Multi-ancestry transcriptome-wide association studies of cognitive function, white matter hyperintensity, and Alzheimer's disease

Chaar, Dima L.[1], Li, Zheng[2], Shang, Lulu[3], Ratliff, Scott M.[1], Mosley, Thomas H.[1,4], Kardia, Sharon L.R. [1], Zhao, Wei[1,5], Zhou, X.[2], Smith, J.A.[1,5]

[1] Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, United States of America
[2] Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan, United States of America
[3] Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America
[4] Memory Impairment and Neurodegenerative Dementia (MIND) Center, University of Mississippi Medical Center, Jackson, Mississippi
[5] Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, United States of America

*Corresponding author:

Jennifer A. Smith, PhD, MPH
Department of Epidemiology
University of Michigan
1415 Washington Heights, #2631
Ann Arbor, MI 48109-2029
Phone: 734-615-9455
Fax: 734-764-3192
Email: smjenn@umich.edu
ORCID: 0000-0002- 3575-5468

Author Summary

Transcriptome-wide association studies (TWAS) can be used to understand the mechanisms of gene expression that underly disease etiology. However, to date, TWAS methods have mostly been used in a single ancestry group, especially European ancestry (EA), and few TWAS have focused on cognitive function or structural brain measures. We used a newly developed TWAS method called the Multi-ancEstry TRanscriptOme-wide analysis (METRO) to incorproate gene expression data from 801 EA and 1,032 African ancestry (AA) adults to identify genes associated with general cognitive function, structural brain changes called white matter hyperintensities (WMH) that predispose people to vascular dementia, and another form of dementia called Alzheimer's disease (AD). We found that reduced gene expression of *ICA1L* was associated with more WMH and with AD, indicating its potential contribution to overlapping AD and vascular dementia neuropathologies. To our knowledge, our study is the first TWAS of cognitive function and neurocognitive disorders using multiple ancestries. This work may expand the benefits of TWAS studies beyond a single ancestry group and help to identify gene targets for pharmaceutical or preventative treatment for dementia.

Abstract

Genetic variants increase the risk of neurocognitive disorders in later life including Vascular Dementia (VaD) and Alzheimer's disease (AD), but the precise relationships between genetic risk factors and underlying disease etiology are not well understood. Transcriptome-wide association studies (TWAS) can be leveraged to better characterize the genes and biological pathways underlying genetic influences on disease. To date, almost all existing TWAS have been conducted using expression studies from individuals of a single genetic ancestry, primarily European. Using the joint likelihood-based inference framework in Multi-ancEstry TRanscriptOme-wide analysis (METRO), we leveraged gene expression data from European (EA) and African ancestries (AA) to identify genes associated with general cognitive function, white matter hyperintensity (WMH), and AD. Regions were fine-mapped using Fine-mapping Of CaUsal gene Sets (FOCUS). We identified 266, 23, 69, and 2 genes associated with general cognitive function, WMH, AD (using EA GWAS summary statistics), and AD (using AA GWAS), respectively (Bonferroni-corrected alpha=$P<2.9x10^{-6}$), some of which were previously identified. Enrichment analysis showed that many of the identified genes were in pathways related to innate immunity, vascular dysfunction, and neuroinflammation. Further, downregulation of *ICA1L* was associated with higher WMH and with AD, indicating its potential contribution to overlapping AD and VaD neuropathology. To our knowledge, our study is the first TWAS of cognitive function and neurocognitive disorders that used expression mapping studies in multiple ancestries. This work may expand the benefits of TWAS studies beyond a single ancestry group and help to identify gene targets for pharmaceutical or preventative treatment for dementia.

Introduction

Adult-onset dementia is comprised of a group of aging-related neurocognitive disorders caused by the gradual degeneration of neurons and the loss of brain function. These changes lead to a decline in cognitive abilities and impairment of daily activities and independent function. In the United States, Alzheimer's disease (AD), the most common cause of dementia, affects 6.8 million adults age 65 and older (1). The second most common form of dementia is vascular dementia (VaD), which often co-occurs with AD and is underdiagnosed (1,2). VaD is often difficult to distinguish from AD because these diseases share cognitive symptoms including noticeable impairment in episodic and semantic memory. While AD and VaD often co-occur, each form of dementia has differing pathophysiology that may precede the illness decades prior.

AD is characterized by aggregation of amyloid-beta protein and neurofibrillary tangles in brain tissue (3,4), while VaD may be caused by reduced blood flow to the brain as a result of small vessel disease (SVD) or stroke and is commonly seen in people with hypertension (5). AD is diagnosed based on a battery of memory tests, brain-imaging tests for degeneration of brain cells and laboratory tests to assess the presence of amyloid and tau proteins in cerebrospinal fluid (6). SVD is primarily detected on magnetic resonance imaging (MRI) as white matter hyperintensities (WMH). It has been hypothesized that vascular and neurodegenerative changes in the brain may interact in ways that increase the likelihood of cognitive impairment. A further challenge in the field is distinguishing between individuals who are aging normally from those with dementia pathology.

A greater understanding of the pathological processes that influence cognitive function in older adults is critical for early intervention during the long preclinical or prodromal phase prior to dementia onset, especially in vulnerable populations (7,8). For example, individuals of African

115 ancestry (AA) have a greater burden of and risk for developing dementia compared to Non-
116 Hispanic Whites (9–12). Differences in gene expression, which are influenced by both genetic
117 and non-genetic factors, likely play a role in shaping racial/ethnic health disparities in
118 neurological outcomes. However, the underlying molecular and environmental mechanisms that
119 influence gene expression are not fully understood, especially in populations with non-European
120 ancestries. Given the multifactorial and complex nature of dementia, multi-omic data integration
121 across ancestry groups may lend insight into these disparities, allowing the identification of
122 targets for intervention and treatment in populations that are most at risk (13).

123 Genome-wide association studies (GWAS) have identified genetic variants associated
124 with cognitive function and dementia; however, most GWAS variants are located in non-coding
125 regions so their functional consequences are difficult to characterize (14). Transcriptome-wide
126 association studies (TWAS) utilize gene expression and genetic data to increase power for
127 identifying gene-trait associations and characterizing transcriptomic mechanisms underlying
128 complex diseases. To date, however, few TWAS have been conducted on cognitive or structural
129 brain measures. Further, previous TWAS have primarily been conducted in populations of
130 European ancestry (EA), but these results cannot always be generalized to other genetic
131 ancestries due to differences in allele frequencies, patterns of linkage disequilibrium (LD), and
132 relationships between SNPs and gene expression between populations (15–18). To better identify
133 gene-trait associations in non-EA ancestries, it is necessary to incorporate results from recent
134 expression quantitative trait locus (eQTL) mapping studies, which identify genetic variants that
135 explain variations in gene expression levels, conducted in different ancestry groups.(19)

136 Multi-ancEstry TRanscriptOme-wide analysis (METRO)(20) is a TWAS method that
137 uses a joint likelihood-based inference framework to borrow complementary information across

138 multiple ancestries to increase TWAS power. In this study, we used genotype and gene
139 expression data from 1,032 AA and 801 EA from the Genetic Epidemiology Network of
140 Arteriopathy (GENOA) and summary statistics from published GWAS (21–24) to identify genes
141 associated with general cognitive function, white matter hyperintensity, and AD. We then
142 examined the contribution of different ancestry-dependent transcriptomic profiles on the gene-
143 trait associations. Greater knowledge of the underlying molecular mechanisms of dementia that
144 are generalizable to both EA and AA is a critical step in evaluating potential causal variants and
145 genes that could be targeted for pharmaceutical development.
146
147
148 **2. Materials and Methods**
149
150 ***2.1. Sample***

151 *The Genetic Epidemiology Network of Arteriopathy (GENOA)*

152 The GENOA study is a community-based longitudinal study aimed at examining the
153 genetic effects of hypertension and related target organ damage (25). EA and AA hypertensive
154 sibships were recruited if at least 2 siblings were clinically diagnosed with hypertension before
155 age 60. All other siblings were invited to participate, regardless of their hypertension status.
156 Exclusion criteria included secondary hypertension, alcoholism or drug abuse, pregnancy,
157 insulin-dependent diabetes mellitus, active malignancy, or serum creatinine levels >2.5mg/dL. In
158 Phase I (1996-2001), 1,854 AA participants (Jackson, MS) and 1,583 EA participants
159 (Rochester, MN) were recruited (25). In Phase II (2000-2004), 1,482 AA and 1,239 EA
160 participants were successfully followed up, and their potential target organ damage from
161 hypertension was measured. Demographics, medical history, clinical characteristics, information
162 on medication use, and blood samples were collected in each phase. After data cleaning and

163 quality control, a total of 1,032 AA and 801 EA with genotype and gene expression data were
164 available for analysis. Written informed consent was obtained from all participants, and approval
165 was granted by participating institutional review boards (University of Michigan, University of
166 Mississippi Medical Center, and Mayo Clinic).
167
168 ***2.2. Measures***
169 **2.2.1. Genetic Data**
170 AA and EA blood samples were genotyped using the Affymetrix® Genome-Wide
171 Human SNP Array 6.0 or the Illumina 1M Duo. We followed the procedures outlined by Shang
172 et al.(18) for data processing. For each platform, samples and SNPs with a call rate <95%,
173 samples with mismatched sex, and duplicate samples were excluded. After removing outliers
174 identified from genetic principal component analysis, there were 1,599 AA and 1,464 EA with
175 available genotype data. Imputation was performed using the Segmented HAPlotype Estimation
176 & Imputation Tool (SHAPEIT) v.2.r(26) and IMPUTE v.2(27) using the 1000 Genomes project
177 phase I integrated variant set release (v.3) in NCBI build 37 (hg19) coordinates (released in
178 March 2012). Imputation for each genotyping platform was performed separately and then
179 combined. The final set of genotype data included 30,022,375 and 26,079,446 genetic variants
180 for AA and EA, respectively. After removing genetic variants with MAF ≤ 0.01, imputation
181 quality score (INFO score) ≤ 0.4 in any platform-based imputation, and indels, a total of
182 13,793,193 SNPs in AA and 7,727,215 SNPs in EA were available for analysis. We used the
183 GENESIS package(28) in R to infer population structure in the analytic sample, and the PC-AiR
184 function was used to extract the first five genotype PCs which were subsequently used to adjust
185 for population structure.

186 **2.2.2. Gene Expression Data**

187 Gene expression levels were measured from Epstein-Barr virus (EBV) transformed B-
188 lymphoblastoid cell lines (LCLs) created from blood samples from a subset of GENOA AA
189 (n=1,233) and EA (n=919). Gene expression levels of AA samples were measured using the
190 Affymetrix Human Transcriptome Array 2.0, while gene expression levels of EA samples were
191 measured using Affymetrix Human Exon 1.0 ST Array. We followed the procedures outlined by
192 Shang et al.(18) In particular, the Affymetrix Expression Console was used for quality control
193 and all array images passed visual inspection. In AA, 28 samples were removed due to either low
194 signal-to-noise ratio (n=1), abnormal polyadenylated RNA spike-in controls (Lys < Phe < Thr <
195 Dap; n=24), sample mislabeling (n=2), or low RNA integrity (n=1), resulting in a total of
196 n=1,205 AA samples for analysis. In EA, duplicated samples (n=31), control samples (n=11) and
197 sex mismatch samples (n=2) were removed, resulting in n=875 EA samples for analysis. We
198 processed data in each population separately. Raw intensity data were processed using the
199 Affymetrix Power Tool software (29). AffymetrixCEL files were normalized using the Robust
200 Multichip Average (RMA) algorithm which included background correction, quantile
201 normalization, $\log_2$-transformation, and probe set summarization.(30) The algorithm also
202 includes GC correction (GCCN), signal space transformation (SST), and gain lock (value=0.75)
203 to maintain linearity. The Brainarray custom CDF(31) v.19 was used to map the probes to genes.
204 This custom CDF uses updated genomic annotations and multiple filtering steps to ensure that
205 the probes used are specific for the intended gene cluster. Specifically, it removes probes with
206 non-unique matching cDNA/EST sequences that can be assigned to more than one gene cluster.
207 As a result, gene expression data processed using custom CDF are expected to be largely free of
208 mappability issues. After mapping, ComBat(32) was used to remove batch effects. For each

209 gene, we applied a linear regression model to adjust for age, sex, and first five genotype principal
210 components (PCs). We then extracted the residuals and quantile normalized residuals across all
211 samples. We analyzed a common set of 17,238 protein coding genes that were annotated in
212 GENCODE (release 12) (33).
213
214 **2.2.3. GWAS summary statistics**
215 We used summary statistics from GWAS for general cognitive function (21), WMH (22),
216 AD in EA(23), and AD in AA(24) as input for METRO. Three of the GWASs, Davies et al.
217 (2018), Sargurupremraj et al. (2020), and Bellenguez et al. (2022), were selected because they
218 are the largest meta-analyses to date with publicly available summary statistics; however, we
219 note that all three were conducted in primarily EA samples. We also selected the Kunkle et al.
220 (2021) GWAS because it is the largest meta-analysis to date with public available summary
221 statistics in primarily AA samples. Below, we describe each GWAS and also provide
222 information about the corresponding TWAS analyses that were reported in two of the input
223 GWAS (WMH(22) and AD in EA(23)) which use the same GWAS summary statistics as our
224 analysis but different gene expression data.
225
226 **General cognitive function**
227 We obtained GWAS summary statistics for general cognitive function from a meta-
228 analysis by Davies et al. (2018) that includes the Cohorts for Heart and Aging Research in
229 Genomic Epidemiology (CHARGE), the Cognitive Genomics Consortium (COGENT) consortia
230 and the UK Biobank (UKB; Table 1) (21). This study included 300,486 EA individuals with ages
231 between 16 and 102 years from 57 population-based cohorts. This is the largest available GWAS

232 for general cognitive function, and there are currently no large-scale GWAS studies available in
233 non-EA. General cognitive function was constructed from a number of cognitive tasks. Each
234 cohort was required to have tasks that tested at least three different cognitive domains. Principal
235 component (PC) analysis was performed on the cognitive tests scores within each cohort, and the
236 first unrotated component was extracted as the measure of general cognitive function. Models
237 performed within each cohort were adjusted for age, sex, and population stratification. Exclusion
238 criteria included clinical stroke (including self-reported stroke) or prevalent dementia.
239
240 **White matter hyperintensity**
241 We obtained the GWAS summary statistics for WMH from a meta-analysis conducted by
242 Sargurupremraj et al. (2020) that included 48,454 EA and 2,516 AA with mean age of 66.0
243 (SD=7.5) years from 23 population-based studies from the CHARGE consortium and UKB
244 (Table 1) (22). We obtained publicly available GWAS summary statistics from only EA
245 individuals. Summary statistics for only EA are publicly available for this GWAS. WMH was
246 measured from MRI scans obtained from scanners with field strengths ranging from 1.5 to 3.0
247 Tesla and interpreted using a standardized protocol blinded to clinical or demographic features.
248 In addition to T1 and T2 weighted scans, some cohorts included fluid-attenuated inversion
249 recovery (FLAIR) and/or proton density (PD) sequences to measure WMH from cerebrospinal
250 fluid. WMH volume measures were inverse normal transformed, and models adjusted for sex,
251 age, genetic PCs and intracranial volume (ICV). Exclusion criteria included history of stroke or
252 other pathologies that influence measurement of WMH at the time of MRI.
253 To functionally characterize and prioritize individual WMH genomic risk loci,
254 Sargurupremraj et al.(22) (2020) conducted TWAS using TWAS-Fusion(34) with summary

255 statistics from the WMH SNP-main effects (EA only) analysis and weights from gene expression
256 reference panels from blood (Netherlands Twin Registry; Young Finns Study), arterial
257 (Genotype-Tissue Expression, GTEx), brain (GTEx, CommonMind Consortium) and peripheral
258 nerve tissue (GTEx). This study did not perform fine-mapping following TWAS analysis.
259

260 **Alzheimer's disease (GWAS in EA)**

261 We obtained the EA GWAS summary statistics for Alzheimer's disease from stage I
262 meta-analysis by Bellenguez et al. (2022) that included EA from the European Alzheimer and
263 Dementia Biobank (EADB), GR@ACE, EADI, GERAD/PERADES, DemGene, Bonn, the
264 Rotterdam study, CCHS study, NxC and the UKB (Table 1) (23). The meta-analysis was
265 performed on 39,106 clinically diagnosed AD cases, 46,828 proxy-AD and related dementia
266 (ADD) cases, and 401,577 controls. AD cases were clinically diagnosed in all cohorts except
267 UKB, where individuals were identified as proxy-ADD cases if their parents had dementia.
268 Participants without the clinical diagnosis of AD, or those without any family history of
269 dementia, were used as controls. Models performed within each cohort were adjusted for PCs
270 and genotyping centers, when necessary.

271 To examine the downstream effects of new AD-associated variants on molecular
272 phenotypes in various AD-relevant tissues, Bellenguez et al. (2022) conducted a TWAS with
273 stage I AD GWAS results. The TWAS was performed by training functional expression and
274 splicing reference panels based on the Accelerating Medicines Partnership (AMP)-AD bulk brain
275 and EADB lymphoblastoid cell lines (LCL) cohorts, while leveraging pre-calculated reference
276 panel weights(35) for the GTEx dataset(36) in tissues and cells of interest. TWAS associations
277 were then fine-mapped using Fine-mapping Of CaUsal gene Sets (FOCUS) (37).

278

279 **Alzheimer's disease (GWAS in AA)**

280 We also acquired the AA GWAS summary statistics for Alzheimer's disease from meta-
281 analysis by Kunkle et al. (2021)(24) that included individuals of African American ancestry from
282 15 cohort studies from the AD Genetics Consortium (ADGC; Table 1). The meta-analysis was
283 performed on 2,748 clinically diagnosed AD cases and 5,222 controls with a mean age of 74.2
284 years (SD=13.6). Models performed within each cohort were adjusted for age, sex, and PCs for
285 population substructure.

286

287 ***2.3 Statistical Methods***

288 **2.3.1. Multi-ancestry transcriptome-wide association study**

289 Using the Multi-ancEstry TRanscriptOme-wide analysis (METRO) (20), we conducted a
290 high-powered TWAS with calibrated type I error control to identify the key gene-trait
291 associations and transcriptomic mechanisms underlying general cognitive function, WMH and
292 AD. Since gene expression prediction models constructed in different ancestries may contain
293 complementary information, even when the input GWAS was conducted in a single ancestry
294 (20), we used METRO to model gene expression from EA and AA simultaneously. METRO
295 uses a joint-likelihood framework that accounts for SNP effect size heterogeneity and LD
296 differences across ancestries. The framework selectively upweights information from the
297 ancestry that has greater certainty in the gene expression prediction model, increasing power and
298 allowing characterization of the relative contribution of each ancestry to the TWAS results.

