# Genetic architecture of brain age and its casual relations with brain and mental disorders

Esten H. Leonardsen, MS[1,2], Didac Vidal-Piñeiro, PhD[1], James M. Roe, PhD[1], Oleksandr Frei, PhD[2], Alexey A. Shadrin, PhD[2,3], Olena Iakunchykova, PhD[1], Ann-Marie G. de Lange, PhD[4,5,6], Tobias Kaufmann, PhD[2,7], Bernd Taschler, PhD[8], Stephen M. Smith, PhD[8], Ole A. Andreassen, MD, PhD[2,3], Thomas Wolfers, PhD[2,4,7], Lars T. Westlye, PhD[2,3,4], Yunpeng Wang, PhD[1,*]

1 Center for Lifespan Changes in Brain and Cognition (LCBC), Department of Psychology, University of Oslo, 0317 Oslo, Norway

2 NORMENT, Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, 0317 Oslo, Norway

3 K.G. Jebsen Centre for Neurodevelopmental disorders, University of Oslo, Oslo, Norway

4 Department of Psychology, University of Oslo, 0317 Oslo, Norway

5 LREN, Centre for Research in Neurosciences, Department of Clinical Neurosciences, Lausanne University Hospital (CHUV) and University of Lausanne, 1015 Lausanne, Switzerland

6 Department of Psychiatry, University of Oxford, OX1 2JD Oxford, UK

7 Department of Psychiatry and Psychotherapy, Tübingen Center for Mental Health, University of Tübingen, 72074 Tübingen, Germany

8 Wellcome Centre for Integrative Neuroimaging (WIN FMRIB), University of Oxford, OX3 9DU Oxford, United Kingdom

[*] Corresponding author:

Yunpeng Wang,

Department of Psychology, University of Oslo,

Forskningsveien 3A, Harald Schjelderups hus, 0373 Oslo, Norway.

Email: yunpeng.wang@psykologi.uio.no.

Tel: +47 4655 9652

Running title: Brain age and mental disorders

**Abstract (**178 words**)**

The difference between chronological age and the apparent age of the brain estimated from brain imaging data — the brain age gap (BAG) — is widely considered a general indicator of brain health. Converging evidence supports that BAG is sensitive to an array of genetic and non-genetic traits and diseases, yet few studies have examined the genetic architecture and its corresponding causal relationships with common brain disorders. Here, we estimate BAG using state-of-the-art neural networks trained on brain scans from 53,542 individuals (age range 3-95 years). A genome-wide association analysis across 28,104 individuals (40-84 years) from the UK Biobank revealed eight independent genomic regions significantly associated with BAG ($p<5x10^{-8}$) implicating neurological, metabolic, and immunological pathways – among which seven are novel. No significant genetic correlations or causal relationships with BAG were found for Parkinson's disease, major depressive disorder, or schizophrenia, but two-sample Mendelian randomization indicated a causal influence of AD ($p=7.9x10^{-4}$) and bipolar disorder ($p=1.35x10^{-2}$) on BAG. These results emphasize the polygenic architecture of brain age and provide insights into the causal relationship between selected neurological and neuropsychiatric disorders and BAG.

46 **Introduction**

47 Over the last decade, brain age has emerged as a promising measure of overall brain health[1,
48 2]. To estimate brain age, machine learning models are applied to brain imaging data to learn
49 visual patterns characteristic of different ages[3, 4]. The difference between predicted brain age
50 and chronological age is termed the brain age gap (BAG) and indicates deviation from a
51 normative aging trajectory, a potential health indicator. Earlier studies have found a large
52 variation in the predicted brain age of individuals with the same chronological age, and that
53 these interindividual variations correlate with neurological and mental disorders[5-7], such as
54 dementia[6, 8], schizophrenia (SCZ)[9, 10], major depressive disorder (MDD)[11], and also mortality[7,
55 12]. In addition, biological, environmental and lifestyle factors associated with these disorders
56 have been reported to correlate with BAG, for example infections[13, 14], smoking[5], physical
57 activity[15], and education level[16].

58 Genetic differences have been shown to explain a sizeable portion of interindividual
59 variation in BAG. Twin-based heritability for BAG has been estimated to be as high as 0.66[17],
60 and single nucleotide polymorphism (SNP)-based heritability estimates are also relatively high
61 — around 0.2 [6, 18]. Earlier gene-discovery efforts investigating genetic associations with BAG
62 have found and examined two genomic loci in detail: one on chromosome 1 containing the
63 potassium channel gene, *KCNK2*, and one in the chromosome 17 inversion region
64 (17q21.31)[18, 19]; Genetic variants in these two regions together explain a negligible fraction of
65 estimated SNP-heritability[18]. These results suggest that existing GWAS were potentially
66 underpowered to fully characterize the genetic architecture, supported by studies using a
67 conditional false discovery rate-based models yielding a larger set of associations[6].

68 Furthermore, Smith *et al.*[20] found a rich set of associations when investigating different facets
69 of a multimodal brain age, suggesting that the interplay between genetic variants is complex.

