

# Genome-wide association study of Alzheimer's disease brain imaging biomarkers and neuropsychological phenotypes in the EMIF-AD Multimodal Biomarker Discovery dataset

1 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>,  
2 Rik Vandenberghe<sup>4,5</sup>, Silvy Gabel<sup>4</sup>, Philip Scheltens<sup>3</sup>, Charlotte E. Teunissen<sup>6</sup>, Sebastiaan  
3 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  
4 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  
5 Martinez-Lage<sup>17</sup>, Mikel Tainta<sup>17</sup>, Richard J. B. Dobson<sup>18</sup>, Cristina Legido-Quigley<sup>19,20</sup>, Kristel  
6 Slegers<sup>21,22</sup>, Christine Van Broeckhoven<sup>21,22</sup>, Michael Wittig<sup>23</sup>, Andre Franke<sup>23</sup>, Christina M.  
7 Lill<sup>1</sup>, Kaj Blennow<sup>24,25</sup>, Henrik Zetterberg<sup>24,25,26,27</sup>, Simon Lovestone<sup>28</sup>, Johannes Streffer<sup>29,30</sup>,  
8 Mara ten Kate<sup>3,31</sup>, Stephanie J. B. Vos<sup>2</sup>, Frederik Barkhof<sup>31,32</sup>, Pieter Jelle Visser<sup>2,3,33</sup>, and Lars  
9 Bertram<sup>1,34\*</sup>

10 § These authors contributed equally

11 <sup>1</sup>Lübeck Interdisciplinary Platform for Genome Analytics (LIGA), University of Lübeck, Lübeck,  
12 Germany

13 <sup>2</sup>Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience,  
14 Alzheimer Centrum Limburg, Maastricht University, Maastricht, the Netherlands

15 <sup>3</sup>Alzheimer Center Amsterdam, Department of Neurology, Amsterdam University Medical Centers,  
16 Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

17 <sup>4</sup>Laboratory for Cognitive Neurology, Department of Neurosciences, KU Leuven, Leuven, Belgium

18 <sup>5</sup>Neurology Service, University Hospital Leuven, Leuven, Belgium

19 <sup>6</sup>Neurochemistry Laboratory, Department of Clinical Chemistry, Amsterdam Neuroscience,  
20 Amsterdam University Medical Centers, Vrije Universiteit, Amsterdam, the Netherlands

21 <sup>7</sup>Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium

22 <sup>8</sup>Department of Neurology and Center for Neurosciences, Universitair Ziekenhuis Brussel and Vrije  
23 Universiteit Brussel (VUB), Brussels, Belgium

24 <sup>9</sup>University of Geneva, Geneva, Switzerland

25 <sup>10</sup>IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy

26 <sup>11</sup>AIX Marseille University, INS, Ap-hm, Marseille, France

27 <sup>12</sup>Neurosciences Therapeutic Area, GlaxoSmithKline R&D, Stevenage, UK

28 <sup>13</sup>University of Lille, Inserm, CHU Lille, Lille, France

29 <sup>14</sup>Memory Unit, Neurology Department. Hospital de Sant Pau, Barcelona and Centro de  
30 Investigación Biomédica en Red en enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain

31 <sup>15</sup>Department of Geriatric Psychiatry, University Hospital of Psychiatry Zurich, Zürich, Switzerland

32 <sup>16</sup>Old Age Psychiatry, Department of Psychiatry, University Hospital of Lausanne, Lausanne,  
33 Switzerland

34 <sup>17</sup>Department of Neurology, Center for Research and Advanced Therapies, CITA-Alzheimer  
35 Foundation, San Sebastian, Spain

36 <sup>18</sup>Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and  
37 Neuroscience, King's College London, London, UK

38 <sup>19</sup>Steno Diabetes Center, Copenhagen, Denmark

39 <sup>20</sup>Institute of Pharmaceutical Sciences, King's College London, London, UK

40 <sup>21</sup>Complex Genetics of Alzheimer's Disease Group, Center for Molecular Neurology, VIB, Antwerp,  
41 Belgium

42 <sup>22</sup>Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium

43 <sup>23</sup>Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany

44 <sup>24</sup>Institute of Neuroscience and Physiology, Department of Psychiatry and Neurochemistry, The  
45 Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

46 <sup>25</sup>Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden

47 <sup>26</sup>Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, Queen  
48 Square, London, UK

49 <sup>27</sup>UK Dementia Research Institute at UCL, London, UK

50 <sup>28</sup>Department of Psychiatry, University of Oxford, Oxford, UK (currently employed at Janssen-Cilag)

51 <sup>29</sup>Reference Center for Biological Markers of Dementia (BIODEM), Institute Born-Bunge,  
52 University of Antwerp, Antwerp, Belgium

53 <sup>30</sup>AC Immune SA, EPFL Innovation Park, Lausanne, Switzerland

54 <sup>31</sup>Department of Radiology and Nuclear Medicine, Amsterdam University Medical Centers,  
55 Amsterdam Neuroscience, Amsterdam, the Netherlands

56 <sup>32</sup>Institutes of Neurology and Healthcare Engineering, University College London, London, UK

57 <sup>33</sup>Department of Neurobiology, Care Sciences and Society, Division of Neurogeriatrics, Karolinska  
58 Institutet, Stockholm, Sweden

59 <sup>34</sup>Department of Psychology, University of Oslo, Oslo, Norway

60 \* **Correspondence:**

61 Prof. Lars Bertram, Lübeck Interdisciplinary Platform for Genome Analytics (LIGA), University of  
62 Lübeck, Ratzeburger Allee 160, V50-319, 23562 Lübeck, Germany, phone: +49 451 3101 7491, e-  
63 mail: lars.bertram@uni-luebeck.de

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

66 **Abstract**

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

90

91 **1 Introduction**

92 Alzheimer's disease is the most common neurodegenerative disease in humans and the most common  
 93 form of dementia. In 2018, estimates were published that 50 million dementia patients exist  
 94 worldwide, about two-third of whom were diagnosed with AD (Patterson, 2018). Pathologically, AD  
 95 is characterized by the accumulation of extracellular amyloid  $\beta$  ( $A\beta$ ) peptide deposits ("plaques") and  
 96 intracellular hyperphosphorylated tau protein aggregates ("tangles") in the brain, leading to synaptic  
 97 dysfunction, neuroinflammation, neuronal loss, and, ultimately, onset of cognitive decline (Sperling  
 98 et al, 2014; Mattsson et al., 2015). Genetically, AD is a heterogeneous disorder with both monogenic  
 99 and polygenic forms. The former is caused by highly penetrant but rare mutations in three genes  
 100 encoding the amyloid beta precursor protein (*APP*) and presenilins 1 and 2 (*PSEN1/PSEN2*), which  
 101 only make up a small fraction ( $\ll 5\%$ ) of all AD cases (Cacace et al., 2016). Most patients, however,  
 102 suffer from "polygenic AD", which is determined by the action (and interaction) of numerous  
 103 independent genomic variants, likely in concert with nongenetic factors, such as environmental  
 104 exposures (e.g., head trauma) and lifestyle choices (e.g., alcohol consumption and cigarette smoking)  
 105 (Bertram and Tanzi, 2020). Based on results from the currently most recent and largest genome-wide  
 106 association study (GWAS) performed in AD, there are now 38 independent loci showing genome-  
 107 wide significant association with disease risk (Wightman et al., 2021). The most strongly and most  
 108 consistently associated AD risk gene is *APOE*, which encodes apolipoprotein E, a cholesterol  
 109 transport protein that has been implicated in numerous amyloid-specific pathways, including amyloid  
 110 trafficking, as well as plaque clearance (Holtzman et al., 2012). Although the heritability of

111 polygenic AD is estimated to be around 60-80% (Gatz et al., 2016), *APOE* and the other currently  
 112 known 37 independent risk loci explain only part of the disease’s phenotypic variance (Wightman et  
 113 al., 2021). While most AD GWAS only consider clinically diagnosed “probable AD” cases and  
 114 cognitively unimpaired controls, involving a risk for mis-diagnosis of patients and inclusion of  
 115 preclinical AD cases as controls, additional information about the genetic architecture of AD and  
 116 additional statistical power is also afforded by using “endophenotypes” related to AD, ideally  
 117 measured on a quantitative scale such as biomarker data, imaging, or neurocognitive performance  
 118 (Gottesman and Gould, 2003; MacRae and Vasani, 2011; Zhang et al., 2020).

