https://clinicaltrials.gov/

REFERENCES

1. Gaugler J, James B, Johnson T, Reimer J, Weuve J. Alzheimer's disease facts and figures. *Alzheimers Dement*. 2021;17(3):327-406. https://doi.org/10.1002/alz.12328
2. Knopman DS, Amieva H, Petersen RC, et al. Alzheimer disease. *Nat Rev Dis Prim*. 2021;7(1):1-21. https://doi.org/10.1038/s41572-021-00269-y
3. Villemagne VL, Burnham S, Bourgeat P, et al. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *Lancet Neurol*. 2013;12(4):357-367. https://doi.org/10.1016/S1474-4422(13)70044-9
4. Zetterberg H, Bendlin BB. Biomarkers for Alzheimer's disease – preparing for a new era of disease-modifying therapies. *Mol Psychiatry*. 2021;26(1):296-308. https://doi.org/10.1038/s41380-020-0721-9
5. Folstein MF, Folstein SE, McHugh PR. Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12(3):189-198. https://doi.org/10.1016/0022-3956(75)90026-6
6. Kim S, Swaminathan S, Shen L, et al. Genome-wide association study of CSF biomarkers Aβ1-42, t-tau, and p-tau181p in the ADNI cohort. *Neurology*. 2011;76(1):69-79. https://doi.org/10.1212/WNL.0b013e318204a397
7. Cruchaga C, Kauwe JSK, Harari O, et al. GWAS of cerebrospinal fluid tau levels identifies risk variants for alzheimer's disease. *Neuron*. 2013;78(2):256-268. https://doi.org/10.1016/j.neuron.2013.02.026
8. 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(5):839-856. https://doi.org/10.1007/s00401-017-1685-y
9. Hong S, Prokopenko D, Dobricic V, et al. Genome-wide association study of Alzheimer's disease CSF biomarkers in the EMIF-AD Multimodal Biomarker Discovery dataset. *Transl Psychiatry*. 2020;10(1):403. https://doi.org/10.1038/s41398-020-01074-z
10. Ramirez A, van der Flier WM, Herold C, et al. SUCLG2 identified as both a determinator of CSF Aβ1-42 levels and an attenuator of cognitive decline in Alzheimer's disease. *Hum Mol Genet*. 2014;23(24):6644-6658. 10.1093/hmg/ddu372
11. Chung J, Wang X, Maruyama T, et al. Genome-wide association study of Alzheimer's disease endophenotypes at prediagnosis stages. *Alzheimers Dement*. 2018;14(5):623-633. https://doi.org/10.1016/j.jalz.2017.11.006
12. Jansen IE, van der Lee SJ, Gomez-Fonseca D, et al. Genome-wide meta-analysis for Alzheimer's disease cerebrospinal fluid biomarkers. *Acta Neuropathol*. 2022;144(5):821-842. https://doi.org/10.1007/s00401-022-02454-z
13. Hong S, Dobricic V, Ohlei O, et al. TMEM106B and CPOX are genetic determinants of cerebrospinal fluid Alzheimer's disease biomarker levels. *Alzheimers Dement*. 2021;17(10):1628-1640. https://doi.org/10.1002/alz.12330
14. Niu L-D, Xu W, Li J-Q, et al. Genome-wide association study of cerebrospinal fluid neurofilament light levels in non-demented elders. *Ann Transl Med*. 2019;7(22):657-657. https://doi.org/10.21037/atm.2019.10.66
15. van der Meer D, Rokicki J, Kaufmann T, et al. Brain scans from 21,297 individuals reveal the genetic architecture of hippocampal subfield volumes. *Mol Psychiatry*. 2020;25(11):3053-3065. https://doi.org/10.1038/s41380-018-0262-7
16. Bis JC, DeCarli C, Smith AV, et al. Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nat Genet*. 2012;44(5):545-551. https://doi.org/10.1038/ng.2237
17. Grasby KL, Jahanshad N, Painter JN, et al. The genetic architecture of the human cerebral cortex. *Science (80-)*. 2020;367(6484):eaay6690. https://doi.org/10.1126/science.aay6690
18. 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. https://doi.org/10.1038/s41467-020-18367-y
19. Fornage M, Debette S, Bis JC, et al. Genome-wide association studies of cerebral white matter lesion burden. *Ann Neurol*. 2011;69(6):928-939. https://doi.org/10.1002/ana.22403
20. Sargurupremraj M, Suzuki H, Jian X, et al. Cerebral small vessel disease genomics and its implications across the lifespan. *Nat Commun*. 2020;11(1):6285. https://doi.org/10.1038/s41467-020-19111-2
21. Simino J, Wang Z, Bressler J, et al. Whole exome sequence-based association analyses of plasma amyloid-β in African and European Americans; the Atherosclerosis Risk in Communities-Neurocognitive