299 METRO is described in Li et al.(20) Briefly, each gene is examined separately using gene
300 expression data from M different genetic ancestries. $\boldsymbol{Z_m}$ is the n_m-vector of gene expression

301 measurements on n_m individuals in the mth ancestry with $m \in \{1,\ldots,M\}$. For the gene of interest,
302 all *cis*-SNPs (p), which are in potential linkage disequilibrium (LD) with each other, were
303 extracted as predictors for gene expression. $\boldsymbol{G_m}$ is denoted as the n_m *x* p genotype matrix for these
304 *cis*-SNPs. Besides the gene expression data, we also used GWAS summary statistics from n
305 individuals for an outcome trait of interest. $\boldsymbol{\gamma}$ is the n-vector of outcome measurements in the
306 GWAS data and $\boldsymbol{G}$ is the corresponding n *x* p genotype matrix on the same set of p *cis*-SNPs.
307 The expression vector $\boldsymbol{z_m}$, the outcome vector $\boldsymbol{\gamma}$, and each column of the genotype matrixes are
308 centered and standardized. $\boldsymbol{G_m}$ and $\boldsymbol{G}$ have a mean of zero and variance of one. For each TWAS,
309 we used GWAS summary statistics in the form of marginal z-scores and a SNP-SNP correlation
310 (LD) matrix estimated with genotype data from our GENOA sample that correspond with the
311 ancestry of the GWAS (EA or AA). The following equations describe the relationships between
312 the SNPs, gene expression and the outcome:

$$\boldsymbol{z}_m = \boldsymbol{G}_m \boldsymbol{\beta}_m + \boldsymbol{\epsilon}_m, \qquad \text{(Equation 1)}$$

313 $$\boldsymbol{\gamma} = \alpha(\boldsymbol{G}\boldsymbol{\beta}) + \boldsymbol{\epsilon}_\gamma, \qquad \text{(Equation 2)}$$

314 Equation (1) describes the relationship between gene expression and the *cis*-SNP
315 genotypes in the gene expression study in GENOA for the mth ancestry (EA or AA). β_m is a p
316 vector of the *cis*-SNP effects on the gene expression in the mth ancestry and ε_m is an n_m-vector of
317 residual errors with each element following an independent and normal distribution N(0, σ^2_m)
318 with an ancestry specific variance σ^2_m. Equation (2) describes the relationship between the
319 genetically regulated gene expression (GReX), calculated from estimated SNP prediction
320 weights, and the outcome trait (general cognitive function, WMH or AD) from the GWAS.
321 There, $G\beta$ denotes an n-vector of GReX constructed for the GWAS individuals, where $\beta = \Sigma_m$

322 $w_m\beta_m$ is a p-vector of SNP effects on the gene expression in the GWAS data, where the weights
323 $\Sigma^M_{m=1} w_m=1$ and $w_m \geq 0$. The alpha value (α) is the effect of GReX constructed for the GWAS
324 individuals on the outcome trait, and ε_y is an n_m-vector of residual errors with each element
325 following an independent and normal distribution N(0, σ^2_m). Both equations, specified based on
326 separate studies, are connected through the predictive SNP effects on the gene expression (β_m
327 and β). A key assumption made is that the SNP effects on the gene expression in the GWAS, β,
328 can be expressed as a weighted summation of the SNP effects on gene expression in the
329 expression studies conducted across ancestries.

330 We derived the overall GReX effect α and the contribution weight of each ancestry (w_1
331 for AA and w_2 for EA) to infer the extent and contribution of the two genetic ancestries in
332 informing the GReX-trait association. The joint model defined in Equations 1 and 2 allows us to
333 borrow association strength across multiple ancestries to enable powerful inference of GReX-
334 trait associations for general cognitive function, WMH and AD. We declared the gene to be
335 significant if the p-value was below the corresponding Bonferroni corrected threshold for the
336 number of tested genes ($P<0.05/17,238 = 2.90\times10^{-6}$). Manhattan plots and quantile-quantile (QQ)
337 plots were generated using the *qqman*(38) R package.

338

339 **2.3.2 Fine-mapping analysis**

340 Since genes residing in the same genomic region may share eQTLs or contain eQTL
341 SNPs in LD with each other, TWAS test statistics for genes in the same region can be highly
342 correlated, making it difficult to identify the true biologically relevant genes among them. To
343 prioritize the putatively causal genes identified by METRO for general cognitive function,
344 WMH, and AD, we conducted TWAS fine-mapping using FOCUS (Fine-mapping Of CaUsal

345 gene Sets) (37). To identify a genomic region with at least one significant gene detected by
346 METRO, we obtained a set of independent, non-overlapping genomic regions, or LD blocks,
347 using Ldetect (39). In each analyzed genomic block, using a standard Bayesian approach, we
348 assigned a posterior inclusion probability (PIP) for each gene to be causal, given the observed
349 TWAS statistics. We used gene-level Z scores, created from p-values using the inverse
350 cumulative distribution function (CDF) of a standard normal distribution, as input into FOCUS.
351 We then ranked the PIPs and computed the 90%-credible set that contains the causal gene with
352 90% probability. In the FOCUS analysis, a null model which assumes none of the genes in the
353 region are causally associated with the trait is also considered as a possible outcome and may be
354 included in the credible set. Through fine-mapping, we narrowed down significantly associated
355 genes identified by METRO to a shorter list of putatively true associations.
356
357 **2.3.3 Characterization of identified genes**
358 To interpret our TWAS findings, both before and after fine-mapping, we further
359 examined whether the genes identified by METRO overlapped with those previously identified
360 by their corresponding input GWAS. We created a set of Venn diagrams of overlapping genes
361 identified using METRO with those from the SNP-based GWAS association results(21–24)
362 mapped to the nearest gene using the *VennDiagram* R package (40). We then constructed a
363 second set of Venn diagrams showing overlapping genes identified using METRO with genes
364 identified by gene-based association analyses in each of the input GWAS studies. The gene-
365 based analyses were conducted using MAGMA(41) (general cognitive function,(21) WMH(22),
366 and AD (AA GWAS)(24)) or gene prioritization tests (AD (EA GWAS)(23)). Finally, we
367 created a set of Venn diagrams comparing genes identified using METRO with those identified

368 in the TWAS that were conducted as part of the WMH(22) and AD (EA GWAS)(23) input
369 GWAS studies. We used the *geneSynonym* R package(42) to ensure that genes named differently
370 across studies were captured.
371
372 **2.3.4 Functional Enrichment Analysis**
373 To characterize the biological function of the identified genes by METRO for general
374 cognitive function, WMH and AD, we performed gene set enrichment analysis. Specifically, we
375 used the g:GOSt(43) tool on the web software g:Profiler and mapped the genes to known
376 functional informational sources, including Gene Ontology (GO): molecular function (MF), GO:
377 biological process (BP), GO: cellular component (CC), Kyoto Encyclopedia of Genes and
378 Genomes (KEGG), Reactome (REAC), WikiPathways (WP), Transfac (TF), MiRTarBase
379 (MIRNA), Human Protein Atlas (HPA), CORUM protein complexes, and Human Phenotype
380 Ontology (HP). In this analysis, we used the default option g:SCS method (Set Counts and Sizes)
381 in g:Profiler for multiple testing correction and presented pathways identified with an adjusted p-
382 value < 0.05. Driver terms in GO are highlighted using a two-stage algorithm for filtering GO
383 enrichment results, providing a more efficient and reliable approach compared to traditional
384 clustering methods. This feature groups significant terms into sub-ontologies based on their
385 relations, and the second stage identifies leading gene sets that give rise to other significant
386 functions in the same group of terms. This method uses a greedy search strategy that recalculates
387 hypergeometric p-values and results in the consideration of multiple leading terms in a
388 component, rather than selection of terms with the highest significance level.
389
390 **Results**

391 In Table 1, we provide descriptive statistics for the samples used in the eQTL mapping
392 study (e.g., 1,032 AA and 801 EA from GENOA) and the four input GWAS.(21–23) The
393 GENOA eQTL study included participants with a mean age of 56.9 (SD=10.0) years. More than
394 half of participants were female (65.6%). Mean age of participants was 56.9 (SD=7.8) years in
395 the general cognitive function GWAS(21), and 64.2 years in the WMH GWAS.(22) In the AD
396 GWAS in EA (23), mean age was 73.6 (SD=8.1) years for cases and 67.9 (SD=8.6) years for
397 controls. In the AD GWAS in AA (24), mean age was 74.2 (SD=13.6) years for all participants.
398 Using METRO, we identified 602 genes associated with general cognitive function, 45
399 genes associated with WMH, 231 genes associated with AD (EA GWAS), and 9 genes
400 associated with AD (AA GWAS) that were significant at the Bonferroni corrected alpha level
401 ($P<2.90x10^{-6}$; Figure 1, Tables S1-3). Genomic inflation factors for the TWAS p-values ranged
402 from 1.00 to 2.55 (Figure 2). Among the three neurocognitive outcomes, prior to fine-mapping,
403 METRO TWAS identified the *ICA1L* gene overlapping between WMH and AD (from EA
404 GWAS); the *FMNL1* gene overlapping between WMH and general cognitive function; and 22
405 genes enriched in AD-related pathways and functions overlapping between general cognitive
406 function and AD (from EA GWAs) (Figure 3a; Figure S1). After fine-mapping, the only
407 overlapping gene that remained was *ICA1L* between WMH and AD (Figure 3b). The METRO
408 TWAS for AD (AA GWAS) identified 9 genes overlapping with those identified in AD (EA
409 GWAS); however, following fine-mapping, only *TOMM40* overlapped between the two AD
410 TWASs (Figure 3b).
411 For all identified genes, we also examined the contribution weights of expression
412 prediction models for the EA and AA ancestries, prior to fine-mapping ($P<2.90x10^{-6}$; Figure 4).
413 For the WMH TWAS, we found that a large proportion of genes had stronger contributions from

414 EA weights than AA weights (65.2%). This is consistent with Li et al. (2022)(20) who found that
415 the gene expression prediction models constructed in the same ancestry as the input GWAS, in
416 this case EA, often have larger contribution weights than those constructed in other ancestries.
417 However, for both general cognitive function and AD (EA and AA GWAS), the contributions
418 from EA and AA weights were similar, which likely increased power to identify genes relevant
419 to AA.

420 After fine-mapping, there were 266 genes in the 90%-credible set across 172 different
421 genomic regions for general cognitive function. This gene set included 82 genes that were not
422 previously identified in the SNP-based GWAS results (mapped to the nearest gene) or the gene-
423 based analysis results from Davies et al. (2018)(21) (Figure 5, Table S1); however, it is likely
424 that some of these genes are in broader genomic regions tagged by the GWAS-identified SNPs.
425 Specifically, there were 126 and 168 overlapping genes between METRO and the SNP-based
426 and gene-based associations from Davies et al. (2018)(21), respectively (Figure 5). The 266
427 METRO-identified genes were enriched in regulatory pathways involved in protein binding (p_{adj}
428 $= 1.17 \times 10^{-5}$), developmental cell growth ($p_{adj} = 3.33 \times 10^{-5}$), and protein metabolic process (p_{adj}
429 $= 7.18 \times 10^{-4}$), as well as neurodevelopmental processes such as neuron to neuron synapse ($p_{adj} =$
430 1.22×10^{-3}) and neuron projection ($p_{adj} = 7.14 \times 10^{-3}$; Figure 6). The 82 genes that were not
431 previously identified in Davies et al. (2018)(21) were enriched for positive regulation of
432 biological process ($p_{adj} = 1.77 \times 10^{-2}$), proteasome activator complex ($p_{adj} = 1.00 \times 10^{-2}$),
433 nucleoplasm ($p_{adj} = 1.29 \times 10^{-2}$) and chromatin ($p_{adj} = 4.71 \times 10^{-5}$; Figure S2).

434 After fine-mapping, there were 23 genes in the 90%-credible set across 15 genomic
435 regions for WMH, including 12 genes that were not previously identified in the SNP-based
436 GWAS results mapped to the nearest gene or the gene-based analysis results from

437 Sargurupremraj et al. (2020)(22) (Figure 7, Table S2). Specifically, there were 7 and 12
438 overlapping genes between METRO and the SNP-based and gene-based associations from
439 Sargurupremraj et al. (2020)(22), respectively (Figure 7). The 23 METRO-identified genes were
440 enriched for zinc finger motif ($p_{adj} = 1.27 \times 10^{-2}$), miRNA has-212-5p ($p_{adj} = 1.94 \times 10^{-2}$) and
441 retinal inner plexiform layer ($p_{adj} = 3.86 \times 10^{-2}$; Figure 8). The 12 genes associated with WMH
442 that were previously not identified by Sargurupremraj et al. (2020)(22) were enriched for DNA
443 binding domain Zinc Finger Protein 690 (ZNF690; $p_{adj} = 2.52 \times 10^{-3}$) and ClpX protein
444 degradation complex ($p_{adj} = 4.97 \times 10^{-2}$; Figure S3).

445 After fine-mapping, there were 69 genes in the 90%-credible set across 56 genomic
446 regions associated with AD (EA GWAS), including 45 genes that were not previously identified
447 in the SNP-based GWAS results mapped to the nearest gene or the gene prioritization analysis
448 results from Bellenguez et al. (2022)(23) (Figure 9, Table S3). Specifically, there were 16 and 14
449 overlapping genes between METRO and the SNP-based and gene prioritization test results from
450 Bellenguez et al. (2022)(23), respectively (Figure 9). The 69 METRO-identified genes were
451 enriched for AD-associated processes including regulation of amyloid fibril formation (p_{adj} =
452 1.87×10^{-3}), amyloid-beta clearance ($p_{adj} = 1.90 \times 10^{-3}$), microglial cell activation ($p_{adj} = 5.79 \times$
453 10^{-3}), amyloid-beta metabolic process ($p_{adj} = 1.07 \times 10^{-2}$), and neurofibrillary tangle ($p_{adj} = 2.80 \times$
454 10^{-4}; Figure 10). The 45 genes associated with AD that were previously not identified by
455 Bellenguez et al. (2022)(23) were enriched for hematopoietic cell lineage ($p_{adj} = 1.73 \times 10^{-3}$) and
456 neurofibrillary tangle ($p_{adj} = 9.13 \times 10^{-3}$; Figure S4).

457 We identified 2 genes, *APOE* and *PVRL2*, in the 90%-credible set associated with AD
458 (AA GWAS) (Table S4). After fine-mapping, none of these genes overlapped with SNP-based
459 GWAS results mapped to the nearest gene or the gene-based analysis results from Kunkle et al.

460 (2021), since they are both in the broader *APOE* region, which was the only identified gene in
461 Kunkle et al (2021). The 2 METRO-identified genes were enriched for coreceptor-mediated
462 virion attachment to host ($p_{adj} = 4.96 \times 10^{-2}$; Figure 11).
463 We compared the genes identified by METRO before and after fine-mapping with those
464 identified by TWAS studies in Sargurupremraj et al. (2020)(22) and Bellenguez et al. (2022)(23)
465 which used TWAS-Fusion (Figure 12). For WMH, there were 16 and 10 genes identified both by
466 METRO before and after fine-mapping and by the TWAS-Fusion analysis conducted by
467 Sargurupremraj et al. (2020)(22), respectively (Table 2). For AD, there were 24 and 10 genes
468 identified both by METRO before and after fine-mapping and by the TWAS-Fusion followed by
469 FOCUS fine-mapping analysis conducted by Bellenguez et al. (2022)(23) (Table 3). *ICA1L* was
470 the only gene overlapping between all four AD and WMH TWAS association results.
471
472 **Discussion**
473 While previous studies have identified genes associated with cognitive function, WMH,
474 and AD, there are few TWAS that utilize genetic and gene expression data from multiple
475 ancestries to elucidate gene-trait associations and molecular mechanisms underlying the
476 etiologies of cognitive function and neurocognitive disorders. Using the METRO method in
477 GWAS consisting primarily of EA followed by FOCUS fine-mapping, we identified 266, 23, and
478 69 genes associated with general cognitive function, WMH, and AD, respectively, with 82, 12
479 and 45 of them not previously identified in the original GWAS. In addition, using an AA
480 GWAS, we identified 2 fine-mapped genes associated with AD, both of which are proximal to
481 *APOE*. Studying the transcriptomic mechanisms underlying cognitive function, WMH and
482 dementia using both EA and AA expression data may enhance our understanding of cognitive

483 health prior to and following the onset of dementia and further allow us to generalize findings
484 from large scale EA GWAS to other ancestries.

485 AD and SVD have overlapping features that contribute to dementia neuropathology
486 including breakdown of the blood-brain barrier(44) and the presence of small cortical and
487 subcortical infarcts, microbleeds, perivascular spacing, and WMH in brain tissue(45). After fine-
488 mapping, Islet Cell Autoantigen 1 Like (*ICA1L*) was identified in both the WMH and AD
489 TWAS. This is as a highly plausible prioritized gene that is likely to modulate the metabolism of
490 amyloid precursor protein (APP)(23) and increase risk of AD. *ICA1L* encodes a protein whose
491 expression is activated by type IV collagen and plays a crucial role in myelination(46). Increased
492 *ICA1L* expression is also associated with lower risk of AD(47–49) and small vessel strokes
493 (SVS), the acute outcomes of cerebral SVD, which may lead to VaD(50). Consistent with these
494 studies, our TWAS found that decreased expression of *ICA1L* is associated with increased risk of
495 AD and WMH, a subclinical indicator of SVD. Single-cell RNA-sequencing has shown *ICA1L*
496 expression to be enriched in cortical glutamatergic excitatory neurons, which are crucial
497 components in neural development and neuropathology through their role in cell proliferation,
498 differentiation, survival, neural network formation and cell death(51,52). *ICA1L* has been
499 examined as a possible drug target for SVD, AD, and other neurodegenerative diseases(50,53);
500 however, it is not recommended as a prioritized drug at this time due to potential side effects
501 including increased risk of coronary artery disease and myocardial infarction as well as lower
502 diastolic blood pressure(53). Nevertheless, *ICA1L* may contribute to overlapping AD and VaD
503 neuropathology, and it could be a potential target for therapeutics and/or preventative treatments
504 for AD and VaD in the future if adverse events can be reduced.