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

138

## 139 **2 Materials and Methods**

### 140 **2.1 Sample description**

141 Analyses were based on the EMIF-AD MBD dataset which was collected across eleven different  
 142 European study centers (Bos et al., 2018). In total, this dataset included 1221 (563 [46%] female;  
 143 mean age = 67.9 years, SD=8.3) individuals from three diagnostic stages: normal controls (NC),  
 144 subjects with mild cognitive impairment (MCI) and subjects with a clinical diagnosis of AD. An  
 145 overview of the quantitative phenotypes investigated in this study is provided in Table 1. Due to  
 146 partially missing phenotype data (in the neurocognitive domain), the effective sample sizes vary for  
 147 the different GWAS analyses (see Table 1). The local medical ethical review boards in each  
 148 participating recruitment center had approved the study prior to commencement. Furthermore, all  
 149 subjects had provided written informed consent at the time of inclusion in the cohort for use of data,  
 150 samples and scans (Bos et al., 2018).

### 151 **2.2 MRI phenotypes description**

152 The five MRI phenotypes were collected for 862 subjects. Brain MRIs were used to assess  
 153 hippocampal volume (mm<sup>3</sup>, left and right hemisphere, and sum of both; all adjusted for intracranial

154 volume), whole brain cortical thickness (in mm), and white matter lesions (WML; using the Fazekas  
 155 scale) (Ten Kate et al., 2018). The Fazekas scale categorizes WMLs into 4 categories: Level 0 (no or  
 156 almost no lesion), level 1 (multiple punctate lesions), level 2 (early confluent WML), and level 3  
 157 (presence of large confluent WML). Details on the scanning procedures and data harmonization  
 158 across centers can be found in Bos et al. (2018) and Ten Kate et al. (2018).

### 159 **2.3 Neuropsychological phenotypes description**

160 Cross-sectional (and follow-up) data were available for the following seven neuropsychological  
 161 domains within the EMIF-AD MBD dataset: global cognition (Mini Mental State Examination,  
 162 MMSE), attention, executive function, language, memory (immediate and delayed) and  
 163 visuoconstruction. For each cognitive domain, a primary test was selected by Bos et al. (2018). If the  
 164 preferred test were not available, an alternative priority test from the same cognitive domain was  
 165 chosen. More details on the neuropsychological tests used for generating these phenotypes can be  
 166 found in Bos et al. (2018). Raw data on these tests were normalized with the help of a z-  
 167 transformation, so that the data were comparable within a cognitive domain despite representing  
 168 partially different tests across centers. For the cross-sectional GWAS analyses, the z-scores derived  
 169 from baseline data were used. The number of subjects used for each test can be found in the  
 170 Supplementary Material. For all seven neuropsychological domains, follow-up data from at least one  
 171 additional time point were available for each individual and used to construct a longitudinal  
 172 phenotype using the following formula (which estimates the relative change in cognitive performance  
 173 per time interval [here: years]):

$$174 \quad \frac{\text{Score}_{\text{last}} - \text{Score}_{\text{first}}}{\frac{\text{Score}_{\text{last}} + \text{Score}_{\text{first}}}{2} * \text{interval}}$$

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

### 180 **2.4 DNA extraction, genotype imputation and quality control**

181 A detailed description of the genotyping procedures, quality control (QC) and subsequent data  
 182 processing can be found in Hong et al. (2020) and in the Supplementary Material. Here, the same  
 183 genotype data were used for the GWAS analyses. Briefly, 936 DNA samples were subjected to  
 184 genome-wide SNP genotyping using the Infinium Global Screening Array (GSA) with Shared  
 185 Custom Content (Illumina Inc.). Imputation was then performed using Minimac3 (Das et al., 2016).  
 186 Extensive post-imputation QC resulted in 7,464,105 autosomal SNPs with a minor allele frequency  
 187 (MAF)  $\geq 0.01$  in 888 individuals of European ancestry. More details can be found in the  
 188 Supplementary Material.

189 For the X chromosome, QC was performed separately for male and female subjects for non-  
 190 pseudoautosomal regions, using slightly different criteria compared to the autosomes (see  
 191 Supplementary Material). In contrast, pseudoautosomal regions (PAR1 and PAR2) were treated  
 192 analogously to the autosomal SNPs. After QC, imputations were performed on the Sanger Institute  
 193 imputation server (<https://imputation.sanger.ac.uk/>) using the extended HRC reference panel  
 194 (McCarthy et al., 2016). After imputation, we used the same QC criteria as for the autosomal SNPs

195 but performed these separately for female and male data sets, except the HWE test ( $P < 1.0E-4$ ) which  
 196 was performed on all samples combined as recommended previously (Graffelman and Weir, 2016)  
 197 and implemented in PLINK2. For males, markers were coded as 0 vs. 2 (instead of 0 vs. 1), to adjust  
 198 for the missing second X chromosome (as recommended in Smith et al. [2021]).

## 199 2.5 GWAS and post-GWAS analyses

200 SNP-based association analyses were performed assuming an additive linear model (command: --  
 201 glm) using allele dosages (to account for imputation uncertainty) in PLINK2 (Purcell et al., 2007).  
 202 The covariates included in the analyses were sex, diagnostic status and the first three principal  
 203 components from a principal component analysis (PCA) to adjust for population-specific differences.  
 204 Generally, we excluded SNPs from the GWAS analyses with  $MAF < 0.01$ . However, due to  
 205 differences in the effective sample sizes across phenotypes this threshold was adapted upward (up to  
 206 0.04) to prevent inflation of test statistics owing to low frequency SNPs (see Table 1 for more  
 207 details). Diagnostic status was coded with two dummy variables as follows: NC = (0,0), MCI = (0,1),  
 208 AD = (1,1). For four longitudinal cognitive phenotypes an additional dummy variable was introduced  
 209 to code for the neuropsychological test used, in cases where two different tests were used for  
 210 generating these phenotypes. Details can be found in the Supplementary Material.

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

216 The FUMA platform (<http://fuma.ctglab.nl/>; Watanabe et al., 2017) was used for post-GWAS  
 217 analyses, including gene-based association analyses (via MAGMA (de Leeuw et al., 2015)) and to  
 218 annotate and visualize the GWAS results. To this end, we defined genome-wide significance at  
 219  $\alpha < 5.0E-08$  for the SNP-based analyses while genome-wide suggestive evidence was set at  $\alpha < 1.0E-$   
 220 05. For the gene-based analyses, we adjusted for the number of protein-coding genes examined  
 221 (19,485) using the Bonferroni method, resulting in a threshold of  $\alpha < 2.566E-06$ .

222 In FUMA, both the SNP annotation and the Combined Annotation Dependent Depletion (CADD)  
 223 score (Rentzsch et al., 2021) are provided. The main GWAS results are reported only for  
 224 "independent significant" SNPs, as defined by FUMA. These represent SNPs that are not highly  
 225 correlated with one another using a threshold of  $r^2 < 0.6$  (using reference data from the 1000 Genomes  
 226 Project).

