RESEARCH ARTICLE

TMEM106B and CPOX are genetic determinants of cerebrospinal fluid Alzheimer's disease biomarker levels

Shengjun Hong[1] | **Valerija Dobricic**[1] | **Olena Ohlei**[1] | **Isabelle Bos**[2,3] | **Stephanie J. B. Vos**[2,3] | **Dmitry Prokopenko**[4] | **Betty M. Tijms**[3] | **Ulf Andreasson**[5,6] | **Kaj Blennow**[5,6] | **Rik Vandenberghe**[7,8] | **Silvy Gabel**[7] | **Philip Scheltens**[3] | **Charlotte E. Teunissen**[9] | **Sebastiaan Engelborghs**[10,11] | **Giovanni Frisoni**[12,13] | **Olivier Blin**[14] | **Jill C. Richardson**[15] | **Regis Bordet**[16] | **Alzheimer's Disease Neuroimaging Initiative (ADNI)**[*] | **Alberto Lleó**[17] | **Daniel Alcolea**[17] | **Julius Popp**[18,19] | **Christopher Clark**[18] | **Gwendoline Peyratout**[19] | **Pablo Martinez-Lage**[20] | **Mikel Tainta**[20] | **Richard J. B. Dobson**[21,22,23,24,25] | **Cristina Legido-Quigley**[26] | **Kristel Sleegers**[27,28] | **Christine Van Broeckhoven**[27,28] | **Rudolph E. Tanzi**[4] | **Mara ten Kate**[29,30] | **Michael Wittig**[31] | **Andre Franke**[31] | **Christina M. Lill**[1,32] | **Frederik Barkhof**[33] | **Simon Lovestone**[34] | **Johannes Streffer**[35,36] | **Henrik Zetterberg**[5,6,37,38] | **Pieter Jelle Visser**[2,3,39] | **Lars Bertram**[1,40]

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

[2] Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Alzheimer Centrum Limburg, Maastricht University, Maastricht, the Netherlands

[3] Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, the Netherlands

[4] Genetics and Aging Unit and McCance Center for Brain Health, Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA

[5] Institute of Neuroscience and Physiology, Department of Psychiatry and Neurochemistry, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[6] Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden

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

[8] Neurology Service, University Hospital Leuven, Leuven, Belgium

[9] Neurochemistry Laboratory, Department of Clinical Chemistry, Amsterdam Neuroscience, Amsterdam University Medical Centers, Vrije Universiteit, Amsterdam, the Netherlands

[10] Department of Biomedical Sciences, Institute Born-Bunge, University of Antwerp, Antwerp, Belgium

[11] Department of Neurology and Center for Neurosciences, UZ Brussel and Vrije Universiteit Brussel (VUB), Brussels, Belgium

[12] University of Geneva, Geneva, Switzerland

[13] IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy

[14] AIX Marseille University, INS, Ap-hm, Marseille, France

[15] Neurosciences Therapeutic Area, GlaxoSmithKline R&D, Stevenage, UK

[16] Inserm, CHU Lille, University of Lille, Lille, France

[17] Memory Unit, Neurology Department. Hospital de Sant Pau, Barcelona and Centro de Investigación Biomédica en Red en enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain

[18] Centre for Gerontopsychiatric Medicine, Department of Geriatric Psychiatry, University Hospital of Psychiatry Zurich, Zürich, Switzerland

[19] Old Age Psychiatry, Department of Psychiatry, University Hospital of Lausanne, Lausanne, Switzerland

[20] Department of Neurology, Center for Research and Advanced Therapies, CITA-Alzheimer Foundation, San Sebastian, Spain

[21] Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

[22] NIHR BioResource Centre Maudsley, NIHR Maudsley Biomedical Research Centre (BRC) at South London and Maudsley NHS Foundation Trust (SLaM) & Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, London, UK

[23] Health Data Research UK London, University College London, London, UK

[24] Institute of Health Informatics, University College London, London, UK

[25] The National Institute for Health Research University College London Hospitals Biomedical Research Centre, University College London, London, UK

[26] Steno Diabetes Center, Copenhagen, Denmark and Institute of Pharmaceutical Sciences, King's College London, London, UK

[27] Neurodegenerative Brain Diseases Group, Center for Molecular Neurology, VIB, Antwerp, Belgium

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

[29] Alzheimer Center and Department of Neurology, Amsterdam University Medical Centers, Amsterdam Neuroscience, Amsterdam, the Netherlands

[30] Department of Radiology and Nuclear Medicine, Amsterdam University Medical Centers, Amsterdam Neuroscience, Amsterdam, the Netherlands

[31] Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany

[32] Ageing Epidemiology Research Unit, School of Public Health, Imperial College, London, United Kingdom

[33] Institutes of Neurology and Healthcare Engineering, University College London, London, UK

[34] Department of Psychiatry, University of Oxford, Oxford, UK

[35] Reference Center for Biological Markers of Dementia (BIODEM), Institute Born-Bunge, University of Antwerp, Antwerp, Belgium

[36] Translational Medicine Neuroscience, UCB Biopharma SPRL, Braine l'Alleud, Belgium

[37] Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, Queen Square, London, UK

[38] UK Dementia Research Institute at UCL, London, UK

[39] Department of Neurobiology, Care Sciences and Society, Division of Neurogeriatrics, Karolinska Instutet, Stockholm, Sweden

[40] Department of Psychology, University of Oslo, Oslo, Norway

Correspondence
Prof. Lars Bertram, Lübeck Interdisciplinary Platform for Genome Analytics, (LIGA), University of Lübeck, Ratzeburger Allee 160, Haus V50, 23562 Lübeck, Germany.
E-mail: lars.bertram@uni-luebeck.de

Simon Lovestone: currently employed at Janssen-Cilag.
*Alzheimer's Disease Neuroimaging Initiative (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 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: Neurofilament light (NfL), chitinase-3-like protein 1 (YKL-40), and neurogranin (Ng) are biomarkers for Alzheimer's disease (AD) to monitor axonal damage, astroglial activation, and synaptic degeneration, respectively.

Methods: We performed genome-wide association studies (GWAS) using DNA and cerebrospinal fluid (CSF) samples from the EMIF-AD Multimodal Biomarker Discovery study for discovery, and the Alzheimer's Disease Neuroimaging Initiative study for validation analyses. GWAS were performed for all three CSF biomarkers using linear regression models adjusting for relevant covariates.

Results: We identify novel genome-wide significant associations between DNA variants in *TMEM106B* and CSF levels of NfL, and between *CPOX* and YKL-40. We confirm previous work suggesting that YKL-40 levels are associated with DNA variants in *CHI3L1*.

Discussion: Our study provides important new insights into the genetic architecture underlying interindividual variation in three AD-related CSF biomarkers. In particular, our data shed light on the sequence of events regarding the initiation and progression of neuropathological processes relevant in AD.

KEYWORDS
Alzheimer's disease, biomarker, cerebrospinal fluid, chitinase-3-like protein 1, genome-wide association study, neurofilament light, neurogranin

1 | BACKGROUND

Elucidation of the genetic architecture underlying Alzheimer's disease (AD) susceptibility has recently seen substantial progress using the genome-wide association study (GWAS) approach. Based on results from the two most recent and largest GWAS in the field,[1,2] there are now >30 independent loci showing genome-wide significant association with AD risk.[3] In contrast, the genetic underpinnings determining interindividual variation in levels of molecular AD biomarkers are less well known. Apart from the two "core" AD proteins, that is amyloid beta 42 (Aβ42) and tau, there are currently only a few GWAS shedding light on the genetic factors determining other blood or cerebrospinal fluid (CSF) AD biomarkers levels. In an effort to close this knowledge gap, we combined CSF and genome-wide single-nucleotide polymorphism (SNP) genotyping data generated in the EMIF-AD Multimodal Biomarker Discovery dataset[4] and performed the first bona fide GWAS on CSF levels of neurofilament light chain (NfL), chitinase-3-like protein 1 (YKL-40), and neurogranin (Ng). Specifically, our analyses assessed the hypothesis that variation in these three biomarker outcomes is (at least partially) genetically determined and that the identification of the underlying genetic factors will eventually yield new insights into the pathophysiology of cognitive decline and AD.