505 Our TWAS of AD (from EA GWAS) identified 45 genes that were not identified in the
506 SNP-based GWAS results mapped to the nearest gene or the gene-based analysis reported in
507 Bellenguez et al. (2022)(23). The 45 genes were enriched for hematopoietic cell lineage, which
508 are progenitors of red and white blood cells including those related to immunity (e.g., natural
509 killer cells, T- and B- lymphocytes and other types of leukocytes)(54–60). Our TWAS identified
510 genes that have been previously associated with AD, including *APOE, TOMM40, APOC4, CLU,*
511 *PICALM* and *CR2,* among others(23,61,62). While we identified *APOE,* the strongest genetic
512 risk factor for AD in most populations, after fine-mapping, we did not identify *ABCA7* which
513 confers an equal or even greater risk for AD in AA(63−65). This finding is perhaps not
514 surprising considering that our TWAS was conducted using an EA GWAS, and the strength of
515 association between *ABCA7* and AD is comparatively weaker in EA than in AA.(65) To identify
516 genes associated with AD risk in AA populations, specifically, it would be beneficial to perform
517 a TWAS utilizing a well-powered AD GWAS in AA.
518 Our TWAS of AD (from AA GWAS) (24) identified only *TOMM40* and *PVRL2*, both
519 proximal to *APOE*. *TOMM40* was also identified in our TWAS of AD (from EA GWAS), as
520 well as other AD GWAS (23,61,62). *PVRL2* has been associated with metabolic syndrome,
521 diabetic dyslipidemia, and AD(66,67). One study found that polymorphisms in *PVRL2* interact
522 with variants in *TOMM40* to increase AD risk through pathways related to amyloid-beta
523 metabolism in older Chinese adults(67). As larger AD GWAS in AA become available, we may
524 be able to identify additional genes associated with disease while leveraging transcriptomic data
525 from EA and AA.
526 In our AD EA TWAS, we also identified genes associated with other neurological and
527 autoimmune diseases including Parkinson's disease (*CYB561*(68) and *SLC25A39*(69)), Crohn's

528 disease (*ATG16L1*(70)), Amyotrophic lateral sclerosis (*SIGLEC9*(71)), and Riboflavin Transport
529 Deficiency (*SLC52A1*(72)). These diseases have in common the progressive peripheral and
530 cranial degeneration of neurons that impact processes such as voluntary muscle movement,
531 vision, hearing and sensation. Although not explicitly identified in Bellenguez et al. (2022)(23),
532 we also identified genes that were associated with AD in other studies including *RIN3* that is
533 implicated in tau-mediated pathology, the *MS4A* (*4A* and *6A*) locus associated with mast cell
534 activation, *TP53INP1* and *ZYX* that have been linked to myeloid enhancer activity(73), and
535 *APOC4*, which is located proximal to *APOE*(74). We also identified additional genes involved in
536 B cell autoimmunity (*HLA-DQA2*(75,76) and *CSTF1*(77)), neurodegenerative processes
537 (*SUPT4H1*(78), *C6orf10*(79), *IKZF1*(80), and *DEDD*(81)), and neuronal growth (*IKZF1*(80) and
538 *STYX*(82)). Our findings support the hypothesis that chronic activation of immune cells resident
539 in the brain and peripheral nervous system appear to play a critical role in neuroinflammatory
540 responses that drive the progression of neurodegeneration in AD.(83) Further, consistent with
541 findings that AD and VaD often co-exist, our AD TWAS identified genes that were associated
542 with lacunar and ischemic strokes as well as cerebral small vessel disease in other studies,
543 including *SLC39A13*(84), *RAPSN*(84), *MAF1*(85), and *MME*(86,87).
544 Although our WMH TWAS identified 12 genes that were not included in the SNP-based
545 GWAS results mapped to the nearest gene or the gene-based analysis reported in Sargurupremraj
546 et al. (2020)(22), other studies found associations between *MAP1LC3B*(88), *ARMS2*(89,90) and
547 *HTRA1*(84) with ischemic stroke, lacunar stroke, and cerebral SVD. The WMH TWAS also
548 identified genes associated with AD (*ARMS2*)(91), atrial fibrillation (*NEURL*(92) and
549 *GJC1*(93)), innate immunity (*EFTUD2*(94)) and apoptosis and neurodevelopment (*PDCD7*(95),
550 *FBXO31*(96), and *ClpX*(97)). The 12 unique genes identified for WMH were enriched for DNA

551 binding domain Zinc Finger Protein 690(98), which plays an essential role in gene regulation,
552 transcription and various cellular processes, and ClpX protein degradation complex(99), which
553 maintains protein homeostasis. Our findings were consistent with studies that showed
554 neuroinflammation to be an immunological cascade reaction by glial cells of the central nervous
555 system where innate immunity resides.

556 While our TWAS for general cognitive function did not show overlapping genes between
557 the TWAS for AD and VaD, we identified genes associated with general cognitive function that
558 were not explicitly identified by Davies et al. (2018)(21) which were associated with pre-clinical
559 AD and VaD risk factors including cardiovascular diseases, immunity and Alzheimer's
560 neuropathology. Our TWAS also identified genes previously associated with cognitive domains,
561 neuropathology, and psychiatric illness including reading-related skills and neural structures
562 (*SEMA6D*(100) and *SETBP1*(101)), working memory tasks (*CDH13*(102)) and Schizophrenia
563 (*HP*(103,104)*, C18orf1*(105) and *TMEM180*(106)). There are likely also distinct transcriptomic
564 mechanisms that differentiate cognitive function and normal age-related brain changes from
565 pathways related to dementia. Individuals who never develop dementia or significant cognitive
566 decline still experience brain deterioration in normal aging that includes gray and white matter
567 loss and ventricular enlargement which is accompanied by memory decline(107). Further,
568 previous GWAS for general cognitive function and AD have shown few overlapping
569 loci(21,108). In addition, studies of older individuals who are cognitively "resilient" with intact
570 cognitive function, despite the presence of AD neuropathology, have found the genetic
571 architecture of cognitive resilience to be distinct from that of AD(109). As such, relatively little
572 is known about the pathways underlying cognitive aging in those without dementia. Thus,

573 studying transcriptomic mechanisms that affect general cognitive function before development of
574 dementia may shed light on cognitive aging without dementia.

575 We also compared genes identified by METRO after fine-mapping with those identified
576 by TWAS-Fusion in Sargurupremraj et al. (2020)(22) and Bellenguez et al. (2022)(23). Among
577 the 92 genes associated with WMH in Sargurupremraj et al. (2020)(22) and 23 genes identified
578 by METRO, 10 genes overlapped. We note that the Sargurupremraj et al. (2020)(22) did not
579 perform fine-mapping of their TWAS results, which is likely why we identified substantially
580 fewer genes. There were also 10 overlapping genes among the 66 genes associated with AD in
581 Bellenguez et al. (2022)(23) and 69 genes identified by METRO. For both TWAS comparisons,
582 a relatively small number of genes overlap likely due to differences in eQTL prediction
583 modeling. Sargurupremraj et al. (2020)(22) and Bellenguez et al. (2022)(23) used eQTL data
584 from brain tissue, while we used eQTL data from transformed beta lymphocytes in blood tissue.
585 While brain tissue is more relevant to WMH and AD phenotypes, blood cells do touch every cell
586 bed that affects the brain, and are related to chronic inflammation, immunity, and oxidative
587 stress, which are linked to cognitive performance and dementia. TWAS results from blood tissue
588 in multiple ancestries provide complementary information to those reported in the GWAS.

589 Several limitations in the present study should be noted. First, our gene expression levels
590 were measured using transformed B-lymphocytes from immortalized cell lines in GENOA.
591 While transformed B-lymphocytes are a convenient source of DNA from blood tissue, we lack
592 eQTL data for tissues that may be most relevant for AD and WMH (e.g., brain tissue, small brain
593 vessels, and microglia). However, B-lymphocytes provide a unique and efficient way to examine
594 the functional effects of genetic variations on gene expression that minimizes environmental
595 influences(110). Second, METRO follows the standard TWAS approach of analyzing one gene

596 at a time. Since genes residing in the same genomic region may share eQTLs or contain eQTL
597 SNPs that are in LD with each other, the TWAS test statistics of genes in the same region may be
598 highly correlated. To that end, it may be challenging to identify the truly biologically relevant
599 genes among them(37,111). As such, we paired METRO with FOCUS to allow us to narrow
600 down the list of potential causal genes for AD, VaD, and cognitive decline(37,112). Lastly, we
601 primarily utilized EA GWAS that were publicly available with large sample sizes for general
602 cognitive function, WMH, and AD. As expected, the gene expression prediction models
603 constructed in the same ancestry as the GWAS (EA) tended to have larger contribution weights
604 than AA. While we conducted a TWAS of AD in AA, the sample size of the AA GWAS likely
605 did not allow us to properly power our TWAS. As such, a future direction would be to conduct
606 TWAS of these traits using summary statistics from well-powered GWAS with AA ancestry or
607 multiple ancestries as they become available.

608 Our study also has notable strengths. To our knowledge, our study is the first TWAS
609 using expression mapping studies in multiple ancestries (EA and AA) to identify genes
610 associated with cognitive function and neurocognitive disorders. By leveraging the
611 complementary information in gene expression prediction models constructed in EA and AA, as
612 well as the uncertainty in SNP prediction weights, we were able to conduct a highly powered
613 TWAS to identify important gene-trait associations and transcriptomic mechanisms related to
614 innate immunity, vascular dysfunction and neuroinflammation underlying AD, VaD, and general
615 cognitive function. Using METRO, we were also able to estimate the ancestry contribution
616 weights for specific genes and identify the extent to which a gene in EA or AA may contribute to
617 the trait. However, it is noteworthy that the larger the contribution of the expression prediction
618 models in the same ancestry as the GWAS (primarily EA, in this study) may allow for better

predictive performance in the same ancestry. We also conducted FOCUS fine-mapping to narrow in on a list of putatively causal genes among multiple significant genes in a region. Our results suggest that there are similar pathways that contribute to healthy cognitive aging and progression of dementia, as well as distinct pathways that are unique to each neuropathology. By understanding overlapping and unique genes and transcriptomic mechanisms associated with each outcome, we may identify possible targets for prevention and/or treatments for cognitive aging and dementia.

Conclusion

In the present study, we conducted a multi-ancestry TWAS in EA and AA to identify genes associated with general cognitive function, WMH and AD. We identified genes associated with innate immunity, vascular dysfunction, and neuroinflammation. The WMH and AD TWAS also indicated that downregulation of *ICA1L* may contribute to overlapping AD and VaD neuropathology. To our knowledge, this study is the first TWAS analysis using expression mapping studies in multiple ancestries to identify genes associated with cognitive function and neurocognitive disorders, which may help to identify gene targets for pharmaceutical or preventative treatment for dementia.

Author Contributions: Conceptualization, D.L.C., J.A.S.; methodology, D.L.C., J.A.S., X.Z.; software, formal analysis, and investigation, D.L.C., Z.L., L.S., S.M.R.; resources, T.H.M. and S.L.R.K.; writing—original draft preparation, D.L.C., J.A.S.; writing—review and editing, D.L.C., S.M.R., W.Z., T.H.M., S.L.R.K., X.Z., J.A.S.; visualization, D.L.C., Z.L.; supervision, J.A.S.; funding acquisition, S.L.R.K. and J.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: Support for the Genetic Epidemiology Network of Arteriopathy (GENOA) was provided by the National Heart, Lung and Blood Institute (NHLBI, U01HL054457, RC1HL100185, R01HL087660, R01HL119443, R01HL133221) and the National Institute of Neurological Disorders and Stroke (NINDS, R01NS041558) of the NIH.

Informed Consent Statement: Written informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The phenotype data and *APOE* genotypes used in the current study are available upon reasonable request to J.A.S. and S.L.R.K., and with a completed data use agreement (DUA). All other genotype data are available from the Database of Genotypes and Phenotypes (dbGaP): phs001401.v2.p1. Methylation and gene expression data are available from the Gene Expression Omnibus (GEO): GSE210256 and GSE138914. Due to IRB restriction, mapping of the sample IDs between genotype data (dbGaP) and methylation data (GEO) cannot be provided publicly, but is available upon written request to J.A.S. and S.L.R.K.

Acknowledgments: The authors wish to thank the staff and participants of the GENOA study.

Conflicts of Interest: The authors declare no conflict of interest.

663

Table 1. Sample characteristics of expression quantitative trait locus (eQTL) mapping study and genome-wide association studies (GWAS) participants

eQTL mapping study: Genetic Epidemiology Network of Arteriopathy (GENOA)	
	Mean (SD) or N (%) or N
N	N=1833
Age (years)	56.85 (10.0)
Female	1202 (65.6%)
Race/Ethnicity	
African Americans	1032 (56.3%)
European Americans	801 (43.7%)

General cognitive function GWAS: CHARGE, COGENT, UKB[a] (Davies et al., 2018)	
	Mean (SD) or N (%) or N
N	300,486
Age (years)	56.91 (7.8)
Female	52.20%
Excluded for dementia and/or stroke diagnosis	N=4919

White matter hyperintensity (WMH) GWAS: CHARGE and UKB[a] (Sargurupremraj et al., 2020)	
	Mean (SD) or N (%) or N
N	48,454
Age (years)	64.17
Female	29215 (57.6%)
WMH volume (cm^3)	7.06 (8.8)
Excluded for stroke or pathologies	N=1572

EA Alzheimer's Disease (AD) GWAS: EADB, GR@ACE, EADI, GERAD/PERADES, DemGene, Bonn, the Rotterdam study, the CCHS study, NxC and the UKB[a] (Bellenguez et al., 2022)	
	Mean (SD) or N (%) or N
Discovery sample	
AD cases	N=39,106
AD proxy cases	N=46,828
Controls	N=401,577
Age (years)	
AD cases or proxy cases	73.55 (8.1)
controls	67.86 (8.6)
Female	
AD cases or proxy cases	62.90%
controls	56.10%

AA Alzheimer's Disease (AD) GWAS: AD Genetics Consortium[b] (Kunkle et al., 2021)	
	Mean (SD) or N (%) or N
N	7,970
AD cases	2,748 (34.5%)
controls	5,222 (65.5%)
Age (years)	74.2 (13.6)
Female	
AD cases	1,944 (69.8%)
controls	3,743 (71.7%)

664 Abbreviations: EA, European ancestry; AA, African ancestry
665 [a] GWAS include only European ancestry participants.
666 [b] GWAS includes only African ancestry participants.
667

668

Table 2. Genes for WMH identified both by METRO followed by fine-mapping with FOCUS and by TWAS-Fusion conducted by Sargurupremraj et al. (2020)

Gene	ENSG	Chr	Start	End	Gene Name	Accession number
CALCRL	ENSG00000064989	2	188206691	188313187	calcitonin receptor like receptor	Source: HGNC Symbol; Acc: HGNC:16709
DCAKD	ENSG00000172992	17	43100706	43138499	dephospho-CoA kinase domain containing	Source: HGNC Symbol; Acc: HGNC: 26238
EFEMP1	ENSG00000115380	2	56093102	56151274	EGF containing fibulin extracellular matrix protein 1	Source: HGNC Symbol; Acc: HGNC: 3218
GJC1	ENSG00000182963	17	42875816	42908184	gap junction protein gamma 1	Source: HGNC Symbol; Acc: HGNC: 4280
ICA1L	ENSG00000163596	2	203637873	203736489	islet cell autoantigen 1 like	Source: HGNC Symbol; Acc: HGNC: 14442
KLHL24	ENSG00000114796	3	183353398	183402307	kelch like family member 24	Source: HGNC Symbol; Acc: HGNC: 25947
NBEAL1	ENSG00000144426	2	203879331	204091101	neurobeachin like 1	Source: HGNC Symbol; Acc: HGNC: 20681
NEURL	ENSG00000107954	10	105253462	105352303	neuralized E3 ubiquitin protein ligase 1	Source: HGNC Symbol; Acc: HGNC: 7761
NMT1	ENSG00000136448	17	43035360	43186384	N-myristoyltransferase 1	Source: HGNC Symbol; Acc: HGNC: 7857
WBP2	ENSG00000132471	17	73841780	73852588	WW domain binding protein 2	Source: HGNC Symbol; Acc: HGNC: 12738

669 Abbreviations: HGNC, Human Genome Organisation Gene Nomenclature Committee

670

671 **Table 3. Genes for AD identified both by METRO followed by fine-mapping with FOCUS and by**
672 **TWAS-Fusion followed by fine-mapping with FOCUS conducted by Bellenguez et al. (2022)**

Gene	ENSG	Chr	Start	End	Gene Name	Accession number
BLNK	ENSG00000095585	10	97948927	98031344	B cell linker	Source:HGNC Symbol;Acc:HGNC:14211
CPSF3	ENSG00000119203	2	9563780	9613230	cleavage and polyadenylation specific factor 3	Source:HGNC Symbol;Acc:HGNC:2326
DDX54	ENSG00000123064	12	113594978	113623284	DEAD-box helicase 54	Source:HGNC Symbol;Acc:HGNC:20084
GRN	ENSG00000030582	17	42422614	42430474	granulin precursor	Source:HGNC Symbol;Acc:HGNC:4601
ICA1L	ENSG00000163596	2	203637873	203736489	islet cell autoantigen 1 like	Source:HGNC Symbol;Acc:HGNC:14442
KLF16	ENSG00000129911	19	1852398	1863578	KLF transcription factor 16	Source:HGNC Symbol;Acc:HGNC:16857
LACTB	ENSG00000103642	15	63414032	63434260	lactamase beta	Source:HGNC Symbol;Acc:HGNC:16468
PPP4C	ENSG00000149923	16	30087299	30096697	protein phosphatase 4 catalytic subunit	Source:HGNC Symbol;Acc:HGNC:9319
SHARPIN	ENSG00000179526	8	145153536	145163027	SHANK associated RH domain interactor	Source:HGNC Symbol;Acc:HGNC:25321
TBX6	ENSG00000149922	16	30097114	30103245	T-box transcription factor 6	Source:HGNC Symbol;Acc:HGNC:11605

673 Abbreviations: HGNC, Human Genome Organisation Gene Nomenclature Committee
674
675

676
677 **References**
678

679 1. 2023 Alzheimer's disease facts and figures. Alzheimer's & Dementia. 2023 Apr;19(4):1598–
680 695.

681 2. De Reuck J, Maurage CA, Deramecourt V, Pasquier F, Cordonnier C, Leys D, et al. Aging
682 and cerebrovascular lesions in pure and in mixed neurodegenerative and vascular dementia
683 brains: a neuropathological study. Folia Neuropathol. 2018;56(2):81–7.

684 3. Hof PR, Glannakopoulos P, Bouras C. The neuropathological changes associated with
685 normal brain aging. Histol Histopathol. 1996 Oct;11(4):1075–88.

686 4. Perl DP. Neuropathology of Alzheimer's Disease. Mt Sinai J Med. 2010;77(1):32–42.

687 5. Cohuet G, Struijker-Boudier H. Mechanisms of target organ damage caused by hypertension:
688 Therapeutic potential. Pharmacology & Therapeutics. 2006 Jul 1;111(1):81–98.

689 6. Weller J, Budson A. Current understanding of Alzheimer's disease diagnosis and treatment.
690 F1000Res. 2018 Jul 31;7:F1000 Faculty Rev-1161.

691 7. Hubbard BM, Fenton GW, Anderson JM. A quantitative histological study of early clinical
692 and preclinical Alzheimer's disease. Neuropathol Appl Neurobiol. 1990 Apr;16(2):111–21.

693 8. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining
694 the preclinical stages of Alzheimer's disease: recommendations from the National Institute
695 on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's
696 disease. Alzheimers Dement. 2011 May;7(3):280–92.

697 9. Mayeda ER, Glymour MM, Quesenberry CP, Whitmer RA. Inequalities in dementia
698 incidence between six racial and ethnic groups over 14 years. Alzheimers Dement. 2016
699 Mar;12(3):216–24.

700 10. Barnes LL, Bennett DA. Alzheimer's disease in African Americans: risk factors and
701 challenges for the future. Health Aff (Millwood). 2014 Apr;33(4):580–6.