227 Subsequently, the top SNPs, i.e., those with the smallest  $P$  values per respective phenotype, were  
 228 examined in more detail using additional tools. First, the Variant Effect Predictor on Ensembl (VEP,  
 229 <http://grch37.ensembl.org/Tools/VEP>; McLaren et al., 2016) was used to determine a possibly  
 230 functional effect due to changes in the coding sequence, e.g. missense variants. Second, SNPs were  
 231 examined using data from the RegulomeDB database (<https://regulomedb.org/regulome-search>;  
 232 Boyle et al., 2012) to assess possible effects on gene expression. Third, we used data from the  
 233 Genotype-Tissue Expression (GTEx, V8) project portal (<https://www.gtexportal.org/home/>; Lonsdale  
 234 et al., 2013) to assess whether SNPs represent expression / splicing quantitative trait loci  
 235 (eQTLs/sQTLs). While GTEx provides data on gene expression in 54 tissues, we laid particular  
 236 emphasis in genes expressed in brain. Lastly, we interrogated the GWAS catalogue  
 237 (<https://www.ebi.ac.uk/gwas/home>; Buniello et al., 2019) to assess whether any of the top SNPs were

238 previously reported to show association with other phenotypes by GWAS. To this end, we considered  
 239 genes and loci within a 1 Mb region ( $\pm 500,000$  bp) around the SNP of interest. In case SNPs not  
 240 identical to our “top SNP” were reported to show association with an AD-relevant phenotype (brain  
 241 imaging, cognition, etc.), the LDlink platform (Machiela and Chanock, 2015) was used to determine  
 242 pairwise LD to top SNPs (<https://ldlink.nci.nih.gov/?tab=ldpair>). In this context we defined relevant  
 243 LD using a threshold of  $r^2 > 0.6$ .

## 244 2.6 Polygenic score (PGS) analysis

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

260

## 261 3 Results

### 262 3.1 GWAS on MRI phenotypes

263 The genomic inflation factor  $\lambda$  ranged between 1.004 and 1.012 in all five SNP-based analyses,  
 264 indicating that the results of the MRI GWAS analyses were not affected by substantial inflation of  
 265 the test statistics. In the actual association analyses of the five quantitative MRI phenotypes, we  
 266 identified no genome-wide significant ( $P < 5.0E-08$ ) signals but observed 385 variants with at least  
 267 suggestively significant ( $P < 1.0E-05$ ) evidence of association (Supplementary Tables 15-19). The  
 268 lowest  $P$  value was observed with SNP rs16829761 for the Fazekas phenotype ( $P = 5.08E-08$ ;  
 269 Supplementary Figure 31), which only fell slightly above the genome-wide significance threshold.  
 270 According to VEP (McLaren et al., 2016), this variant is located in an intron of the genes *IQCJ*  
 271 (protein: IQ motif containing J) and *SCHIP1* (protein: schwannomin-interacting protein 1). In the  
 272 GTEx database (Lonsdale et al., 2013), the lead-SNP identified here (rs16829761) is not listed as  
 273 eQTL or sQTL, which may be due to the comparatively low MAF (0.01). The CADD score, i.e., the  
 274 *in silico* predicted deleteriousness, of rs16829761 is also low at approximately 0.074. In addition,  
 275 none of the gene-based GWAS analyses using MAGMA revealed any genome-wide significant  
 276 signals ( $P < 2.566E-06$ ) using the MRI traits analyzed. The genomic inflation factor  $\lambda$  ranged between  
 277 0.984 and 1.060 in these five gene-based analyses.

### 278 3.2 GWAS on neuropsychological phenotypes

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

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

297 *Analyses of longitudinal data.* The strongest signal in the longitudinal analyses was elicited by a  
 298 locus on chromosome 6q27 (rs73045836;  $P=7.50E-11$ ; Table 2; Figure 1B; Supplementary Figure  
 299 25) in the analysis using an immediate memory recall paradigm. This SNP is located in an intron of  
 300 *SMOC2* coding for secreted modular calcium-binding protein 2, which, among other functions,  
 301 promotes extracellular matrix assembly (Gao et al., 2019). It needs to be noted that with an MAF  
 302 ~2% this SNP is rather infrequent which may increase the possibility of representing a false-positive  
 303 finding. Perhaps more interesting is the association signal observed near SNP rs5943462 (MAF  
 304 ~0.05) and the visuoconstruction phenotype on the X chromosome ( $P=1.06E-09$ ; Table 2; Figure 1B;  
 305 Supplementary Figure 29). This SNP is an intronic variant located in *IL1RAPL1* encoding interleukin  
 306 1 receptor accessory protein-like 1, which belongs to a class of molecules that regulate synapse  
 307 formation (Montani et al., 2019). The third highlighted signal in this domain relates to the genome-  
 308 wide significant variant rs74381761 (MAF ~0.05) on chromosome 8p23.1 ( $P=1.89E-08$ ; Table 2;  
 309 Figure 1B; Supplementary Figure 5) which shows association with the longitudinal MMSE  
 310 phenotype. The lead SNP is located in an intergenic region near *TNKS* (gene-based  $P=4.87E-04$ ;  
 311 Table 2). This gene encodes the protein tankyrase, which belongs to a class of poly (ADP-ribose)  
 312 polymerases and is involved in various processes in the body, such as telomere length regulation, the  
 313 Wnt/ $\beta$ -catenin signaling pathway, or glucose transport (Damale et al., 2020). The last featured signal  
 314 relates to the association observed near SNP rs9652864 (MAF ~0.22) on chromosome 17q24.1  
 315 ( $P=3.20E-08$ ; Table 2; Figure 1B; Supplementary Figure 21) and the delayed recall test. This variant  
 316 is located in an intron of *CEP112*, which encodes centrosomal protein 112. Overall, there were eight  
 317 correlated SNPs in this locus all showing strongly association (Supplementary Table 10).

318 *Comparison of cross-sectional vs. longitudinal GWAS results.* After completion of the separate  
 319 GWAS on cross-sectional and longitudinal outcomes, we assessed whether the results of these two  
 320 analysis arms showed any overlap. To this end, we followed two approaches: First, we performed a  
 321 look-up of top results from one paradigm in the equivalent other. Specifically, we checked whether a  
 322 genome-wide significant SNP from the cross-sectional analyses also had a low  $P$  value in the  
 323 corresponding longitudinal GWAS and vice versa. The lowest corresponding  $P$  value was 0.015 (at  
 324 baseline) for rs73045828, which attained  $P=5.65E-09$  in the longitudinal GWAS for immediate



325 memory (Supplementary Table 21). No further signal overlaps were observed across corresponding  
 326 cross-sectional and longitudinal phenotypes. Second, we took a more comprehensive approach by  
 327 comparing a larger set of SNPs across both phenotypic domains. To this end, we constructed PGS  
 328 from the summary statistics of the cross-sectional GWAS (as an approximate measure of “aggregated  
 329 SNP effects”) and used these PGS as independent variables in a linear model predicting longitudinal  
 330 outcomes. Effectively, this allowed us to determine how much phenotypic variance in the  
 331 longitudinal data can be explained by top SNPs of the matching cross-sectional GWAS. Overall,  
 332 these analyses did not reveal a substantial correlation in genetic results for corresponding phenotypes  
 333 (Supplementary Table 22), in agreement with the look up of individual SNPs (see above). The best  
 334 model fit was observed with the PGS for executive function and visuoconstruction, where the GWAS  
 335 top SNPs from the cross-sectional data used in the PGS explained 4-9% of the phenotypic variance of  
 336 the corresponding longitudinal outcomes, respectively (Supplementary Table 22). We note, however,  
 337 that the PGS method was not designed for computing genetic correlations of non-independent  
 338 samples (as is the case here), so this analysis must be considered “exploratory”, and the reported  
 339 results represent no more than “upper bounds” of the potential genetic correlations.

### 340 **3.3 Role *APOE* in GWAS on MRI and neuropsychological performance**

341 Given the substantial role that variants in *APOE* play in the genetic architecture of AD, we present  
 342 findings for this locus separately, i.e. the results for SNP rs429358 (which defines the  $\epsilon 4$  allele) and  
 343 rs7412 (which defines the  $\epsilon 2$  allele). In relation to the common genotype  $\epsilon 3/\epsilon 3$ , the risk to develop  
 344 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  
 345 ORs around 10-12 when compared to normal controls (Neu et al., 2017). The minor allele at  
 346 rs429358 ( $\epsilon 4$ ) is overrepresented in the EMIF-AD MBD dataset with an MAF  $\sim 29\%$  (the MAF in the  
 347 general Northern European population is  $\sim 16\%$ ), which is due to the special design of participant  
 348 recruitment (see Bos et al. (2018)). For the neuropsychological phenotypes, the  $P$  values of rs429358  
 349 are unremarkable except for the domain “delayed memory”, where  $P$  values of 0.0005 and 0.0042  
 350 were observed for the baseline and longitudinal analyses, respectively (Supplementary Table 20). In  
 351 the MRI analyses, the only association signal observed with rs429358 was with hippocampal volume  
 352 (Supplementary Table 20). Interestingly, this was driven by an association with the volume of the left  
 353 ( $P=0.0002$ ) hippocampus, while no association was observed with the corresponding data of the right  
 354 hemisphere ( $P=0.2956$ ). We note that for both traits, i.e. delayed memory and left hippocampal  
 355 volume, the effect direction the corresponding  $\beta$  coefficient is consistent with the deleterious effect of  
 356 the minor ( $T/\epsilon 4$ ) allele at rs429358 known from the literature (Neu et al., 2017). For the minor allele  
 357 at rs7412 ( $\epsilon 2$ ) we observed no noteworthy association signals in any of the analyses performed in this  
 358 study (Supplementary Table 20), although power for this variant was much reduced owing to its  
 359 lower MAF (4.6% here, 7.5% in the general western European control population).

