



OPEN ACCESS

**Edited by:**

Can Martin Zhang,  
Massachusetts General Hospital  
and Harvard Medical School,  
United States

**Reviewed by:**

Davide Chiasserini,  
University of Perugia, Italy  
Xia Li,  
Shanghai Jiao Tong University, China

**\*Correspondence:**

Lars Bertram  
lars.bertram@uni-luebeck.de

† These authors have contributed  
equally to this work

**\*Present address:**

Johannes Streffer,  
AC Immune SA, EPFL Innovation  
Park, Lausanne, Switzerland

**Specialty section:**

This article was submitted to  
Alzheimer's Disease and Related  
Dementias,  
a section of the journal  
Frontiers in Aging Neuroscience

**Received:** 21 December 2021

**Accepted:** 15 February 2022

**Published:** 21 March 2022

**Citation:**

Homann J, Osburg T, Ohlei O,  
Dobricic V, Deecke L, Bos I,  
Vandenberghe R, Gabel S,  
Scheltens P, Teunissen CE,  
Engelborghs S, Frisoni G, Blin O,  
Richardson JC, Bordet R, Lleó A,  
Alcolea D, Popp J, Clark C,  
Peyratout G, Martinez-Lage P,  
Tainta M, Dobson RJB,  
Legido-Quigley C, Slegers K,  
Van Broeckhoven C, Wittig M,  
Franke A, Lill CM, Blennow K,  
Zetterberg H, Lovestone S, Streffer J,  
ten Kate M, Vos SJB, Barkhof F,  
Visser PJ and Bertram L (2022)  
Genome-Wide Association Study  
of Alzheimer's Disease Brain Imaging  
Biomarkers and Neuropsychological  
Phenotypes in the European Medical  
Information Framework  
for Alzheimer's Disease Multimodal  
Biomarker Discovery Dataset.  
*Front. Aging Neurosci.* 14:840651.  
doi: 10.3389/fnagi.2022.840651

# Genome-Wide Association Study of Alzheimer's Disease Brain Imaging Biomarkers and Neuropsychological Phenotypes in the European Medical Information Framework for Alzheimer's Disease Multimodal Biomarker Discovery Dataset

Jan Homann<sup>1†</sup>, Tim Osburg<sup>1†</sup>, Olena Ohlei<sup>1</sup>, Valerija Dobricic<sup>1</sup>, Laura Deecke<sup>1</sup>, Isabelle Bos<sup>2,3</sup>, Rik Vandenberghe<sup>4,5</sup>, Silvy Gabel<sup>4</sup>, Philip Scheltens<sup>3</sup>, Charlotte E. Teunissen<sup>6</sup>, Sebastiaan Engelborghs<sup>7,8</sup>, Giovanni Frisoni<sup>9,10</sup>, Olivier Blin<sup>11</sup>, Jill C. Richardson<sup>12</sup>, Regis Bordet<sup>13</sup>, Alberto Lleó<sup>14</sup>, Daniel Alcolea<sup>14</sup>, Julius Popp<sup>15,16</sup>, Christopher Clark<sup>15</sup>, Gwendoline Peyratout<sup>16</sup>, Pablo Martinez-Lage<sup>17</sup>, Mikel Tainta<sup>17</sup>, Richard J. B. Dobson<sup>18,19,20,21,22</sup>, Cristina Legido-Quigley<sup>23,24</sup>, Kristel Slegers<sup>25,26</sup>, Christine Van Broeckhoven<sup>26,27</sup>, Michael Wittig<sup>28</sup>, Andre Franke<sup>28</sup>, Christina M. Lill<sup>1,29</sup>, Kaj Blennow<sup>30,31</sup>, Henrik Zetterberg<sup>30,31,32,33</sup>, Simon Lovestone<sup>34</sup>, Johannes Streffer<sup>35,36†</sup>, Mara ten Kate<sup>4,37</sup>, Stephanie J. B. Vos<sup>2</sup>, Frederik Barkhof<sup>37,38</sup>, Pieter Jelle Visser<sup>2,4,39</sup> and Lars Bertram<sup>1,40\*</sup>

<sup>1</sup> Lübeck Interdisciplinary Platform for Genome Analytics (LIGA), University of Lübeck, Lübeck, Germany, <sup>2</sup> Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Alzheimer Centrum Limburg, Maastricht University, Maastricht, Netherlands, <sup>3</sup> Department of Neurology, Alzheimer Center Amsterdam, Amsterdam University Medical Centers, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, <sup>4</sup> Laboratory for Cognitive Neurology, Department of Neurosciences, KU Leuven, Leuven, Belgium, <sup>5</sup> Neurology Service, University Hospital Leuven, Leuven, Belgium, <sup>6</sup> Neurochemistry Laboratory, Department of Clinical Chemistry, Amsterdam Neuroscience, Amsterdam University Medical Centers, Vrije Universiteit, Amsterdam, Netherlands, <sup>7</sup> Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium, <sup>8</sup> Department of Neurology and Center for Neurosciences, Universitair Ziekenhuis Brussel and Vrije Universiteit Brussel (VUB), Brussels, Belgium, <sup>9</sup> Department of Psychiatry, University of Geneva, Geneva, Switzerland, <sup>10</sup> IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy, <sup>11</sup> Institut Neurosciences Timone, Aix Marseille University, Marseille, France, <sup>12</sup> Neurosciences Therapeutic Area, GlaxoSmithKline R&D, Stevenage, United Kingdom, <sup>13</sup> Lille Neuroscience and Cognition, University of Lille, Inserm, CHU Lille, Lille, France, <sup>14</sup> Memory Unit, Neurology Department, Hospital de Sant Pau, Barcelona and Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain, <sup>15</sup> Department of Geriatric Psychiatry, University Hospital of Psychiatry Zurich, Zurich, Switzerland, <sup>16</sup> Old Age Psychiatry, Department of Psychiatry, University Hospital of Lausanne, Lausanne, Switzerland, <sup>17</sup> Department of Neurology, Center for Research and Advanced Therapies, CITA-Alzheimer Foundation, Donostia-San Sebastian, Spain, <sup>18</sup> Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, London, United Kingdom, <sup>19</sup> NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, London, United Kingdom, <sup>20</sup> Health Data Research UK London, University College London, London, United Kingdom, <sup>21</sup> Institute of Health Informatics, University College London, London, United Kingdom, <sup>22</sup> NIHR Biomedical Research Centre at University College London Hospitals NHS Foundation Trust, London, United Kingdom, <sup>23</sup> Steno Diabetes Center, Copenhagen, Denmark, <sup>24</sup> King's College London, Institute of Pharmaceutical Sciences, London, United Kingdom, <sup>25</sup> Complex Genetics of Alzheimer's Disease Group, Center for Molecular Neurology, VIB, Antwerp, Belgium, <sup>26</sup> Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium, <sup>27</sup> Neurodegenerative Brain Diseases Group, Center for Molecular Neurology, VIB, Antwerp, Belgium, <sup>28</sup> Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany, <sup>29</sup> Ageing Epidemiology Research Unit, School of Public Health, Imperial College London, London, United Kingdom, <sup>30</sup> Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, <sup>31</sup> Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden, <sup>32</sup> Department of Neurodegenerative Disease, University College London, Queen Square Institute of Neurology, Queen Square, London, United Kingdom, <sup>33</sup> UK Dementia Research Institute at University College London, London,

United Kingdom, <sup>34</sup> Department of Psychiatry, University of Oxford, Oxford, United Kingdom, <sup>35</sup> Reference Center for Biological Markers of Dementia (BIODEM), Institute Born-Bunge, University of Antwerp, Antwerp, Belgium, <sup>36</sup> Janssen R&D, LLC, Beerse, Belgium, <sup>37</sup> Department of Radiology and Nuclear Medicine, Amsterdam University Medical Centers, Amsterdam Neuroscience, Amsterdam, Netherlands, <sup>38</sup> Institutes of Neurology and Healthcare Engineering, University College London, London, United Kingdom, <sup>39</sup> Division of Neurogeriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden, <sup>40</sup> Department of Psychology, University of Oslo, Oslo, Norway

Alzheimer's disease (AD) is the most frequent neurodegenerative disease with an increasing prevalence in industrialized, aging populations. AD susceptibility has an established genetic basis which has been the focus of a large number of genome-wide association studies (GWAS) published over the last decade. Most of these GWAS used dichotomized clinical diagnostic status, i.e., case vs. control classification, as outcome phenotypes, without the use of biomarkers. An alternative and potentially more powerful study design is afforded by using quantitative AD-related phenotypes as GWAS outcome traits, an analysis paradigm that we followed in this work. Specifically, we utilized genotype and phenotype data from  $n = 931$  individuals collected under the auspices of the European Medical Information Framework for Alzheimer's Disease Multimodal Biomarker Discovery (EMIF-AD MBD) study to perform a total of 19 separate GWAS analyses. As outcomes we used five magnetic resonance imaging (MRI) traits and seven cognitive performance traits. For the latter, longitudinal data from at least two timepoints were available in addition to cross-sectional assessments at baseline. Our GWAS analyses revealed several genome-wide significant associations for the neuropsychological performance measures, in particular those assayed longitudinally. Among the most noteworthy signals were associations in or near *EHBP1* (EH domain binding protein 1; on chromosome 2p15) and *CEP112* (centrosomal protein 112; 17q24.1) with delayed recall as well as *SMOC2* (SPARC related modular calcium binding 2; 6p27) with immediate recall in a memory performance test. On the X chromosome, which is often excluded in other GWAS, we identified a genome-wide significant signal near *IL1RAPL1* (interleukin 1 receptor accessory protein like 1; Xp21.3). While polygenic score (PGS) analyses showed the expected strong associations with SNPs highlighted in relevant previous GWAS on hippocampal volume and cognitive function, they did not show noteworthy associations with recent AD risk GWAS findings. In summary, our study highlights the power of using quantitative endophenotypes as outcome traits in AD-related GWAS analyses and nominates several new loci not previously implicated in cognitive decline.