70 Although BAG has been frequently associated with mental disorders, the underlying
71 genetic basis for the observed association has seldom been investigated, possibly due to
72 incomplete knowledge of the genetic architecture of the former. Furthermore, the causal
73 relationships between BAG and brain disorders remain untapped. Mendelian randomization
74 (MR) has become an attractive tool to interrogate cause-effect relationships between risk
75 factors or disorders[21]. Two-sample MR models have been used to infer causal relations
76 between hundreds of traits or diseases[22]. However, MR analyses targeting the causal relations
77 between BAG and brain disorders and associated traits have been lacking[23].

78 In the present work, we improve the yield of genetic associations for BAG using three
79 strategies: First, we estimate brain age using a state-of-the-art deep neural network
80 architecture (SFCN-reg) trained on one of the largest samples assembled to date[5]. Then we
81 perform a GWAS for BAG on out-of-sample predictions for a portion of the UK Biobank v3
82 data containing 28,104 unrelated individuals, about eight thousand more than earlier studies.
83 Finally, we use two-sample MR to assess the genetic and causal relations between BAG and
84 SCZ, bipolar disorder (BIP), Alzheimer's disease (AD), MDD, and Parkinson's disease (PD).

85

## Methods

### Sample

All datasets used in the present study have been obtained from previously published studies which have been approved by their respective institutional review board, research ethic committee or other relevant ethic organizations.

We used UK Biobank imaging data (UKB, accession number 27412) released in 2019 in combination with a pre-compiled dataset from various sources (**Supplementary Table S1**) for brain age model training and estimation. For the downstream genetic analyses, we started with the initial 40,330 UKB participants that had undergone at least one brain scan (using baseline scans where more were available). We excluded those with recorded brain injury or neurological or psychiatric conditions, those failing standard image quality checks[5], those whose genetic data failed quality check by the UKB team, in addition to participants who withdrew consent. Among related participants (UKB-reported kinship coefficient >0), only one was included in the study. In total, 28,104 unique participants remained.

### Brain age estimation

A minimal preprocessing protocol was applied to all raw T1-weighted brain MRI images before brain age estimation[5]: The auto-recon pipeline from FreeSurfer 5.3[24] was used to remove non-brain tissue. The resulting volumes were reoriented to the standard FSL[25] orientation using fslreorient2std, and linearly registered to the 1 mm FSL (version 6.0) MNI152 template using FLIRT[26], with 6 degrees of freedom. For efficiency, during model fitting, we cropped a central cube spanning the voxels 6:173, 2:214, and 0:160 in the sagittal, coronal, and axial dimensions,

respectively. Before modelling, all voxel intensities were normalized by a constant factor to produce values in the range [0, 1].

The data from all sources (**Supplementary Table S1** and UKB) were split into five equally-sized and disjoint folds with comparable age-ranges and sex distributions. Four of these folds were used for fitting the brain age model, and out-of-sample estimates were computed for the remaining fold. This procedure was repeated five times, resulting in out-of-sample brain age estimates for all participants. Next, BAG was calculated by subtracting chronological age from estimated brain age. The subsequent analyses were performed on the out-of-sample estimates of the UKB data (**Supplementary Table S2**).

**Genome-wide association study**

Imputed genotypes for the 28,104 participants were obtained from UKB (Category 100314, for further details see[27]). We excluded SNPs based on missing rate (>0.02), the Hardy-Weinberg Equilibrium test ($p<10^{-6}$) and minor allele frequency (MAF; < 0.01). In total, ≈8.6 million SNPs were analyzed. Since we have observed apparent differences in predicted brain age across folds (**Supplementary Fig. S1**), a GWAS was performed on each hold-out fold separately using PLINK 1.90 beta[28]. The additive genetic model was assumed, and chronological age, sex and the top ten principal components were included as covariates, accounting for population structure. Association results for each hold-out fold of UKB along with distributions of BAG are shown in **Supplementary Figure S1**. These association results were then meta-analyzed using the inverse variance weighted model implemented in PLINK to identify SNPs that are associated with BAG. **Supplementary Figure S2** shows the association QQ plot which indicated no noticeable genomic inflation.

132 **Associated regions and genes**

133 Association results were 'clumped' by the FUMA[29] web-service using the linkage
134 disequilibrium (LD) structure from the 1000 Genomes projects phase 3 EUR dataset (1KGp3),
135 with parameters *–clump-p 5e-8 –clump-2 1e-6 –clump-r2 0.1.* The standard 250 kilo-bases (kb)
136 were used as the inter-region distance threshold. Genes whose genomic coordinates located
137 within the boundaries of each region were assigned to the corresponding region. SNPs with
138 the smallest association p-values were taken as the lead SNPs for the corresponding regions.
139 In addition, the gene that is closest to each lead SNP by genomic position was annotated using
140 the Ensembl tool VEP[30] (**Table 1**).