NfL is one type of four different neurofilament subunits which function as structural components of the neural cytoskeleton[5] performing essential roles in axon development[6] and synaptic function.[6] As such, NfL is considered one of several "core" biomarkers of axonal injury and neurodegeneration across neurological diseases.[7,8] In addition, other recent data suggest that changes in NfL serum levels can predict disease onset and progression of brain neurodegeneration at very early, pre-symptomatic stages of familial AD.[9] YKL-40 is a glycoprotein produced in several inflammatory conditions and cancers,[10] and was classified as an "emerging" AD biomarker in a recent meta-analysis.[7] While its precise physiological role remains elusive, it appears that in AD, YKL-40 is predominantly expressed in astrocytes and likely plays a role in the inflammatory response occurring near Aβ plaques.[10,11] Finally, Ng is a neuron-specific protein, mainly expressed in the cortex and hippocampus, where it is involved in synaptic long-term potentiation and learning.[12–14] In AD, CSF Ng was proposed to represent a marker of synaptic degeneration and was recently reported to correlate with cognitive decline.[15] To the best of our knowledge, our study represents the first bona fide GWAS on these biomarker traits with the exception of two small (n = 133 and n = 265, respectively) CSF protein quantitative trait loci (pQTL) GWAS on YKL-40 in people of Asian descent[34] and a GWAS on NfL in the subset of non-demented elderly from the Alzheimer Disease Neuroimaging Initiative (ADNI) cohort.[35]

2 | METHODS

2.1 | Sample description

The ascertainment procedures for the EMIF-AD Multimodal Biomarker Discovery (MBD) dataset are described elsewhere.[16] In brief, the dataset includes 1221 elderly individuals (years of age:

RESEARCH IN CONTEXT

1. **Systematic Review**: Despite >30 independent loci reported to show genome-wide significant association with AD risk from GWAS, the genetic underpinnings determining interindividual variation in levels of molecular AD biomarkers are less well known. We searched PubMed, Google Scholar, and the GWAS catalog for GWAS articles on CSF neurofilament light chain (NfL), chitinase-3-like protein 1 (YKL-40), neurogranin (Ng) and their association with Alzheimer's disease (AD) with no language restrictions from database inception up to November 30, 2019, using the combination (AND/OR) of the following terms: "neurofilament light/NF-L/NfL," "chitinase-3-like protein 1/CHI3L1/YKL-40," "neurogranin," "cerebrospinal fluid biomarkers/CSF," "Alzheimer's disease/AD," "genetic association," "genome-wide association study/GWAS," considering only studies reporting the results of genetic association analyses for these three CSF biomarkers. For CSF levels of NfL, YKL-40, or Ng we only identified two small previous GWAS, one investigating CSF YKL-40 in 133 Japanese individuals and one investigating CSF NfL in 265 non-demented elderly.

2. **Interpretation**: Our analyses yielded novel genome-wide significant associations with markers in *TMEM106B* and CSF levels of NfL in GWAS analyses using the EMIF-AD Multimodal Biomarker Discovery dataset, a finding that was independently replicated in the ADNI study. We also uncovered novel genetic association signals with DNA variants in *CPOX* and cerebrospinal fluid (CSF) levels of YKL-40. Moreover, we confirmed previous studies reporting that variants in *CHI3L1* act as *cis* protein quantitative trait loci (pQTL) for CSF YKL-40.

3. **Future directions**: Our study sheds light on the sequence of events regarding the initiation and progression of neuropathological processes relevant in AD. Further independent studies of comparable or larger sample size are needed to validate our results and to elucidate the molecular mechanisms underlying the observed associations.

mean = 67.9, standard deviation [SD] = 8.3; 667 females, 554 males) with different cognitive diagnoses at baseline (NC = normal cognition; MCI = mild cognitive impairment; AD = AD-type dementia). Clinical follow-up data were available for 759 individuals. Independent CSF biomarker and genotyping data used in the replication phase of this study were obtained from the ADNI database (URL: http://adni.loni.usc.edu). ADNI was launched in 2003 as a public–private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial

TABLE 1 Demographic information and summary of CSF traits used in GWAS analyses of the EMIF-AD MBD and ADNI datasets

CSF biomarker	CSF biomarker description	Detection method	N	N male/female	Mean age ± SD (range)	N Controls	N MCI	N AD
EMIF-AD MBD dataset (discovery)								
NfL	Neurofilament light	ELISA (UmanDiagnostics)	671	319/352	69.49 ± 8.35 (45.94-92.29)	122	395	154
YKL-40	Chitinase-3-like protein 1	ELISA (R&D systems)	677	323/354	69.42 ± 8.31 (45.94-92.29)	122	401	154
Ng	Neurogranin	Immunoassay (in-house)	672	319/353	69.50 ± 8.34 (45.94-92.29)	122	398	152
ADNI dataset (replication)								
NfL	Neurofilament light	ELISA (UmanDiagnostics)	308	185/123	74.80 ± 6.98 (58.40-89.60)	87	151	70
YKL-40	Chitinase-3-like protein 1	Electrochemiluminescence (Meso Scale Discovery)	131	85/46	74.88 ± 6.09 (58.40-89.60)	49	70	12
Ng	Neurogranin	ELISA (Quidel Corp)	308	185/123	74.80 ± 6.98 (58.40-89.60)	87	151	70

Abbreviations: AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; CSF, cerebrospinal fluid; ELISA, enzyme-linked immunosorbent assay; GWAS, genome-wide association studies; MBD, multimodal biomarker discovery; MCI, mild cognitive impairment; NfL, neurofilament light chain; Ng, neurogranin; SD, standard deviation; YKL-40, chitinase-3-like protein 1.

magnetic resonance imaging (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. The demographic information for the three quantitative CSF phenotypes in the EMIF-AD MBD and ADNI datasets used in this study is summarized in Table 1.

2.2 | DNA extraction, genotype imputation, and quality control

A detailed account of the genotyping procedures in EMIF-AD MBD and subsequent bioinformatic workflows can be found in the supporting information and Hong et al.[17] In brief, a total of 936 DNA samples were sent for genome-wide SNP genotyping using the Infinium Global Screening Array (GSA) with Shared Custom Content (Illumina Inc.). After extensive quality control (QC) and imputation, a total of 7,778,465 autosomal SNPs with minor allele frequency (MAF) $\geq$0.01 were retained in 898 individuals of European ancestry for the downstream genome-wide association analyses. For more details see supporting information and Hong et al.[17]

2.3 | CSF biomarkers

Details of the CSF biomarker measurements can be found in Bos et al.[4] In brief, the CSF specimens were collected individually at each of the 11 EMIF-AD MDB participating sites. CSF samples were shipped to the Department of Psychiatry and Neurochemistry at University of Gothenburg, Sweden. Relevant to the analyses presented here, NfL levels were measured using a commercial enzyme-linked immunosorbent assay (ELISA; NF-light ELISA, UmanDiagnostics; Zetterberg et al.[18]).

Ng levels were measured using an in-house immunoassay for Ng.[19] YKL-40 levels were measured using a human chitinase-3 quantikine ELISA kit (R&D Systems, Inc.; Olsson et al.[20]). To reduce the skewness of phenotype distributions, data for all three CSF variables were log-transformed prior to analysis (Figure S1 in supporting information).