**Keywords:** genome-wide association study, GWAS, X chromosome, Alzheimer's disease (AD), MRI, imaging, cognitive function

## INTRODUCTION

Alzheimer's disease is the most common neurodegenerative disease in humans and the most common form of dementia. In 2018, estimates were published that 50 million dementia patients exist worldwide, about two-third of whom were diagnosed with AD (Patterson, 2018). Pathologically, AD is characterized by the accumulation of extracellular amyloid  $\beta$  ( $A\beta$ ) peptide deposits ("plaques") and intracellular hyperphosphorylated tau protein aggregates ("tangles") in the brain, leading to synaptic dysfunction, neuroinflammation, neuronal loss, and, ultimately, onset of cognitive decline (Sperling et al., 2014; Mattsson et al., 2015). Genetically, AD is a heterogeneous

disorder with both monogenic and polygenic forms. The former is caused by highly penetrant but rare mutations in three genes encoding the amyloid beta precursor protein (*APP*) and presenilins 1 and 2 (*PSEN1/PSEN2*), which only make up a small fraction (<<5%) of all AD cases (Cacace et al., 2016). Most patients, however, suffer from "polygenic AD," which is determined by the action (and interaction) of numerous independent genomic variants, likely in concert with non-genetic factors, such as environmental exposures (e.g., head trauma) and lifestyle choices (e.g., alcohol consumption and cigarette smoking) (Bertram and Tanzi, 2020). Based on results from the currently most recent and largest genome-wide association study (GWAS) performed in AD, there are

now 38 independent loci showing genome-wide significant association with disease risk (Wightman et al., 2021). The most strongly and most consistently associated AD risk gene is *APOE*, which encodes apolipoprotein E, a cholesterol transport protein that has been implicated in numerous amyloid-specific pathways, including amyloid trafficking, as well as plaque clearance (Holtzman et al., 2012). Although the heritability of polygenic AD is estimated to be around 60–80% (Gatz et al., 2006), *APOE* and the other currently known 37 independent risk loci explain only part of the disease's phenotypic variance (Wightman et al., 2021). While most AD GWAS only consider clinically diagnosed “probable AD” cases and cognitively unimpaired controls, involving a risk for mis-diagnosis of patients and inclusion of preclinical AD cases as controls, additional information about the genetic architecture of AD and additional statistical power is also afforded by using “endophenotypes” related to AD, ideally measured on a quantitative scale such as biomarker data, imaging, or neurocognitive performance (Gottesman and Gould, 2003; MacRae and Vasan, 2011; Zhang et al., 2020).

In our study, we expand earlier work from our group (Hong et al., 2020, 2021) derived from European Medical Information Framework Alzheimer's Disease Multimodal Biomarker Discovery (EMIF-AD MBD) sample (Bos et al., 2018). Specifically, in two previous GWAS we set out to identify variants underlying variation in several cerebrospinal fluid (CSF) phenotypes, such as levels of CSF A $\beta$  and tau protein (Hong et al., 2020), or neurofilament light (NfL) chain, chitinase-3-like protein 1 (YKL-40), and neurogranin (Ng), which reflect axonal damage, astroglial activation, and synaptic degeneration, respectively (Hong et al., 2021). However, the EMIF-AD MBD dataset features several other quantitative phenotypes, including cross-sectional MRI measurements and cross-sectional and longitudinal neuropsychological tests, which are used as outcome traits in the current study. Thus, while using the same individuals and identical genome-wide SNP genotype data as in Hong et al. (2020, 2021), we substantially extended our previous work by focusing on entirely novel phenotypic domains available in EMIF-AD MBD. Specifically, we performed GWAS and polygenic score (PGS) analyses on seven neuropsychological (using both cross-sectional and longitudinal data) and five brain imaging phenotypes (using cross-sectional data from MRI scans). In the 19 performed GWAS scans (which also included the X chromosome), we identified a total of 13 genome-wide significant loci highlighting several novel genes showing association with the analyzed traits. While we do not see a noteworthy overlap in the genetic architectures underlying our “endophenotypes” and AD by polygenic score (PGS) analysis, we did observe significant correlations in PGS constructed from earlier GWAS on hippocampal volume (Hibar et al., 2017) and general cognitive function (Davies et al., 2018) with the respective phenotypes in EMIF-AD MBD. Taken together, our novel results pinpoint

several new genetic loci potentially involved in AD-related pathophysiology.

## MATERIALS AND METHODS

### Sample Description

Analyses were based on the EMIF-AD MBD dataset which was collected across 11 different European study centers (Bos et al., 2018). In total, this dataset included 1,221 [563 (46%) female; mean age = 67.9 years, *SD* = 8.3] individuals from three diagnostic stages: normal controls (NC), subjects with mild cognitive impairment (MCI) and subjects with a clinical diagnosis of AD. A diagnosis of AD was based on National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer's Disease and Related Disorders Association criteria (NINCDS-ADRDA) (McKhann et al., 1984) while MCI was diagnosed using criteria of Petersen (2004) in nine centers, while two centers used the criteria by Winblad et al. (2004). Individuals showing normal performance on neuropsychological assessment (within 1.5 *SD* of the average for age, gender, and education) at baseline were classified as NC (Bos et al., 2018). An overview of the quantitative phenotypes investigated in this study is provided in **Table 1**. Due to partially missing phenotype data (in the neurocognitive domain), the effective sample sizes vary for the different GWAS analyses (see **Table 1**). The local medical ethical review boards in each participating recruitment center had approved the study prior to commencement. Furthermore, all subjects had provided written informed consent at the time of inclusion in the cohort for use of data, samples and scans (Bos et al., 2018).

### Magnetic Resonance Imaging Phenotypes Description

The five MRI phenotypes were collected for 862 subjects. Brain MRIs were used to assess hippocampal volume (mm<sup>3</sup>, left and right hemisphere, and sum of both; all adjusted for intracranial volume), whole brain cortical thickness (in mm), and white matter lesions (WML; using the Fazekas scale) (Ten Kate et al., 2018). The Fazekas scale categorizes WMLs into 4 categories: Level 0 (no or almost no lesion), level 1 (multiple punctate lesions), level 2 (early confluent WML), and level 3 (presence of large confluent WML). Details on the scanning procedures and data harmonization across centers can be found in Bos et al. (2018) and Ten Kate et al. (2018).

### Neuropsychological Phenotypes Description

Cross-sectional (and follow-up) data were available for the following seven neuropsychological domains within the EMIF-AD MBD dataset: global cognition (Mini Mental State Examination, MMSE), attention, executive function, language, memory (immediate and delayed) and visuoconstruction (for a detailed description of all neuropsychological tests see **Supplementary Material**). For each cognitive domain, a primary

**TABLE 1** | Description of EMIF-AD MBD datasets analyzed per phenotype.

Category	Phenotype	Baseline		Longitudinal	
		Sample size	MAF filter	Sample size	MAF filter
Neuropsychological	MMSE	867	0.02	520	0.02
	Attention	806	0.01	402	0.02
	Executive functioning	686	0.01	234	0.02
	Language	849	0.01	409	0.02
	Memory Delayed	729	0.01	337	0.02
	Memory Immediate	797	0.01	345	0.02
	Visuoconstruction	429	0.02	149	0.04
MRI	Fazekas score	606	0.01	n.a.	n.a.
	Cortical thickness	560	0.01	n.a.	n.a.
	Left Hippocampus volume	605	0.01	n.a.	n.a.
	Right Hippocampus volume	605	0.01	n.a.	n.a.
	Summed Hippocampus volume	605	0.01	n.a.	n.a.

"MAF filter" denotes the applied MAF filter for each GWAS. For cross-sectional MMSE we used an MAF threshold of 0.02 due to residual inflation of the GWAS test statistics. Information on tests used for generating baseline and longitudinal phenotypes can be found in **Supplementary Material**. MMSE, Mini Mental State Examination. "n.a.", not available. MRI, Magnetic Resonance Imaging. MAF, Minor Allele Frequency.

test was selected by Bos et al. (2018). If the preferred test were not available, an alternative priority test from the same cognitive domain was chosen. More details on the neuropsychological tests used for generating these phenotypes can be found in Bos et al. (2018). Raw data on these tests were normalized with the help of a  $z$ -transformation, so that the data were comparable within a cognitive domain despite representing partially different tests across centers. For the cross-sectional GWAS analyses, the  $z$ -scores derived from baseline data were used. The number of subjects used for each test can be found in **Supplementary Material**. For all seven neuropsychological domains, follow-up data from at least one additional time point were available for each individual and used to construct a longitudinal phenotype using the following formula [which estimates the relative change in cognitive performance per time interval (here: years)]:

$$\frac{\text{Score}_{\text{last}} - \text{Score}_{\text{first}}}{\frac{\text{Score}_{\text{last}} + \text{Score}_{\text{first}}}{2} * \text{interval}}$$

When calculating longitudinal phenotypes, this formula was applied separately for each neuropsychological test. Outlying scores were determined using a false discovery rate (FDR) threshold of 0.05 and were excluded from all subsequent analyses. Only the most frequently used tests per cognitive domain were included in the final phenotypes (for more information, see **Supplementary Material**). Both baseline and longitudinal phenotypes were adjusted for age at baseline.

## DNA Extraction, Genotype Imputation and Quality Control

A detailed description of the genotyping procedures, quality control (QC) and subsequent data processing can be found in Hong et al. (2020) (**Supplementary Material**). Here, the same genotype data were used for the GWAS analyses. Briefly, 936 DNA samples were subjected to genome-wide SNP genotyping

using the Infinium Global Screening Array (GSA) with Shared Custom Content (Illumina Inc.). Imputation was then performed using Minimac3 (Das et al., 2016). Extensive post-imputation QC resulted in 7,464,105 autosomal SNPs with a minor allele frequency (MAF)  $\geq 0.01$  in 888 individuals of European ancestry. More details can be found in **Supplementary Material**.