702 11. Brewster P, Barnes L, Haan M, Johnson JK, Manly JJ, Nápoles AM, et al. Progress and
703 future challenges in aging and diversity research in the United States. Alzheimers Dement.
704 2019 Jul;15(7):995–1003.

705 12. Rajan KB, Weuve J, Barnes LL, Wilson RS, Evans DA. Prevalence and incidence of
706 clinically diagnosed Alzheimer's disease dementia from 1994 to 2012 in a population
707 study. Alzheimers Dement. 2019 Jan;15(1):1–7.

708 13. Cacabelos R, Torrellas C. Epigenetics of Aging and Alzheimer's Disease: Implications for
709 Pharmacogenomics and Drug Response. International Journal of Molecular Sciences. 2015
710 Dec;16(12):30483–543.

711 14. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-
712 wide association studies. Nat Rev Genet. 2019 Aug;20(8):467–84.

713 15. Lappalainen T, Sammeth M, Friedländer MR, ‘t Hoen PA, Monlong J, Rivas MA, et al.
714 Transcriptome and genome sequencing uncovers functional variation in humans. Nature.
715 2013 Sep 26;501(7468):506–11.

716 16. Mogil LS, Andaleon A, Badalamenti A, Dickinson SP, Guo X, Rotter JI, et al. Genetic
717 architecture of gene expression traits across diverse populations. PLOS Genetics. 2018 Aug
718 10;14(8):e1007586.

719 17. Wojcik G, Graff M, Nishimura K, Tao R, Haessler J, Gignoux C, et al. Genetic analyses of
720 diverse populations improves discovery for complex traits. Nature. 2019;570(7762):514–8.

721 18. Shang L, Smith JA, Zhao W, Kho M, Turner ST, Mosley TH, et al. Genetic Architecture of
722 Gene Expression in European and African Americans: An eQTL Mapping Study in
723 GENOA. Am J Hum Genet. 2020 Apr 2;106(4):496–512.

724 19. Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. Philos Trans
725 R Soc Lond B Biol Sci. 2013 Jun 19;368(1620):20120362.

726 20. Li Z, Zhao W, Shang L, Mosley TH, Kardia SLR, Smith JA, et al. METRO: Multi-ancestry
727 transcriptome-wide association studies for powerful gene-trait association detection. The
728 American Journal of Human Genetics. 2022 May 5;109(5):783–801.

729 21. Davies G, Lam M, Harris SE, Trampush JW, Luciano M, Hill WD, et al. Study of 300,486
730 individuals identifies 148 independent genetic loci influencing general cognitive function.
731 Nat Commun. 2018 May 29;9(1):2098.

732 22. Sargurupremraj M, Suzuki H, Jian X, Sarnowski C, Evans TE, Bis JC, et al. Cerebral small
733 vessel disease genomics and its implications across the lifespan. Nature Communications.
734 2020 Dec 8;11(1):6285.

735 23. Bellenguez C, Küçükali F, Jansen IE, Kleineidam L, Moreno-Grau S, Amin N, et al. New
736 insights into the genetic etiology of Alzheimer’s disease and related dementias. Nat Genet.
737 2022 Apr 4;1–25.

738 24. Kunkle BW, Schmidt M, Klein HU, Naj AC, Hamilton-Nelson KL, Larson EB, et al. Novel
739 Alzheimer Disease Risk Loci and Pathways in African American Individuals Using the
740 African Genome Resources Panel: A Meta-analysis. JAMA Neurol. 2021 Jan 1;78(1):102–
741 13.

742 25. Daniels PR, Kardia SLR, Hanis CL, Brown CA, Hutchinson R, Boerwinkle E, et al. Familial
743 aggregation of hypertension treatment and control in the Genetic Epidemiology Network of
744 Arteriopathy (GENOA) study. Am J Med. 2004 May 15;116(10):676–81.

745 26. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and
746 population genetic studies. Nat Methods. 2013 Jan;10(1):5–6.

747 27. Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method
748 for the Next Generation of Genome-Wide Association Studies. PLOS Genetics. 2009 Jun
749 19;5(6):e1000529.

750 28. Gogarten SM, Sofer T, Chen H, Yu C, Brody JA, Thornton TA, et al. Genetic association
751 testing using the GENESIS R/Bioconductor package. Bioinformatics. 2019;35(24):5346–8.

752 29. Lockstone HE. Exon array data analysis using Affymetrix power tools and R statistical
753 software. Briefings in Bioinformatics. 2011 Nov 1;12(6):634–44.

754 30. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix
755 GeneChip probe level data. Nucleic Acids Research. 2003 Feb 15;31(4):e15.

756 31. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, et al. Evolving gene/transcript
757 definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Research.
758 2005 Nov 1;33(20):e175.

759 32. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using
760 empirical Bayes methods. Biostatistics. 2007 Jan 1;8(1):118–27.

761 33. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al.
762 GENCODE: the reference human genome annotation for The ENCODE Project. Genome
763 research. 2012;22(9):1760–74.

764 34. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for
765 large-scale transcriptome-wide association studies. Nat Genet. 2016 Mar;48(3):245–52.

766 35. Barbeira AN, Bonazzola R, Gamazon ER, Liang Y, Park Y, Kim-Hellmuth S, et al.
767 Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. Genome
768 Biology. 2021 Jan 26;22(1):49.

769 36. THE GTEX CONSORTIUM. The GTEx Consortium atlas of genetic regulatory effects
770 across human tissues. Science. 2020 Sep 11;369(6509):1318–30.

771 37. Mancuso N, Freund MK, Johnson R, Shi H, Kichaev G, Gusev A, et al. Probabilistic fine-
772 mapping of transcriptome-wide association studies. Nature genetics. 2019;51(4):675–82.

773 38. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan
774 plots [Internet]. bioRxiv; 2014 [cited 2023 Jun 6]. p. 005165. Available from:
775 https://www.biorxiv.org/content/10.1101/005165v1

776 39. Berisa T, Pickrell JK. Approximately independent linkage disequilibrium blocks in human
777 populations. Bioinformatics. 2016;32(2):283.

778 40. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable
779 Venn and Euler diagrams in R. BMC Bioinformatics. 2011 Jan 26;12(1):35.

780 41. de Leeuw CA, Stringer S, Dekkers IA, Heskes T, Posthuma D. Conditional and interaction
781 gene-set analysis reveals novel functional pathways for blood pressure. Nat Commun. 2018
782 Sep 14;9(1):3768.

783 42. Mancarcı BO. Gene Synonym [Internet]. 2022 [cited 2023 Jul 6]. Available from:
784 https://github.com/oganm/geneSynonym

785 43. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web
786 server for functional enrichment analysis and conversions of gene lists (2019 update).
787 Nucleic Acids Research. 2019 Jul 2;47(W1):W191–8.

788 44. Chagnot A, Barnes SR, Montagne A. Magnetic Resonance Imaging of Blood–Brain Barrier
789 permeability in Dementia. Neuroscience. 2021 Oct 15;474:14–29.

790 45. Kalaria RN, Sepulveda-Falla D. Cerebral Small Vessel Disease in Sporadic and Familial
791 Alzheimer Disease. The American Journal of Pathology. 2021 Nov 1;191(11):1888–905.

792 46. Zhang C, Qin F, Li X, Du X, Li T. Identification of novel proteins for lacunar stroke by
793 integrating genome-wide association data and human brain proteomes. BMC Medicine.
794 2022 Jun 23;20(1):211.

795 47. Bai B, Vanderwall D, Li Y, Wang X, Poudel S, Wang H, et al. Proteomic landscape of
796 Alzheimer's Disease: novel insights into pathogenesis and biomarker discovery. Mol
797 Neurodegener. 2021 Aug 12;16:55.

798 48. Wingo AP, Liu Y, Gerasimov ES, Gockley J, Logsdon BA, Duong DM, et al. Integrating
799 human brain proteomes with genome-wide association data implicates new proteins in
800 Alzheimer's disease pathogenesis. Nat Genet. 2021 Feb;53(2):143–6.

801 49. Ou YN, Yang YX, Deng YT, Zhang C, Hu H, Wu BS, et al. Identification of novel drug
802 targets for Alzheimer's disease by integrating genetics and proteomes from brain and blood.
803 Mol Psychiatry. 2021 Oct;26(10):6065–73.

804 50. Cullell N, Gallego-Fábrega C, Cárcel-Márquez J, Muiño E, Llucià-Carol L, Lledós M, et al.
805 ICA1L Is Associated with Small Vessel Disease: A Proteome-Wide Association Study in
806 Small Vessel Stroke and Intracerebral Haemorrhage. Int J Mol Sci. 2022 Mar
807 15;23(6):3161.

808 51. Liu DZ, Sharp FR. Excitatory and Mitogenic Signaling in Cell Death, Blood–brain Barrier
809 Breakdown, and BBB Repair after Intracerebral Hemorrhage. Transl Stroke Res. 2012 Jul
810 1;3(1):62–9.

811 52. Lai TW, Zhang S, Wang YT. Excitotoxicity and stroke: Identifying novel targets for
812 neuroprotection. Progress in Neurobiology. 2014 Apr 1;115:157–88.

813 53. Ge YJ, Ou YN, Deng YT, Wu BS, Yang L, Zhang YR, et al. Prioritization of Drug Targets
814 for Neurodegenerative Diseases by Integrating Genetic and Proteomic Data From Brain and
815 Blood. Biological Psychiatry. 2023 May 1;93(9):770–9.

816 54. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating
817 viruses and cellular organisms. Nucleic Acids Research. 2021 Jan 8;49(D1):D545–51.

818 55. De Bruijn MF, Speck NA. Core-binding factors in hematopoiesis and immune function.
819 Oncogene. 2004 May 24;23(24):4238–48.

820 56. Maillard I, Adler SH, Pear WS. Notch and the Immune System. Immunity. 2003
821 Dec;19(6):781–91.

822 57. Radtke F, Wilson A, Mancini SJC, MacDonald HR. Notch regulation of lymphocyte
823 development and function. Nat Immunol. 2004 Mar;5(3):247–53.

824 58. Rothenberg EV, Taghon T. MOLECULAR GENETICS OF T CELL DEVELOPMENT.
825 Annu Rev Immunol. 2005 Apr 1;23(1):601–49.

826 59. Shapiro-Shelef M, Calame K. Regulation of plasma-cell development. Nat Rev Immunol.
827 2005 Mar;5(3):230–42.

828 60. Spits H. Development of αβ T cells in the human thymus. Nat Rev Immunol. 2002
829 Oct;2(10):760–72.

830 61. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-
831 analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease.
832 Nat Genet. 2013 Dec;45(12):1452–8.

833 62. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM, et al. Common
834 variants in ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with
835 Alzheimer's disease. Nat Genet. 2011 May;43(5):429–35.

836 63. Reitz C, Mayeux R. Genetics of Alzheimer's Disease in Caribbean Hispanic and African
837 American Populations. Biological Psychiatry. 2014 Apr;75(7):534–41.

838 64. Reitz C, Jun G, Naj A, Rajbhandary R. Variants in the ATP-Binding Cassette Transporter
839 (ABCA7), Apolipoprotein E ϵ4,and the Risk of Late-Onset Alzheimer Disease in African
840 Americans. JAMA. 2013 Apr 10;309(14):1483–92.

841 65. De Roeck A, Van Broeckhoven C, Sleegers K. The role of ABCA7 in Alzheimer's disease:
842 evidence from genomics, transcriptomics and methylomics. Acta Neuropathol. 2019 Aug
843 1;138(2):201–20.

844 66. Guardiola M, Muntané G, Martínez I, Martorell L, Girona J, Ibarretxe D, et al. Metabolic
845 Overlap between Alzheimer's Disease and Metabolic Syndrome Identifies the PVRL2 Gene
846 as a New Modulator of Diabetic Dyslipidemia. Int J Mol Sci. 2023 Apr 18;24(8):7415.

847 67. Liang X, Liu C, Liu K, Cong L, Wang Y, Liu R, et al. Association and interaction of
848 TOMM40 and PVRL2 with plasma amyloid-β and Alzheimer's disease among Chinese
849 older adults: a population-based study. Neurobiol Aging. 2022 May;113:143–51.

850 68. Chatterjee P, Roy D, Bhattacharyya M, Bandyopadhyay S. Biological networks in
851 Parkinson's disease: an insight into the epigenetic mechanisms associated with this disease.
852 BMC Genomics. 2017 Sep 12;18(1):721.

853 69. Gialluisi A, Reccia MG, Modugno N, Nutile T, Lombardi A, Di Giovannantonio LG, et al.
854 Identification of sixteen novel candidate genes for late onset Parkinson's disease. Mol
855 Neurodegener. 2021 Jun 21;16(1):35.

856 70. Salem M, Ammitzboell M, Nys K, Seidelin JB, Nielsen OH. ATG16L1: A multifunctional
857 susceptibility factor in Crohn disease. Autophagy. 2015 Apr 23;11(4):585–94.

858 71. Hansson O, Kumar A, Janelidze S, Stomrud E, Insel PS, Blennow K, et al. The genetic
859 regulation of protein expression in cerebrospinal fluid. EMBO Mol Med. 2023 Jan
860 11;15(1):e16359.

861 72. Cali E, Dominik N, Manole A, Houlden H. Riboflavin Transporter Deficiency. In: Adam
862 MP, Mirzaa GM, Pagon RA, Wallace SE, Bean LJ, Gripp KW, et al., editors.
863 GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993 [cited
864 2023 Jul 31]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK299312/

865 73. Novikova G, Kapoor M, Tcw J, Abud EM, Efthymiou AG, Chen SX, et al. Integration of
866 Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory
867 elements and genes. Nat Commun. 2021 Mar 12;12(1):1610.

868 74. Allan CM, Walker D, Segrest JP, Taylor JM. Identification and characterization of a new
869 human gene (APOC4) in the apolipoprotein E, C-I, and C-II gene locus. Genomics. 1995
870 Jul 20;28(2):291–300.

871 75. Wang P, Luo M, Zhou W, Jin X, Xu Z, Yan S, et al. Global Characterization of Peripheral B
872 Cells in Parkinson's Disease by Single-Cell RNA and BCR Sequencing. Front Immunol.
873 2022;13:814239.

874 76. Lp Y, Mj S. The cryptic HLA-DQA2 ("DX alpha") gene is expressed in human B cell lines.
875 Journal of immunology (Baltimore, Md : 1950) [Internet]. 1991 Dec 15 [cited 2023 Aug
876 4];147(12). Available from: https://pubmed.ncbi.nlm.nih.gov/1753107/

877 77. Li GJ, Yang QH, Yang GK, Yang G, Hou Y, Hou LJ, et al. MiR-125b and SATB1-AS1
878 might be shear stress-mediated therapeutic targets. Gene. 2023 Mar 20;857:147181.

879 78. Naguib A, Sandmann T, Yi F, Watts RJ, Lewcock JW, Dowdle WE. SUPT4H1 Depletion
880 Leads to a Global Reduction in RNA. Cell Rep. 2019 Jan 2;26(1):45-53.e4.

881 79. Greaves CV, Rohrer JD. An update on genetic frontotemporal dementia. J Neurol. 2019
882 Aug;266(8):2075–86.

883 80. Ballasch I, García-García E, Vila C, Pérez-González A, Sancho-Balsells A, Fernández J, et
884 al. Ikzf1 as a novel regulator of microglial homeostasis in inflammation and
885 neurodegeneration. Brain Behav Immun. 2023 Mar;109:144–61.

886 81. Schutte B, Henfling M, Ramaekers FCS. DEDD association with cytokeratin filaments
887 correlates with sensitivity to apoptosis. Apoptosis. 2006 Sep;11(9):1561–72.

888 82. Dahal A, Hinton SD. Antagonistic roles for STYX pseudophosphatases in neurite outgrowth.
889 Biochem Soc Trans. 2017 Apr 15;45(2):381–7.

890 83. Burgaletto C, Munafò A, Di Benedetto G, De Francisci C, Caraci F, Di Mauro R, et al. The
891 immune system on the TRAIL of Alzheimer's disease. Journal of Neuroinflammation. 2020
892 Oct 13;17(1):298.

893 84. Traylor M, Persyn E, Tomppo L, Klasson S, Abedi V, Bakker MK, et al. Genetic basis of
894 lacunar stroke: a pooled analysis of individual patient data and genome-wide association
895 studies. Lancet Neurol. 2021 May;20(5):351–61.

896 85. Chen K, Zhu L, Guo L, Pan YB, Feng DF. Maf1 regulates dendritic morphogenesis and
897 influences learning and memory. Cell Death Dis. 2020 Jul 30;11(7):606.

898 86. Kumar AA, Yeo N, Whittaker M, Attra P, Barrick TR, Bridges LR, et al. Vascular Collagen
899 Type-IV in Hypertension and Cerebral Small Vessel Disease. Stroke. 2022
900 Dec;53(12):3696–705.

901 87. Miners S, van Helmond Z, Barker R, Passmore PA, Johnston JA, Todd S, et al. Genetic
902 variation in MME in relation to neprilysin protein and enzyme activity, Aβ levels, and
903 Alzheimer's disease risk. Int J Mol Epidemiol Genet. 2012;3(1):30–8.

904 88. Chen G, Li L, Tao H. Bioinformatics Identification of Ferroptosis-Related Biomarkers and
905 Therapeutic Compounds in Ischemic Stroke. Front Neurol. 2021;12:745240.

906 89. Pan Y, Fu Y, Baird PN, Guymer RH, Das T, Iwata T. Exploring the contribution of ARMS2
907 and HTRA1 genetic risk factors in age-related macular degeneration. Prog Retin Eye Res.
908 2022 Dec 27;101159.

909 90. Thee EF, Colijn JM, Cougnard-Grégoire A, Meester-Smoor MA, Verzijden T, Hoyng CB, et
910 al. The Phenotypic Course of Age-Related Macular Degeneration for ARMS2/HTRA1: The
911 EYE-RISK Consortium. Ophthalmology. 2022 Jul;129(7):752–64.

912 91. Gatta LB, Vitali M, Zanola A, Venturelli E, Fenoglio C, Galimberti D, et al. Polymorphisms
913 in the LOC387715/ARMS2 putative gene and the risk for Alzheimer's disease. Dement
914 Geriatr Cogn Disord. 2008;26(2):169–74.

915 92. Wang P, Qin W, Wang P, Huang Y, Liu Y, Zhang R, et al. Genomic Variants in NEURL,
916 GJA1 and CUX2 Significantly Increase Genetic Susceptibility to Atrial Fibrillation. Sci
917 Rep. 2018 Feb 19;8(1):3297.

918 93. Li RG, Xu YJ, Ye WG, Li YJ, Chen H, Qiu XB, et al. Connexin45 (GJC1) loss-of-function
919 mutation contributes to familial atrial fibrillation and conduction disease. Heart Rhythm.
920 2021 May;18(5):684–93.

921 94. Zhu C, Xiao F, Lin W. EFTUD2 on innate immunity. Oncotarget. 2015 Oct 20;6(32):32313–
922 4.