2.4 | GWAS and post-GWAS analyses

Linear regression (using mach2qtl[21,22]) was used to perform SNP-based association analyses using imputation-derived allele dosages as independent variables and the log-transformed concentrations of CSF NfL, CSF Ng, and CSF YKL-40 as dependent outcome variables. Covariates included into the regression models were sex, age at examination, diagnostic groups (coded as AD = 3, MCI = 2, controls = 1), and ancestry-specific principal components (PCs; here the first five were used). The genomic inflation factor was calculated in R using the "GenABEL" package.[23] The FUMA platform (http://fuma.ctglab.nl/)[24] was used for post-GWAS analyses, including gene-based analyses.[25] Genome-wide significance was defined as follows for the two types of genome-wide analyses performed: 5E-08, a widely accepted threshold for the SNP-based analyses,[1,2] and 2.651E-6 for the gene-based analyses, corresponding to testing 18,862 protein coding genes (as suggested on FUMA). Overall, the effective sample sizes for the GWAS analyses were n = 671, n = 677, and n = 672 in EMIF-AD MBD for CSF NfL, YKL-40, and Ng, respectively (Table 1; see below for corresponding numbers in ADNI). In addition to running separate GWAS for EMIF-AD and ADNI, we combined results from both datasets by meta-analysis using inverse variance weighted fixed-effect models as implemented in METAL (http://csg.sph.umich.edu/abecasis/Metal/).

2.5 | Polygenic risk score analysis

Summary statistics of the two largest and most recent AD case-control GWAS[1,2] were used for calculating polygenic risk score (PRS) for each individual in EMIF-AD MDB. PRS were constructed for 11 different *P*-value thresholds using PLINK v1.9[26] after additional QC (supporting information). The resulting PRSs were used as independent variables in the regression models adjusting for sex, age, diagnosis, and PC1-5 as covariates as in the primary GWAS analyses.

2.6 | Replication analyses in ADNI

For replication purposes, we used CSF biomarker and genotyping data from the ADNI study to assess replicability of the EMIF-AD MBD findings. The ADNI sample used here comprises n = 131 (CSF YKL-40) or n = 308 (CSF NfL and Ng) subjects with available microarray-based genotyping and relevant biomarker phenotype data (Table 1; for more details on ADNI see http://adni.loni.usc.edu/). In brief, CSF NfL levels in ADNI were measured with the same ELISA as for EMIF-AD MBD (NF-light ELISA, UmanDiagnostics). Ng levels were measured using a monoclonal antibody specific for neurogranin (Ng7) with electro-chemiluminescence technology (Meso Scale Discovery). YKL-40 levels were measured with the MicroVue YKL-40 ELISA (Quidel Corp). ADNI genotyping data was processed using the same workflow as for the EMIF-AD MBD analyses (i.e., same QC and imputation pipeline, same regression models) with one exception: we applied a slightly more conservative MAF threshold (i.e., 0.03 instead of 0.01), as variants in the lower allele spectrum tend to produce less reliable results in small datasets. Accordingly, this led to a lower overall number of SNPs available for GWAS (5,858,348) in this dataset.

A detailed description of all methods and procedure applied in this study can be found in the supporting information.

3 | RESULTS

3.1 | GWAS analyses using CSF neurofilament light levels

The GWAS analyses using CSF NfL as outcome yielded five SNPs showing genome-wide significant association in the EMIF-AD MBD dataset (Table 2, Figure S2A, and Table S1 in supporting information). Three of these SNPs are located in two distinct loci on chromosome 7 (i.e., on chr. 7q36.1 [rs111748411, rs3094407] and on 7p15.3[rs77589784]), while the other two are located on chromosomes 1p36.12 (rs4654961) and 10q26.3 (rs138898705; Table S1 and Figure S2A). MAFs for SNPs in all but the 7q36.1 locus were ≈1% complicating any inferences and functional interpretations of these variants given the limited size of our dataset. For the two common SNPs in the chr. 7q36.1 locus (i.e., rs111748411, rs3094407), post-GWAS variant annotation in FUMA suggested no obvious functional consequences (Table S2 in supporting information). Owing to their low allele frequency, none of these variants were assessed in the ADNI dataset, so these CSF NfL findings should be viewed with caution until independent replication in sufficiently sized samples is provided.

In contrast to the SNP-based results, the gene-based analyses using MAGMA elicited a third locus on chromosome 7 (7p21.3) at transmembrane protein 106B (*TMEM106B*) showing genome-wide significant association with CSF NfL levels (Table 2, Figure S2A, and Table S1). This gene, which is an established genetic risk modifier for frontotemporal lobar degeneration (FTLD),[27,28] contains 187 SNPs of which the majority (n = 124 SNPs) are in strong linkage disequilibrium (LD; $r^2 > 0.6$) with the lead variant in *TMEM106B*, that is, rs1548884 showing evidence for genome-wide suggestive association in EMIF-AD MBD (*P* = 2.62E-07, Table 2 and Table S1). Given its MAF of ≈0.42, this SNP could be assessed in the independent ADNI dataset where it showed nominal evidence of association with consistent effect direction (*P* = .0026, Table 2, Figure S2B, Table S1). Combining the EMIF-AD MBD and ADNI GWAS results for this SNP by meta-analysis yielded association evidence that exceeded the threshold for genome-wide significance (P_{meta} = 3.85E-09; Figure 1A, Table 2). In addition to rs1548884, >80 variants in the *TMEM106B* region elicited genome-wide significant association in the meta-analyses (best P_{meta} = 2.27E-09 at rs7797705; Figure 1A, Table S1). Gene-based analyses using MAGMA in ADNI also showed independent association between *TMEM106B* and CSF NfL (*P* = .00128; Table 2, Figure S2B), increasing the overall gene-based evidence of meta-analyses across both EMIF-AD MBD and ADNI datasets (P_{meta} = 1.32E-08; Figure 1B, Table 2, Figure S2C and D, Table S1).

FUMA-based functional annotations show one non-synonymous variant (rs3173615; in nearly perfect LD [$r^2 > 0.99$] with rs1548884, Tables S2 & S3 in supporting information) eliciting a Thr185Ser change with a CADD score of 21.4, and a predicted "moderate" impact by ENSEMBL's variant effect predictor (VEP) algorithm. In addition to possibly exerting an effect on protein function by directly altering the amino-acid sequence of *TMEM106B*, the same variant is also reported as a modest expression quantitative trait locus (eQTL) in cortical brain samples of the Genotype Tissue Expression (GTEx, v8) project (*P* = 3.8E-05; Table S3). Interestingly, the lead variant in our CSF NfL GWAS (rs1548884) is in strong LD (r^2 = 0.98; Table S2) with the SNP originally identified and subsequently replicated to be associated with FTLD (rs1990622, P_{meta} = 4.21E-09, Table S1; Van Deerlin et al.[27]).