141 Associated regions were fine-mapped using the FINEMAP[31] program. The LD structure
142 from 1KGp3 was also used in this analysis. The default settings of FINEMAP were used, which
143 compares causal models assuming one causal variant in each region to that assuming two,
144 based on the estimated posterior probabilities (PP_1 versus PP_2). FINEMAP ranks all possible
145 configurations in each model presented as 95% credible sets. The confidence of a variant
146 belonging to a set was evaluated by posterior probabilities of inclusion (PPI). In the case of
147 assuming one causal variant, each single variant was assigned a PPI. In the two causal variants
148 cases, each pair of variants was assigned a PPI.

149

150 **Post-GWAS annotations**

151 Both FUMA and Garfield[32] were used for annotating associated SNPs. First, SNPs were
152 assigned to genic elements (*e.g.*, exon, intron, 3' and 5' untranslated regions, intergenic
153 regions, etc.), and the enrichment of this assignment was tested by hypergeometric test
154 (FUMA) or logistic regression models (Garfield). Expression levels of annotated genes to the
155 associated SNPs were inspected in each of the 54 tissue types from the GTEx v8 dataset[33].

156 Detailed biological functions for proteins coded by these genes were manually searched in
157 the NCBI Entrez Gene database[34] and the UniProtKB database[35].
158
159 **Genetic correlations between BAG and disorders**
160 GWAS summary data for four disorders (SCZ[36], BIP[37], MDD[38], and AD[39]) were obtained from
161 the Psychiatric Genomics Consortium (PGC, https://med.unc.edu/pgc/download-results). For
162 each GWAS, the association results for European ancestral samples excluding samples from
163 23andMe were used (SCZ: n case=67,390, n control=94,015; BIP: n case=41,917, n
164 control=371,549; MDD: n case=59,851, n control=113,154; AD: n case=71,880, n
165 control=383,378). The PD GWAS results were obtained from the fixed-effect meta-analysis
166 performed by the International Parkinson Disease Genomics Consortium (IPDGC, n
167 case=33674, n control=449056)[40].
168 Before post-GWAS analysis we processed the results from all GWAS using a standard
169 protocol. Specifically, SNPs having a MAF <0.05, or imputation INFO <0.5, or ambiguous allelic
170 coding (A/T, or C/G) were removed from subsequent analyses. The LD score model (ldsc)[41]
171 was applied to estimate SNP-heritability and genetic correlations between BAG and disorders.
172 Only high-quality SNPs published in the HapMap3 dataset were used for estimation. The LD
173 score derived from the 1KGp3 was used as input to ldsc. The Benjamini-Hochberg False
174 Discovery Rate (FDR) procedure was used to correct for multiple testing across disorders
175 (FDR-corrected p<0.05 was considered statistically significant).
176 To visualize polygenic enrichment, conditional QQ plots[42, 43] were made for BAG versus
177 each disorder. In these plots, the QQ curves for the association statistics (-log10 p-values) for
178 BAG were stratified by the corresponding association strength for the conditioned disorder.
179 As the association strength to the conditioned disorder increases, a successive leftward

180 deflation in these curves indicates polygenic enrichment. Similarly, conditional QQ for each
181 disorder versus BAG shows polygenic enrichment in the reverse direction.
182
183 **Two-sample Mendelian randomization**
184 To study the cause-effect relations between BAG and the five disorders, two sets of MR
185 analyses were performed. The first set, using standard models, included the inverse-variance
186 weighted model (IVW)[44], weighted median (wMed)[45], Egger regression (Egger)[46] and MR-
187 PRESSO (PRESSO)[47]. For these analyses, only genome wide significant SNPs ($p<5x10^{-8}$) to the
188 exposure traits or disorders were used as potential instruments. The PLINK program and the
189 LD structure of 1KGp3 dataset were used to select instruments with the following parameters,
190 *--clump-kb 500 kb*, *--clump-p1 5x10$^{-8}$*, and *--clump-r2 0.01*. The TwoSampleMR package[19] was
191 used for data harmonization and causal inference for the IVW, wMed and Egger models. The
192 same harmonized datasets were used as input to the MR-PRESSO software to assess outliers
193 that may artificially affect MR estimates, *i.e.*, SNPs that show horizontal pleiotropy to both
194 BAG and disorders. Harmonized instrumental SNPs are shown in **Supplementary Tables S6-**
195 **S15.**
196 The second set of models included the robust adjusted profile score (RAPS)[48] and the
197 CAUSE models[49]. These models can make use of SNPs that show a suggestive level of
198 association ($p<10^{-3}$) with exposure to increase statistical power without incurring weak
199 instrument bias in estimation. While both models control for horizontal pleiotropy, CAUSE
200 directly tests for a shared (correlated horizontal pleiotropy) versus a causal model for each
201 relation[49]. Instead of estimating causal effects, CAUSE provides z scores for such tests;
202 negative z scores suggest the existence of a causal relation and positive suggest otherwise.