### 360 **3.4 Polygenic score (PGS) analyses using published GWAS results**

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

364 *PGS analyses with MRI phenotypes.* As expected, the strongest overlap was observed with a prior  
 365 GWAS also using MRI outcomes. Specifically, we used GWAS results by the ENIGMA group  
 366 (Hibar et al., 2017) who studied 26 imaging traits in  $n=33,536$  individuals. Here, the best overlap was  
 367 seen with each of the three hippocampal MRI traits (up to 2.7% variance explained,  $P=6.0E-06$ ;  
 368 Table 3; Supplementary Table 24). In contrast, in PGS analyses using SNPs associated with AD risk

369 (Jansen et al., 2019), we found only one moderate correlation with white matter damage (measured  
 370 by the Fazekas score). For this trait the AD SNPs explained 1.4% variance ( $P=3.7E-03$ ; Table 3;  
 371 Supplementary Table 24).

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

#### 387 388 **4 Discussion**

389 This study extends previous GWAS analyses from our group utilizing phenotypic data from the  
 390 EMIF-AD MBD study (Hong et al., 2020; Hong et al., 2021). The overarching goal of this work was  
 391 to decipher the genetic architecture of AD-related MRI and neuropsychological (endo)phenotypes to  
 392 better understand AD pathophysiology. Both previous EMIF-AD MBD GWAS focused on AD  
 393 biomarkers measured in CSF and, among other findings, identified variants in *TMEM106B* as *trans-*  
 394 *p*QTLs of CSF neurofilament light (NfL) levels (Hong et al., 2021). Interestingly, the same locus was  
 395 subsequently highlighted as a novel AD risk locus in a GWAS on >1.1 million individuals  
 396 (Wightman et al. 2021), showcasing the power of the quantitative biomarker GWAS approach that  
 397 was also followed in this study. In the current work, we focused on biomarkers / phenotypes derived  
 398 from brain imaging and neuropsychological testing in the same EMIF-AD MBD individuals. Overall,  
 399 we performed 19 individual GWAS and identified a total of 13 genome-wide significant loci  
 400 highlighting several novel genes that are potentially involved in contributing to AD pathophysiology.  
 401 Our study represents one of few GWAS in the literature to also include the X chromosome, where we  
 402 identified a genome-wide significant association between markers near *ILIRAPLI* and longitudinal  
 403 visuoconstructive ability. Interestingly, neither *APOE* nor the other recently described AD GWAS  
 404 loci appear to have a major impact on the traits analyzed in our study. In summary, our extensive  
 405 genome-wide analyses nominate several novel loci potentially involved in neurocognitive  
 406 functioning. Some of these may prove informative to better understand the genetic forces underlying  
 407 AD and related phenotypes.

408 In the remainder of this section, we discuss the potential role of five loci, which we consider the most  
 409 interesting findings of our study. The strongest GWAS signal was elicited by SNP rs73045836  
 410 ( $P=7.50E-11$ ; Table 2; Figure 1B; Supplementary Figure 25) showing genome-wide significant  
 411 association with the longitudinal data of the immediate recall memory phenotype. The gene  
 412 annotated to the associated region on chromosome 6q27, *SMOC2*, encodes secreted modular  
 413 calcium-binding protein 2. *SMOC2* is an extracellular matrix protein from the secreted protein, acidic

414 and rich in cysteine (SPARC) family (Gao et al., 2019) recently linked to age-dependent bone loss in  
 415 humans (Morkmued et al., 2020). Despite performing careful literature and database searches, we  
 416 could not pinpoint any obvious mechanistic connection of this locus to cognitive functioning or other  
 417 AD-relevant phenotypes.

418 The second strongest association signal was observed near SNP rs5943462 (MAF ~0.05) on the X  
 419 chromosome ( $P=1.06E-09$ ; Table 2; Figure 1B; Supplementary Figure 29) with the longitudinal data  
 420 of the visuoconstruction phenotype. The SNP is located in an intron of *ILIRAPLI*. This gene encodes  
 421 interleukin 1 receptor accessory protein-like 1, which belongs to a class of molecules that regulate  
 422 synapse formation. *ILIRAPLI* is mostly expressed in brain areas that are involved in memory  
 423 development, such as hippocampus, dentate gyrus, and entorhinal cortex, suggesting that the protein  
 424 may have a specialized role in physiological processes underlying memory and learning abilities  
 425 (Montani et al., 2019). Even small changes in the expression and function of these proteins can  
 426 provoke major alterations in synaptic connectivity, resulting in cognitive damage (Montani et al.,  
 427 2019). Moreover, *ILIRAPLI* was nominated as a candidate gene for X-linked mental retardation  
 428 (Raymond, 2006). Although the GWAS on longitudinal visuoconstruction included only 149  
 429 individuals, we believe this signal to be plausible and very interesting because of the well-established  
 430 role of *ILIRAPLI* on human brain function.

431  
 432 The third highlighted signal relates to the association between variant rs74381761 (MAF ~0.05) on  
 433 chromosome 8p23.1 ( $P=1.89E-08$ ; Table 2; Figure 1B; Supplementary Figure 5) and longitudinal  
 434 MMSE measurements. This SNP is located in an intergenic region near *TNKS*. This gene encodes the  
 435 protein tankyrase, which belongs to a class of poly (ADP-ribose) polymerases and is involved in  
 436 various processes in the body, such as telomere regulation, Wnt/ $\beta$ -catenin signaling pathway or  
 437 glucose transport (Damale et al., 2020). According to GTEx (Lonsdale et al., 2013), *TNKS* is highly  
 438 expressed in brain (mostly in cerebellum). Moreover, SNPs annotated to *TNKS* were associated with  
 439 brain white matter hyperintensity (WMH) measurements (Armstrong et al., 2020; Sargurupremraj et  
 440 al., 2020; Zhao et al., 2021) and cortical surface area measurements (Grasby et al., 2020) according  
 441 to the GWAS catalog (Buniello et al., 2019). With a gene-based  $P$  value of  $4.87E-04$  and the strong  
 442 functional link to brain function, we consider the signal around *TNKS* as plausible and very  
 443 interesting.

444 The last highlighted finding from the longitudinal analyses relates to the genome-wide significant  
 445 association observed between SNP rs9652864 and the delayed recall memory phenotype ( $P=3.20E-$   
 446  $08$ ; Table 2; Figure 1B; Supplementary Figure 21). This variant (MAF=0.218) attained a  $P$  value of  
 447  $6.73E-04$  in the GWAS of Davies (2018) on cross-sectional cognitive performance, lending  
 448 additional support to our finding. The SNPs is located in an intron of *CEP112* which encodes  
 449 centrosomal protein 112. Centrosomal proteins are known as the components of the centrosome  
 450 involved in centriole biogenesis, cell cycle progression, and spindle-kinetochore assembly control  
 451 (Mazaheri Moghaddam et al., 2021). Despite showing only low levels of expression in the central  
 452 nervous system (CNS) according to GTEx, SNPs in this gene have been associated with cortical  
 453 surface area by neuroimaging in two independent GWAS (Grasby et al., 2020; van der Meer et al.,  
 454 2020) according to the GWAS catalog (Buniello et al., 2019). However, none of these neuroimaging  
 455 SNPs is in relevant LD ( $r^2>0.6$ ) to the lead variant identified here. Notwithstanding, given that  
 456 variants in this gene have shown genetic links to both cognitive function and structural brain  
 457 imaging, we consider this finding as plausible and highly interesting.