For the X chromosome, QC was performed separately for male and female subjects for non-pseudoautosomal regions, using slightly different criteria compared to the autosomes (see **Supplementary Material**). In contrast, pseudoautosomal regions (PAR1 and PAR2) were treated analogously to the autosomal SNPs. After QC, imputations were performed on the Sanger Institute imputation server<sup>1</sup> using the extended HRC reference panel (McCarthy et al., 2016). After imputation, we used the same QC criteria as for the autosomal SNPs but performed these separately for female and male data sets, except the HWE test ( $P < 1.0E-4$ ) which was performed on all samples combined as recommended previously (Graffelman and Weir, 2016) and implemented in PLINK2. For males, markers were coded as 0 vs. 2 (instead of 0 vs. 1), to adjust for the missing second X chromosome (as recommended in Smith et al., 2021).

## Genome-Wide Association Studies and Post-Genome-Wide Association Studies Analyses

SNP-based association analyses were performed assuming an additive linear model (command: `-glm`) using allele dosages (to account for imputation uncertainty) in PLINK2 (Purcell et al., 2007). The model is equivalent to a test for a dose-response relationship between allele dose (i.e., one or two copies vs. reference genotype) on the outcome trait. The covariates included in the analyses were sex, diagnostic status and the first three principal components from a principal component analysis (PCA) to adjust for population-specific differences. Generally,

<sup>1</sup><https://imputation.sanger.ac.uk/>

we excluded SNPs from the GWAS analyses with  $MAF < 0.01$ . However, due to differences in the effective sample sizes across phenotypes this threshold was adapted upward (up to 0.04) to prevent inflation of test statistics owing to low frequency SNPs (see **Table 1** for more details). Diagnostic status was coded with two dummy variables as follows: NC = (0, 0), MCI = (0, 1), AD = (1, 1). For four longitudinal cognitive phenotypes an additional dummy variable was introduced to code for the neuropsychological test used, in cases where two different tests were used for generating these phenotypes. Details can be found in **Supplementary Material**.

To explore associations on the X chromosome that were potentially driven by genetic sex, we additionally conducted the analyses separately in females and males. We then combined these two additional sets of results in a meta-analysis using Stouffer's method as implemented in METAL (Willer et al., 2010). As we found no noteworthy differences in the results using Stouffer's method, only the results from the linear regression analysis in the combined sample are shown.

The FUMA platform<sup>2</sup> (Watanabe et al., 2017) was used for post-GWAS analyses, including gene-based association analyses (via MAGMA, de Leeuw et al., 2015) and to annotate and visualize the GWAS results. To this end, we defined genome-wide significance at  $\alpha = 5.0E-08$  for the SNP-based analyses while genome-wide suggestive evidence was set at  $\alpha = 1.0E-05$ . For the gene-based analyses, we adjusted for the number of protein-coding genes examined (19,485) using the Bonferroni method, resulting in a threshold of  $\alpha = 2.566E-06$ .

In FUMA, both the SNP annotation and the Combined Annotation Dependent Depletion (CADD) score (Rentzsch et al., 2021) are provided. The main GWAS results are reported only for "independent significant" SNPs, as defined by FUMA. These represent SNPs that are not highly correlated with one another using a threshold of  $r^2 < 0.6$  (using reference data from the 1,000 Genomes Project).

Subsequently, the top SNPs, i.e., those with the smallest  $P$ -values per respective phenotype, were examined in more detail using additional tools. First, the Variant Effect Predictor on Ensembl (VEP<sup>3</sup>, McLaren et al., 2016) was used to determine a possibly functional effect due to changes in the coding sequence, e.g., missense variants. Second, SNPs were examined using data from the RegulomeDB database<sup>4</sup> (Boyle et al., 2012) to assess possible effects on gene expression. Third, we used data from the Genotype-Tissue Expression (GTEx, V8) project portal<sup>5</sup> (Lonsdale et al., 2013) to assess whether SNPs represent expression / splicing quantitative trait loci (eQTLs/sQTLs). While GTEx provides data on gene expression in 54 tissues, we laid particular emphasis in genes expressed in brain. Lastly, we interrogated the GWAS catalog<sup>6</sup> (Buniello et al., 2019) to assess whether any of the top SNPs were previously reported to show association with other phenotypes by GWAS. To this end, we

considered genes and loci within a 1 Mb region ( $\pm 500,000$  bp) around the SNP of interest. In case SNPs not identical to our "top SNP" were reported to show association with an AD-relevant phenotype (brain imaging, cognition, etc.), the LDlink platform (Machiela and Chanock, 2015) was used to determine pairwise LD to top SNPs.<sup>7</sup> In this context we defined relevant LD using a threshold of  $r^2 > 0.6$ .

## Polygenic Score Analysis

In addition to the primary GWAS analyses described above, we also calculated polygenic scores (PGS) to estimate the extent of genetic correlation with the GWAS results for three other phenotypes. To this end, we used the summary statistics of a GWAS on AD risk (Jansen et al., 2019) as comparison to both phenotypic domains (MRI and neurocognitive performance) of our study, and the GWAS on general cognitive function (Davies et al., 2018) as comparison to the GWAS on neuropsychological phenotypes. Finally, the GWAS on hippocampal volume (Hibar et al., 2017) served as comparison to our GWAS analyses on MRI phenotypes. PGS calculations were performed using PRSice-2 software (Choi and O'Reilly, 2019). Statistical analyses fitted general linear regression models with PGS as predictor adjusting for the same covariates as in the primary GWAS analyses: sex, diagnostic status, and PC1-3 (and type of cognitive test, where applicable). To adjust for multiple testing of this arm of our study, we used a conservative threshold based on Bonferroni adjustment ( $\alpha = 5.0E-03 = 0.05/(5*2)$  for the MRI phenotypes, and  $\alpha = 1.8E-03 = 0.05/(14*2)$  for the neuropsychological phenotypes). However, given the (at least partial) correlation between phenotypes, we note that the true threshold is likely somewhere between 0.05 and these Bonferroni-adjusted values.

## RESULTS

### Genome-Wide Association Studies on Magnetic Resonance Imaging Phenotypes

The genomic inflation factor  $\lambda$  ranged between 1.004 and 1.012 in all five SNP-based analyses, indicating that the results of the MRI GWAS analyses were not affected by substantial inflation of the test statistics. In the actual association analyses of the five quantitative MRI phenotypes, we identified no genome-wide significant ( $P < 5.0E-08$ ) signals but observed 385 variants with at least suggestively significant ( $P < 1.0E-05$ ) evidence of association (**Supplementary Tables 15–19**). The lowest  $P$ -value was observed with SNP rs16829761 for the Fazekas phenotype ( $P = 5.08E-08$ ; **Supplementary Figure 31**), which only fell slightly above the genome-wide significance threshold. According to VEP (McLaren et al., 2016), this variant is located in an intron of the genes *IQCJ* (protein: IQ motif containing J) and *SCHIP1* (protein: schwannomin-interacting protein 1). In the GTEx database (Lonsdale et al., 2013), the lead-SNP identified here (rs16829761) is not listed as eQTL or sQTL, which may be due

<sup>2</sup><http://fuma.ctglab.nl/>

<sup>3</sup><http://grch37.ensembl.org/Tools/VEP>

<sup>4</sup><https://regulomedb.org/regulome-search>

<sup>5</sup><https://www.gtexportal.org/home/>

<sup>6</sup><https://www.ebi.ac.uk/gwas/home>

<sup>7</sup><https://ldlink.nci.nih.gov/?tab=ldpair>

to the comparatively low MAF (0.01). The CADD score, i.e., the *in silico* predicted deleteriousness, of rs16829761 is also low at approximately 0.074. In addition, none of the gene-based GWAS analyses using MAGMA revealed any genome-wide significant signals ( $P < 2.566E-06$ ) using the MRI traits analyzed. The genomic inflation factor  $\lambda$  ranged between 0.984 and 1.060 in these five gene-based analyses.

## Genome-Wide Association Studies on Neuropsychological Phenotypes

Across the 14 GWAS performed on cross-sectional and longitudinal neuropsychological phenotypes available in EMIF-AD MBD, there were a total of 13 genome-wide significant loci, two of which were identified via the gene-based analyses using MAGMA (de Leeuw et al., 2015). Three of the genome-wide significant signals were observed in the analyses of cross-sectional phenotype data and 10 with longitudinal outcomes. Overall, none of the sets of GWAS results in this arm of our study appeared to be strongly affected by inflation of the genome-wide test statistics as evidenced by genomic inflation factors near 1 (range: 0.969–1.012 in the SNP-based analyses and 0.922–1.036 in the gene-based analyses). **Table 2** provides a detailed summary of these genome-wide significant loci, and **Figure 1** shows multi-trait Manhattan (MH) plots of the SNP-based GWAS results for cross-sectional (**Figure 1A**) and longitudinal (**Figure 1B**) analyses (for corresponding QQ plots: see **Supplementary Figures 1, 2**). The following two paragraphs highlight the most interesting results in either the analyses of cross-sectional or longitudinal neuropsychological traits.

### Analyses of Cross-Sectional Data

The most interesting finding in this domain was elicited by markers in *EHBPI* which showed genome-wide significant evidence of association with the delayed recall memory phenotype in the gene-based analysis ( $P = 1.17E-07$ ; **Table 2** and **Supplementary Figure 20**). The lead SNP (rs6705798) in this region only missed the genome-wide significance threshold by a small margin ( $P = 8.78E-08$ ; **Table 2** and **Figure 1A**). *EHBPI* is located on chromosome 2p15 and encodes EH domain binding protein 1.