3.2 | GWAS analyses using CSF YKL-40 levels

The SNP-based GWAS using CSF YKL-40 levels yielded one genome-wide significantly associated locus on chromosome 1q32.1. This signal was driven by three independent SNPs (i.e., rs7551263, rs1417152, and rs10399931; Table 2, Figure S3A, Table S4 in supporting information) and also represents the single most significant GWAS signal in the EMIF-AD MBD dataset of this study (P = 4.79E-11 for rs10399931). Unlike the GWAS results for the other two CSF markers analyzed here, the strongest results were observed with relatively common variants showing allele frequencies between ≈16% and 21% in people of European ancestry (Table S5a in supporting information). Association analyses in the ADNI dataset showed independent evidence of association with the same direction of effect for two of the three SNPs (*P* = .041, *P* = .5523, *P* = 9.19E-07, for rs7551263, rs1417152, and rs10399931, respectively; Table 2, Figure S3B, Table S4). Accordingly, the

TABLE 2 Top results from GWAS analyses of three CSF traits measured in the EMIF-AD MBD and ADNI datasets

Trait	ID	Chr	Pos	Nearest gene	Alleles	MAF	Results for EMIF-AD MDB dataset			Results for ADNI dataset			Results for meta-analysis		
							Beta	P (SNP)	P (Gene)	Beta	P (SNP)	P (Gene)	Beta	P (SNP)	P (Gene)
CSF NfL	rs4654961	1p36.12	21865495	ALPL	A/G	0.01	−0.359	**3.143E-08**	0.20497	n.a	n.a	0.13704	n.a	n.a	0.069226
	rs77589784	7p15.3	25481515	n.a.	G/A	0.0118	−0.367	**1.643E-08**	n.a.	n.a	n.a	n.a	n.a	n.a	n.a
	rs1548884	7p21.3	12279761	TMEM106B	C/A	0.4283	0.067	2.622E-07	**2.6072E-07**	0.101	0.002621	0.0012809	0.0715	**3.85E-09**	**1.3381E-08**
	rs111748411	7q36.1	152254171	AC104843.4	A/C	0.0167	−0.403	**2.285E-09**	n.a.	n.a	n.a	n.a	n.a	n.a	n.a
	rs3094407	7q36.1	152373705	XRCC2	T/G	0.027	0.236	**4.071E-08**	0.21394	n.a	n.a	0.70739	n.a	n.a	0.211
	rs138898705	10q26.3	132563077	n.a.	T/C	0.0106	−0.389	**5.238E-09**	n.a.	n.a	n.a	n.a	n.a	n.a	n.a
CSF YKL-40	rs7551263	1q32.1	203150756	CHI3L1	C/T	0.1726	−0.166	**1.567E-10**	**2.5151E-08**	−0.111	0.041	0.28397	−0.1559	**3.09E-11**	**4.1383E-09**
	rs10399931	1q32.1	203156080	CHI3L1	C/T	0.2381	−0.143	**4.789E-11**	**2.5151E-08**	−0.224	9.19E-07	0.28397	−0.1583	**1.42E-15**	**4.1383E-09**
	rs58943879	3q11.2-12.1	98280429	CPOX	C/T	0.4967	0.096	6.102E-07	**8.7456E-09**	−0.014	0.7385	0.69852	0.0781	7.09E-06	**1.2043E-06**
CSF Ng	rs10052776	5p15.2	11178648	CTNND2	T/C	0.393	0.157	1.022E-07	0.041368	0.122	0.2516	0.058017	−0.1366	2.25E-06	0.0091132
	rs8111341	19q13.33	48889558	KDELR1	G/A	0.156	−0.167	1.17E-05	2.29E-05	−0.27	0.06313	0.16703	−0.1736	2.34E-06	8.0186E-06

Abbreviations: AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; CSF, cerebrospinal fluid; ELISA, enzyme-linked immunosorbent assay; GWAS, genome-wide association studies; MBD, multimodal biomarker discovery; MCI, mild cognitive impairment; n.a., not available; NfL, neurofilament light chain; Ng, neurogranin; SD, standard deviation; SNP, single-nucleotide polymorphism; YKL-40, chitinase-3-like protein 1.

Notes: More results from these analyses can be found in the supporting information tables.

Bold font indicates genome-wide significant (on SNP- or gene-level) results (see Methods section for details).

Chr and Pos according to GRCh37/hg19.