923 95. Abo Elwafa R, Gamaleldin M, Ghallab O. The clinical and prognostic significance of FIS1,
924 SPI1, PDCD7 and Ang2 expression levels in acute myeloid leukemia. Cancer Genet. 2019
925 Apr;233–234:84–95.

926 96. Vadhvani M, Schwedhelm-Domeyer N, Mukherjee C, Stegmüller J. The centrosomal E3
927 ubiquitin ligase FBXO31-SCF regulates neuronal morphogenesis and migration. PLoS One.
928 2013;8(2):e57530.

929 97. Serpeloni F, Nätt D, Assis SG de, Wieling E, Elbert T. Experiencing community and
930 domestic violence is associated with epigenetic changes in DNA methylation of BDNF and
931 CLPX in adolescents. Psychophysiology. 2020;57(1):e13382.

932 98. ZSCAN29 zinc finger and SCAN domain containing 29 [Homo sapiens (human)] - Gene -
933 NCBI [Internet]. [cited 2023 Aug 1]. Available from:
934 https://www.ncbi.nlm.nih.gov/gene/146050#summary

935 99. Baker TA, Sauer RT. ClpXP, an ATP-powered unfolding and protein-degradation machine.
936 Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2012 Jan 1;1823(1):15–
937 28.

938 100. Thomas T, Perdue MV, Khalaf S, Landi N, Hoeft F, Pugh K, et al. Neuroimaging genetic
939 associations between SEMA6D, brain structure, and reading skills. J Clin Exp
940 Neuropsychol. 2021 Apr;43(3):276–89.

941 101. Perdue MV, Mascheretti S, Kornilov SA, Jasińska KK, Ryherd K, Einar Mencl W, et al.
942 Common variation within the SETBP1 gene is associated with reading-related skills and
943 patterns of functional neural activation. Neuropsychologia. 2019 Jul;130:44–51.

944 102. Ziegler GC, Ehlis AC, Weber H, Vitale MR, Zöller JEM, Ku HP, et al. A Common CDH13
945 Variant Is Associated with Low Agreeableness and Neural Responses to Working Memory
946 Tasks in ADHD. Genes. 2021 Sep;12(9):1356.

947 103. Trubetskoy V, Pardiñas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB, et al.
948 Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature.
949 2022 Apr;604(7906):502–8.

950 104. Beeri MS, Lin HM, Sano M, Ravona-Springer R, Liu X, Bendlin BB, et al. Association of
951 the Haptoglobin Gene Polymorphism With Cognitive Function and Decline in Elderly
952 African American Adults With Type 2 Diabetes: Findings From the Action to Control
953 Cardiovascular Risk in Diabetes-Memory in Diabetes (ACCORD-MIND) Study. JAMA
954 Netw Open. 2018 Nov 2;1(7):e184458.

955 105. Kikuchi M, Yamada K, Toyota T, Yoshikawa T. C18orf1 located on chromosome 18p11.2
956 may confer susceptibility to schizophrenia. J Med Dent Sci. 2003 Sep;50(3):225–9.

957 106. Zhang Z, Shang J, Dai Z, Yao Y, Shi Y, Zhong D, et al. Transmembrane Protein 170B is a
958 Prognostic Biomarker and Associated With Immune Infiltrates in Pancreatic
959 Adenocarcinoma. Front Genet. 2022;13:848391.

960 107. Ge Y, Grossman RI, Babb JS, Rabin ML, Mannon LJ, Kolson DL. Age-Related Total Gray
961 Matter and White Matter Changes in Normal Adult Brain. Part I: Volumetric MR Imaging
962 Analysis. American Journal of Neuroradiology. 2002 Sep 1;23(8):1327–33.

963 108. Davies G, Armstrong N, Bis JC, Bressler J, Chouraki V, Giddaluru S, et al. Genetic
964 contributions to variation in general cognitive function: a meta-analysis of genome-wide
965 association studies in the CHARGE consortium (N=53 949). Mol Psychiatry. 2015
966 Feb;20(2):183–92.

967 109. Dumitrescu L, Mahoney ER, Mukherjee S, Lee ML, Bush WS, Engelman CD, et al.
968 Genetic variants and functional pathways associated with resilience to Alzheimer's disease.
969 Brain. 2020 Aug 25;143(8):2561–75.

970 110. Fridman AL, Tainsky MA. Critical pathways in cellular senescence and immortalization
971 revealed by gene expression profiling. Oncogene. 2008 Oct;27(46):5975–87.

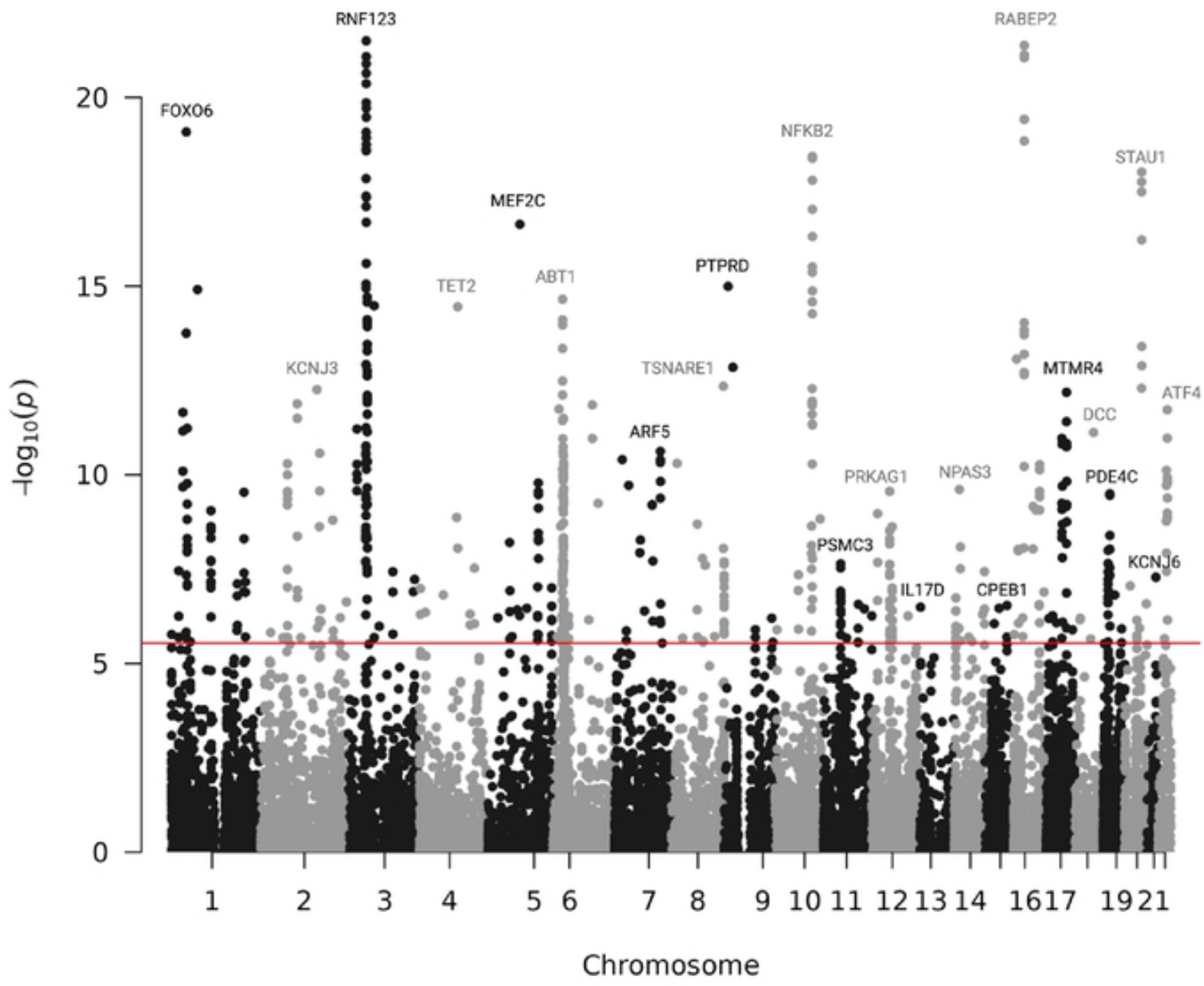
972 111. Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, et
973 al. Opportunities and challenges for transcriptome-wide association studies. Nat Genet.
974 2019 Apr;51(4):592–9.

975 112. Wu C, Pan W. A powerful fine-mapping method for transcriptome-wide association
976 studies. Hum Genet. 2020 Feb;139(2):199–213.

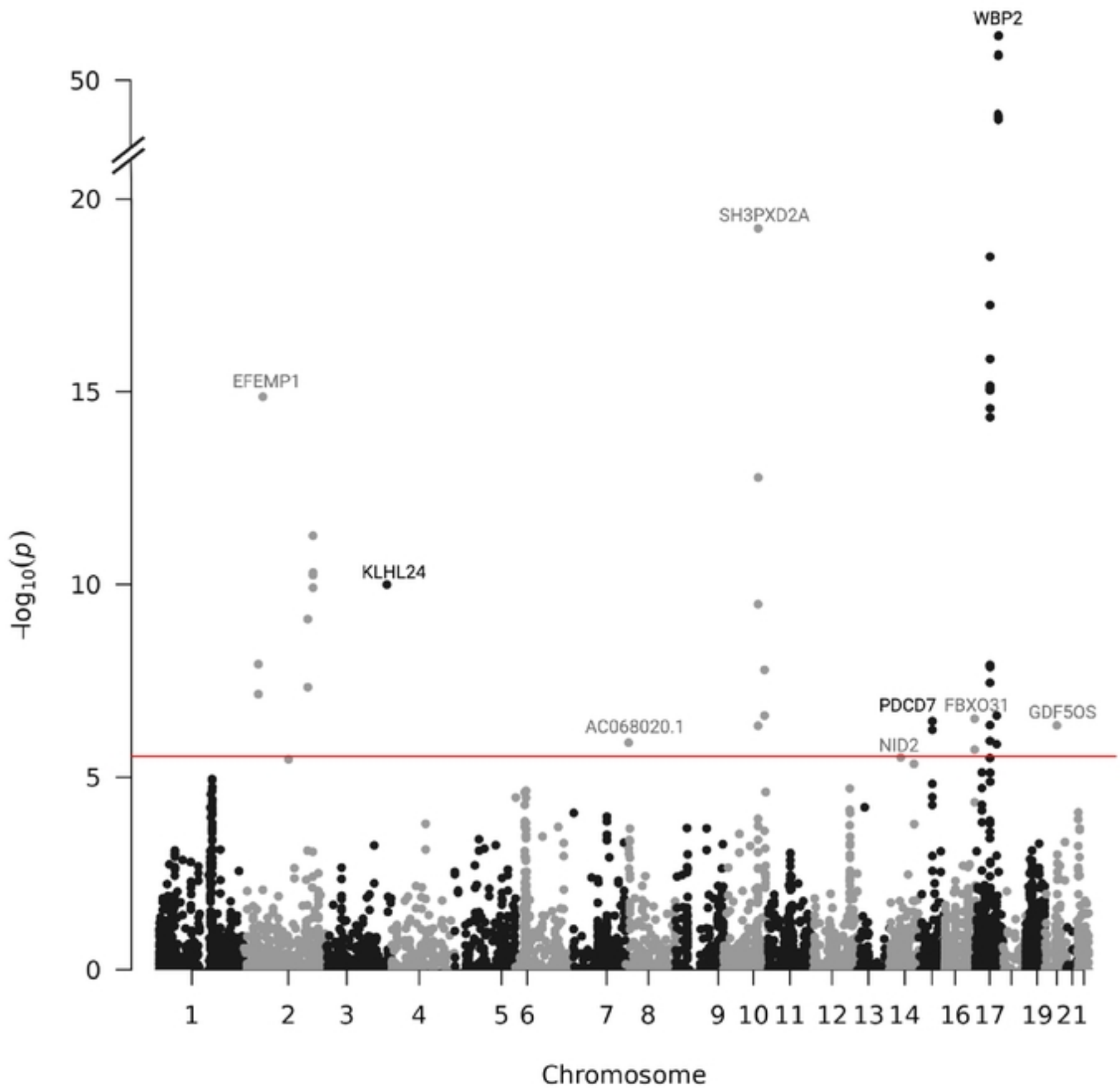
977

978

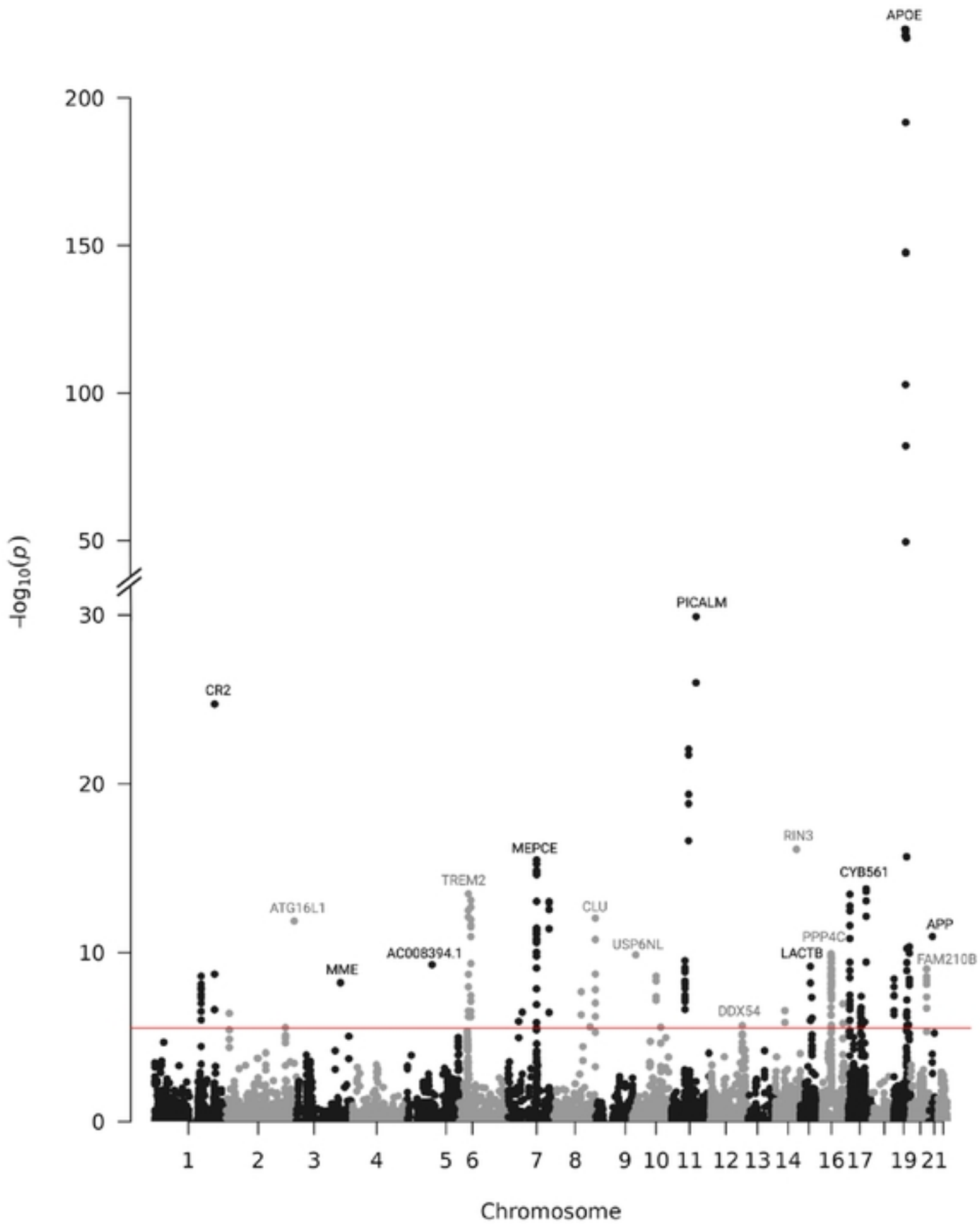
(a) General cognitive function from Davies et al. (2018) GWAS



(b) White matter hyperintensity from Sargurupremraj et al. (2020) GWAS



(c) Alzheimer's Disease from Bellenguez et al. (2022) (EA GWAS sample)



(d) Alzheimer's Disease from Kunkle et al. (2021) (AA GWAS sample)

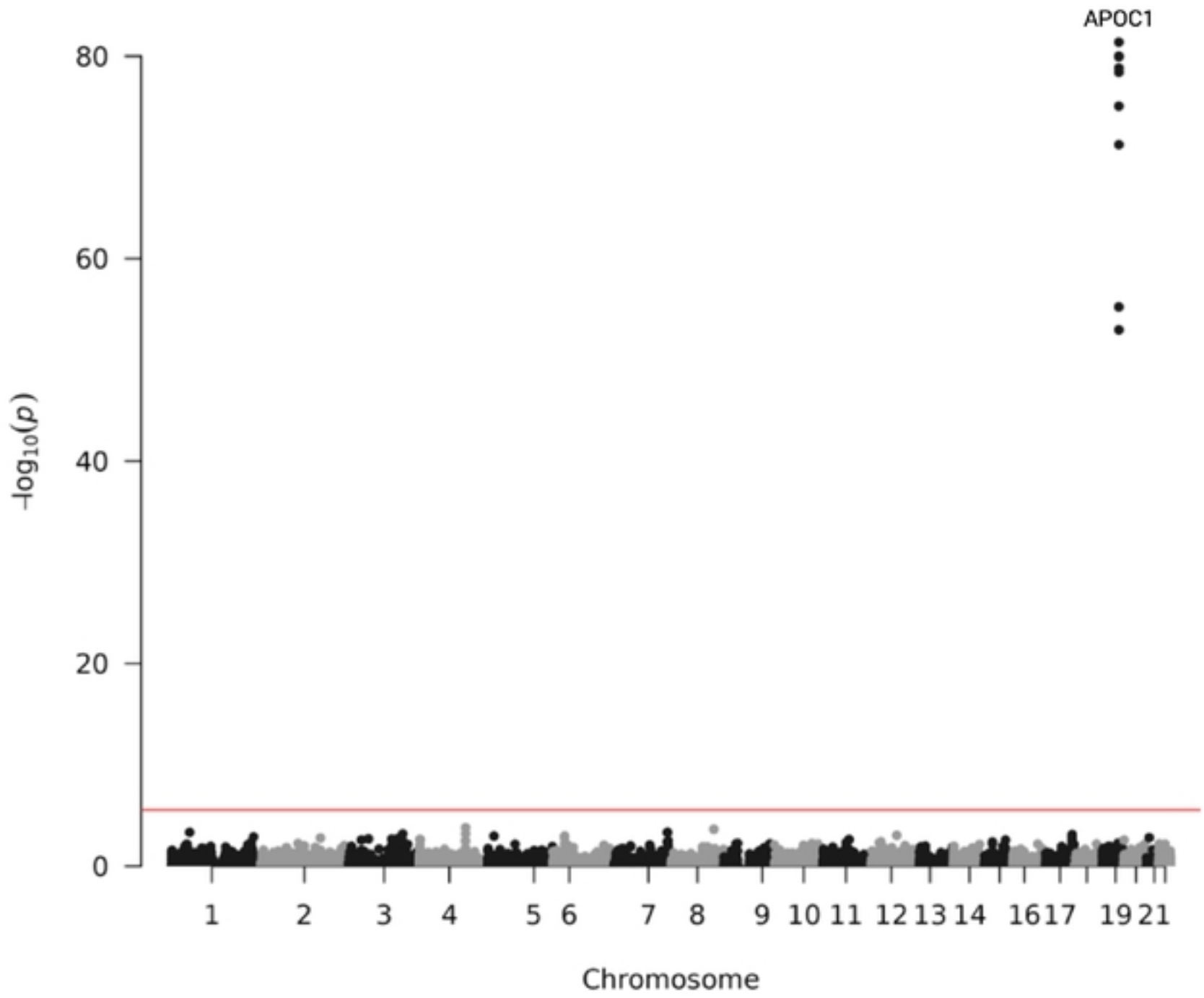


Figure 1. Manhattan plots of -$\log_{10}$ p-values for gene-trait associations in METRO. Manhattan plots of the association between genes and (a) general cognitive function using summary statistics from Davies et al. (2018), (b) White matter hyperintensity from Sargurupremraj et al. (2020), (c) Alzheimer's disease from Bellenguez et al. (2022) (EA GWAS sample) and (d) Alzheimer's disease from Kunkle et al. (2021) (AA GWAS sample) using GENOA gene expression data. The red line indicates significance after Bonferroni correction ($P<2.90\times10^{-6}$).