458 In the GWAS analyses of the cross-sectional neurocognitive phenotypes, we observed three genome-  
 459 wide significant signals, of which we consider the gene-based association with *EHBPI* as the most

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

471 In addition to searching for novel genetic determinants of the neuroimaging and neurocognitive traits  
 472 analyzed in this study, we also investigated the overlap with known GWAS findings. First and  
 473 foremost, this relates to two commonly studied alleles in the *APOE* gene, which appear to only play a  
 474 minor role in this setting. Specifically, SNP rs429358, which defines the  $\epsilon 4$  allele in *APOE*, does not  
 475 even reach genome-wide suggestive significance ( $P<1.0E-05$ ) in any of the 19 GWAS investigated  
 476 here. The strongest associations with this allele were seen with MRI-based hippocampus volume (left  
 477 volume  $P=0.0002$ , summed volume  $P=0.0005$ ; Supplementary Table 20) and with the delayed recall  
 478 memory test (baseline  $P=0.0005$ , longitudinal  $P=0.0042$ ; Supplementary Table 20). The effect  
 479 directions of these associations are consistent with the deleterious influence of rs429358 on AD (Neu  
 480 et al., 2017). These at best marginal associations are in line with the literature: In the GWAS on  
 481 cognitive function by Davies (2018), *APOE*  $\epsilon 4$  also only showed marginal association ( $P=2.2E-04$ ),  
 482 while it was not reported to show any evidence of association with hippocampal volume in the  
 483 GWAS by Hibar (2017). These findings are different from our earlier GWAS in the EMIF-AD MBD  
 484 dataset, where the  $\epsilon 4$  allele showed very pronounced evidence of association in CSF and imaging  
 485 markers related to  $A\beta 42$  (Hong et al., 2020). Extending the comparison to additional genetic variants  
 486 associated with AD risk in the GWAS by Jansen et al. (2019) did also not show any noteworthy or  
 487 consistent overlap with the GWAS results generated in this study. In contrast, highly significant  
 488 overlaps by PGS analysis were observed upon using GWAS results from Davies et al. (2018) for the  
 489 neuropsychological and Hibar et al. (2017) for the MRI phenotypes, which is not surprising given  
 490 that very similar neuropsychological and neuroimaging traits were used as outcomes in these studies.  
 491 Collectively, the PGS results of this and previous work show that there is only very limited overlap in  
 492 the genetic architecture (at least when studying common SNPs) between AD on the one and  
 493 neuropsychological performance or structural brain imaging on the other hand. We note that this does  
 494 not preclude the possibility that certain molecular pathways targeted by the genes highlighted in this  
 495 GWAS may be shared with AD pathophysiology.

496 While our study has several noteworthy strengths (e.g. the use of highly standardized procedures in  
 497 generating and harmonizing both the genotype and phenotype data of our study, use of both cross-  
 498 sectional and longitudinal neurocognitive performance data, inclusion of the X chromosome in the  
 499 GWAS), it may also have been negatively affected by some limitations. First and foremost, we note  
 500 that the sample size used for the present analyses is comparatively small for “GWAS standards” and  
 501 was well under 1,000 in some instances (Table 1). Accordingly, the statistical power of these  
 502 analyses was low. This limitation is at least partially countered by the quantitative nature of nearly all  
 503 analyzed phenotypes: it is well established that quantitative trait association analyses are more  
 504 powerful than those using binary phenotypes, e.g. in a case-control setting (Bush and Moore, 2012).  
 505 Second, in addition to resulting in low power, small sample sizes also increase the possibility of  
 506 false-positive findings, especially for infrequent variants (i.e. those with an MAF  $<5\%$ ). In this

507 context we note that eight of our thirteen genome-wide significant signals were elicited by such  
 508 variants. Thus, independent replication – ideally in larger datasets – is needed to confirm the main  
 509 findings of our GWAS before any further-reaching conclusions can be reached. Third, we note that  
 510 the phenotype data used as outcome traits in our GWAS analyses were collected at different  
 511 participating centers at times using different types of examinations (e.g. different tests to study the  
 512 same overarching neuropsychological domain). To alleviate potential bias resulting from this  
 513 inherent phenotypic heterogeneity, all clinical data were processed, quality-controlled and  
 514 harmonized (e.g. by normalizing most variables within centers) centrally by an experienced team of  
 515 researchers (see Bos et al. (2018) for more details). We emphasize that this potential heterogeneity  
 516 does not apply to the genetic data as these were generated in one laboratory experiment and  
 517 subsequently processed jointly in one analytical framework, minimizing the emergence of potential  
 518 batch effects. Last but not least, we emphasize that owing to its particular ascertainment design (Bos  
 519 et al., 2018) the EMIF-AD MBD dataset does not (attempt to) constitute a representative sample  
 520 from the “general population”. Accordingly, the results presented here cannot be generalized to the  
 521 general population. We note that the same is true for many GWAS in this and other fields, which  
 522 typically use clinic-based ascertainment which is not representative of the population as a whole.

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

## 528 **Conflict of Interest**

529 HZ has served at scientific advisory boards and/or as a consultant for Abbvie, Alector, Annexon,  
 530 Artery Therapeutics, AZTherapies, CogRx, Denali, Eisai, Nervgen, Pinteon Therapeutics, Red  
 531 Abbey Labs, Passage Bio, Roche, Samumed, Siemens Healthineers, Triplet Therapeutics, and Wave,  
 532 has given lectures in symposia sponsored by Cellectricon, Fujirebio, Alzecure, Biogen, and Roche,  
 533 and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the  
 534 GU Ventures Incubator Program. FB is supported by the NIHR biomedical research centre at UCLH.  
 535 JP received consultation honoraria from Nestle Institute of Health Sciences, Ono Pharma, OM  
 536 Pharma, and Fujirebio, unrelated to the submitted work. CET has a collaboration contract with ADx  
 537 Neurosciences, Quanterix and Eli Lilly, performed contract research or received grants from AC-  
 538 Immune, Axon Neurosciences, Biogen, Brainstorm Therapeutics, Celgene, EIP Pharma, Eisai,  
 539 PeopleBio, Roche, Toyama, Vivoryon. She serves on editorial boards of *Medidact*  
 540 *Neurologie/Springer*, *Alzheimer Research and Therapy*, *Neurology: Neuroimmunology &*  
 541 *Neuroinflammation*, and is editor of a *Neuromethods* book Springer. CET also holds a speaker’s  
 542 contract with Roche, Inc. KB has served as a consultant, at advisory boards, or at data monitoring  
 543 committees for Abcam, Axon, BioArctic, Biogen, JOMDD/Shimadzu, Julius Clinical, Lilly, MagQu,  
 544 Novartis, Pharmatrophix, Prothena, Roche Diagnostics, and Siemens Healthineers, and is a co-  
 545 founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures  
 546 Incubator Program, outside the work presented in this paper. SL is an employee of Janssen-Cilag. JS  
 547 is an employee and chief medical officer of AC Immune SA. The other authors declare that the  
 548 research was conducted in the absence of any commercial or financial relationships that could be  
 549 construed as a potential conflict of interest.

## 550 **Author Contributions**

551 JH and TO performed all the analyses and together with LB interpreted the data and wrote the  
 552 manuscript. OO performed X chromosomal analyses. VD was responsible for EMIF-AD MBD DNA  
 553 sample preparation and DNA extraction. LD contributed to the interpretation of the data and  
 554 visualization of the results. IB, SV, and PJV coordinated the collection and harmonization of  
 555 phenotypes and biosamples in EMIF-AD MBD. MW and AF supervised the genotyping experiments.  
 556 RV, SG, JS, PS, CT, SE, GF, OB, JR, RB, AL, DA, JP, CC, GP, PML, MT, RD, CLQ, KS, CVB,  
 557 CL, KB, HZ, MTK, and FB contributed to sample and phenotype data collection. JS, PJV and SL are  
 558 the lead PIs for the EMIF-AD MBD study as a whole and were in charge of designing and managing  
 559 the platform. LB designed and supervised the genomics portion of the EMIF-AD MBD project and  
 560 co-wrote all drafts of the manuscript. All authors critically revised all manuscripts drafts, read and  
 561 approved the final manuscript.