### Analyses of Longitudinal Data

The strongest signal in the longitudinal analyses was elicited by a locus on chromosome 6q27 (rs73045836;  $P = 7.50E-11$ ; **Table 2**, **Figure 1B**, and **Supplementary Figure 25**) in the analysis using an immediate memory recall paradigm. This SNP is located in an intron of *SMOC2* coding for secreted modular calcium-binding protein 2, which, among other functions, promotes extracellular matrix assembly (Gao et al., 2019). It needs to be noted that with an MAF  $\sim 2\%$  this SNP is rather infrequent which may increase the possibility of representing a false-positive finding. Perhaps more interesting is the association signal observed near SNP rs5943462 (MAF  $\sim 0.05$ ) and the visuoconstruction phenotype on the X chromosome ( $P = 1.06E-09$ ; **Table 2**, **Figure 1B**, and **Supplementary Figure 29**). This SNP is an intronic variant located in *IL1RAPL1* encoding interleukin 1 receptor accessory protein-like 1, which belongs to a class of molecules that

regulate synapse formation (Montani et al., 2019). The third highlighted signal in this domain relates to the genome-wide significant variant rs74381761 (MAF  $\sim 0.05$ ) on chromosome 8p23.1 ( $P = 1.89E-08$ ; **Table 2**, **Figure 1B**, and **Supplementary Figure 5**) which shows association with the longitudinal MMSE phenotype. The lead SNP is located in an intergenic region near *TNKS* (gene-based  $P = 4.87E-04$ ; **Table 2**). This gene encodes the protein tankyrase, which belongs to a class of poly (ADP-ribose) polymerases and is involved in various processes in the body, such as telomere length regulation, the Wnt/ $\beta$ -catenin signaling pathway, or glucose transport (Damale et al., 2020). The last featured signal relates to the association observed near SNP rs9652864 (MAF  $\sim 0.22$ ) on chromosome 17q24.1 ( $P = 3.20E-08$ ; **Table 2**, **Figure 1B**, and **Supplementary Figure 21**) and the delayed recall test. This variant is located in an intron of *CEP112*, which encodes centrosomal protein 112. Overall, there were eight correlated SNPs in this locus all showing strongly association (**Supplementary Table 10**).

### Comparison of Cross-Sectional vs. Longitudinal Genome-Wide Association Studies Results

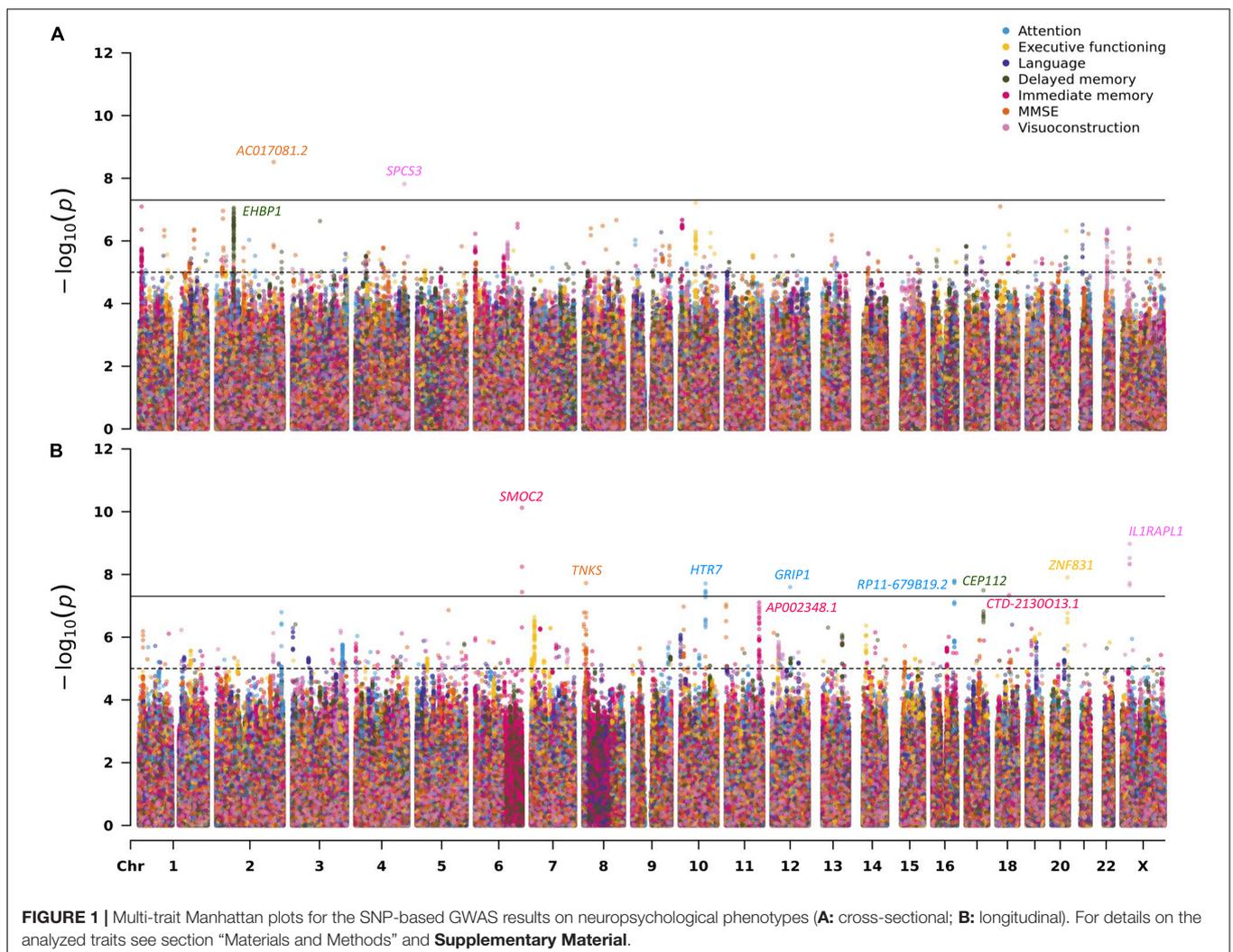
After completion of the separate GWAS on cross-sectional and longitudinal outcomes, we assessed whether the results of these two analysis arms showed any overlap. To this end, we followed two approaches: First, we performed a look-up of top results from one paradigm in the equivalent other. Specifically, we checked whether a genome-wide significant SNP from the cross-sectional analyses also had a low  $P$ -value in the corresponding longitudinal GWAS and vice versa. The lowest corresponding  $P$ -value was 0.015 (at baseline) for rs73045828, which attained  $P = 5.65E-09$  in the longitudinal GWAS for immediate memory (**Supplementary Table 21**). No further signal overlaps were observed across corresponding cross-sectional and longitudinal phenotypes. Second, we took a more comprehensive approach by comparing a larger set of SNPs across both phenotypic domains. To this end, we constructed PGS from the summary statistics of the cross-sectional GWAS (as an approximate measure of “aggregated SNP effects”) and used these PGS as independent variables in a linear model predicting longitudinal outcomes. Effectively, this allowed us to determine how much phenotypic variance in the longitudinal data can be explained by top SNPs of the matching cross-sectional GWAS. Overall, these analyses did not reveal a substantial correlation in genetic results for corresponding phenotypes (**Supplementary Table 22**), in agreement with the look up of individual SNPs (see above). The best model fit was observed with the PGS for executive function and visuoconstruction, where the GWAS top SNPs from the cross-sectional data used in the PGS explained 4–9% of the phenotypic variance of the corresponding longitudinal outcomes, respectively (**Supplementary Table 22**). We note, however, that the PGS method was not designed for computing genetic correlations of non-independent samples (as is the case here), so this analysis must be considered “exploratory,” and the reported results represent no more than “upper bounds” of the potential genetic correlations.

**TABLE 2** | Genome-wide significant associations observed in GWAS of cognitive phenotypes.

Study arm	Phenotype	Lead variant	Chr	Position	Nearest gene	A1	A2	Beta	MAF	P (SNP)	P (Gene)
Cross-sectional	MemoryDelayed	rs6705798	2p15	63,259,881	<i>EHBP1</i>	C	T	-0.32739	0.358	8.78E-08	<b>1.17E-07</b>
	MMSE	rs2122118	2q33.3	207,252,439	<i>AC017081.2</i>	G	A	-2.82266	0.022	<b>3.03E-09</b>	n.a.
	Visuoconstruction	rs113492235	4q34.2	177,252,900	<i>SPCS3</i>	T	C	-2.17956	0.022	<b>1.51E-08</b>	0.15674
Longitudinal	MemoryImmediate	rs73045836	6q27	169,062,739	<i>SMOC2</i>	G	T	-0.36094	0.020	<b>7.50E-11</b>	0.0035373
	MMSE	rs74381761	8p23.1	9,389,761	<i>TNKS</i>	C	G	-0.08453	0.048	<b>1.89E-08</b>	0.00048716
	Attention	rs116900143	10q23.31	92,588,290	<i>HTR7</i>	C	T	-0.35173	0.023	<b>1.95E-08</b>	0.019663
	MemoryImmediate	rs11217863	11q23.3	120,293,138	<i>AP002348.1</i>	A	G	-0.16626	0.080	7.81E-08	<b>8.91E-07</b>
	Attention	rs111959303	12q14.3	66,844,015	<i>GRIP1</i>	T	C	0.37459	0.022	<b>2.52E-08</b>	0.81478
	Attention	rs34736485	16q23.2	79,272,611	<i>RP11-679B19.2</i>	T	G	0.31924	0.022	<b>1.59E-08</b>	n.a.
	MemoryDelayed	rs9652864	17q24.1	63,741,645	<i>CEP112</i>	A	T	0.29184	0.218	<b>3.20E-08</b>	0.016339
	MemoryImmediate	rs146202660	18q21.1	45,022,937	<i>CTD-2130013.1</i>	T	G	-0.29342	0.029	<b>4.63E-08</b>	n.a.
	Executive	rs16982556	20q13.32	57,801,889	<i>ZNF831</i>	T	C	-0.29752	0.062	<b>1.26E-08</b>	0.0025565
	Visuoconstruction	rs5943462	Xp21.3	28,823,154	<i>IL1RAPL1</i>	G	C	-0.14082	0.051	<b>1.06E-09</b>	0.006719

**Bold font indicates genome-wide significant (on SNP- or gene-level) results (see section Materials and Methods for details). "Chr" and "Position" according to GRCh37/hg19. "A1" denotes the effect allele. "P (SNP)" is the P-value of the lead SNP at this locus. "P (Gene)" is the P-value belonging to "Nearest gene." Top results from these GWAS analyses can be found in **Supplementary Tables 1–19**.**

MMSE, Mini Mental State Examination. "n.a.", not available. MRI, Magnetic Resonance Imaging. MAF, Minor Allele Frequency.