Top results from these GWAS analyses can be found in the supporting information tables (except for KDELR1).
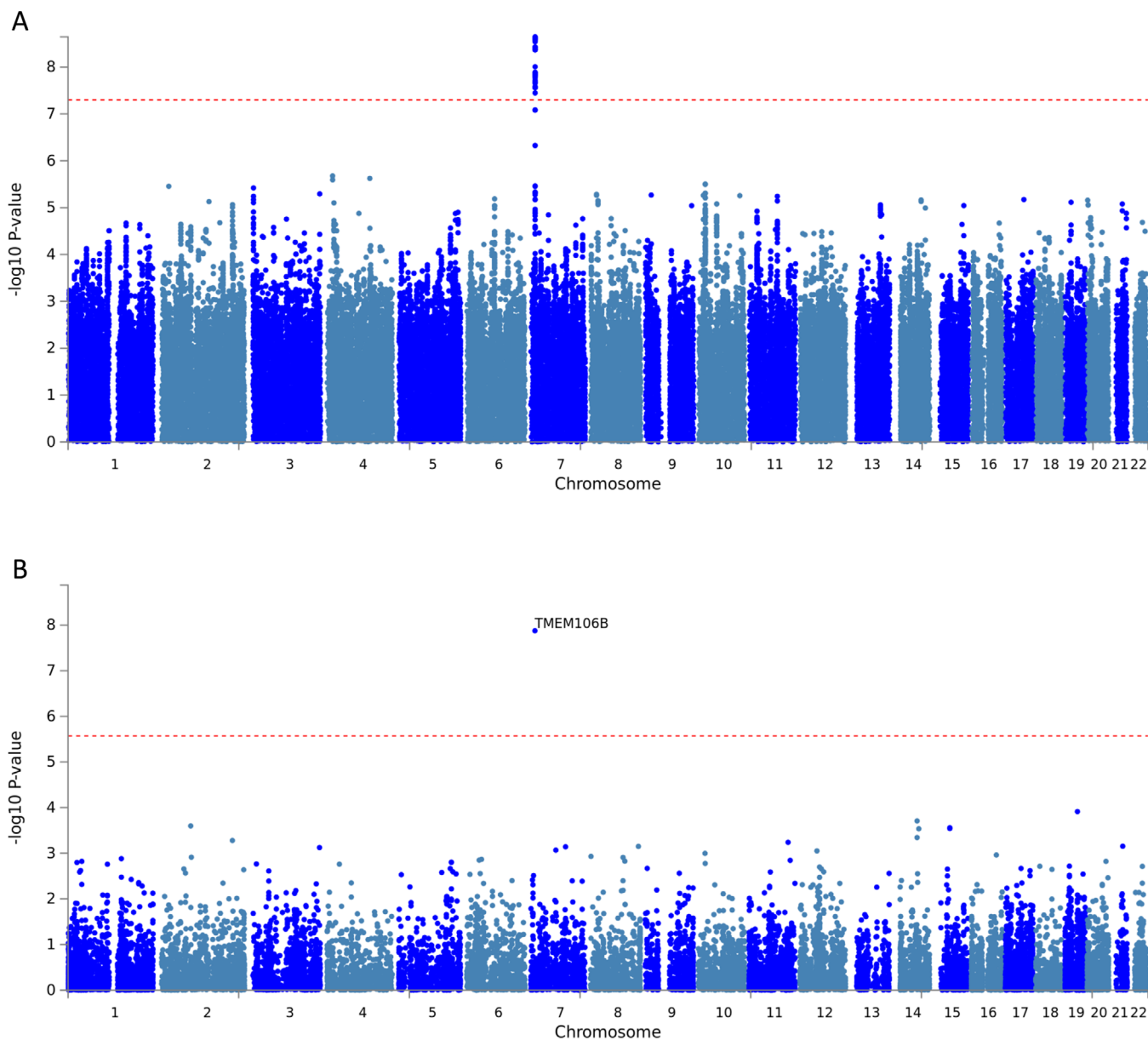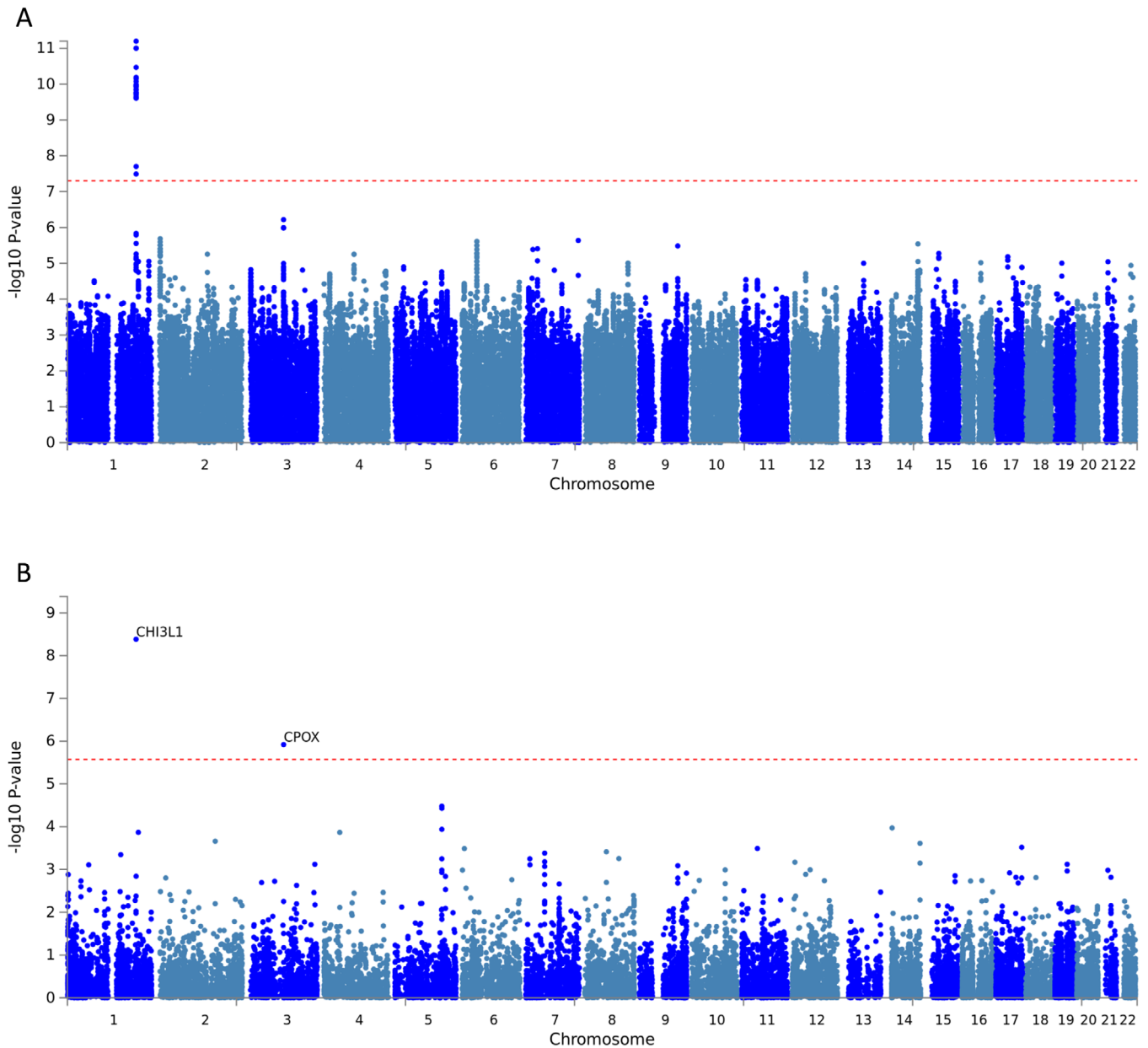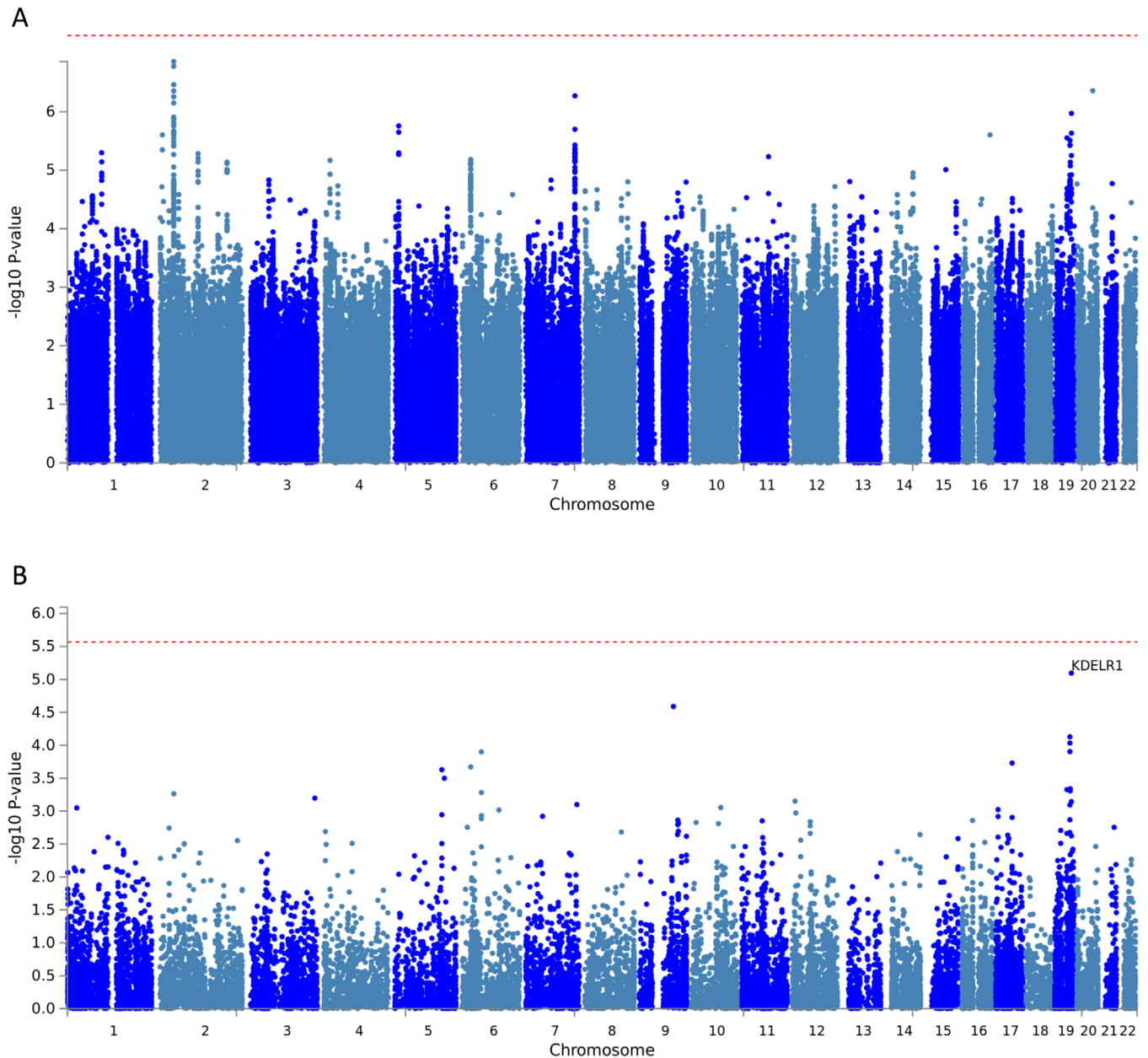
FIGURE 1 Manhattan plots of (A) single-nucleotide polymorphism (SNP) level and (B) gene-level genome-wide association results using log-transformed cerebrospinal fluid neurofilament light chain levels as outcome trait in meta-analyses combining EMIF-AD MBD and ADNI datasets (n = 979). Gene assignments are according to FUMA.[24] Dotted red lines represent the threshold for genome-wide significance, that is $\alpha = 5.0E-08$ for SNP-based (A) and $\alpha = 2.651E-6$ for gene-based (B) analyses (see Methods section). Manhattan plots of dataset specific results can be found in supporting information