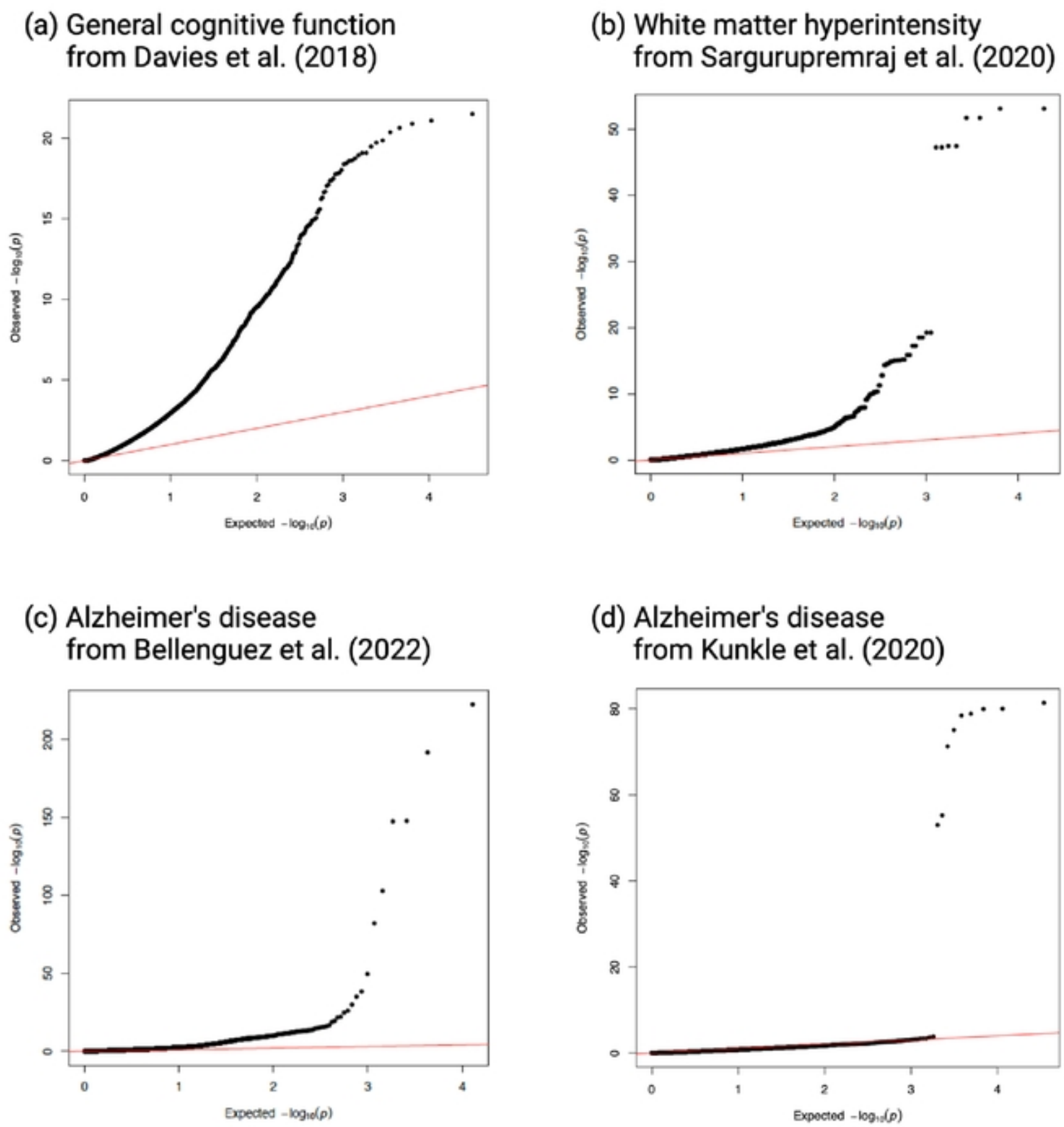
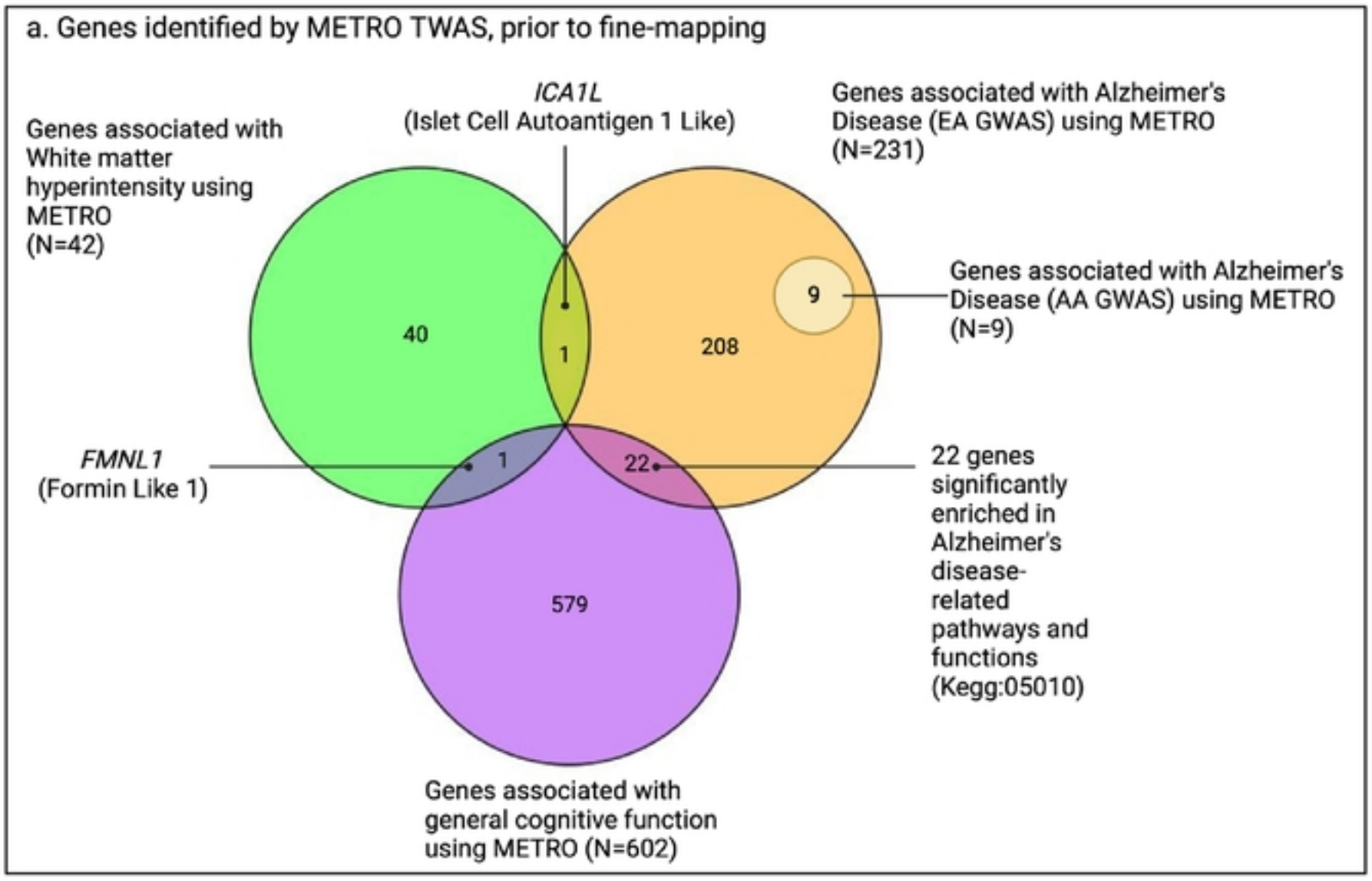


Figure 2. Quantile-quantile plots of -$\log_{10}$ p-values for gene-trait associations in METRO. Q-Q plots of the associations between genes and (a) general cognitive function (λ= 2.55) using summary statistics from Davies et al. (2018), (b) white matter hyperintensity (λ= 1.45) from Sargurupremraj et al. (2020), (c) Alzheimer's disease (λ= 2.09) from Bellenguez et al. (2022) (EA GWAS sample) and (d) Alzheimer's disease (λ= 1.0) from Kunkle et al. (2021)(AA GWAS sample) using GENOA gene expression data.


a. Genes identified by METRO TWAS, prior to fine-mapping
ICA1L
(Islet Cell Autoantigen 1 Like)
Genes associated with White matter hyperintensity using METRO (N=42)
Genes associated with Alzheimer's Disease (EA GWAS) using METRO (N=231)
Genes associated with Alzheimer's Disease (AA GWAS) using METRO (N=9)
40
1
208
9
FMNL1
(Formin Like 1)
1
22
22 genes significantly enriched in Alzheimer's disease-related pathways and functions (Kegg:05010)
579
Genes associated with general cognitive function using METRO (N=602)

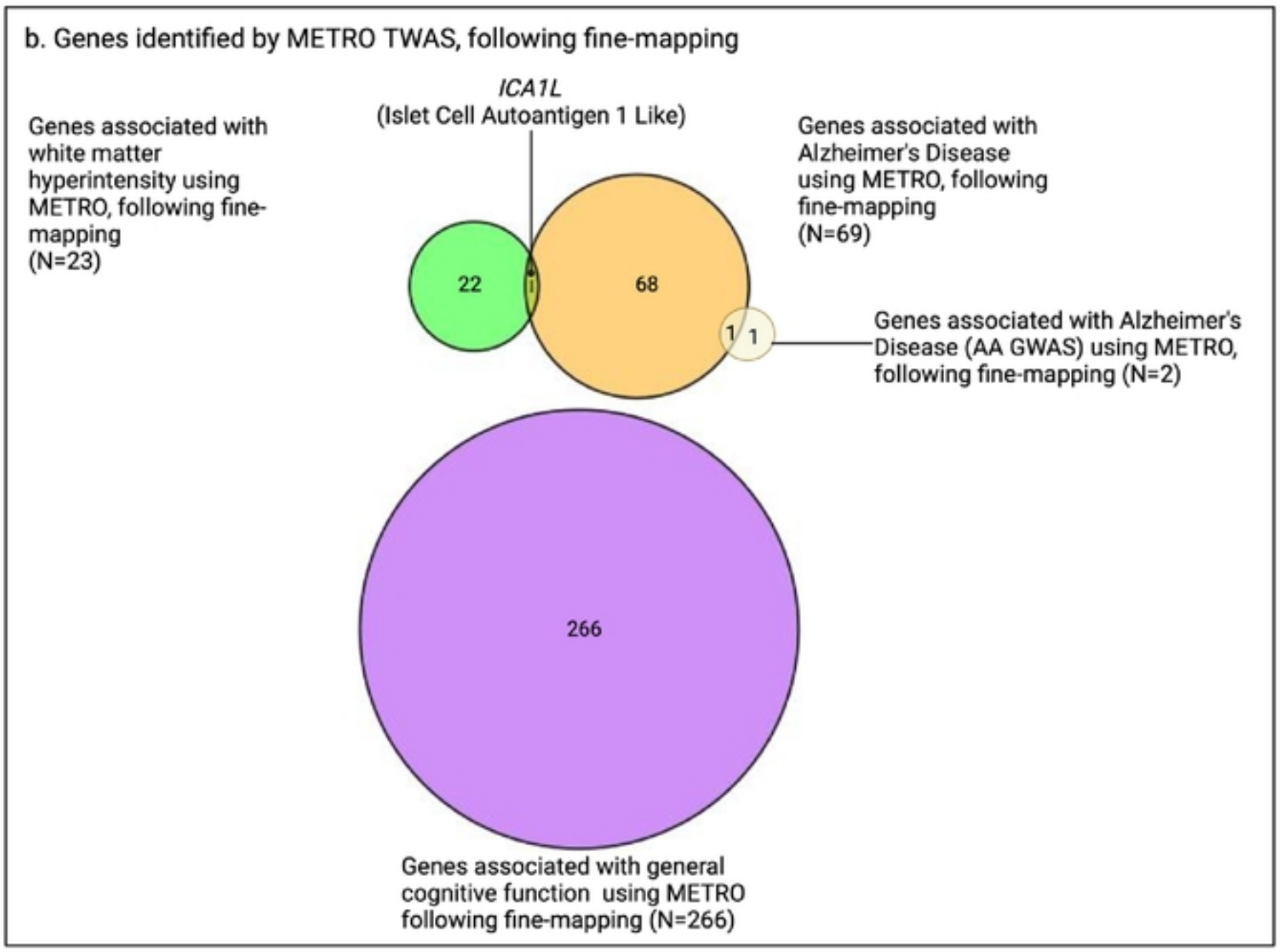

b. Genes identified by METRO TWAS, following fine-mapping
ICA1L
(Islet Cell Autoantigen 1 Like)
Genes associated with white matter hyperintensity using METRO, following fine-mapping (N=23)
Genes associated with Alzheimer's Disease using METRO, following fine-mapping (N=69)
22
1
68
1
1
Genes associated with Alzheimer's Disease (AA GWAS) using METRO, following fine-mapping (N=2)
266
Genes associated with general cognitive function using METRO following fine-mapping (N=266)

Figure 3. Venn diagrams comparing number of genes associated with general cognitive function, white matter hyperintensity and Alzheimer's disease (AD) in European ancestry (EA) and AD in African ancestry (AA) using METRO, prior to and following FOCUS fine-mapping.

Venn diagrams comparing the number of genes associated with general cognitive function (purple; N=266 genes), white matter hyperintensity (WMH; green; N=23 genes) Alzheimer's disease (AD) in EA (orange; N=69 genes), and AD in AA (yellow; N=2 genes), (a) prior to fine-mapping and (b) following FOCUS(5) fine-mapping using METRO and GENOA expression data after Bonferroni correction ($P<2.90\times10^{-6}$), with GWAS summary statistics obtained from the Davies et al. (2018), Sargurupremraj et al. (2020), Bellenguez et al. (2022), and Kunkle et al. (2021).

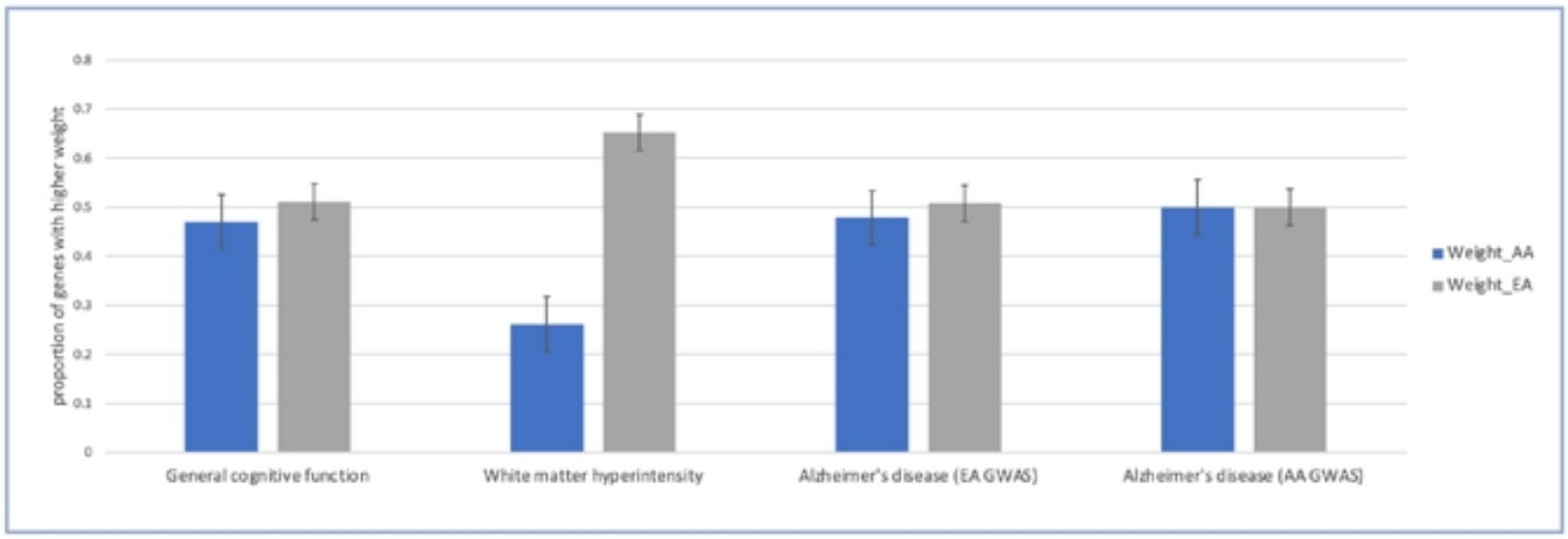


Figure 4. Contribution weights of expression prediction models across all significant fine-mapped genes identified by METRO.
Barplots of general cognitive function, white matter hyperintensity and Alzheimer's disease (AD) in European ancestry (EA) and AD in African ancestry (AA) comparing the proportion of significant genes with higher contribution weights of expression prediction models across all significant genes ($P<2.90\times10^{-6}$). Black bars are the standard errors for the estimated proportions.