**FIGURE 1** | Multi-trait Manhattan plots for the SNP-based GWAS results on neuropsychological phenotypes (**A**: cross-sectional; **B**: longitudinal). For details on the analyzed traits see section "Materials and Methods" and **Supplementary Material**.

## Role of *APOE* in Genome-Wide Association Studies on Magnetic Resonance Imaging and Neuropsychological Performance

Given the substantial role that variants in *APOE* play in the genetic architecture of AD, we present findings for this locus separately, i.e., the results for SNP rs429358 (which defines the  $\epsilon 4$  allele) and rs7412 (which defines the  $\epsilon 2$  allele). In relation to the common genotype  $\epsilon 3/\epsilon 3$ , the risk to develop AD is increased by a factor of  $\sim 3.2$  for genotype  $\epsilon 3/\epsilon 4$ , while two  $\epsilon 4$  alleles (genotype  $\epsilon 4/\epsilon 4$ ) show ORs around 10–12 when compared to normal controls (Neu et al., 2017). The minor allele at rs429358 ( $\epsilon 4$ ) is overrepresented in the EMIF-AD MBD dataset with a MAF  $\sim 29\%$  (the MAF in the general Northern European population is  $\sim 16\%$ ), which is due to the special design of participant recruitment (see Bos et al., 2018). For the neuropsychological phenotypes, the *P*-values of rs429358 are unremarkable except for the domain “delayed memory,” where *P*-values of 0.0005 and 0.0042 were observed for the baseline and longitudinal analyses, respectively (Supplementary Table 20). In the MRI analyses, the only association signal observed with rs429358 was with hippocampal volume (Supplementary Table 20). Interestingly, this was driven by an association with the volume of the left ( $P = 0.0002$ ) hippocampus, while no association was observed with the corresponding data of the right hemisphere ( $P = 0.2956$ ). We note that for both traits, i.e., delayed memory and left hippocampal volume, the effect direction the corresponding  $\beta$  coefficient is consistent with the deleterious effect of the minor (T/ $\epsilon 4$ ) allele at rs429358 known from the literature (Neu et al., 2017). For the minor allele at rs7412 ( $\epsilon 2$ ) we observed no noteworthy association signals in any of the analyses performed in this study (Supplementary Table 20), possibly because power for this variant was much reduced owing to its lower MAF (4.6% here, 7.5% in the general western European control population).

## Polygenic Score Analyses Using Published Genome-Wide Association Studies Results

In these analyses we aimed to estimate the degree of genetic overlap between the MRI and neuropsychological outcomes available in EMIF-AD MBD and other relevant traits from the literature, such as AD risk, using published GWAS summary statistics.

### Polygenic Score Analyses With Magnetic Resonance Imaging Phenotypes

As expected, the strongest overlap was observed with a prior GWAS also using MRI outcomes. Specifically, we used GWAS results by the ENIGMA group (Hibar et al., 2017) who studied 26 imaging traits in  $n = 33,536$  individuals. Here, the best overlap was seen with each of the three hippocampal MRI traits (up to 2.7% variance explained,  $P = 6.0E-06$ ; Table 3 and Supplementary Table 24). In contrast, in PGS analyses using SNPs associated with AD risk (Jansen et al., 2019), we found only one moderate correlation with white matter damage (measured

by the Fazekas score). For this trait the AD SNPs explained 1.4% variance ( $P = 3.7E-03$ ; Table 3 and Supplementary Table 24).

### Polygenic Score Analyses With Neuropsychological Phenotypes

As for the MRI data, the best fit in the PGS analyses with the neuropsychological phenotypes was observed with a GWAS that also used neurocognitive performance as outcome (Davies et al., 2018). Specifically, this study defined a PCA-derived factor for “general cognitive function” which was analyzed in  $> 300,000$  individuals. In EMIF-AD MBD, associations with four of the 14 calculated PGS fell below the multiple testing threshold of  $1.8E-03$  (Table 3). The strongest association was observed with the longitudinal attention function for which the GWAS results from Davies et al. (2018) explained 2.3% of the phenotypic variance ( $P = 1.79E-03$ , Table 3 and Supplementary Table 23). The next best associations were seen with longitudinal executive functioning ( $r^2 = 0.028$ ;  $P = 9.79E-03$ ; Supplementary Table 23) and visuoconstructive abilities ( $r^2 = 0.058$ ;  $P = 3.08E-03$ ; Supplementary Table 23). However, these latter two associations do not survive multiple testing correction (Table 3). Interestingly and similar to the MRI-based results, we did not find strong evidence for a genetic overlap between the neurocognitive outcomes tested here and AD risk based on Jansen et al. (2019) (Supplementary Table 23). This included the various phenotypes measuring components of “memory” performance, regardless of whether or not they were ascertained cross-sectionally or longitudinally.

## DISCUSSION

This study extends previous GWAS analyses from our group utilizing phenotypic data from the EMIF-AD MBD study (Hong et al., 2020, 2021) using different outcome traits hitherto not analyzed by GWAS. The overarching goal of this work was to decipher the genetic architecture of AD-related MRI and neuropsychological (endo)phenotypes to better understand AD pathophysiology. Both previous EMIF-AD MBD GWAS focused on AD biomarkers measured in CSF and, among other findings, identified variants in *TMEM106B* as *trans*-pQTLs of CSF neurofilament light (NfL) levels (Hong et al., 2021). Interestingly, the same locus was subsequently highlighted as a novel AD risk locus in a GWAS on  $> 1.1$  million individuals (Wightman et al., 2021), showcasing the power of the quantitative biomarker GWAS approach that was also followed in this study. In the current work, we focused on biomarkers / phenotypes derived from brain imaging and neuropsychological testing in the same EMIF-AD MBD individuals. Overall, we performed 19 individual GWAS and identified a total of 13 genome-wide significant loci highlighting several novel genes that are potentially involved in contributing to AD pathophysiology. Our study represents one of few GWAS in the literature to also include the X chromosome, where we identified a genome-wide significant association between markers near *ILIRAPLI* and longitudinal visuoconstructive ability. Interestingly, neither *APOE* nor the other recently described AD GWAS loci appear to have a major

**TABLE 3** | Summary of PGS results significant after multiple testing correction.

Prior GWAS	Phenotype	Threshold	Number of SNPs	R <sup>2</sup>	P-value
Hibar et al. (2017)	Hippocampus volume sum	0.0001	127	0.027	6.06E-06
	Hippocampus volume left	0.0001	127	0.026	9.98E-06
	Hippocampus volume right	0.0001	127	0.024	2.48E-05
Jansen et al. (2019)	Fazekas	0.17075	22,269	0.014	3.72E-03
Davies et al. (2018)	Baseline MMSE	0.248	63,792	0.016	1.66E-06
	Baseline executive functioning	0.0014	4,469	0.018	1.98E-05
	Baseline language	0.0061	9,031	0.010	1.25E-03
	Longitudinal attention	5.0E-08	163	0.023	1.79E-03

"Threshold" refers to P-value cut-off used for PGS construction in prior GWAS summary statistics and "Number of SNPs" refers to the LD-pruned SNPs passing this threshold that are included in PGS calculations. "R<sup>2</sup>" denotes the phenotypic variance explained by the SNPs of the prior GWAS in the EMIF-AD MBD dataset. A full listing of results from these PRS analyses can be found in **Supplementary Material**.  
MMSE, Mini Mental State Examination.

impact on the traits analyzed in our study. In summary, our extensive genome-wide analyses nominate several novel loci potentially involved in neurocognitive functioning. Some of these may prove informative to better understand the genetic forces underlying AD and related phenotypes.

In the remainder of this section, we discuss the potential role of five loci, which we consider the most interesting findings of our study. The strongest GWAS signal was elicited by SNP rs73045836 ( $P = 7.50E-11$ ; **Table 2, Figure 1B**, and **Supplementary Figure 25**) showing genome-wide significant association with the longitudinal data of the immediate recall memory phenotype. The gene annotated to the associated region on chromosome 6q27, *SMOC2*, encodes secreted modular calcium-binding protein 2. *SMOC2* is an extracellular matrix protein from the secreted protein, acidic and rich in cysteine (SPARC) family (Gao et al., 2019) recently linked to age-dependent bone loss in humans (Morkmued et al., 2020). In the AD context, it is noteworthy that *SMOC2* was recently found to be altered in CSF samples of early AD in a proteomics profiling study (Whelan et al., 2019). Interestingly, there is similar evidence on a potential link to AD for a *SMOC2* isoform, i.e., *SMOC1* (gene: *SMOC1*, located on chromosome 14q24.2). While variants in this gene did not show strong evidence of genetic association with the traits analyzed here, it is noteworthy that *SMOC1* was recently nominated as a novel AD biomarker in proteomic screens of AD CSF and brain samples in various studies (Bai et al., 2020; Wang et al., 2020; Sathe et al., 2021). In summary, our finding of genome-wide significant association between *SMOC2* and memory performance in the EMIF-AD MBD datasets extends the emerging literature on the role of SPARC protein family members in AD and related traits.