meta-analyses across both datasets yielded strong genome-wide significant support for an association between two of these markers and CSF levels of YKL-40 ($P_{meta} = 3.09E-11$, $P_{meta} = 5.87E-08$, $P_{meta} = 1.42E-15$, for rs7551263, rs1417152, and rs10399931, respectively; Figure 2A, Table 2, Figure S3C and D, Table S4). While FUMA-based gene annotations (Table S5b) highlight up to 26 different gene symbols in the implicated region, the most obvious candidate of likely biological relevance is *CHI3L1* (chitinase-3-like 1), that is, the gene encoding YKL-40 protein. In other words, this GWAS result represent a bona fide *cis* pQTL result. Furthermore, and corresponding to these pQTL results, eQTL annotations summarized by FUMA con-

verge on *CHI3L1* as the most strongly and most significantly associated gene when using the YKL-40–associated SNPs or their proxies as input (see also Table S6 in supporting information). Interestingly, this SNP is also listed as methylation QTL (mQTL) on the mQTL database (http://www.mqtldb.org/).[29]

Gene-based GWAS analyses confirmed the association with *CHI3L1* ($P = 2.52E-08$; Figure S3A, Table S4), and revealed a second, independent locus, CPOX (coproporphyrinogen oxidase; chromosome 3q11.2-12.1), showing even more significant gene-based association with CSF YKL-40 levels ($P = 8.75E-09$; Figure S3A, Table S4). The most significantly associated single variant in *CPOX* was rs58943879 ($P = 6.1E-07$;

FIGURE 2 Manhattan plots of (A) single-nucleotide polymorphism (SNP) level and (B) gene-level genome-wide association results using log-transformed cerebrospinal fluid CHITINASE-3-LIKE PROTEIN 1levels as outcome trait in meta-analyses combining EMIF-AD MBD and ADNI datasets (n = 808). Gene assignments are according to FUMA.[24] Dotted red lines represent the threshold for genome-wide significance, that is, α = 5.0E-08 for SNP-based (A) and α = 2.651E-6 for gene-based (B) analyses (see Methods section). Manhattan plots of dataset specific results can be found in supporting information

Table S4). While the ADNI analyses did not provide independent support for an association between *CPOX* and CSF YKL-40 (Table 2, Figure S3B, Table S4), the gene-based GWAS across both datasets still elicited gene-wide significance, albeit with reduced evidence compared to the EMIF-AD MBD dataset alone (P_{meta} = 1.2043E-06; Figure 2B, Table 2, Table S3C and D, Table S4). Manual lookup on the GTEx portal (v8) revealed no previously observed eQTLs in the brain, despite CPOX's relatively pronounced expression in all brain tissues sampled in GTEx (https://www.gtexportal.org/home/gene/CPOX).

3.3 | GWAS analyses using CSF neurogranin levels

Last, the GWAS using CSF Ng levels yielded no genome-wide significant association in the SNP-based analyses (Table 2, Figure S4A, and Table S7 in supporting information). The top-ranking SNP-based finding was elicited by rs10052776 (P = 1.0E-07, Table S7), located in *CTNND2* mapping to chromosome 5p15.2. Interestingly, SNPs in this gene were previously associated with both late-onset AD and cognitive performance by GWAS[30,31] according to the "GWAS

FIGURE 3 Manhattan plots of (A) single-nucleotide polymorphism (SNP) level and (B) gene-level genome-wide association results using log-transformed cerebrospinal fluid neurogranin levels as outcome trait in meta-analyses combining EMIF-AD MBD and ADNI datasets (n = 980). Gene assignments are according to FUMA.[24] Dotted red lines represent the threshold for genome-wide significance, that is, $\alpha = 5.0E\text{-}08$ for SNP-based (A) and $\alpha = 2.651E\text{-}6$ for gene-based (B) analyses (see Methods section). Manhattan plots of dataset specific results can be found in supporting information

catalog" (https://www.ebi.ac.uk/gwas/genes/CTNND2). However, replication analyses in ADNI did not support these findings, resulting in less pronounced association evidence in the meta-analyses ($P_{meta} = 0.2516$ for rs10052776; Table 2, Figure 3A, Table S7). Gene-based association analyses using MAGMA also did not reveal any genome-wide significant signals with variants annotated to the 18,862 genes used in these analyses, neither in EMIF-AD MBD alone (Figure S4A) nor in the meta-analyses with ADNI (Figure 3B; Figure S4C and D, Table S7). The top-ranking gene-based finding with 21 SNPs was observed with *KDELR1* ($P_{EMIF} = 2.29E\text{-}05$, $P_{meta} = 8.21E\text{-}06$) mapping

to chromosome 19q13.33, a gene hitherto not associated with the traits listed in the GWAS catalog.

3.4 | Genetic correlation analyses

To assess whether and to which degree variation of the three AD CSF biomarkers analyzed in this study show association with "AD-related" variants we calculated PRSs as predictors of CSF biomarker variation (see supporting information for justification and details). Similar

analyses in earlier work[17] had revealed a comparatively strong association with Aβ-derived but not tau-derived biomarkers. Similar to CSF-tau measures, the CSF biomarker phenotype variance in EMIF-AD MBD explained by AD PRS was collectively minor, reaching nominal significance for some CSF phenotypes and *P*-value thresholds using Kunkle et al.,[2] but none from the Jansen et al.[1] data (Table S8 in supporting information). This is in contrast to applying AD GWAS-based PRS to Aβ-related CSF phenotypes in the same EMIF-AD MDB dataset: here, the strongest associations explained 6% to 12% of the phenotypic variance (P = 9E-09 and 1E-07, respectively),[17] suggesting that the genetic architectures underlying AD risk and variation at CSF NfL, YKL-40, and Ng in general do not show any substantial overlap.

4 | DISCUSSION

We performed GWAS analyses for three CSF AD-related biomarkers in the EMIF-AD MDB dataset and identified novel genome-wide significant association with genetic markers in the established FTLD risk gene *TMEM106B* and NfL protein concentrations in the CSF. This finding showed consistent independent replication in comparable CSF data from ADNI. In addition, we detected very strong and genome-wide significant association between markers in *CHI3L1* and CSF levels of YKL-40, representing the only *cis* pQTL finding in our analyses. To the best of our knowledge, our study is the first bona fide GWAS for all three of these CSF biomarkers with the exception of two small (n = 133 and n = 265) CSF pQTL GWAS on YKL-40 in people of Asian descent[34] and a GWAS on NfL in a subset of non-demented elderly from the ADNI cohort.[35] Of note, the former study also identified strong and genome-wide significant *cis* pQTL effects at the *CHI3L1*/YKL-40 locus, corroborating our findings. Other noteworthy results from our study include evidence for several rare-variant associations with CSF NfL levels and an overall lack of AD-related genetic association signals with the CSF biomarkers analyzed here. This latter point explicitly includes genetic variants in or near the apolipoprotein E gene, the strongest currently known genetic AD risk factor,[3] which did not show association with any of the CSF biomarkers analyzed here.