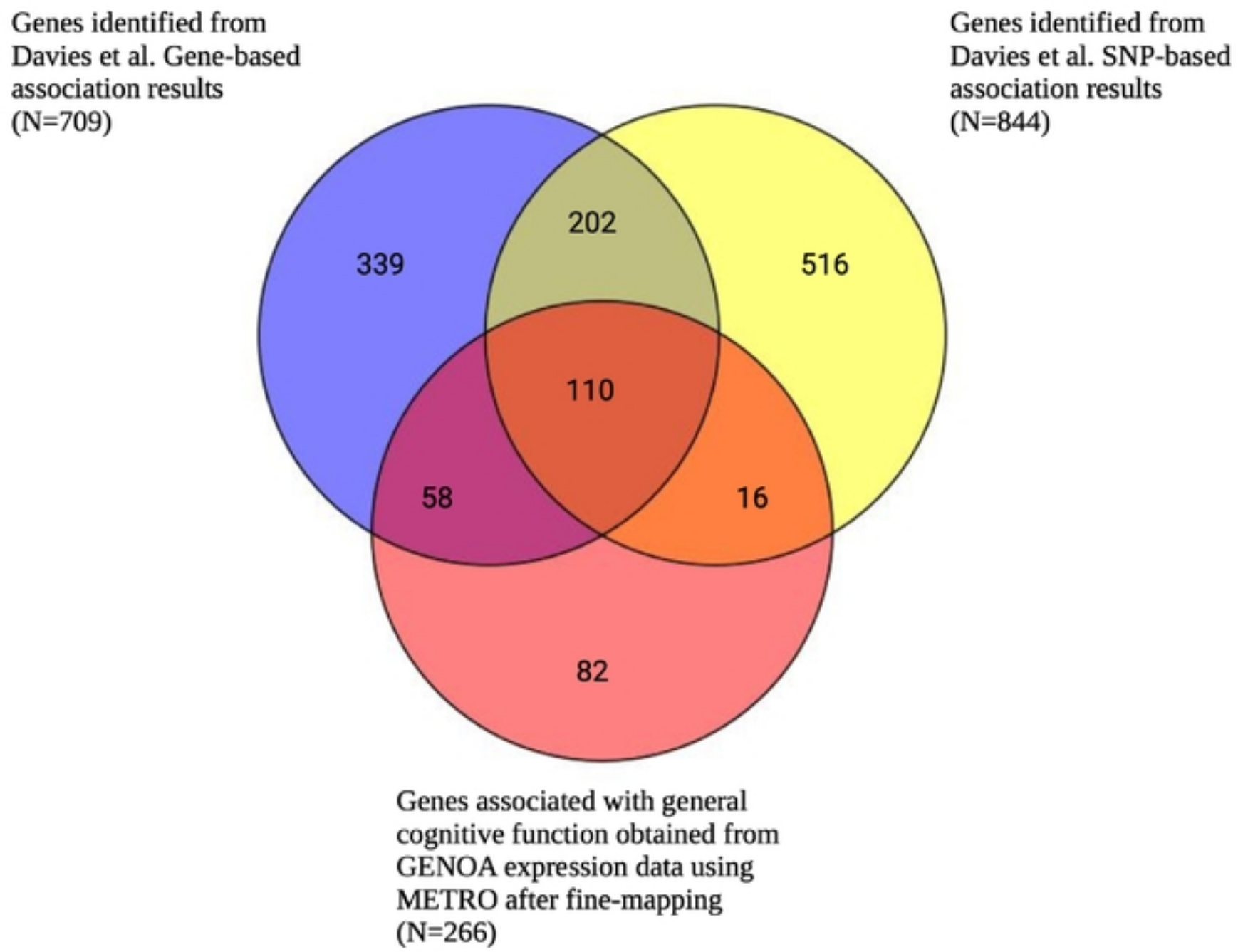
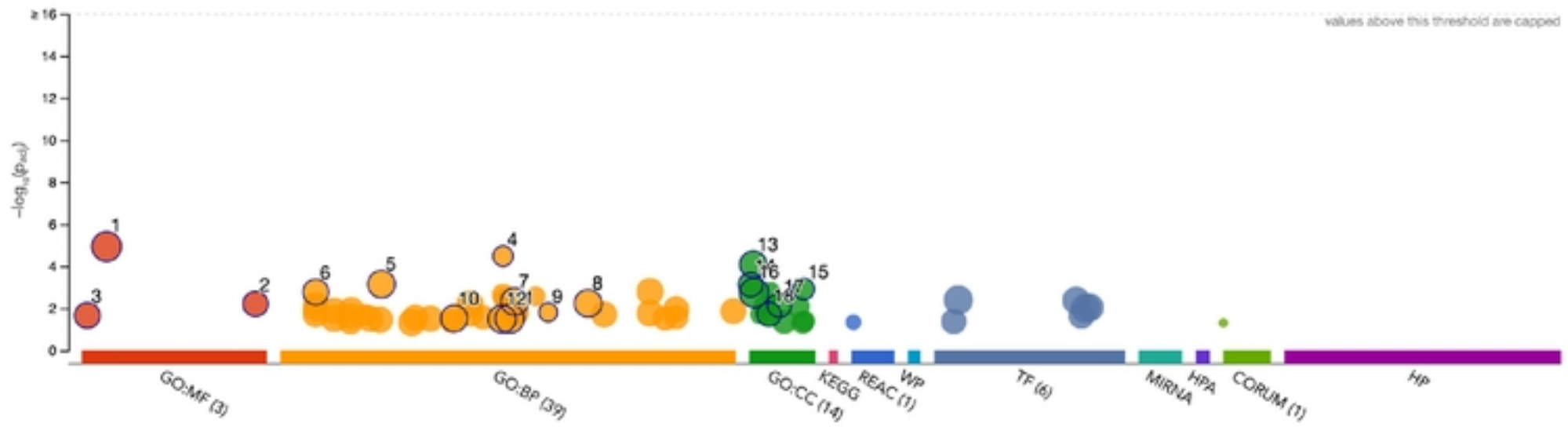


Figure 5. Venn diagram comparing number of METRO-identified genes associated with general cognitive function following FOCUS fine-mapping and genes identified by Davies et al. (2018) gene-based and SNP-based analyses.
Venn diagram comparing the number of genes associated with general cognitive function obtained from METRO using GENOA gene expression data after Bonferroni correction ($P<2.90x10^{-6}$) and FOCUS fine-mapping (red) and Davies et al. (2018). Davies et al. results included SNP-based association results that were mapped to the nearest gene ($P<5x10^{-8}$; yellow), and gene-based association results ($P<2.75x10^{-6}$; blue).



ID	Term ID	Term Name	Adjusted P-value
		GO:MF	
1	GO:0005515	protein binding	1.17E-05
2	GO:0140110	Transcription regulator activity	6.23E-03
3	GO:0003677	DNA binding	2.23E-02
		GO:BP	
4	GO:0048588	developmental cell growth	3.33E-05
5	GO:0019538	Protein metabolic process	7.18E-04
6	GO:0006357	Regulation of transcription by RNA polymerase II	4.79E-03
7	GO:0051171	Regulation of nitrogen compound metabolic processes	4.79E-03
8	GO:0080090	Regulation of primary metabolic process	5.86E-03
9	GO:0061387	Regulation of extent of cell growth	1.56E-02
10	GO:0042221	Response to chemical	3.06E-02
11	GO:0050794	Regulation of cellular process	3.13E-02
12	GO:0048518	Positive regulation of biological process	3.24E-02
		GO:CC	
13	GO:0005654	nucleoplasm	8.23E-05
14	GO:0000785	chromatin	7.56E-04
15	GO:0098984	Neuron to neuron synapse	1.22E-03
16	GO:0005737	cytoplasm	1.84E-03
17	GO:0043005	neuron projection	7.14E-03
18	GO:0031967	organelle envelope	1.96E-02

Figure 6. Functional enrichment analysis on the fine-mapped gene set identified for general cognitive function using METRO TWAS (N=266 genes). The top panel consists of a Manhattan plot that illustrates the enrichment analysis results. The x-axis represents functional terms that are grouped and color-coded by data sources, including Gene Ontology (GO): molecular function (MF; red), GO: biological process (BP; orange), GO: cellular component (CC; dark green), Kyoto Encyclopedia of Genes and Genomes (KEGG; pink), Reactome (REAC; dark blue), WikiPathways (WP; turquoise), Transfac (TF; light blue), MiRTarBase (MIRNA; emerald green), Human Protein Atlas (HPA; dark purple), CORUM protein complexes (light green), and Human Phenotype Ontology (HP; violet), in order from left to right. The y-axis shows the adjusted enriched -$\log_{10}$ p-values <0.05. Multiple testing correction was performed using g:SCS method (Set Counts and Sizes) that takes into account overlapping terms. The top panel highlights driver GO terms identified using the greedy filtering algorithm in g:Profiler. The light circles represent terms that were not significant after filtering. The circle sizes are in accordance with the corresponding term size (i.e., larger terms have larger circles). The number in parentheses following the source name in the x-axis shows how many significantly enriched terms were from this source.

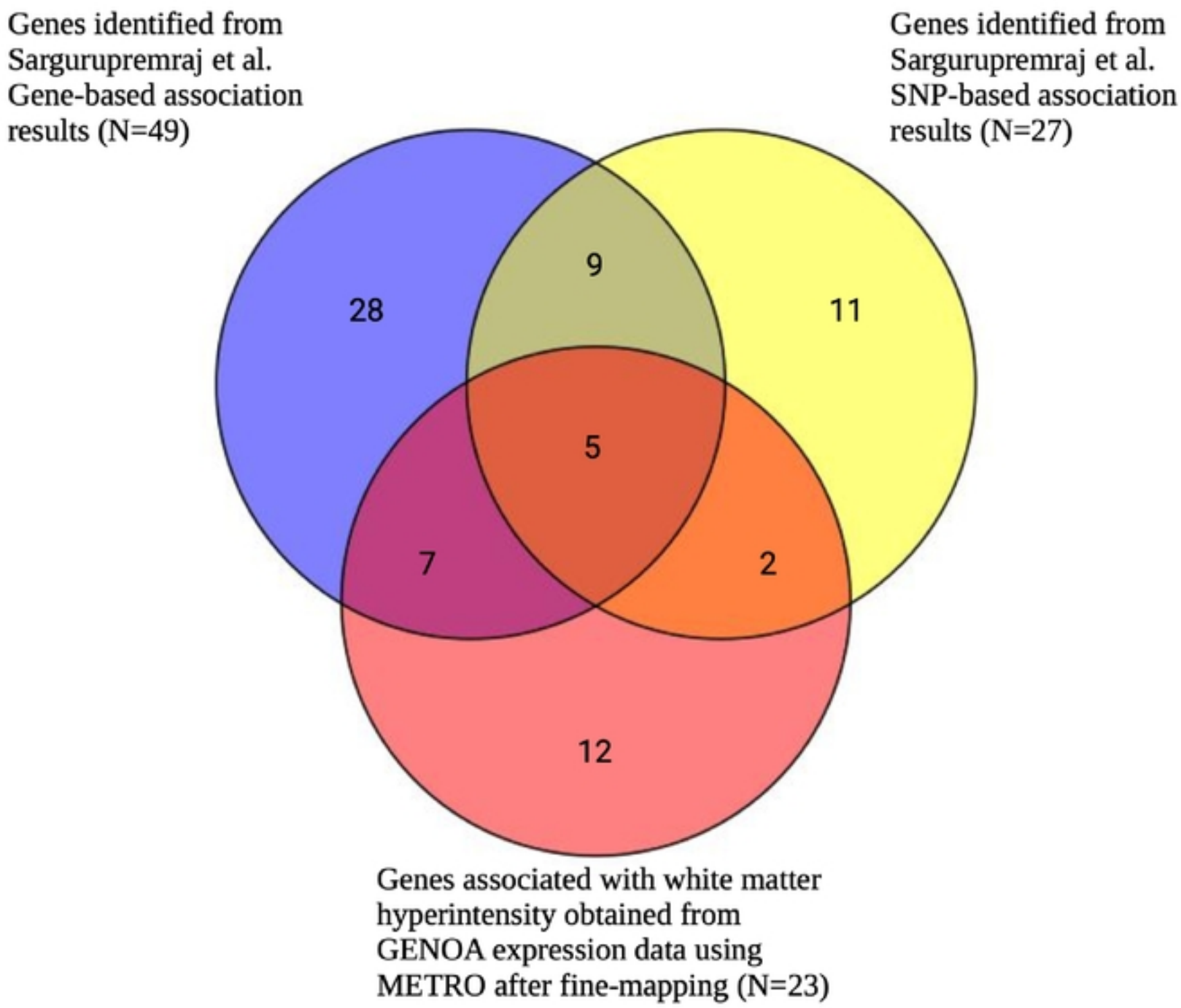
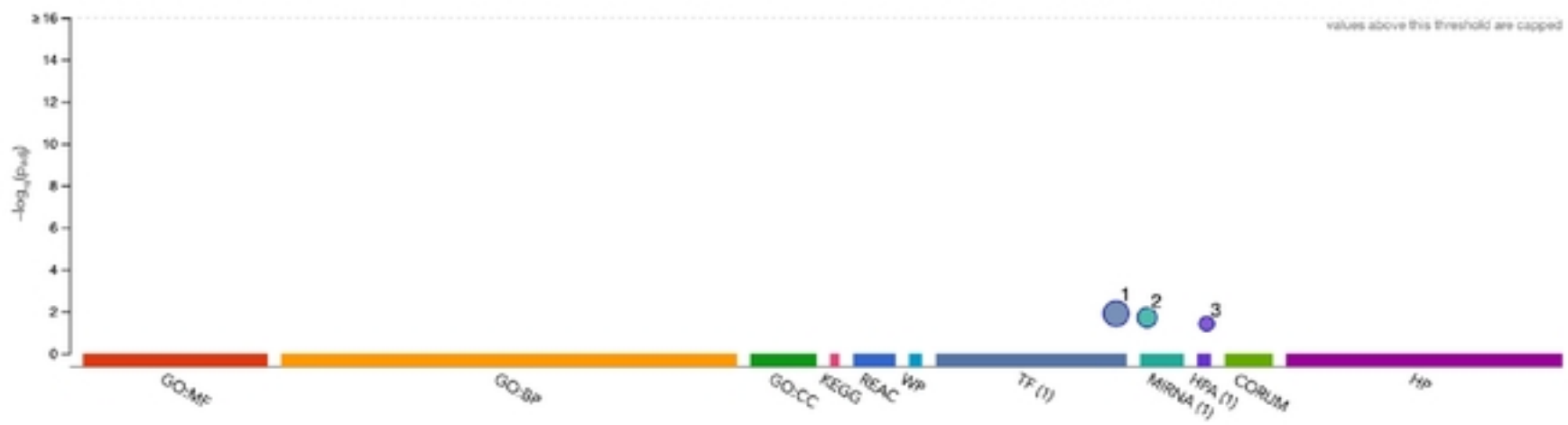


Figure 7. Venn diagram comparing number of METRO-identified genes associated with white matter hyperintensity following FOCUS fine-mapping and genes identified by Sargurupremraj et al. (2020) gene-based and SNP-based analyses.
Venn diagram comparing the number of significantly associated genes associated with white matter hyperintensity (WMH) obtained from METRO using GENOA expression data after Bonferroni correction ($P<2.90x10^{-6}$), and fine-mapping (red) and Sargurupremraj et al. (2020). Sargurupremraj et al. results included SNP-based association results that were mapped to the nearest gene ($P<5x10^{-8}$; yellow), and gene-based association results ($P<2.77x10^{-6}$; blue).



ID	Term ID	Term Name	Adjusted P-value
		TF	
1	TF:M12713	Factor: ZNF690; motif: GTCTACRCNG	1.27E-02
		MIRNA	
2	MIRNA: has-miR-212-5p	hsa-miR-212-5p	1.94E-02
		HPA	
3	HPA:0411242	Retina; inner plexiform layer [≥Medium]	3.86E-02

Figure 8. Functional enrichment analysis on the fine-mapped gene set identified for white matter hyperintensity using METRO TWAS (N=23 genes). The top panel consists of a Manhattan plot that illustrates the enrichment analysis results. The x-axis represents functional terms that are grouped and color-coded by data sources, including Gene Ontology (GO): molecular function (MF; red), GO: biological process (BP; orange), GO: cellular component (CC; dark green), Kyoto Encyclopedia of Genes and Genomes (KEGG; pink), Reactome (REAC; dark blue), WikiPathways (WP; turquoise), Transfac (TF; light blue), MiRTarBase (MIRNA; emerald green), Human Protein Atlas (HPA; dark purple), CORUM protein complexes (light green), and Human Phenotype Ontology (HP; violet), in order from left to right. The y-axis shows the adjusted enriched -$\log_{10}$ p-values < 0.05. Multiple testing correction was performed using g:SCS method (Set Counts and Sizes) that takes into account overlapping terms. The top panel highlights driver GO terms identified using the greedy filtering algorithm in g:Profiler. The light circles represent terms that were not significant after filtering. The circle sizes are in accordance with the corresponding term size (i.e., larger terms have larger circles). The number in parentheses following the source name in the x-axis shows how many significantly enriched terms were from this source.