## 562 **Funding**

563 The present study was conducted as part of the EMIF-AD MBD project, which has received support  
 564 from the Innovative Medicines Initiative Joint Undertaking under EMIF grant agreement No.  
 565 115372, the resources of which are composed of financial contribution from the European Union's  
 566 Seventh Framework Program (FP7/2007-2013) and EFPIA companies' in kind contribution. Parts of  
 567 this study were made possible through support from the German Research Foundation (DFG grant  
 568 FOR2488: Main support by subproject "INF-GDAC" BE2287/7-1 to LB) and the Cure Alzheimer's  
 569 Fund (to LB). RV acknowledges support by the Stichting Alzheimer Onderzoek (#13007, #11020,  
 570 #2017-032) and the Flemish Government (VIND IWT 135043). KB is supported by the Swedish  
 571 Research Council (#2017-00915); the Alzheimer Drug Discovery Foundation (ADDF), USA  
 572 (#RDAPB-201809-2016615); the Swedish Alzheimer Foundation (#AF-742881), Hjärnfonden,  
 573 Sweden (#FO2017-0243); the Swedish state under the agreement between the Swedish government  
 574 and the County Councils; the ALF-agreement (#ALFGBG-715986); European Union Joint Program  
 575 for Neurodegenerative Disorders (JPND2019-466-236); and the Alzheimer's Association 2021  
 576 Zenith Award (ZEN-21-848495). HZ is a Wallenberg Scholar supported by grants from the Swedish  
 577 Research Council (#2018-02532), the European Research Council (#681712), and Swedish State  
 578 Support for Clinical Research (#ALFGBG-720931). SJBV received funding from the Innovative  
 579 Medicines Initiative 2 Joint Undertaking under ROADMAP grant agreement No. 116020 and from  
 580 ZonMw during the conduct of this study. Research at VIB-UAntwerp was in part supported by the  
 581 University of Antwerp Research Fund and SAO-FRA 2018 0016. The Lausanne study was funded by  
 582 a grant from the Swiss National Research Foundation (SNF 320030\_141179) to JP. Research of CET  
 583 is supported by the European Commission (Marie Curie International Training Network, grant  
 584 agreement No 860197 (MIRIAD), and JPND), Health Holland, the Dutch Research Council  
 585 (ZonMW), Alzheimer Drug Discovery Foundation, The Selfridges Group Foundation, Alzheimer  
 586 Netherlands, Alzheimer Association. CET is recipient of ABOARD, which is a public-private  
 587 partnership receiving funding from ZonMW (#73305095007) and Health~Holland, Topsector Life  
 588 Sciences & Health (PPP-allowance; #LSHM20106). More than 30 partners participate in ABOARD.  
 589 ABOARD also receives funding from Edwin Bouw Fonds and Gieskes-Strijbisfonds.

## 590 **Acknowledgments**

591 The authors acknowledge the assistance of Ellen De Roeck, Naomi De Roeck, and Hanne Struyfs  
 592 (UAntwerp) with data collection. We thank Mrs. Tanja Wesse and Mrs. Sanaz Sedghpour Sabet at  
 593 the Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany  
 594 for technical assistance with the GSA genotyping. We thank Dr. Fabian Kilpert for his assistance  
 595 with the QC and genotype imputations and Marcel Schilling for his assistance on visualizing the

596 results. The LIGA team acknowledges computational support from the OMICS compute cluster at  
597 the University of Lübeck.

## 598 **References**

- 599 Armstrong, N. J., Mather, K. A., Sargurupremraj, M., Knol, M. J., Malik, R., Satizabal, C. L., et al.  
600 (2020). Common Genetic Variation Indicates Separate Causes for Periventricular and Deep White  
601 Matter Hyperintensities. *Stroke*, 51:7, 2111–2121. doi:10.1161/STROKEAHA.119.027544
- 602 Bertram, L., and Tanzi, R. E. (2020). Genomic mechanisms in Alzheimer's disease. *Brain pathology*  
603 (*Zurich, Switzerland*), 30:5, 966–977. doi:10.1111/bpa.12882
- 604 Bos, I., Vos, S., Vandenberghe, R., Scheltens, P., Engelborghs, S., Frisoni, G., et al. (2018). The  
605 EMIF-AD Multimodal Biomarker Discovery study: design, methods and cohort  
606 characteristics. *Alzheimer's research and therapy*, 10:1, 64. doi:10.1186/s13195-018-0396-5
- 607 Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., et al. (2012).  
608 Annotation of functional variation in personal genomes using RegulomeDB. *Genome research*, 22:9,  
609 1790–1797. doi:10.1101/gr.137323.112
- 610  
611 Buniello, A., MacArthur, J., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., et al. (2019).  
612 The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and  
613 summary statistics 2019. *Nucleic acids research*, 47:D1, D1005–D1012. doi:10.1093/nar/gky1120
- 614  
615 Bush, W. S., and Moore, J. H. (2012). Chapter 11: Genome-wide association studies. *PLoS*  
616 *computational biology*, 8:12, e1002822. doi:10.1371/journal.pcbi.1002822
- 617  
618 Cacace, R., Slegers, K., and Van Broeckhoven, C. (2016). Molecular genetics of early-onset  
619 Alzheimer's disease revisited. *Alzheimer's & dementia: the journal of the Alzheimer's Association*,  
620 12:6, 733–748. doi:10.1016/j.jalz.2016.01.012
- 621 Choi, S. W., and O'Reilly, P. F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale  
622 data. *GigaScience*, 8(7), giz082. doi:10.1093/gigascience/giz082
- 623 Damale, M. G., Pathan, S. K., Shinde, D. B., Patil, R. H., Arote, R. B., and Sangshetti, J. N. (2020).  
624 Insights of tankyrases: A novel target for drug discovery. *European journal of medicinal chemistry*,  
625 207, 112712. doi:10.1016/j.ejmech.2020.112712
- 626  
627 Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., et al. (2016). Next-generation  
628 genotype imputation service and methods. *Nature genetics*, 48:10, 1284–1287. doi:10.1038/ng.3656
- 629  
630 Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., Hill, W. D., et al. (2018). Study of  
631 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function.  
632 *Nature communications*, 9:1, 2098. doi:10.1038/s41467-018-04362-x
- 633 de Leeuw, C. A., Mooij, J. M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set  
634 analysis of GWAS data. *PLoS computational biology*, 11:4, e1004219.  
635 doi:10.1371/journal.pcbi.1004219