Possibly the most noteworthy novel signal observed in our study relates to the association between *TMEM106B* and CSF NfL. DNA variants in *TMEM106B* have first been implicated in neuropsychiatric research by a GWAS on FTLD with TAR DNA-binding protein (TDP-43) inclusions (FTLD-TDP; Van Deerlin et al.[27]), a finding that was subsequently confirmed in independent datasets (see Pottier et al.[28] for recent GWAS results). Furthermore, a recent meta-analysis revealed that CSF NfL levels are significantly increased in FTLD.[8] In addition to FTLD, the "GWAS catalog" database lists a number of other, mostly neuropsychiatric (e.g., depression, differential brain aging, neuroticism) but also non-neurological (e.g., leukemia, height, high-density lipoprotein levels) traits showing genome-wide association with *TMEM106B* (https://www.ebi.ac.uk/gwas/genes/TMEM106B; Buniello et al.[36]) There is also strong *cis* pQTL GWAS evidence for SNPs in this region (in particular: rs10950398) and *TMEM106B* protein levels in blood.[37] In GTEx, this same SNP is also reported to correlate

with *TMEM106B* mRNA expression in the brain (cortex and cerebellum), albeit at lesser significance (https://www.gtexportal.org/home/snp/rs10950398). Of note, this pQTL/eQTL variant (rs10950398) is in nearly perfect LD with the lead SNP identified here to show association with CSF NfL levels (i.e., rs1548884, r^2 = 0.97; Tables S1 & 2). In summary, there is now convincing evidence converging from multiple lines of independent data and datasets that DNA variants in *TMEM106B* not only show association with several neuropsychiatric and non-neurological phenotypes, but also *cis* (*TMEM106B*, previous work) and *trans* (CSF NfL, this study) pQTL associations with proteins relevant for neuronal function. In the AD field, NfL recently (re)gained interest based on data suggesting that NfL protein dynamics in serum may predict progression and brain neurodegeneration at early pre-symptomatic stages of familial AD,[9] and help track neurodegeneration in sporadic AD.[38] The novel results from our GWAS indicate that DNA sequence variants in *TMEM106B* may be involved in regulating CSF NfL protein levels. Because the same variant(s) are also *cis* eQTLs/pQTLs of *TMEM106B* mRNA/protein levels, it is tempting to speculate that the observed effect on CSF NfL may be mediated by *TMEM106B* mRNA or protein. In line with this hypothesis is the observation that the lead genetic variant in *TMEM106B* highlighted in our analyses (rs1548884) shows some, albeit sub–genome-wide, evidence for association with AD risk in the two largest and most recent GWAS in the field, that is, P = 0.00005 and P = 0.005 in Jansen et al.[1] and Kunkle et al.,[2] respectively. In conclusion, our novel data now provide the genetic foundation for future work aimed at elucidating whether the observed increase in CSF NfL levels represents a "cause" or "effect" of the neurodegenerative processes underlying symptomatic and pre-symptomatic AD. The observation that the same *TMEM106B* variants show association with both AD risk *and* CSF NfL levels in independent datasets provides a first indication that the recently proposed change in "NfL dynamics" may, indeed, contribute to AD neuropathology rather than simply reflect an effect of the same. However, additional work is needed, for example replication of the original NfL dynamics result and approaches applying Mendelian randomization, to address this question more formally.

The only other GWAS investigating CSF NfL was performed on 265 non-demented individuals from the ADNI cohort.[35] These authors highlighted two SNPs (i.e., rs465401 and rs460420) in *ADAMTS1* to show genome-wide significant association with CSF NfL. While, as expected, both markers showed some level of association in the ADNI analyses of our study as well (both with *P*-values = .001), there was no evidence of association in the EMIF-AD MBD dataset (*P*-values = .5485 and .5391, respectively) nor in the meta-analyses across all samples (P_{meta} = 0.56 and 0.55, respectively). Conversely, variants in *TMEM106B* were not highlighted as "peak results" in that study. However, apart from the difference in sample size, there are several additional noteworthy differences in the analysis approach used by us and Niu et al.[35] (e.g., they did not use genotype imputations to increase their coverage of untyped portions of the genome). Future work in more individuals needs to determine whether the association between *TMEM106B* and CSF NfL observed here will continue to prove genuine.

The second novel result emerging from our analyses implies that variants in *CPOX* (encoding coproporphyrinogen oxidase) are

associated with YKL-40 levels in CSF in gene-based analyses of the EMIF-AD MBD dataset. While these findings were not independently replicated in ADNI, we note that data for CSF YKL-40 currently remain particularly scarce (n = 131; Table 1) in this dataset as opposed to CSF NfL and Ng (n = 308). Notwithstanding, gene-based results across both datasets elicited gene-wide significant association between markers in *CPOX* and CSF YKL-40, albeit at reduced significance compared to EMIF-AD MBD alone. *CPOX* is ubiquitously expressed (based on GTEx release [v8]) and encodes an enzyme involved in the heme biosynthetic pathway. Intracellularly, it localizes to the mitochondria and catalyzes the two-step oxidative decarboxylation of the heme precursor copro-porphyrinogen III to protoporphyrinogen IX (https://omim.org/entry/612732). While common variants in this gene have hitherto not been associated with any human trait recorded in the "GWAS catalog," rare mutations in *CPOX* can cause coproporphyria and harderoporphyria (OMIM phenotype ID # 121300), hereditary forms of porphyrias characterized by enzyme deficiencies in the heme biosynthetic pathway. Previous work has suggested that heme has a strong affinity for binding Aβ42 peptide in vitro,[39] leading to speculations that porphyrias could potentially alter the risk and/or course of AD.[40] However, given that these potential links have not hitherto been directly proven in experimental or other work and owing to the current lack of independent validation in ADNI, the potential link between *CPOX* and CSF NfL must be considered preliminary until further replication evidence is accrued.

Finally, the third main finding worth discussing is the *cis* pQTL result linking markers in *CHI3L1* to CSF levels of YKL-40, which represents the strongest and most significant of all association signals in our GWAS. Several prior publications have implicated genetic variants in *CHI3L1* to represent *cis* pQTLs of YKL-40 levels in blood[37,41–43] (for more details, see: https://www.ebi.ac.uk/gwas/genes/CHI3L1). However, to the best of our knowledge, only one prior study has investigated YKL-40 levels in the human CSF.[34] Notwithstanding that study's relatively small sample size (n = 133) and different ethnicity (Japanese), this GWAS also reported very pronounced *cis* pQTL effects of genetic variants in *CHI3L1*. Taken together, there is now compelling converging evidence that expression of YKL-40 in both blood and CSF is regulated by DNA sequence variants located in the very gene encoding this protein. These same variants are also found as eQTL and mQTLs in independent datasets. However, unlike the situation observed for our GWAS results for NfL, the YKL-40 regulatory SNPs do not show any evidence of association with AD risk in the GWAS by Jansen et al.[1] and Kunkle et al.[2] Thus, owing to this general absence of genetic association with AD risk it appears that the observed association between CSF YKL-40 and AD status[7] probably lies downstream of the initiation of AD neuropathology.

As with all GWAS, our study is subject to some limitations. First, while we successfully provided a first line of replication evidence of our main EMIF-AD MBD findings in data from the ADNI project, we note that the currently available sample size for the biomarkers in question in ADNI is comparatively small, especially for YKL-40 (Table 1). Thus, these analyses will need to be repeated when more extensive biomarker assessments become available. Second, although the EMIF-

AD MBD dataset is the first and/or largest to allow GWAS analyses on all three CSF variables covered, the sample size available for analysis is still relatively modest (range: 671 to 677), limiting our power to detect genetic variants of moderate to small effects. Thus, the results of our GWAS likely only represent the "lowest hanging fruit" of the genetic factors underlying the analyzed traits. Third, we note that while both the genome-wide SNP genotypes as well as CSF biomarker concentrations were generated in one run of consecutive experiments in two dedicated laboratories (one for genotyping, one for CSF markers; likely reducing the possibility of batch effects), the CSF specimens were collected individually at each of the 11 EMIF-AD MDB participating sites, sometimes using different CSF collection procedures. While this sampling heterogeneity could have affected our results (although YKL-40 and Ng were recently shown to be quite stable across a range of conditions[44]), we note that CSF drawing was performed independent of genotype, so any batch effect in this particular setting should be minimal. Fourth, to clarify the role of the observed SNP/CSF protein associations for use as potential biomarkers for an early recognition of the development of AD, it would have been useful to separately analyze subjects with other non-AD neurodegenerative disorders. However, these types of diagnoses are currently not available within this dataset. Finally, we note that the EMIF-AD MBD dataset was not designed to be "representative" of the general population but was assembled with the aim to achieve approximately equal proportions of "amyloid-positive" versus "amyloid–negative" individuals across individuals with normal cognition and MCI although this was only achieved for MCI.[16] While this ascertainment strategy does not invalidate our GWAS results, they may not be generalizable to the underlying population as a whole. However, this limitation can affect any study with clinically ascertained participants and, thus, applies to most previously published GWAS in the field, including those performed in ADNI.