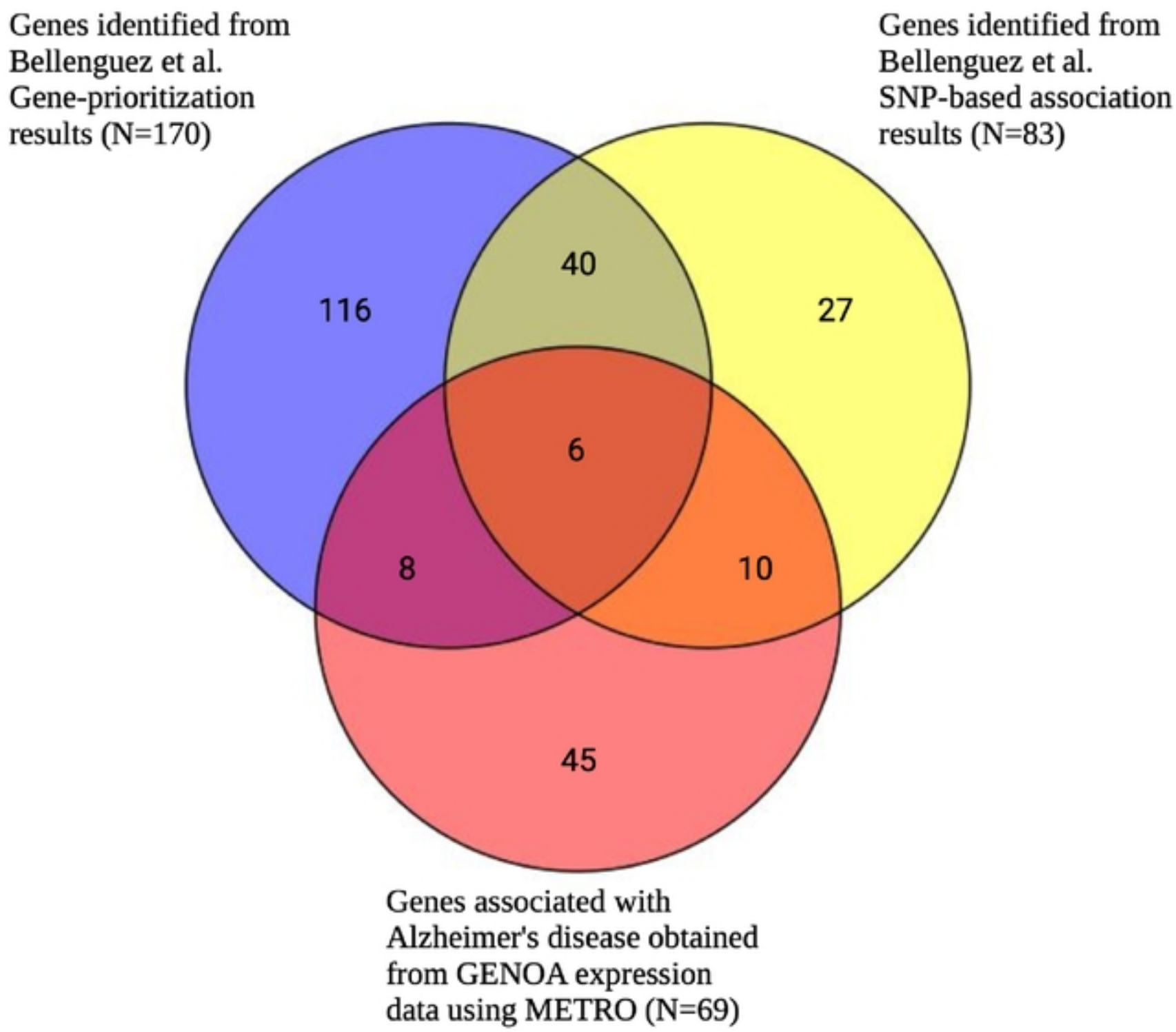
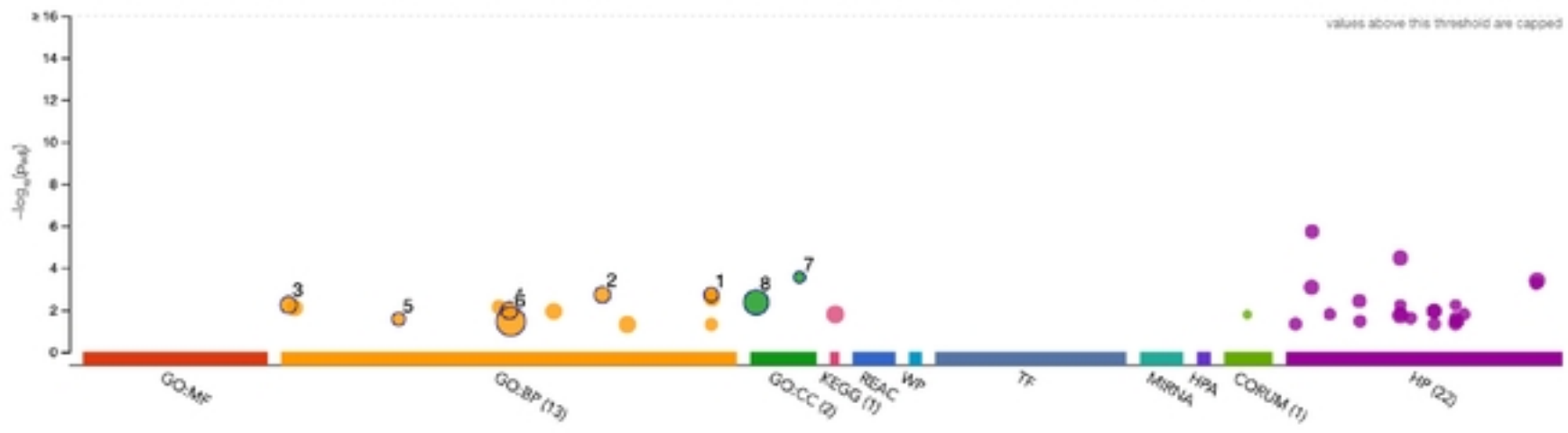
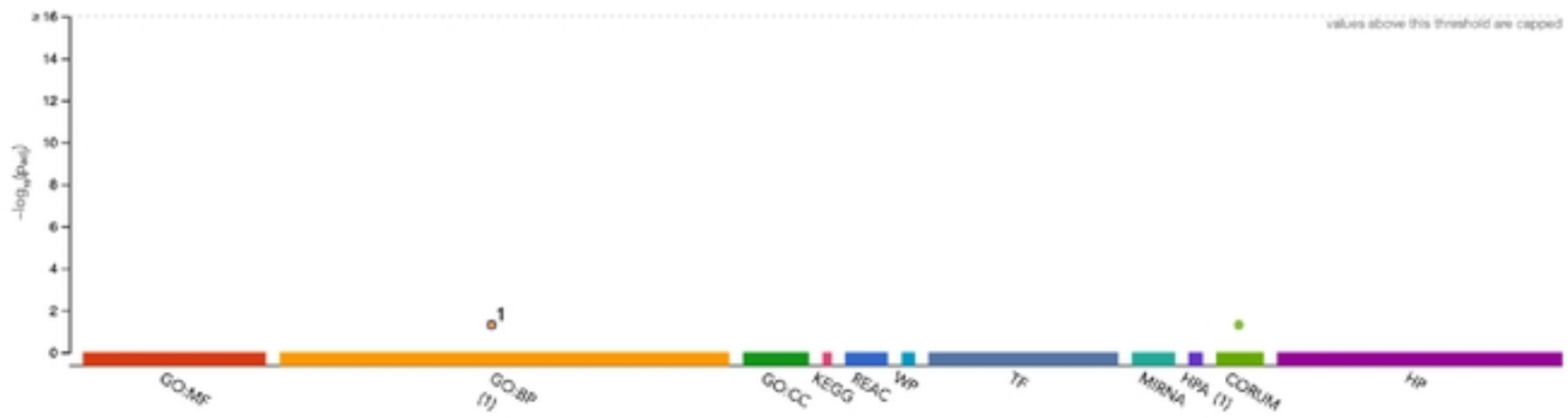


Figure 9. Venn diagram comparing number of METRO-identified genes associated with Alzheimer's disease (EA GWAS) following FOCUS fine-mapping and genes identified by Bellenguez et al. (2022) gene prioritization and SNP-based analyses.
Venn diagram comparing the number of significantly associated genes associated with Alzheimer's disease (from EA GWAS) obtained from METRO using GENOA expression data after Bonferroni correction ($P<2.90x10^{-6}$) and fine-mapping (red) and Bellenguez et al. (2022). Bellenguez et al. results included SNP-based association results that were mapped to the nearest gene ($P<5x10^{-8}$; yellow), and gene prioritization results for the genes in the novel AD risk loci (blue). In the gene prioritization analysis, Bellenguez et al. analyzed the downstream effects of new AD-associated loci on molecular phenotypes (i.e., expression, splicing, protein expression, methylation and histone acetylation) in various *cis*-quantitative trait loci (*cis*-QTL) catalogues from AD-relevant tissues, cell types and brain regions.



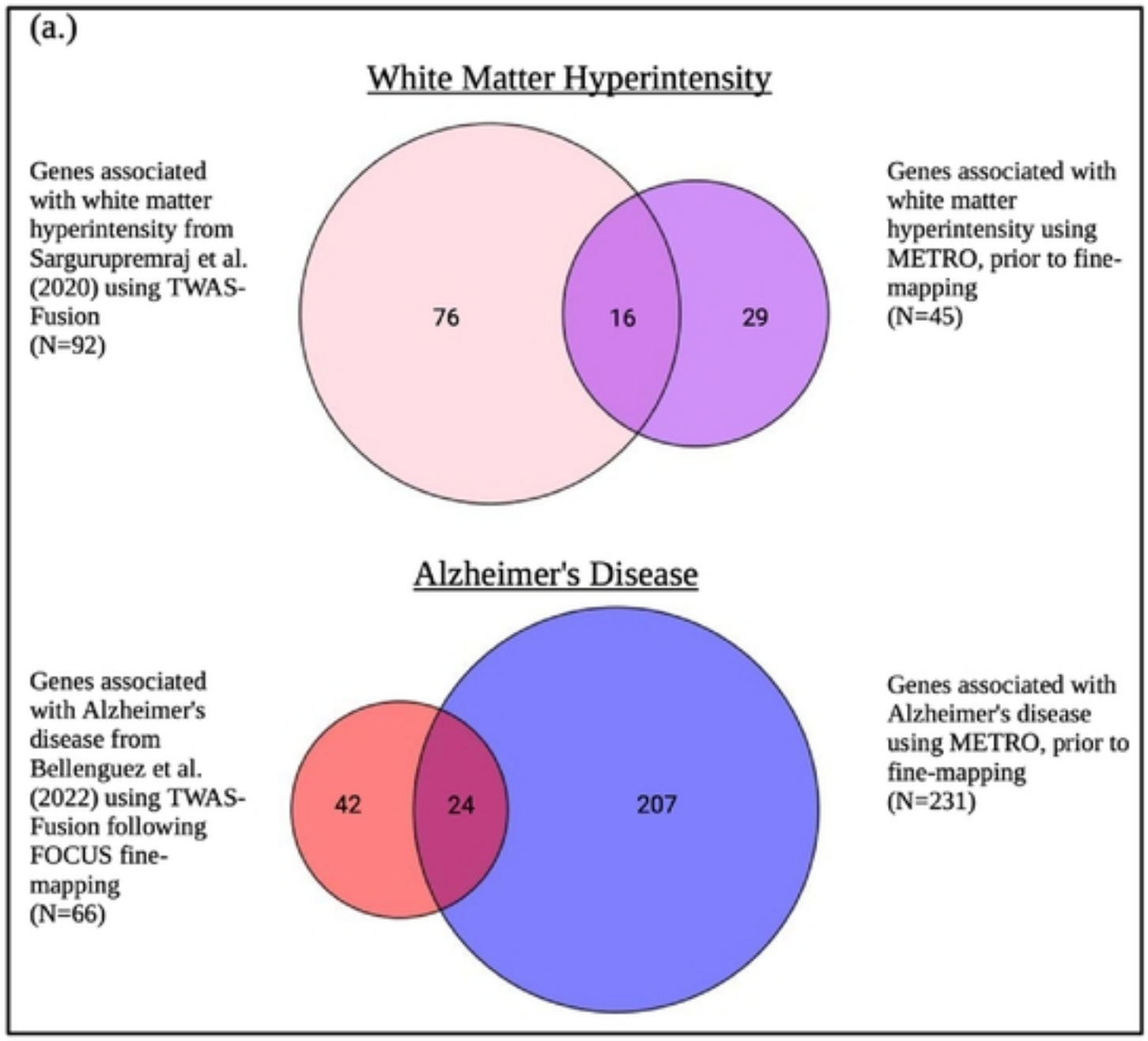
ID	Term ID	Term Name	Adjusted P-value
		GO:BP	
1	GO:1905906	regulation of amyloid fibril formation	1.87E-03
2	GO:0097242	amyloid-beta clearance	1.90E-03
3	GO:0001774	microglial cell activation	5.79E-03
4	GO:0050435	amyloid-beta metabolic process	1.07E-02
5	GO:0030450	regulation of complement activation, classical pathway	2.71E-02
6	GO:0050794	regulation of cellular process	3.60E-02
		GO: CC	
7	GO:0097418	neurofibrillary tangle	2.80E-04
8	GO:0005794	Golgi apparatus	4.39E-03

Figure 10. Functional enrichment analysis on the fine-mapped gene set identified for Alzheimer's disease (EA GWAS) using METRO TWAS (N=69 genes). The top panel consists of a Manhattan plot that illustrates the enrichment analysis results. The x-axis represents functional terms that are grouped and color-coded by data sources, including Gene Ontology (GO): molecular function (MF; red), GO: biological process (BP; orange), GO: cellular component (CC; dark green), Kyoto Encyclopedia of Genes and Genomes (KEGG; pink), Reactome (REAC; dark blue), WikiPathways (WP; turquoise), Transfac (TF; light blue), MiRTarBase (MIRNA; emerald green), Human Protein Atlas (HPA; dark purple), CORUM protein complexes (light green), and Human Phenotype Ontology (HP; violet), in order from left to right. The y-axis shows the adjusted enriched $-\log_{10}$ p-values < 0.05. Multiple testing correction was performed using g:SCS method (Set Counts and Sizes) that takes into account overlapping terms. The top panel highlights driver GO terms identified using the greedy filtering algorithm in g:Profiler. The light circles represent terms that were not significant after filtering. The circle sizes are in accordance with the corresponding term size (i.e., larger terms have larger circles). The number in parentheses following the source name in the x-axis shows how many significantly enriched terms were from this source.



ID	Term ID	Term Name	Adjusted P-value
1	GO:0046814	Coreceptor-mediated virion attachment to host	4.96E-02

Figure 11. Functional enrichment analysis fine-mapped gene set identified for Alzheimer's disease (AA GWAS) using METRO TWAS (N=2 genes). The top panel consists of a Manhattan plot that illustrates the enrichment analysis results. The x-axis represents functional terms that are grouped and color-coded by data sources, including Gene Ontology (GO): molecular function (MF; red), GO: biological process (BP; orange), GO: cellular component (CC; dark green), Kyoto Encyclopedia of Genes and Genomes (KEGG; pink), Reactome (REAC; dark blue), WikiPathways (WP; turquoise), Transfac (TF; light blue), MiRTarBase (MIRNA; emerald green), Human Protein Atlas (HPA; dark purple), CORUM protein complexes (light green), and Human Phenotype Ontology (HP; violet), in order from left to right. The y-axis shows the adjusted enriched -$\log_{10}$ p-values < 0.05. Multiple testing correction was performed using g:SCS method (Set Counts and Sizes) that takes into account overlapping terms. The top panel highlights driver GO terms identified using the greedy filtering algorithm in g:Profiler. The light circles represent terms that were not significant after filtering. The circle sizes are in accordance with the corresponding term size (i.e., larger terms have larger circles). The number in parentheses following the source name in the x-axis shows how many significantly enriched terms were from this source.


(a.)
White Matter Hyperintensity
Genes associated with white matter hyperintensity from Sargurupremraj et al. (2020) using TWAS-Fusion (N=92)
76
16
29
Genes associated with white matter hyperintensity using METRO, prior to fine-mapping (N=45)
Alzheimer's Disease
Genes associated with Alzheimer's disease from Bellenguez et al. (2022) using TWAS-Fusion following FOCUS fine-mapping (N=66)
42
24
207
Genes associated with Alzheimer's disease using METRO, prior to fine-mapping (N=231)

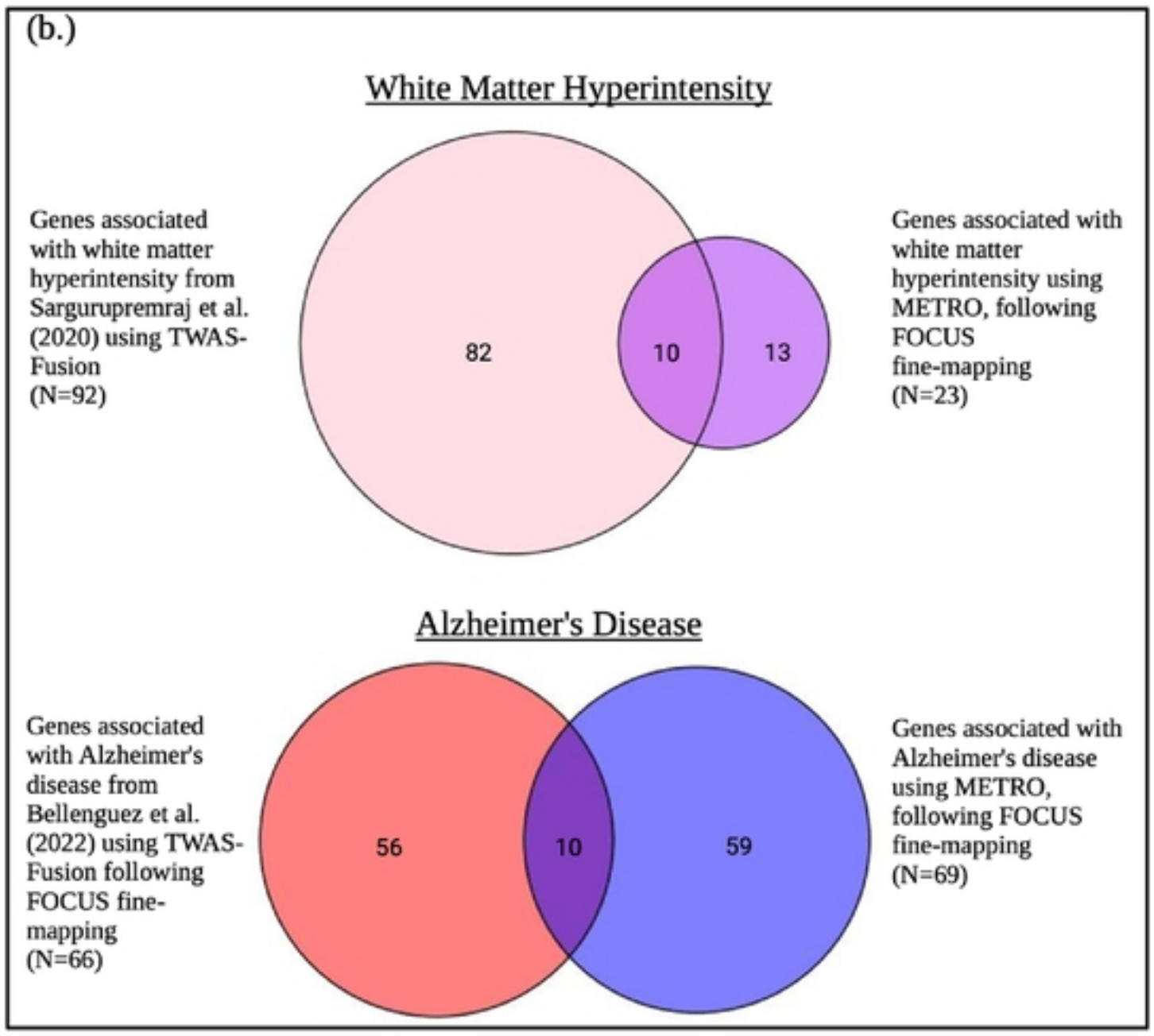

(b.)
White Matter Hyperintensity
Genes associated with white matter hyperintensity from Sargurupremraj et al. (2020) using TWAS-Fusion (N=92)
82
10
13
Genes associated with white matter hyperintensity using METRO, following FOCUS fine-mapping (N=23)
Alzheimer's Disease
Genes associated with Alzheimer's disease from Bellenguez et al. (2022) using TWAS-Fusion following FOCUS fine-mapping (N=66)
56
10
59
Genes associated with Alzheimer's disease using METRO, following FOCUS fine-mapping (N=69)

Figure 12. Venn diagram comparing METRO TWAS results prior to and following FOCUS fine-mapping with TWAS results from Sargurupremraj et al. (2020) and Bellenguez et al. (2022).

Venn diagram comparing METRO TWAS results (a) prior to and (b) following FOCUS fine-mapping with TWAS results using Fusion for white matter hyperintensity from Sargurupremraj et al. (2020) without fine-mapping and Alzheimer's disease from Bellenguez et al. (2022) (EA GWAS) with FOCUS fine-mapping.