- 636 Gao, Q., Mok, H. P., and Zhuang, J. (2019). Secreted modular calcium-binding proteins in  
637 pathophysiological processes and embryonic development. *Chinese medical journal*, 132:20, 2476–  
638 2484. doi:10.1097/CM9.0000000000000472
- 639 Gatz, M., Reynolds, C. A., Fratiglioni, L., Johansson, B., Mortimer, J. A., Berg, S., et al. (2006).  
640 Role of genes and environments for explaining Alzheimer disease. *Archives of general psychiatry*,  
641 63:2, 168–174. doi:10.1001/archpsyc.63.2.168
- 642  
643 Gottesman, I. I., and Gould, T. D. (2003). The endophenotype concept in psychiatry: etymology and  
644 strategic intentions. *The American journal of psychiatry*, 160:4, 636–645.  
645 doi:10.1176/appi.ajp.160.4.636
- 646  
647 Graffelman, J., and Weir, B. S. (2016). Testing for Hardy-Weinberg equilibrium at biallelic genetic  
648 markers on the X chromosome. *Heredity*, 116:6, 558–568. doi:10.1038/hdy.2016.20
- 649  
650 Grasby, K. L., Jahanshad, N., Painter, J. N., Colodro-Conde, L., Bralten, J., Hibar, D. P., et al.  
651 (2020). The genetic architecture of the human cerebral cortex. *Science (New York, N.Y.)*, 367:6484,  
652 eaay6690. doi:10.1126/science.aay6690
- 653 Hibar, D. P., Adams, H., Jahanshad, N., Chauhan, G., Stein, J. L., Hofer, M. A., et al. (2017). Novel  
654 genetic loci associated with hippocampal volume. *Nature communications*, 8, 13624.  
655 doi:10.1038/ncomms13624
- 656  
657 Holtzman, D. M., Herz, J., and Bu, G. (2012). Apolipoprotein E and apolipoprotein E receptors:  
658 normal biology and roles in Alzheimer disease. *Cold Spring Harbor perspectives in medicine*, 2:3,  
a006312. doi:10.1101/cshperspect.a006312
- 659  
660 Hong, S., Prokopenko, D., Dobricic, V., Kilpert, F., Bos, I., Vos, S., et al. (2020). Genome-wide  
661 association study of Alzheimer's disease CSF biomarkers in the EMIF-AD Multimodal Biomarker  
Discovery dataset. *Translational psychiatry*, 10:1, 403. doi:10.1038/s41398-020-01074-z
- 662  
663 Hong, S., Dobricic, V., Ohlei, O., Bos, I., Vos, S., Prokopenko, D., et al. (2021). TMEM106B and  
664 CPOX are genetic determinants of cerebrospinal fluid Alzheimer's disease biomarker  
665 levels. *Alzheimer's and dementia : the journal of the Alzheimer's Association*, 17:10, 1628–1640.  
doi:10.1002/alz.12330
- 666  
667 Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., et al. (2019).  
668 Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's  
disease risk. *Nature genetics*, 51:3, 404–413. doi:10.1038/s41588-018-0311-9
- 669  
670 Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., et al. (2013). The Genotype-  
671 Tissue Expression (GTEx) project. *Nature genetics*, 45:6, 580–585. doi:10.1038/ng.2653
- 672  
673 Machiela, M. J., and Chanock, S. J. (2015). LDlink: a web-based application for exploring  
674 population-specific haplotype structure and linking correlated alleles of possible functional variants.  
675 *Bioinformatics (Oxford, England)*, 31:21, 3555–3557. doi:10.1093/bioinformatics/btv402
- 676  
677 MacRae, C. A., and Vasan, R. S. (2011). Next-generation genome-wide association studies: time to  
678 focus on phenotype?. *Circulation. Cardiovascular genetics*, 4:4, 334–336.  
doi:10.1161/CIRCGENETICS.111.960765



- 679 Martin, P. M., Carnaud, M., Garcia del Caño, G., Irondelle, M., Irinopoulou, T., Girault, J. A., et al.  
680 (2008). Schwannomin-interacting protein-1 isoform IQCJ-SCHIP-1 is a late component of nodes of  
681 Ranvier and axon initial segments. *The Journal of neuroscience : the official journal of the Society*  
682 *for Neuroscience*, 28:24, 6111–6117. doi:10.1523/JNEUROSCI.1044-08.2008
- 683 Mattsson, N., Carrillo, M. C., Dean, R. A., Devous, M. D., Sr, Nikolcheva, T., Pesini, P., et al.  
684 (2015). Revolutionizing Alzheimer's disease and clinical trials through biomarkers. *Alzheimer's &*  
685 *dementia (Amsterdam, Netherlands)*, 1:4, 412–419. doi:10.1016/j.dadm.2015.09.001
- 686 Martin, P. M., Cifuentes-Diaz, C., Devaux, J., Garcia, M., Bureau, J., Thomasseau, S., et al. (2017).  
687 Schwannomin-interacting Protein 1 Isoform IQCJ-SCHIP1 Is a Multipartner Ankyrin- and Spectrin-  
688 binding Protein Involved in the Organization of Nodes of Ranvier. *The Journal of biological*  
689 *chemistry*, 292:6, 2441–2456. doi:10.1074/jbc.M116.758029
- 690  
691 Mazaheri Moghaddam, M., Mazaheri Moghaddam, M., Hamzeiy, H., Baghbanzadeh, A.,  
692 Pashazadeh, F., and Sakhinia, E. (2021). Genetic basis of acephalic spermatozoa syndrome, and  
693 intracytoplasmic sperm injection outcomes in infertile men: a systematic scoping review. *Journal of*  
694 *assisted reproduction and genetics*, 38:3, 573–586. doi:10.1007/s10815-020-02008-w
- 695  
696 McCarthy, S., Das, S., Kretschmar, W., Delaneau, O., Wood, A. R., Teumer, A., et al. (2016). A  
697 reference panel of 64,976 haplotypes for genotype imputation. *Nature genetics*, 48:10, 1279–1283.  
698 doi:10.1038/ng.3643
- 699  
700 McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., et al. (2016). The  
701 Ensembl Variant Effect Predictor. *Genome biology*, 17:1, 122. doi:10.1186/s13059-016-0974-4
- 702  
703 Montani, C., Gritti, L., Beretta, S., VerPELLI, C., and Sala, C. (2019). The Synaptic and Neuronal  
704 Functions of the X-Linked Intellectual Disability Protein Interleukin-1 Receptor Accessory Protein  
705 Like 1 (IL1RAPL1). *Developmental neurobiology*, 79:1, 85–95. doi:10.1002/dneu.22657
- 706  
707 Morkmued, S., Clauss, F., Schuhbauer, B., Fraulob, V., Mathieu, E., Hemmerlé, J., et al. (2020).  
708 Deficiency of the SMOC2 matricellular protein impairs bone healing and produces age-dependent  
709 bone loss. *Scientific reports*, 10:1, 14817. doi:10.1038/s41598-020-71749-6
- 710  
711 Neu, S. C., Pa, J., Kukull, W., Beekly, D., Kuzma, A., Gangadharan, P., et al. (2017). Apolipoprotein  
712 E Genotype and Sex Risk Factors for Alzheimer Disease: A Meta-analysis. *JAMA neurology*, 74:10,  
713 1178–1189. doi:10.1001/jamaneurol.2017.2188
- 714  
715 Omodei, D., Acampora, D., Russo, F., De Filippi, R., Severino, V., Di Francia, R., et al. (2009).  
716 Expression of the brain transcription factor OTX1 occurs in a subset of normal germinal-center B  
717 cells and in aggressive Non-Hodgkin Lymphoma. *The American journal of pathology*, 175:6, 2609–  
718 2617. doi:10.2353/ajpath.2009.090542
- 719  
720 Patterson, C. (2018) World Alzheimer Report 2018. The State of the Art of Dementia Research: New  
721 Frontiers. An Analysis of Prevalence, Incidence, Cost and Trends. *Alzheimer's Disease International*.
- 721  
722 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007).  
723 PLINK: a tool set for whole-genome association and population-based linkage analyses. *American*  
*journal of human genetics*, 81:3, 559–575. doi:10.1086/519795