203 The same instrument selection procedures used in the first set of models were used here,
204 except that $10^{-3}$ was taken as the cut-off for selecting instruments, *i.e. –clump-p1* $10^{-3}$.
205 As each of the six MR models has different assumptions that are difficult to verify in
206 real data, a majority vote ensemble scheme was used to make conclusions for the existence
207 of cause-effect relations: specifically, only when four or more models indicated a cause-effect
208 relation (FDR adjusted $p<0.05$) was such a relation considered causal.
209 In addition to applying multiple MR models, GWAS results for height measured for
210 European samples[50] (n=253,288; https://portals.broadinstitute.org/collaboration/giant/), for
211 AD diagnosed in a Japanese sample[51] (n case=3,962 and n control=4,047) and an African
212 sample[52] (n case=2,784 and n control=5,222) and for BIP diagnosed in a Japanese sample[53](n
213 case=2,964 and n control=61,887) were used to corroborate MR findings. As commonly done
214 in genetic studies, height was used as a negative exposure control to test if population
215 stratification could generate spurious causal effects [54]. The non-European GWAS data were
216 used to test if any observed causal effects generalize across ancestry groups, although with
217 significantly smaller sample sizes.

218 **Results**

219 We obtained accurate brain age estimates; mean absolute errors (MAEs) in each of the five
220 disjoint folds were consistently below 2.5 years (**Supplementary Table S2**). This was consistent
221 when we split the dataset into different subsets based on covariates (MAE = 2.40 in females
222 compared to 2.53 in males; 2.40 in the youngest half compared to 2.52 in the oldest), although
223 we observed a slight age bias (**Supplementary Figure S9**). Based on the meta-analyzed GWAS
224 results, we estimated a SNP heritability of 0.27 (standard error (SE)=0.036) for BAG (**Methods**).

225 Our estimate is comparable to or higher than the two previously reported estimates (0.26,
226 SE=0.044[6]; 0.19, SE=0.02[18]).

227 We identified eight independent genomic loci significantly associated with BAG (**Fig. 1**
228 **a**,**b** and **Table 1**). Associations of lead SNPs in these regions to BAG are highly consistent in
229 directions and effect sizes across the five folds (**Supplementary Table S3**). Among these loci,
230 the one in the inversion region on chromosome 17 (lead SNP rs2106786), including the *MAPT*
231 gene, has been previously reported[18, 19], although indexed by a different SNP (rs2435204). This
232 SNP was also highly significant in our analysis ($p=5.4x10^{-21}$, beta=0.27 years, effective-
233 allele=G). The locus containing the *RUNX2* gene (lead by rs2790102: $p=8.92x10^{-9}$, beta=-0.15
234 years, effective-allele=A), which showed suggestive significance in Jonsson *et al*.[18], was
235 genome-wide significant in the present study. The *RUNX2* gene codes for a master
236 transcription factor which plays a critical role in skeletal development[55]. Among the remaining
237 six novel loci, the rs79107704-A allele showed the largest association with BAG; one copy of
238 this allele was associated with an average increase in brain age of 0.63 years (**Table 1**). This
239 SNP is located 3,405bp downstream of the Betaine-homocystein S-methyltransferase 2 gene
240 (*BHMT2*, **Fig. 1b**), a gene whose product is involved in choline metabolism during
241 development[56]. Other protein-coding genes that are closest to the lead SNPs include those
242 involved in calcium signaling (*CAMK2N2* and *INPP5A*) and metabolism and transcription
243 regulation (*GALC*, *KLF3* and *KLHL38*), both processes are implicated in biological ageing[57]. In
244 **Supplementary Table S5** we present detailed annotations of biological functions to each gene.

245 We further annotated these identified SNPs to nearby genes and regulatory elements
246 (**Methods**). Most of the associated SNPs are in non-coding regions such as intergenic, intron

247 or untranslated regions (**Supplementary Figs. S3** and **S4**). Using the default parameters in
248 FUMA[29], 54 unique genes were found to be implied by these significant associations by
249 genomic position. The expression levels of these genes in the 54 tissue types from the GTEx
250 v8 project[33] showed three remarkable patterns (**Fig. 1c**). The first set of genes expressed highly
251 across almost all 54 tissues; the second set of genes showed low expression levels in most
252 tissue types; and the last set, including eight genes, was highly expressed only in brain tissue
253 types (**Fig. 1c**), for example *MAPT*, *GFAP*, and the Homeobox protein gene *NKX6-2*. These
254 results suggest that BAG encodes coordinated physiological processes implicating both the
255 brain and the peripheral systems.

256 To nominate causal variants in each locus we performed statistical fine-mapping[31] for
257 regions around each lead SNP in **Table 1** (**Methods**). Except for the locus on chromosome 14
258 which was not resolvable, all loci clearly indicated that the 95% credible sets suggest a causal
259 model with one causal SNP, instead of two, *i.e.*, the posterior probability for the 1-SNP set
260 were larger than those of the 2-SNP sets (**Supplementary Table S16**). Furthermore, four
261 credible sets indicated that the lead SNPs were also the causal ones (posterior inclusion
262 probability (PPI_1) > 0.05 and >PPI of the second most probable SNP(PPI_2)) (**Methods**;
263 **Supplementary Table 16**) but identifying the causal SNP for the remaining were difficult. For
264 example, the *MAPT* locus on chromosome 17 and the *RUNX2* locus on chromosome 6 showed
265 two SNPs having almost equal and small PPIs (*i.e.*, <=0.05), indicating that the true causal
266 variants may be some untyped rare ones not investigated in this study. The clearest signal
267 comes from the regions on chromosome 3 and 5, where the PPIs for the lead SNPs were much
268 larger than that for the second most probable causal SNPs.