In conclusion, our GWAS on CSF NfL, YKL-40, and Ng provides important new insights into the genetic architecture underlying interindividual variation in these traits. Together with recent GWAS results on AD risk from case-control studies, especially the *TMEM106B* effects on CSF NfL shed important new light on the sequence of events in relation to the initiation and progression of neuropathological processes relevant in AD. Additional work is needed to set our results onto a broader evidence-based foundation and to clarify the molecular mechanisms underlying the observed associations.

CONFLICTS OF INTEREST

AUTHOR CONTRIBUTIONS

REFERENCES

1. 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.
2. Kunkle BW, Grenier-Boley B, Sims R, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat Genet*. 2019;51:414-430.
3. Bertram L, Tanzi RE. Genomic mechanisms in Alzheimer's Disease. *Brain Pathol*. 2020;30(5):966-977.
4. 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. *Alzheimer's. Dement*. 2019;15:644-654.
5. Yuan A, Nixon RA. Specialized roles of neurofilament proteins in synapses: relevance to neuropsychiatric disorders. *Brain Res Bull*. 2016;126:334-346.
6. Yuan A, Rao MV, Veeranna NixonRA. Neurofilaments at a glance. *J Cell Sci*. 2012;125:3257-3263.
7. Olsson B, Lautner R, Andreasson U, et al. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Lancet Neurol*. 2016;15:673-684.

8. Bridel C, Van Wieringen WN, Zetterberg H, et al. Diagnostic value of cerebrospinal fluid neurofilament light protein in neurology: a systematic review and meta-analysis. *JAMA Neurol*. 2019;76:1035-1048.

9. Preische O, Schultz SA, Apel A, et al. Serum neurofilament dynamics predicts neurodegeneration and clinical progression in presymptomatic Alzheimer's disease. *Nat Med*. 2019;25:277-283.

10. Molinuevo JL, Ayton S, Batrla R, et al. Current state of Alzheimer's fluid biomarkers. *Acta Neuropathol*. 2018;136:821-853.

11. Querol-Vilaseca M, Colom-Cadena M, Pegueroles J, et al. YKL-40 (Chitinase 3-like I) is expressed in a subset of astrocytes in Alzheimer's disease and other tauopathies. *J Neuroinflammation*. 2017;14: 118.

12. Chang JW, Schumacher E, Coulter PM, Vinters HV, Watson JB. Dendritic translocation of RC3/neurogranin mRNA in normal aging, Alzheimer disease and fronto-temporal dementia. *J Neuropathol Exp Neurol*. 1997;56:1105-1118.

13. Scheff SW, Price DA, Schmitt FA, DeKosky ST, Mufson EJ. Synaptic alterations in CA1 in mild Alzheimer's disease and mild cognitive impairment. *Neurology*. 2007;68:1501-1508.

14. Slemmon JR, Feng B, Erhardt JA. Small proteins that modulate calmodulin-dependent signal transduction. *Mol Neurobiol*. 2000;22:099-114.

15. Tarawneh R, D'Angelo G, Crimmins D, et al. Diagnostic and prognostic utility of the synaptic marker neurogranin in Alzheimer's disease. *JAMA Neurol*. 2016;73:561-571.

16. Bos I, Vos S, Vandenberghe R, et al. The EMIF-AD multimodal biomarker discovery study: design, methods and cohort characteristics. *Alzheimers Res Ther*. 2018;10:64.

17. 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.

18. Zetterberg H, Skillbäck T, Mattsson N, et al. Association of cerebrospinal fluid neurofilament light concentration with Alzheimer's disease progression. *JAMA Neurol*. 2016;73:60-67.

19. Portelius E, Zetterberg H, Skillbäck T, et al. Cerebrospinal fluid neurogranin: relation to cognition and neurodegeneration in Alzheimer's disease. *Brain*. 2015;138:3373-3385.

20. Olsson B, Hertze J, Lautner R, et al. Microglial markers are elevated in the prodromal phase of Alzheimer's disease and vascular dementia. *J Alzheimer's Dis*. 2013;33:45-53.

21. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*. 2010;34:816-834.

22. Li Y, Willer C, Sanna S, Abecasis G. Genotype Imputation. *Annu Rev Genomics Hum Genet*. 2009;10:387-406.

23. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 2007;23:1294-1296.

24. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8:1826.

25. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLOS Comput Biol*. 2015;11:e1004219.

26. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559-575.

27. Van Deerlin VM, Sleiman PMA, Martinez-Lage M, et al. Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. *Nat Genet*. 2010;42:234-239.

28. Pottier C, Zhou X, Perkerson RB, et al. Potential genetic modifiers of disease risk and age at onset in patients with frontotemporal lobar degeneration and GRN mutations: a genome-wide association study. *Lancet Neurol*. 2018;17:548-558.

29. Gaunt TR, Shihab HA, Hemani G, et al. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol*. 2016;17:61.

30. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet*. 2018;50:1112-1121.

31. Mez J, Chung J, Jun G, et al. Two novel loci, COBL and SLC10A2, for Alzheimer's disease in African Americans. Alzheimer's. *Dement*. 2017;13:119-129.

32. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47:1236-1241.

33. Byers SA, Price JP, Cooper JJ, Li Q, Price DH. HEXIM2, a HEXIM1-related protein, regulates positive transcription elongation factor b through association with 7SK. *J Biol Chem*. 2005;280:16360-16367.

34. Sasayama D, Hattori K, Ogawa S, et al. Genome-wide quantitative trait loci mapping of the human cerebrospinal fluid proteome. *Hum Mol Genet*. 2017;26:44-51.

35. 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;7:657-657.

36. Buniello A, Macarthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47:D1005-D1012.

37. Emilsson V, Ilkov M, Lamb JR, et al. Co-regulatory networks of human serum proteins link genetics to disease. *Science (80-)*. 2018;36:769-773.

38. Mattsson N, Cullen NC, Andreasson U, Zetterberg H, Blennow K. Association between longitudinal plasma neurofilament light and neurodegeneration in patients with Alzheimer's disease. *JAMA Neurol*. 2019;76:791.

39. Atamna H, Frey WH. A role for heme in Alzheimer's disease: heme binds amyloid β and has altered metabolism. *Proc Natl Acad Sci U S A*. 2004;101:11153-11158.

40. Dwyer BE, Stone M, Zhu X, Perry G, Smith MA. Heme deficiency in Alzheimer's disease: a possible connection to porphyria. *J Biomed Biotechnol*. 2006;2006:24038.

41. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. *Nature*. 2018;558(7708):73-79.

42. Suhre K, Arnold M, Bhagwat AM, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun*. 2017;8:14357.

43. Ober C, Tan Z, Sun Y, et al. Effect of variation in CHI3L1 on serum YKL-40 level, risk of asthma, and lung function. *N Engl J Med*. 2008;358(16):1682-1691.

44. Willemse EAJ, Vermeiren Y, Garcia-Ayllon MS, et al. Pre-analytical stability of novel cerebrospinal fluid biomarkers. *Clin Chim Acta*. 2019;497:204-211.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.