- 724 Quintino-Santos, S., Diniz, B. S., Firmo, J. O., Moriguchi, E. H., Lima-Costa, M. F., and Castro-  
725 Costa, E. (2015). APOE ε4 allele is associated with worse performance in memory dimensions of the  
726 mini-mental state examination: the Bambuí Cohort Study of Aging. *International journal of geriatric*  
727 *psychiatry*, 30:6, 573–579. doi:10.1002/gps.4186  
728
- 729 Rai, A., Bleimling, N., Vetter, I. R., and Goody, R. S. (2020). The mechanism of activation of the  
730 actin binding protein EHBP1 by Rab8 family members. *Nature communications*, 11:1, 4187.  
731 doi:10.1038/s41467-020-17792-3  
732
- 733 Raymond F. L. (2006). X linked mental retardation: a clinical guide. *Journal of medical genetics*,  
734 43:3, 193–200. doi:10.1136/jmg.2005.033043  
735
- 736 Rentzsch, P., Schubach, M., Shendure, J., and Kircher, M. (2021). CADD-Splice-improving genome-  
737 wide variant effect prediction using deep learning-derived splice scores. *Genome medicine*, 13:1, 31.  
738 doi:10.1186/s13073-021-00835-9  
739
- 740 Sargurupremraj, M., Suzuki, H., Jian, X., Sarnowski, C., Evans, T. E., Bis, J. C., et al. (2020).  
741 Cerebral small vessel disease genomics and its implications across the lifespan. *Nature*  
742 *communications*, 11:1, 6285. doi:10.1038/s41467-020-19111-2
- 743 Scheufele, E., Aronzon, D., Coopersmith, R., McDuffie, M. T., Kapoor, M., Uhrich, C. A., et al.  
744 (2014). tranSMART: An Open Source Knowledge Management and High Content Data Analytics  
745 Platform. *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on*  
746 *Translational Science*, 2014, 96–101.
- 747 Smith, S. M., Douaud, G., Chen, W., Hanayik, T., Alfaro-Almagro, F., Sharp, K., et al. (2021). An  
748 expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank.  
749 *Nature neuroscience*, 24:5, 737–745. doi:10.1038/s41593-021-00826-4
- 750 Sperling, R., Mormino, E., and Johnson, K. (2014). The evolution of preclinical Alzheimer's disease:  
751 implications for prevention trials. *Neuron*, 84:3, 608–622. doi: 10.1016/j.neuron.2014.10.038  
752
- 753 Ten Kate, M., Redolfi, A., Peira, E., Bos, I., Vos, S. J., Vandenberghe, R., et al. (2018). MRI  
754 predictors of amyloid pathology: results from the EMIF-AD Multimodal Biomarker Discovery study.  
755 *Alzheimer's research & therapy*, 10:1, 100. doi:10.1186/s13195-018-0428-1  
756
- 757 van der Meer, D., Frei, O., Kaufmann, T., Shadrin, A. A., Devor, A., Smeland, O. B., et al. (2020).  
758 Understanding the genetic determinants of the brain with MOSTest. *Nature communications*, 11:1,  
759 3512. doi:10.1038/s41467-020-17368-1
- 760 Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and  
761 annotation of genetic associations with FUMA. *Nature communications*, 8:1, 1826.  
762 doi:10.1038/s41467-017-01261-5
- 763 Wightman, D. P., Jansen, I. E., Savage, J. E., Shadrin, A. A., Bahrami, S., Holland, D., et al. (2021).  
764 A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's  
765 disease. *Nature genetics*, 53:9, 1276–1282. doi:10.1038/s41588-021-00921-z

766 Willer, C. J., Li, Y., and Abecasis, G. R. (2010). METAL: fast and efficient meta-analysis of  
 767 genomewide association scans. *Bioinformatics (Oxford, England)*, 26:17, 2190–2191.  
 768 doi:10.1093/bioinformatics/btq340

769 Zhang, Q., Cai, Z., Lhomme, M., Sahana, G., Lesnik, P., Guerin, M., et al. (2020). Inclusion of  
 770 endophenotypes in a standard GWAS facilitate a detailed mechanistic understanding of genetic  
 771 elements that control blood lipid levels. *Scientific reports*, 10:1, 18434. doi:10.1038/s41598-020-  
 772 75612-6

773  
 774 Zhao, B., Zhang, J., Ibrahim, J. G., Luo, T., Santelli, R. C., Li, Y., et al. (2021). Large-scale GWAS  
 775 reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive  
 776 and mental health traits (n = 17,706). *Molecular psychiatry*, 26:8, 3943–3955. doi:10.1038/s41380-  
 777 019-0569-z

### 778 **Data Availability Statement**

779 GWAS summary statistics for the top ( $P$  value  $< 1.0E-05$ ) results are listed in the Supplementary  
 780 Tables. Full GWAS summary statistics are available from the authors upon request. Clinical data and  
 781 genome-wide genotyping data are stored on an online data platform using the “tranSMART” data  
 782 warehouse framework. Access to the genome-wide genotyping data can be requested from the  
 783 corresponding author of this study who will forward each request to the EMIF-AD data access team.

784 All scripts used to generate the primary GWAS and PGS analyses are available from the authors  
 785 upon request.

### 786 **Figure legends**

787 **Figure 1:** Multi-trait Manhattan plots for the SNP-based GWAS results on neuropsychological  
 788 phenotypes (A: cross-sectional; B: longitudinal). For details on the analyzed traits see Methods and  
 789 Supplementary Material.

790 **Tables**791 **Table 1:** Description of EMID-AD MBD datasets analyzed per phenotype.

792 "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  
 793 inflation of the GWAS test statistics. Information on tests used for generating baseline and longitudinal phenotypes can be found in the  
 794 Supplementary Material. MMSE = Mini Mental State Examination. "n.a." = not available.

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.

795

796 **Table 2:** Genome-wide significant associations observed in GWAS of cognitive phenotypes.

797 **Bold font** indicates genome-wide significant (on SNP- or gene-level) results (see Methods section for details). "Chr" and "Position"  
 798 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  
 799 belonging to "Nearest gene". Top results from these GWAS analyses can be found in the Supplementary Tables. MMSE = Mini Mental State  
 800 Examination. "n.a." = not available.

Study arm	Phenotype	Lead variant	Chr	Position	Nearest gene	A1	A2	Beta	MAF	P (SNP)	P (Gene)
Cross-sectional	MemoryDelayed	rs6705798	2p15	63259881	<i>EHBPI</i>	C	T	-0.32739	0.358	8.78E-08	<b>1.17E-07</b>
	MMSE	rs2122118	2q33.3	207252439	<i>AC017081.2</i>	G	A	-2.82266	0.022	<b>3.03E-09</b>	n.a.
	Visuoconstruction	rs113492235	4q34.2	177252900	<i>SPCS3</i>	T	C	-2.17956	0.022	<b>1.51E-08</b>	0.15674
Longitudinal	MemoryImmediate	rs73045836	6q27	169062739	<i>SMOC2</i>	G	T	-0.36094	0.020	<b>7.50E-11</b>	0.0035373
	MMSE	rs74381761	8p23.1	9389761	<i>TNKS</i>	C	G	-0.08453	0.048	<b>1.89E-08</b>	0.00048716
	Attention	rs116900143	10q23.31	92588290	<i>HTR7</i>	C	T	-0.35173	0.023	<b>1.95E-08</b>	0.019663
	MemoryImmediate	rs11217863	11q23.3	120293138	<i>AP002348.1</i>	A	G	-0.16626	0.080	7.81E-08	<b>8.91E-07</b>
	Attention	rs111959303	12q14.3	66844015	<i>GRIP1</i>	T	C	0.37459	0.022	<b>2.52E-08</b>	0.81478
	Attention	rs34736485	16q23.2	79272611	<i>RP11-679B19.2</i>	T	G	0.31924	0.022	<b>1.59E-08</b>	n.a.
	MemoryDelayed	rs9652864	17q24.1	63741645	<i>CEP112</i>	A	T	0.29184	0.218	<b>3.20E-08</b>	0.016339
	MemoryImmediate	rs146202660	18q21.1	45022937	<i>CTD-2130O13.1</i>	T	G	-0.29342	0.029	<b>4.63E-08</b>	n.a.
	Executive	rs16982556	20q13.32	57801889	<i>ZNF831</i>	T	C	-0.29752	0.062	<b>1.26E-08</b>	0.0025565
	Visuoconstruction	rs5943462	Xp21.3	28823154	<i>IL1RAPL1</i>	G	C	-0.14082	0.051	<b>1.06E-09</b>	0.006719

801

802 **Table 3:** Summary of PGS results significant after multiple testing correction.

803 "Threshold" refers to *P* value cut-off used for PGS construction in prior GWAS summary statistics and "Number of SNPs" refers to the LD-  
 804 pruned SNPs passing this threshold that are included in PGS calculations. "R2" denotes the phenotypic variance explained by the SNPs of  
 805 the prior GWAS in the EMIF-AD MBD dataset. A full listing of results from these PRS analyses can be found in Supplementary Material.  
 806 MMSE = Mini Mental State Examination.  
 807

Prior GWAS	Phenotype	Threshold	Number of SNPs	R2	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	22269	0.014	3.72E-03
Davies et al. (2018)	Baseline MMSE	0.248	63792	0.016	1.66E-06
	Baseline executive functioning	0.0014	4469	0.018	1.98E-05
	Baseline language	0.0061	9031	0.010	1.25E-03
	Longitudinal attention	5.0E-08	163	0.023	1.79E-03

808