269 We observed nominally significant genetic correlation between BAG and AD that did
270 not survive FDR-correction (r=0.23, SE=0.1, p=0.02, FDR adjusted-p=0.13) and no significant
271 associations with any other of the four disorders (**Fig. 2a**). SNP heritability estimates for the
272 five disorders were all significant but varied greatly; SCZ showed the largest (0.34, SE=0.01)
273 and AD showed the lowest (0.01, SE=0.005) estimates. Bidirectional conditional QQ plots (**Fig.**
274 **2b-d; Supplementary Figs. S6** and **S7**) showed that there was noticeable genetic enrichment
275 for BIP conditional on BAG but not in the reverse direction. For AD and PD, both directions
276 showed clear enrichment, surprisingly for PD which did not show significant genetic
277 correlation with BAG (r=-0.07, p=0.42).

278 We then performed extensive MR analyses using six different models to examine the
279 existence of cause-effect relations between BAG and the five disorders (**Methods**). **Figure 3a**
280 shows that BAG was only causally associated with PD, *i.e.*, four out of the six MR models
281 showed a negative relation with varying effect sizes (all with adjusted p <0.05). One year
282 increase in genetically predicted BAG was estimated to reduce the risk of PD by a log odds
283 ratio from 1.4 (by Egger regression) to 0.02 (by MR-RAPS) (**Supplementary Table S15**). In the
284 reverse direction (*i.e.*, disorders as exposure), increased genetic risk for AD and BIP were
285 causally associated with increased BAG (30 and 55 SNPs used as instruments, respectively);
286 these estimated causal effects on BAG were relatively larger for AD than BIP (**Fig. 3b;**
287 **Supplementary Table S16**).

288 A close investigation into the scatter plots of instrumental SNPs showed that the causal
289 effect of AD on BAG was primarily driven by a SNP (rs59007384) in the APOE region, which
290 was not identified as a horizontal pleiotropic instrument by MR-PRESSO (outlier test p>0.05)

291 (**Fig. 3c**); there were no extreme instruments identified for the BIP to BAG relation by MR-
292 PRESSO (**Fig. 3d**) but Egger-regression indicated existence of horizontal pleiotropy (Egger
293 intercept test: p=0.017). The causal effects of BAG on PD were primarily driven by two SNPs
294 in the inversion region on chromosome 17, effective alleles of these SNPs were associated
295 with higher BAG and lower risk of PD (**Fig. 3e**). SNPs in the same region also drove the negative
296 causal relation (not significant) from PD to BAG (**Fig. 3f**). However, both SNPs were flagged as
297 horizontal pleiotropic instruments by MR-PRESSO (p<0.05) and Egger-regression (Egger
298 intercept test: p=0.03 and 0.008, respectively). Therefore, the observed negative relations
299 between BAG and PD are less likely to be causal.

300 We used the GWAS results for height of European samples and cross-ancestry MR
301 analysis to corroborate the identified causal relations (**Methods**). We found no causal effect
302 between BAG and height with any of the MR methods employed (all p > 0.05). Therefore, our
303 observed AD and BIP to BAG relations are less likely to be driven by population stratification,
304 *i.e.*, both the exposure and outcome data originating from the same ancestry group. There
305 was also no significant cross-ancestral causal effect detected using AD data from Japanese or
306 African samples (IVW p=0.74, 0,85, respectively), and BIP data from the Japanese sample to
307 BAG (beta=0.10, p=0.13).

308

309 **Discussion**

310 Combining the advantages of large samples and advanced models for brain age prediction, we
311 confirmed that BAG is a heritable and polygenic trait, and estimated the genetic pleiotropy
312 and causal genetic relations with major brain and mental disorders. We identified seven novel
313 loci associated with BAG, in addition to confirming the previously reported *MAPT* loci[18, 19].
314 While MR indicated that increased genetic risk for AD or BIP may be causally associated with
315 higher BAG, our results demonstrate that individual variability and previously reported case-
316 control differences in BAG only to a marginal degree should be attributed to the genetic risk
317 of the respective diseases.