Study. Sleegers K (Ed). *PLoS One*. 2017;12(7):e0180046. https://doi.org/10.1371/journal.pone.0180046
22. Malik R, Beaufort N, Frerich S, et al. Whole-exome sequencing reveals a role of HTRA1 and EGFL8 in brain white matter hyperintensities. *Brain*. 2021;144(9):2670-2682. https://doi.org/10.1093/brain/awab253
23. Neumann A, Küçükali F, Bos I, et al. Rare variants in IFFO1, DTNB, NLRC3 and SLC22A10 associate with Alzheimer's disease CSF profile of neuronal injury and inflammation. *Mol Psychiatry*. 2022;27(4):1990-1999. https://doi.org/10.1038/s41380-022-01437-6
24. Bos I, Vos S, Vandenberghe R, et al. The EMIF-AD Multimodal Biomarker Discovery study: design, methods and cohort characteristics. *Alzheimers Res Ther*. 2018;10(1):64. 10.1186/s13195-018-0396-5
25. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology*. 2010;74(3):201-209. https://doi.org/10.1212/WNL.0b013e3181cb3e25
26. Bos I, Vos S, Verhey F, et al. Cerebrospinal fluid biomarkers of neurodegeneration, synaptic integrity, and astroglial activation across the clinical Alzheimer's disease spectrum. *Alzheimers Dement*. 2019;31:1-11. https://doi.org/10.1016/j.jalz.2019.01.004
27. Jack CR, Wiste HJ, Weigand SD, et al. Defining imaging biomarker cut points for brain aging and Alzheimer's disease. *Alzheimers Dement*. 2017;13(3):205-216. https://doi.org/10.1016/j.jalz.2016.08.005
28. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589-595. 10.1093/bioinformatics/btp698
29. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a Mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303. https://doi.org/10.1101/gr.107524.110
30. Cuyvers E, De Roeck A, Van den Bossche T, et al. Mutations in ABCA7 in a Belgian cohort of Alzheimer's disease patients: a targeted resequencing study. *Lancet Neurol*. 2015;14(8):814-822. 10.1016/S1474-4422(15)00133-7
31. De Roeck A, Van den Bossche T, van der Zee J, et al. Deleterious ABCA7 mutations and transcript rescue mechanisms in early onset Alzheimer's disease. *Acta Neuropathol*. 2017;134(3):475-487. https://doi.org/10.1007/s00401-017-1714-x
32. De Roeck A, Van Broeckhoven C, Sleegers K. The role of ABCA7 in Alzheimer's disease: evidence from genomics, transcriptomics and methylomics. *Acta Neuropathol*. 2019;138(2):201-220. https://doi.org/10.1007/s00401-019-01994-1
33. Verheijen J, Van den Bossche T, van der Zee J, et al. A comprehensive study of the genetic impact of rare variants in SORL1 in European early-onset Alzheimer's disease. *Acta Neuropathol*. 2016;132(2):213-224. https://doi.org/10.1007/s00401-016-1566-9
34. Bellenguez C, Charbonnier C, Grenier-Boley B, et al. Contribution to Alzheimer's disease risk of rare variants in TREM2, SORL1, and ABCA7 in 1779 cases and 1273 controls. *Neurobiol Aging*. 2017;59:220.e1-220.e9. https://doi.org/10.1016/j.neurobiolaging.2017.07.001
35. Campion D, Charbonnier C, Nicolas G. SORL1 genetic variants and Alzheimer disease risk: a literature review and meta-analysis of sequencing data. *Acta Neuropathol*. 2019;138(2):173-186. https://doi.org/10.1007/s00401-019-01991-4
36. Holstege H, Hulsman M, Charbonnier C, et al. Exome sequencing identifies rare damaging variants in the ATP8B4 and ABCA1 genes as novel risk factors for Alzheimer's Disease. *MedRxiv*. Published online 2021. https://doi.org/10.1101/2020.07.22.20159251
37. Lee S, Emond MJ, Bamshad MJ, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet*. 2012;91(2):224-237. https://doi.org/10.1016/j.ajhg.2012.06.007
38. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am J Hum Genet*. 2011;88(5):586-598. https://doi.org/10.1016/j.ajhg.2011.04.014
39. Lee S, Teslovich TM, Boehnke M, Lin X. General framework for meta-analysis of rare variants in sequencing association studies. *Am J Hum Genet*. 2013;93(1):42-53. https://doi.org/10.1016/j.ajhg.2013.05.010
40. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826. https://doi.org/10.1038/s41467-017-01261-5
41. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)*. 2005;95(3):221-227. https://doi.org/10.1038/sj.hdy.6800717
42. Althouse AD. Adjust for multiple comparisons? It's not that simple. *Ann Thorac Surg*. 2016;101(5):1644-1645. https://doi.org/10.1016/j.athoracsur.2015.11.024
43. Feise RJ. Do multiple outcome measures require p-value adjustment? *BMC Med Res Methodol*. 2002;2(1):8. https://doi.org/10.1186/1471-2288-2-8
44. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. https://doi.org/10.1038/nature15393
45. Park J, Gupta RS. Adenosine kinase and ribokinase - The RK family of proteins. *Cell Mol Life Sci*. 2008;65(18):2875-2896. https://doi.org/10.1007/s00018-008-8123-1
46. Wei Y, He R. A brief study of the correlation of Urine D-ribose with MMSE scores of patients with Alzheimer' s disease and cognitively normal participants. *Am J Urol Res*. 2019;4:018-023.
47. Zhu X, Wei Y, He Y, He R, Li J. Urine D-ribose levels correlate with cognitive function in community-dwelling older adults. *BMC Geriatr*. 2022;22(1):693. https://doi.org/10.1186/s12877-022-03288-w
48. Yu L, Chen Y, Xu Y, He T, Wei Y, He R. D-ribose is elevated in T1DM patients and can be involved in the onset of encephalopathy. *Aging (Albany NY)*. 2019;11(14):4943-4969. https://doi.org/10.18632/aging.102089
49. Pan X, Chen Z, Fei G, et al. Long-term cognitive improvement after benfotiamine administration in patients with Alzheimer's Disease. *Neurosci Bull*. 2016;32(6):591-596. https://doi.org/10.1007/s12264-016-0067-0
50. Mahoney DE, Hiebert JB, Thimmesch A, et al. Understanding D-ribose and mitochondrial function. *Adv Biosci Clin Med*. 2018;6(1):1-5. https://doi.org/10.7575/aiac.abcmed.v.6n.1p.1
51. Wei Y, Han CS, Zhou J, Liu Y, Chen L, He RQ. D-ribose in glycation and protein aggregation. *Biochim Biophys Acta*. 2012;1820(4):488-494. https://doi.org/10.1016/j.bbagen.2012.01.005
52. Kim B-H, Choi Y-H, Yang J-J, Kim S, Nho K, Lee J-M. Identification of novel genes associated with cortical thickness in Alzheimer's disease: systems biology approach to neuroimaging endophenotype. Brookes K (Ed). *J Alzheimers Dis*. 2020;75(2):531-545. https://doi.org/10.3233/JAD-191175
53. Gallagher MD, Chen-Plotkin AS. The Post-GWAS era: from association to function. *Am J Hum Genet*. 2018;102(5):717-730. https://doi.org/10.1016/j.ajhg.2018.04.002
54. Galesloot TE, van Steen K, Kiemeney LALM, Janss LL, Vermeulen SH. A comparison of multivariate genome-wide association methods. Aulchenko YS (Ed). *PLoS One*. 2014;9(4):e95923. https://doi.org/10.1371/journal.pone.0095923
55. Guerreiro RJ, Baquero M, Blesa R, et al. Genetic screening of Alzheimer's disease genes in Iberian and African samples yields novel mutations in presenilins and APP. *Neurobiol Aging*. 2010;31(5):725-731. https://doi.org/10.1016/j.neurobiolaging.2008.06.012
56. Park J, An SSA, Van GiauV, et al. Identification of a novel PSEN1 mutation (Leu232Pro) in a Korean patient with early-onset Alzheimer's disease and a family history of dementia. *Neurobiol Aging*.

2017;56:212.e11-212.e17. https://doi.org/10.1016/j.neurobiolaging.2017.04.012

57. Apostolova LG, Risacher SL, Duran T, et al. Associations of the top 20 Alzheimer disease risk variants with brain amyloidosis. *JAMA Neurol*. 2018;46202:1-14. https://doi.org/10.1001/jamaneurol.2017.4198

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: Küçükali F, Neumann A, Van Dongen J. et al., Whole-exome rare-variant analysis of Alzheimer's disease and related biomarker traits. *Alzheimer's Dement*. 2023;19:2317–2331. https://doi.org/10.1002/alz.12842