The second strongest association signal was observed near SNP rs5943462 (MAF  $\sim 0.05$ ) on the X chromosome ( $P = 1.06E-09$ ; **Table 2; Figure 1B** and **Supplementary Figure 29**) with the longitudinal data of the visuoconstruction phenotype. The SNP is located in an intron of *ILIRAPL1*. This gene encodes interleukin 1 receptor accessory protein-like 1, which belongs to a class of molecules that regulate synapse formation. *ILIRAPL1* is mostly expressed in brain areas that are involved in memory development, such as hippocampus, dentate gyrus, and entorhinal cortex, suggesting that the

protein may have a specialized role in physiological processes underlying memory and learning abilities (Montani et al., 2019). Even small changes in the expression and function of these proteins can provoke major alterations in synaptic connectivity, resulting in cognitive damage (Montani et al., 2019). Moreover, *ILIRAPL1* was nominated as a candidate gene for X-linked mental retardation (Raymond, 2006). Although the GWAS on longitudinal visuoconstruction included only 149 individuals, we believe this signal to be plausible and very interesting because of the well-established role of *ILIRAPL1* on human brain function.

The third highlighted signal relates to the association between variant rs74381761 (MAF  $\sim 0.05$ ) on chromosome 8p23.1 ( $P = 1.89E-08$ ; **Table 2, Figure 1B**, and **Supplementary Figure 5**) and longitudinal MMSE measurements. This SNP is located in an intergenic region near *TNKS*. This gene encodes the protein tankyrase, which belongs to a class of poly (ADP-ribose) polymerases and is involved in various processes in the body, such as telomere regulation, Wnt/ $\beta$ -catenin signaling pathway or glucose transport (Damale et al., 2020). According to GTEx (Lonsdale et al., 2013), *TNKS* is highly expressed in brain (mostly in cerebellum). Moreover, SNPs annotated to *TNKS* were associated with brain white matter hyperintensity (WMH) measurements (Armstrong et al., 2020; Sargurupremraj et al., 2020; Zhao et al., 2021) and cortical surface area measurements (Grasby et al., 2020) according to the GWAS catalog (Buniello et al., 2019). With a gene-based P-value of 4.87E-04 and the strong functional link to brain function, we consider the signal around *TNKS* as plausible and very interesting.

The last highlighted finding from the longitudinal analyses relates to the genome-wide significant association observed between SNP rs9652864 and the delayed recall memory phenotype ( $P = 3.20E-08$ ; **Table 2, Figure 1B**, and **Supplementary Figure 21**). This variant (MAF = 0.218) attained a P-value of 6.73E-04 in the GWAS of Davies et al. (2018) on cross-sectional cognitive performance, lending additional support to our finding. The SNP is located in an intron of *CEP112* which encodes centrosomal protein 112. Centrosomal proteins are known as the components of the centrosome involved in centriole biogenesis, cell cycle progression, and spindle-kinetochore assembly control (Mazaheri Moghaddam et al., 2021). Despite showing only low levels of expression in

the central nervous system (CNS) according to GTEx, SNPs in this gene have been associated with cortical surface area by neuroimaging in two independent GWAS (Grasby et al., 2020; van der Meer et al., 2020) according to the GWAS catalog (Buniello et al., 2019). However, none of these neuroimaging SNPs is in relevant LD ( $r^2 > 0.6$ ) to the lead variant identified here. Notwithstanding, given that variants in this gene have shown genetic links to both cognitive function and structural brain imaging, we consider this finding as plausible and highly interesting.

In the GWAS analyses of the cross-sectional neurocognitive phenotypes, we observed three genome-wide significant signals, of which we consider the gene-based association with *EHBPI* as the most interesting finding ( $P = 1.17E-07$ ; **Table 2**, **Figure 1A**, and **Supplementary Figure 20**). This protein interacts with Eps15-homology domain-containing protein 1/2 (EHD1/2) that plays a central role in GLUT4 transport and couples endocytic vesicles to the actin cytoskeleton (Rai et al., 2020). It is highly expressed in many tissues, including the brain, according to GTEx (Lonsdale et al., 2013). While there does not appear to be an obvious link between *EHBPI* and brain function in the literature (e.g., in the GWAS catalog, Buniello et al., 2019), we note that this gene is located within 5 kb of *OTX1* (orthodenticle homeobox 1; gene-based  $P = 1.19E-05$ ), which acts as transcription factor and plays a role in brain and sensory organ development in *Drosophila* and vertebrates, including humans (Omodei et al., 2009). Our lead SNP in this region, i.e., rs6705798, falls just short of attaining genome-wide significance ( $P = 8.78E-08$ ; **Table 2** and **Figure 1A**) and is reported to represent an eQTL of both *OTX1* and *EHBPI* in various human tissues according to GTEx (Lonsdale et al., 2013).

In addition to searching for novel genetic determinants of the neuroimaging and neurocognitive traits analyzed in this study, we also investigated the overlap with known GWAS findings. First and foremost, this relates to two commonly studied alleles in the *APOE* gene, which appear to only play a minor role in this setting. Specifically, SNP rs429358, which defines the  $\epsilon 4$  allele in *APOE*, does not even reach genome-wide suggestive significance ( $P < 1.0E-05$ ) in any of the 19 GWAS investigated here. The strongest associations with this allele were seen with MRI-based hippocampus volume (left volume  $P = 0.0002$ , summed volume  $P = 0.0005$ ; **Supplementary Table 20**) and with the delayed recall memory test (baseline  $P = 0.0005$ , longitudinal  $P = 0.0042$ ; **Supplementary Table 20**). The effect directions of these associations are consistent with the deleterious influence of rs429358 on AD (Neu et al., 2017). While this at best modest degree of association between rs429358 and the traits investigated here could be due to a number of study-specific aspects, e.g., insufficient power, aspects of sample ascertainment, it is in general agreement with the literature: In the GWAS on cognitive function by Davies et al. (2018), *APOE*  $\epsilon 4$  also showed only marginal association ( $P = 2.2E-04$ ), and it was not reported to show genome-wide significant evidence of association ( $P = 4.1E-07$ ) with hippocampal volume in the GWAS by Hibar et al. (2017). Lastly, while rs429358 did show evidence of genome-wide significant association with several imaging traits in the

more recent GWAS on brain imaging phenotypes in the United Kingdom biobank, none of the hippocampus-based measurements exceeded that study's multiple testing threshold (Smith et al., 2021). We note, that our current findings are different from our earlier GWAS analyses in the EMIF-AD MBD dataset, where the  $\epsilon 4$  allele showed very pronounced evidence of association in CSF and imaging markers related to A $\beta$ 42 (Hong et al., 2020). Taken together, there is now converging evidence from our and previous studies that despite *APOE*  $\epsilon 4$ 's role in contributing to AD risk and A $\beta$ 42-related phenotypes, the same allele does not appear to show a comparably strong influence on variation of other AD-related traits, including the ones studied here.

Extending the comparison to additional genetic variants associated with AD risk in the GWAS by Jansen et al. (2019) did also not show any noteworthy or consistent overlap with the GWAS results generated in this study. In contrast, highly significant overlaps by PGS analysis were observed upon using GWAS results from Davies et al. (2018) for the neuropsychological and Hibar et al. (2017) for the MRI phenotypes, which is not surprising given that very similar neuropsychological and neuroimaging traits were used as outcomes in these studies. Collectively, the PGS results of this and previous work show that there is only very limited overlap in the genetic architecture (at least when studying common SNPs) between AD on the one and neuropsychological performance or structural brain imaging on the other hand. We note that this does not preclude the possibility that certain molecular pathways targeted by the genes highlighted in this GWAS may be shared with AD pathophysiology.

While our study has several noteworthy strengths (e.g., the use of highly standardized procedures in generating and harmonizing both the genotype and phenotype data of our study, use of both cross-sectional and longitudinal neurocognitive performance data, inclusion of the X chromosome in the GWAS), it may also have been negatively affected by some limitations. First and foremost, we note that the sample size used for the present analyses is comparatively small for "GWAS standards" and was well under 1,000 in some instances (**Table 1**). Accordingly, the statistical power of these analyses was low. This limitation is at least partially countered by the quantitative nature of nearly all analyzed phenotypes: it is well established that quantitative trait association analyses are more powerful than those using binary phenotypes, e.g., in a case-control setting (Bush and Moore, 2012). Second, in addition to resulting in low power, small sample sizes also increase the possibility of false-positive findings, especially for infrequent variants (i.e., those with an MAF  $< 5\%$ ). In this context we note that eight of our thirteen genome-wide significant signals were elicited by such variants. Thus, independent replication—ideally in larger datasets—is needed to confirm the main findings of our GWAS before any further-reaching conclusions can be reached. Third, we note that the phenotype data used as outcome traits in our GWAS analyses were collected at different participating centers at times using different types of examinations (e.g., different tests to study the same overarching neuropsychological domain). To

alleviate potential bias resulting from this inherent phenotypic heterogeneity, all clinical data were processed, quality-controlled and harmonized (e.g., by normalizing most variables within centers) centrally by an experienced team of researchers (see Bos et al., 2018 for more details). We emphasize that this potential heterogeneity does not apply to the genetic data as these were generated in one laboratory experiment and subsequently processed jointly in one analytical framework, minimizing the emergence of potential batch effects. Last but not least, we emphasize that owing to its particular ascertainment design (Bos et al., 2018) the EMIF-AD MBD dataset does not (attempt to) constitute a representative sample from the “general population.” Accordingly, the results presented here cannot be generalized to the general population. We note that the same is true for many GWAS in this and other fields, which typically use clinic-based ascertainment which is not representative of the population as a whole.

## CONCLUSION

In conclusion, our study delivers an entirely novel set of GWAS results from participants of the EMIF-AD MBD dataset. We nominate several novel and functionally interesting genetic association signals with phenotypes related to neurocognitive function and structural brain imaging. Even though independent replication is still needed, our results may prove informative to better understand the genetic forces underlying AD and related phenotypes.