318 Functional annotation of the genes linked to the identified loci confirms that
319 deviations in BAG are linked to complex processes encompassing multiple biological systems[20].
320 While earlier work observed this variety when investigating different multimodal aspects of
321 imaging data linked to brain aging, our findings suggest it also exists when looking at a singular
322 BAG computed from only T1-weighted structural imaging. Our coarse division of the implied
323 54 genes into three groups indicates that only eight genes are specifically expressed in brain
324 tissues. The remaining genes were either expressed in abundance across all tissue types
325 tested, including the brain, or expressed at very low levels across all tissues. Nonetheless, the
326 proteins coded by these non-brain specific genes have been implicated in brain-related
327 disorders or traits (**Supplementary Table S5**). For example, among the genes we found to be
328 expressed across all tissue types (group 1), mutations in *AP2M1* have been linked to epilepsy,
329 intellectual developmental disorder, and seizures[58]; among the genes expressed in low levels
330 across tissues (group 2), *STH* has been associated with frontotemporal dementia and

331 17q21.31 duplication syndrome[59, 60]. In addition, although our analysis revealed no significant
332 pathway enrichment, these 54 genes contribute to biological functions that include calcium
333 signaling, protein metabolism, DNA damage repair and general innate immune defense. Thus,
334 our analyses highlight the role of these diverse sets of processes affecting the brain
335 throughout life.

336 Prior work has shown higher BAG in patients with a multitude of disorders compared
337 to healthy controls[5, 6, 8, 9, 11], and has documented partly overlapping genetic associations
338 between BAG and clinical conditions[6] . However, the causal effects have remained unclear.
339 Our MR approach revealed that genetically predicted risks for AD or BIP were causally
340 associated with increased BAG. However, these relations were only weakly supported by
341 genetic correlation analysis. One possible explanation for this weaker support from genome
342 wide signals (genetic correlation) in contrast to MR (significant associations only) might be
343 due to heterogeneous genetic correlations across the genome, *i.e.* some genomic regions
344 show positive correlations while others show negative correlations[61, 62]. In such a scenario,
345 the net genetic correlations between the two traits are expected to be lower than regional
346 correlations.

347 The causal effect of genetically predicted risk to AD on BAG was small but consistent
348 in directions across the six MR models, four of which were significant after multiple-testing
349 correction. For BIP, although four models showed significant effects, the CAUSE model
350 suggested an opposite direction of effect to the other five models. Thus, we advise careful
351 interpretation of this result. Our attempts of testing across ancestral group causal relations

352 led to largely null fundings for the AD to BAG relations. We believe these non-significant
353 fundings are largely due to the lack of statistical power in the non-European GWAS[51-53].

354 The observed causal relations between genetically predicted risk of brain disorders
355 and BAG are intriguing. One possible interpretation is that overt changes in the brain incurred
356 by the disorders contribute to accelerated aging. Another possibility may be that lifestyle and
357 health-related behaviors of patients with clinical conditions such as AD and BIP, *e.g.*,
358 medication[63], may increase brain age. Yet another is that genetic variation drives early life
359 changes in brain structure which contribute to both phenotypes later in life. Given the
360 comparable sample sizes to the GWAS of AD and BIP and the widely observed clinical
361 correlations, surprisingly, no genetic nor causal relations of SCZ and MDD with BAG were
362 found. On the one hand this may suggest that previously reported case-control differences do
363 not reflect causal relations, but rather a combination of indirect and confounding factors. For
364 example, smoking and physical exercise have been associated both with MDD and SCZ[64-67] and
365 brain age[5, 66]. Alternatively, it has been shown that both BAG[20] and psychiatric disorders are
366 highly heterogeneous phenotypes[68, 69], and thus further identification and characterization of
367 the causal relations may require even larger, and carefully screened, samples. It is also worth
368 noting that while the sample sizes for the disorders are large, the one underlying our BAG
369 GWAS is relatively small. Thus, our null findings in the direction from BAG to disorders may be
370 due to too weak instruments[70].

371 Our initial results showed weak evidence of a causal relation between BAG and PD,
372 corroborating two recent studies which reported a weaker correlation between BAG and PD
373 [71, 72]. Striking patterns of enrichment between the two were shown in the conditional QQ plots

374 and four out of six MR models indicated that genetically predicted BAG may have protective
375 effect from PD. However, we found that these relations were completely caused by the *MAPT*
376 gene region on chromosome 17: after removing chromosome 17 from our analyses, no
377 enrichment was observed in either direction (**Supplementary Fig. S8**). In addition,
378 instrumental SNPs in this region were detected by MR-PRESSO as horizontal pleiotropic SNPs,
379 *i.e.*, affecting BAG and PD through independent biological pathways. Reperforming MR
380 analysis excluding these outlier SNPs confirmed null causal relations. Thus, we conclude that
381 we found no evidence for causal relations between genetically predicted risk for PD and BAG.
382 Our analytic procedures also highlight the importance of triangulation and converging
383 evidence in causal inference analysis[73].

384 While the present study advances current knowledge regarding the genetic
385 architecture of and causal contributions to BAG, the results should be interpreted with caution.
386 Although we confirm previously reported genetic associations with BAG, e.g., the *MAPT* gene
387 locus[18, 19], our sample overlaps with previous ones—which were also based on UK Biobank
388 data. We attempted replicating our findings in three independent but small samples (n ranges
389 from 321 to 702; **Supplementary Analysis** and **Table S4**) but no clear replications were
390 achieved. Therefore, independent large-scale samples are needed for replication. We used a
391 simple voting schema across six different MR models to infer causal relations between
392 genetically predicted BAG and brain disorders. Furthermore, as we only identified eight
393 independent loci showing significant associations to BAG, other models[74] that require large
394 number of genome-wide significant instruments were considered not applicable. However, it
395 should be noted our simple voting approach may not be the most efficient strategy for
396 identifying causal effects. Formal development of ensemble methods, such as bagging[75], may

397 provide better grounds for precise interpretation. Furthermore, our BAG GWAS is still smaller
398 than GWAS performed for the disorders, which may partly explain the lack of causal effects of
399 BAG on brain disorders. Therefore, to increase our confidence in the identified relations, large-
400 scale data for BAG, and replications in independent datasets are needed. Relatedly, our
401 estimation of brain age was based on cross-sectional samples, which makes its interpretation
402 non-trivial[76], and studies built on longitudinal data could help disentangle its complexities.
403 Finally, although we refer to our brain age estimation in general terms, it is important to
404 reaffirm that it is based on T1-weighted imaging data. The brain is a complex and
405 heterogeneous organ, and different imaging modalities are known to capture different
406 aspects of the naturally occurring variation. Thus, studies relying on other modalities, either
407 independently or in combination, could reveal a broader set of associations[77].

408 In conclusion, the present study increases the yield of genetic associations with brain
409 age to eight genomic loci; implicated genes indicate involvement of calcium signaling, DNA
410 damage repair, protein metabolism, and general innate immune defense. Our analysis did not
411 provide evidence of a causal relationship between BAG and the included clinical conditions,
412 and their interactions remain unclear.

413

414 **Conflict of Interests**

415 All authors declared no conflict of interests.

416 **Funding**

417 This study is supported by Norwegian Research Council (No. 223273, 298646, 300767,
418 302854), the UiO:Life Science Convergence environment, University of Oslo, Norway

(4MENT), the South-Eastern Norway Regional Health Authority (2019101), KG Jebsen Stiftelsen, and the European Research Council under the European Union's Horizon 2020 research and Innovation program (ERC StG, Grant 802998).

**Author contributions**

EL, TW, and YW designed this study; EL, DVP, OF, AAS, OI, AMG and YW performed data analysis; EL, DVP, JMR, OI, TK, LTW, OOA, TW, BT, SMS, and YW interpreted the results; EL, DVP, JMR, OI, YW prepared the first draft of the manuscripts; All authors contributed and approved the final draft.

**Acknowledgement**

This research has been conducted using the UK Biobank Resource under Application Number 27412. We thank the computational resources provided by UNINETT Sigma2-the National Infrastructure for High Performance Computing and Data Storage in Norway – with project no. (nn9769k/ns9769k)**,** the PGC consortium and the International Parkinson Disease Genomics Consortium (IPDGC) for sharing of their GWAS results.

## References


1. Franke K, Gaser C. Ten Years of BrainAGE as a Neuroimaging Biomarker of Brain Aging: What Insights Have We Gained? *Frontiers in Neurology* 2019; **10:** 789.

2. Cole JH, Franke K. Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers. *Trends in neurosciences* 2017; **40**(12)**:** 681-690.

3. Franke K, Luders E, May A, Wilke M, Gaser C. Brain maturation: Predicting individual BrainAGE in children and adolescents using structural MRI. *NeuroImage* 2012; **63**(3)**:** 1305-1312.

4. Smith SM, Vidaurre D, Alfaro-Almagro F, Nichols TE, Miller KL. Estimation of brain age delta from brain imaging. *NeuroImage* 2019; **200:** 528-539.

5. Leonardsen EH, Peng H, Kaufmann T, Agartz I, Andreassen OA, Celius EG *et al.* Deep neural networks learn general and clinically relevant representations of the ageing brain. *NeuroImage* 2022; **256:** 119210.

6. Kaufmann T, van der Meer D, Doan NT, Schwarz E, Lund MJ, Agartz I *et al.* Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nature Neuroscience* 2019; **22**(10)**:** 1617-1623.

7. Cole JH, Ritchie SJ, Bastin ME, Valdés Hernández MC, Muñoz Maniega S, Royle N *et al.* Brain age predicts mortality. *Molecular Psychiatry* 2018; **23**(5)**:** 1385-1392.

8. Gaser C, Franke K, Klöppel S, Koutsouleris N, Sauer H, Alzheimer's Disease Neuroimaging I. BrainAGE in Mild Cognitive Impaired Patients: Predicting the Conversion to Alzheimer’s Disease. *PloS one* 2013; **8**(6)**:** e67346.

9. Schnack HG, van Haren NEM, Nieuwenhuis M, Hulshoff Pol HE, Cahn W, Kahn RS. Accelerated Brain Aging in Schizophrenia: A Longitudinal Pattern Recognition Study. *American Journal of Psychiatry* 2016; **173**(6)**:** 607-616.