## DATA AVAILABILITY STATEMENT

GWAS summary statistics for the top ( $P$ -value  $< 1E-05$ ) results are listed in the Supplementary Tables. Full GWAS summary statistics are available from the authors upon request. Clinical data and genome-wide genotyping data are stored on an online data platform using the “tranSMART” data warehouse framework. Access to the genome-wide genotyping data can be requested from the corresponding author of this study who will forward each request to the EMIF-AD data access team.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committees of all participating centers of this study (for a full list see **Supplementary Table 4** of primary publication Bos et al., 2018). The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

JH and TO performed all the analyses and together with LB interpreted the data and wrote the manuscript. OO performed X chromosomal analyses. VD was responsible for EMIF-AD MBD DNA sample preparation and DNA extraction. LD contributed

to the interpretation of the data and visualization of the results. IB, SV, and PV coordinated the collection and harmonization of phenotypes and biosamples in EMIF-AD MBD. MW and AF supervised the genotyping experiments. RV, SG, JS, PS, CT, SE, GF, OB, JR, RB, AL, DA, JP, CC, GP, PM-L, MT, RD, CL-Q, KS, CV, CL, KB, HZ, MK, and FB contributed to sample and phenotype data collection. JS, PV, and SL were the lead PIs for the EMIF-AD MBD study as a whole and were in charge of designing and managing the platform. LB designed and supervised the genomics portion of the EMIF-AD MBD project and co-wrote all drafts of the manuscript. All authors critically revised all manuscripts drafts, read and approved the final manuscript.

## FUNDING

The present study was conducted as part of the EMIF-AD MBD project, which has received support from the Innovative Medicines Initiative Joint Undertaking under EMIF grant agreement No. 115372, the resources of which were composed of financial contribution from the European Union’s Seventh Framework Program (FP7/2007-2013) and EFPIA companies’ in kind contribution. Parts of this study were made possible through support from the German Research Foundation (DFG grant FOR2488: Main support by subproject “INF-GDAC” BE2287/7-1 to LB) and the Cure Alzheimer’s Fund (to LB). RV acknowledges support by the Stichting Alzheimer Onderzoek (#13007, #11020, #2017-032) and the Flemish Government (VIND IWT 135043). KB was supported by the Swedish Research Council (#2017-00915), the Alzheimer Drug Discovery Foundation (ADDF), United States (#RDAPB-201809-2016615), the Swedish Alzheimer Foundation (#AF-742881), Hjärnfonden, Sweden (#FO2017-0243), the Swedish state under the agreement between the Swedish Government and the County Councils, the ALF-agreement (#ALFGBG-715986), European Union Joint Program for Neurodegenerative Disorders (JPND2019-466-236), and the Alzheimer’s Association 2021 Zenith Award (ZEN-21-848495). HZ was a Wallenberg Scholar supported by grants from the Swedish Research Council (#2018-02532), the European Research Council (#681712), and Swedish State Support for Clinical Research (#ALFGBG-720931). SV received funding from the Innovative Medicines Initiative 2 Joint Undertaking under ROADMAP grant agreement No. 116020 and from ZonMw during the conduct of this study. Research at VIB-UAntwerp was in part supported by the University of Antwerp Research Fund and SAO-FRA 2018 0016. The Lausanne study was funded by a grant from the Swiss National Research Foundation (SNF 320030\_141179) to JP. Research of CT was supported by the European Commission [Marie Curie International Training Network, grant agreement No 860197 (MIRIADE), and JPND], Health Holland, the Dutch Research Council (ZonMW), Alzheimer Drug Discovery Foundation, the Selfridges Group Foundation, Alzheimer Netherlands, Alzheimer Association. CT was recipient of ABOARD, which was a public-private partnership receiving funding from ZonMW (#73305095007) and Health~Holland, Topsector Life Sciences and Health (PPP-allowance; #LSHM20106). More than 30 partners participate in

ABOARD. ABOARD also received funding from Edwin Bouw Fonds and Gieskes-Strijbisfonds. RD was supported by the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, London, United Kingdom; Health Data Research UK, which was funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome Trust; the BigData@Heart Consortium, funded by the Innovative Medicines Initiative-2 Joint Undertaking under grant agreement No. 116074. This Joint Undertaking received support from the European Union's Horizon 2020 Research and Innovation Programme and EFPIA; it is chaired by DE Grobbee and SD Anker, partnering with 20 academic and industry partners and ESC; the National Institute for Health Research University College London Hospitals Biomedical Research Centre; the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London; the UK Research and Innovation London Medical Imaging and Artificial Intelligence Centre for Value-Based Healthcare; and the National Institute for Health

Research (NIHR) Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust. We acknowledge financial support by Land Schleswig-Holstein within the funding programme Open Access Publikationsförderung.

## ACKNOWLEDGMENTS

We acknowledge the assistance of Ellen De Roeck, Naomi De Roeck, and Hanne Struyfs (UAntwerp) with data collection. We thank Tanja Wesse and Sanaz Sedghpour Sabet at the Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany for technical assistance with the GSA genotyping. We thank Fabian Kilpert for his assistance with the QC and genotype imputations and Marcel Schilling for his assistance on visualizing the results. The LIGA team acknowledges computational support from the OMICS compute cluster at the University of Lübeck.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnagi.2022.840651/full#supplementary-material>

## REFERENCES

- Armstrong, N. J., Mather, K. A., Sargurupremraj, M., Knol, M. J., Malik, R., Satizabal, C. L., et al. (2020). Common genetic variation indicates separate causes for periventricular and deep white matter hyperintensities. *Stroke* 51, 2111–2121. doi: 10.1161/STROKEAHA.119.027544
- Bai, B., Wang, X., Li, Y., Chen, P. C., Yu, K., Dey, K. K., et al. (2020). Deep multilayer brain proteomics identifies molecular networks in Alzheimer's disease progression. *Neuron* 105, 975–991.e7. doi: 10.1016/j.neuron.2019.12.015
- Bertram, L., and Tanzi, R. E. (2020). Genomic mechanisms in Alzheimer's disease. *Brain Pathol. (Zurich, Switzerland)* 30, 966–977. doi: 10.1111/bpa.12882
- Bos, I., Vos, S., Vandenbergh, R., Scheltens, P., Engelborghs, S., Frisoni, G., et al. (2018). The EMIF-AD multimodal biomarker discovery study: design, methods and cohort characteristics. *Alzheimers Res. Ther.* 10:64. doi: 10.1186/s13195-018-0396-5
- Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 22, 1790–1797. doi: 10.1101/gr.137323.112
- Buniello, A., MacArthur, J., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2019). The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47, D1005–D1012. doi: 10.1093/nar/gky1120
- Bush, W. S., and Moore, J. H. (2012). Chapter 11: genome-wide association studies. *PLoS Comput. Biol.* 8:e1002822. doi: 10.1371/journal.pcbi.1002822
- Cacace, R., Slegers, K., and Van Broeckhoven, C. (2016). Molecular genetics of early-onset Alzheimer's disease revisited. *Alzheimers Dement.* 12, 733–748. doi: 10.1016/j.jalz.2016.01.012
- Choi, S. W., and O'Reilly, P. F. (2019). PRSice-2: polygenic Risk Score software for biobank-scale data. *GigaScience* 8:giz082. doi: 10.1093/gigascience/giz082
- Damale, M. G., Pathan, S. K., Shinde, D. B., Patil, R. H., Arote, R. B., and Sangshetti, J. N. (2020). Insights of tankyrases: a novel target for drug discovery. *Eur. J. Med. Chem.* 207:112712. doi: 10.1016/j.ejmech.2020.112712
- Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., et al. (2016). Next-generation genotype imputation service and methods. *Nat. Genet.* 48, 1284–1287. doi: 10.1038/ng.3656
- Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., Hill, W. D., et al. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* 9:2098. doi: 10.1038/s41467-018-04362-x
- de Leeuw, C. A., Mooij, J. M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* 11:e1004219. doi: 10.1371/journal.pcbi.1004219
- Gao, Q., Mok, H. P., and Zhuang, J. (2019). Secreted modular calcium-binding proteins in pathophysiological processes and embryonic development. *Chin. Med. J.* 132, 2476–2484. doi: 10.1097/CM9.0000000000000472
- Gatz, M., Reynolds, C. A., Fratiglioni, L., Johansson, B., Mortimer, J. A., Berg, S., et al. (2006). Role of genes and environments for explaining Alzheimer disease. *Arch. Gen. Psychiatry* 63, 168–174. doi: 10.1001/archpsyc.63.2.168
- Gottesman, I. I., and Gould, T. D. (2003). The endophenotype concept in psychiatry: etymology and strategic intentions. *Am. J. Psychiatry* 160, 636–645. doi: 10.1176/appi.ajp.160.4.636
- Graffelman, J., and Weir, B. S. (2016). Testing for Hardy-Weinberg equilibrium at biallelic genetic markers on the X chromosome. *Heredity* 116, 558–568. doi: 10.1038/hdy.2016.20
- Grasby, K. L., Jahanshad, N., Painter, J. N., Colodro-Conde, L., Bralten, J., Hibar, D. P., et al. (2020). The genetic architecture of the human cerebral cortex. *Science (New York, N.Y.)* 367:eaay6690. doi: 10.1126/science.aay6690
- Hibar, D. P., Adams, H., Jahanshad, N., Chauhan, G., Stein, J. L., Hofer, M. A., et al. (2017). Novel genetic loci associated with hippocampal volume. *Nat. Commun.* 8:13624. doi: 10.1038/ncomms13624
- Holtzman, D. M., Herz, J., and Bu, G. (2012). Apolipoprotein E and apolipoprotein E receptors: normal biology and roles in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* 2:a006312. doi: 10.1101/cshperspect.a006312
- Hong, S., Dobricic, V., Ohlei, O., Bos, I., Vos, S., Prokopenko, D., et al. (2021). TMEM106B and CPOX are genetic determinants of cerebrospinal fluid Alzheimer's disease biomarker levels. *Alzheimers Dement.* 17, 1628–1640. doi: 10.1002/alz.12330

- Hong, S., Prokopenko, D., Dobricic, V., Kilpert, F., Bos, I., Vos, S., et al. (2020). Genome-wide association study of Alzheimer's disease CSF biomarkers in the EMIF-AD multimodal biomarker discovery dataset. *Transl. Psychiatry* 10:403. doi: 10.1038/s41398-020-01074-z
- Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* 51, 404–413. doi: 10.1038/s41588-018-0311-9
- Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., et al. (2013). The genotype-tissue expression (GTEx) project. *Nat. Genet.* 45, 580–585. doi: 10.1038/ng.2653
- Machiela, M. J., and Chanock, S. J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics (Oxf. Engl.)* 31, 3555–3557. doi: 10.1093/bioinformatics/btv402
- MacRae, C. A., and Vasan, R. S. (2011). Next-generation genome-wide association studies: time to focus on phenotype? *Circ. Cardiovasc. Genet.* 4, 334–336. doi: 10.1161/CIRCGENETICS.111.960765
- Mattsson, N., Carrillo, M. C., Dean, R. A., Devous, M. D. Sr., Nikolcheva, T., Pesini, P., et al. (2015). Revolutionizing Alzheimer's disease and clinical trials through biomarkers. *Alzheimers Dement. (Amst.)* 1, 412–419. doi: 10.1016/j.dadm.2015.09.001
- Mazaheri Moghaddam, M., Mazaheri Moghaddam, M., Hamzei, H., Baghbanzadeh, A., Pashazadeh, F., and Sakhinia, E. (2021). Genetic basis of acephalic spermatozoa syndrome, and intracytoplasmic sperm injection outcomes in infertile men: a systematic scoping review. *J. Assist. Reprod. Genet.* 38, 573–586. doi: 10.1007/s10815-020-02008-w
- McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* 48, 1279–1283. doi: 10.1038/ng.3643
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., and Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology* 34, 939–944. doi: 10.1212/wnl.34.7.939
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., et al. (2016). The ensembl variant effect predictor. *Genome Biol.* 17:122. doi: 10.1186/s13059-016-0974-4
- Montani, C., Gritti, L., Beretta, S., Verpelli, C., and Sala, C. (2019). The synaptic and neuronal functions of the X-Linked intellectual disability protein interleukin-1 receptor accessory protein like 1 (ILIRAPL1). *Dev. Neurobiol.* 79, 85–95. doi: 10.1002/dneu.22657
- Morkmued, S., Clauss, F., Schuhbaur, B., Fraulob, V., Mathieu, E., Hemmerlé, J., et al. (2020). Deficiency of the SMOC2 matricellular protein impairs bone healing and produces age-dependent bone loss. *Sci. Rep.* 10:14817. doi: 10.1038/s41598-020-71749-6
- Neu, S. C., Pa, J., Kukull, W., Beekly, D., Kuzma, A., Gangadharan, P., et al. (2017). Apolipoprotein E genotype and sex risk factors for Alzheimer disease: a meta-analysis. *JAMA Neurol.* 74, 1178–1189. doi: 10.1001/jamaneurol.2017.2188
- Omodei, D., Acampora, D., Russo, F., De Filippi, R., Severino, V., Di Francia, R., et al. (2009). Expression of the brain transcription factor OTX1 occurs in a subset of normal germinal-center B cells and in aggressive Non-Hodgkin Lymphoma. *Am. J. Pathol.* 175, 2609–2617. doi: 10.2353/ajpath.2009.090542
- Patterson, C. (2018). *World Alzheimer Report 2018. The State of the Art of Dementia Research: New Frontiers. An Analysis of Prevalence, Incidence, Cost and Trends.* London: Alzheimer's Disease International.
- Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *J. Intern. Med.* 256, 183–194. doi: 10.1111/j.1365-2796.2004.01388.x
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Rai, A., Bleimling, N., Vetter, I. R., and Goody, R. S. (2020). The mechanism of activation of the actin binding protein EHB1 by Rab8 family members. *Nat. Commun.* 11:4187. doi: 10.1038/s41467-020-17792-3
- Raymond, F. L. (2006). X linked mental retardation: a clinical guide. *J. Med. Genet.* 43, 193–200. doi: 10.1136/jmg.2005.033043
- Rentzsch, P., Schubach, M., Shendure, J., and Kircher, M. (2021). CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med.* 13:31. doi: 10.1186/s13073-021-00835-9
- Sargurupremraj, M., Suzuki, H., Jian, X., Sarnowski, C., Evans, T. E., Bis, J. C., et al. (2020). Cerebral small vessel disease genomics and its implications across the lifespan. *Nat. Commun.* 11:6285. doi: 10.1038/s41467-020-19111-2
- Sathe, G., Albert, M., Darrow, J., Saito, A., Troncoso, J., Pandey, A., et al. (2021). Quantitative proteomic analysis of the frontal cortex in Alzheimer's disease. *J. Neurochem.* 156, 988–1002. doi: 10.1111/jnc.15116
- Smith, S. M., Douaud, G., Chen, W., Hanayik, T., Alfaro-Almagro, F., Sharp, K., et al. (2021). An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat. Neurosci.* 24, 737–745. doi: 10.1038/s41593-021-00826-4
- Sperling, R., Mormino, E., and Johnson, K. (2014). The evolution of preclinical Alzheimer's disease: implications for prevention trials. *Neuron* 84, 608–622. doi: 10.1016/j.neuron.2014.10.038
- Ten Kate, M., Redolfi, A., Peira, E., Bos, I., Vos, S. J., Vandenberghe, R., et al. (2018). MRI predictors of amyloid pathology: results from the EMIF-AD multimodal biomarker discovery study. *Alzheimers Res. Ther.* 10:100. doi: 10.1186/s13195-018-0428-1
- van der Meer, D., Frei, O., Kaufmann, T., Shadrin, A. A., Devor, A., Smeland, O. B., et al. (2020). Understanding the genetic determinants of the brain with MOSTest. *Nat. Commun.* 11:3512. doi: 10.1038/s41467-020-17368-1
- Wang, H., Dey, K. K., Chen, P. C., Li, Y., Niu, M., Cho, J. H., et al. (2020). Integrated analysis of ultra-deep proteomes in cortex, cerebrospinal fluid and serum reveals a mitochondrial signature in Alzheimer's disease. *Mol. Neurodegen.* 15:43. doi: 10.1186/s13024-020-00384-6
- Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* 8:1826. doi: 10.1038/s41467-017-01261-5
- Whelan, C. D., Mattsson, N., Nagle, M. W., Vijayaraghavan, S., Hyde, C., Janelidze, S., et al. (2019). Multiplex proteomics identifies novel CSF and plasma biomarkers of early Alzheimer's disease. *Acta Neuropathol. Commun.* 7:169. doi: 10.1186/s40478-019-0795-2
- Wightman, D. P., Jansen, I. E., Savage, J. E., Shadrin, A. A., Bahrami, S., Holland, D., et al. (2021). A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat. Genet.* 53, 1276–1282. doi: 10.1038/s41588-021-00921-z
- Willer, C. J., Li, Y., and Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxf. Engl.)* 26, 2190–2191. doi: 10.1093/bioinformatics/btq340
- Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L. O., et al. (2004). Mild cognitive impairment—beyond controversies, towards a consensus: report of the international working group on mild cognitive impairment. *J. Intern. Med.* 256, 240–246. doi: 10.1111/j.1365-2796.2004.01380.x
- Zhang, Q., Cai, Z., Lhomme, M., Sahana, G., Lesnik, P., Guerin, M., et al. (2020). Inclusion of endophenotypes in a standard GWAS facilitate a detailed mechanistic understanding of genetic elements that control blood lipid levels. *Sci. Rep.* 10:18434. doi: 10.1038/s41598-020-75612-6
- Zhao, B., Zhang, J., Ibrahim, J. G., Luo, T., Santelli, R. C., Li, Y., et al. (2021). Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n=17,706). *Mol. Psychiatry* 26, 3943–3955. doi: 10.1038/s41380-019-0569-z

**Conflict of Interest:** HZ has served at scientific advisory boards and/or as a consultant for Abbvie, Alector, Annexon, Artery Therapeutics, 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, Biogen, and Roche, and was a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which was a part of the GU Ventures Incubator Program. FB is supported by the NIHR biomedical research centre at UCLH. JP received consultation honoraria from Nestle Institute of Health Sciences, Ono Pharma, OM Pharma, and Fujirebio, unrelated to the submitted work. CT has a collaboration contract with ADx Neurosciences, Quanterix and Eli Lilly, performed contract research or received grants from AC-Immune, Axon Neurosciences, Biogen, Brainstorm Therapeutics,

Celgene, EIP Pharma, Eisai, PeopleBio, Roche, Toyama, Vivoryon. She serves on editorial boards of *Medicinal Neurology/Springer*, *Alzheimer Research and Therapy*, *Neurology: Neuroimmunology and Neuroinflammation*, and was editor of a *Neuroinformatics* book Springer. CT also holds a speaker's contract with Roche, Inc. KB has served as a consultant, at advisory boards, or at data monitoring committees for Abcam, Axon, BioArctic, Biogen, JOMDD/Shimadzu, Julius Clinical, Lilly, MagQu, Novartis, Pharmatrophix, Prothena, Roche Diagnostics, and Siemens Healthineers, and was a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which was a part of the GU Ventures Incubator Program, outside the work presented in this paper. SL is currently an employee at Janssen Medical UK. JS was an employee of Janssen R&D, LLC., and is currently an employee and chief medical officer of AC Immune SA. JR was an employee of Neurosciences Therapeutic Area, GlaxoSmithKline R&D, Stevenage, UK.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2022 Homann, Osburg, Ohlei, Dobricic, Deecke, Bos, Vandenberghe, Gabel, Scheltens, Teunissen, Engelborghs, Frisoni, Blin, Richardson, Bordet, Lleó, Alcolea, Popp, Clark, Peyratout, Martinez-Lage, Tainta, Dobson, Legido-Quigley, Slegers, Van Broeckhoven, Wittig, Franke, Lill, Blennow, Zetterberg, Lovestone, Streffer, ten Kate, Vos, Barkhof, Visser and Bertram. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*