10. Constantinides C, Han LK, Alloza C, Antonucci L, Arango C, Ayesa-Arriola R *et al.* Brain ageing in schizophrenia: evidence from 26 international cohorts via the ENIGMA Schizophrenia consortium. *medRxiv* 2022**:** 2022.2001.2010.21267840.

11. Han LKM, Dinga R, Hahn T, Ching CRK, Eyler LT, Aftanas L *et al.* Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. *Molecular Psychiatry* 2021; **26**(9)**:** 5124-5139.

12. Elliott ML, Belsky DW, Knodt AR, Ireland D, Melzer TR, Poulton R *et al.* Brain-age in midlife is associated with accelerated biological aging and cognitive decline in a longitudinal birth cohort. *Molecular Psychiatry* 2021; **26**(8)**:** 3829-3838.

13. Kuhn T, Kaufmann T, Doan NT, Westlye LT, Jones J, Nunez RA *et al.* An augmented aging process in brain white matter in HIV. *Human brain mapping* 2018; **39**(6)**:** 2532-2540.

14. Cole JH, Underwood J, Caan MWA, De Francesco D, van Zoest RA, Leech R *et al.* Increased brain-predicted aging in treated HIV disease. *Neurology* 2017; **88**(14)**:** 1349-1357.

15. Steffener J, Habeck C, O'Shea D, Razlighi Q, Bherer L, Stern Y. Differences between chronological and brain age are related to education and self-reported physical activity. *Neurobiology of Aging* 2016; **40:** 138-144.

16. Wrigglesworth J, Ward P, Harding IH, Nilaweera D, Wu Z, Woods RL *et al.* Factors associated with brain ageing - a systematic review. *BMC neurology* 2021; **21**(1)**:** 312.

17. Cole JH, Poudel RPK, Tsagkrasoulis D, Caan MWA, Steves C, Spector TD *et al.* Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage* 2017; **163:** 115-124.

18. Jonsson BA, Bjornsdottir G, Thorgeirsson TE, Ellingsen LM, Walters GB, Gudbjartsson DF *et al.* Brain age prediction using deep learning uncovers associated sequence variants. *Nature Communications* 2019; **10**(1)**:** 5409.

19. Ning K, Zhao L, Matloff W, Sun F, Toga AW. Association of relative brain age with tobacco smoking, alcohol consumption, and genetic variants. *Scientific Reports* 2020; **10**(1)**:** 10.

20. Smith SM, Elliott LT, Alfaro-Almagro F, McCarthy P, Nichols TE, Douaud G *et al.* Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations. *eLife* 2020; **9:** e52677.

21. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International journal of epidemiology* 2003; **32**(1)**:** 1-22.

22. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* 2018; **7**.

23. Kolbeinsson A, Filippi S, Panagakis Y, Matthews PM, Elliott P, Dehghan A *et al.* Accelerated MRI-predicted brain ageing and its associations with cardiometabolic and brain disorders. *Scientific Reports* 2020; **10**(1)**:** 19940.

24. Ségonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK *et al.* A hybrid approach to the skull stripping problem in MRI. *NeuroImage* 2004; **22**(3)**:** 1060-1075.

25. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *NeuroImage* 2012; **62**(2)**:** 782-790.

26. Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis* 2001; **5**(2)**:** 143-156.

27. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; **562**(7726)**:** 203-209.

28. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; **4:** 7.

29. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* 2017; **8**(1)**:** 1826.

30. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A *et al.* The Ensembl Variant Effect Predictor. *Genome Biology* 2016; **17**(1)**:** 122.

31. Benner C, Spencer CC, Havulinna AS, Salomaa V, Ripatti S, Pirinen M. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* 2016; **32**(10)**:** 1493-1501.

32. Yap CX, Henders AK, Alvares GA, Wood DLA, Krause L, Tyson GW *et al.* Autism-related dietary preferences mediate autism-gut microbiome associations. *Cell* 2021; **184**(24)**:** 5916-5931.e5917.

33. Consortium G. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015; **348**(6235)**:** 648-660.

34. Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Research* 2011; **39**(suppl_1)**:** D52-D57.

35. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. UniProtKB/Swiss-Prot. In: Edwards D (ed). *Plant Bioinformatics: Methods and Protocols*. Humana Press: Totowa, NJ, 2007, pp 89-112.

36. Trubetskoy V, Pardiñas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* 2022;**604**(7906):502-508.

37. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nature Genetics* 2021; **53**(6)**:** 817-829.

38. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics* 2018; **50**(5)**:** 668-681.

39. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics* 2019; **51**(3)**:** 404-413.

40. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet Neurology* 2019; **18**(12)**:** 1091-1102.

41. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics* 2015; **47**(11)**:** 1236-1241.

42. Lo MT, Wang Y, Kauppi K, Sanyal N, Fan CC, Smeland OB *et al.* Modeling prior information of common genetic variants improves gene discovery for neuroticism. *Human Molecular Genetics* 2017; **26**(22)**:** 4530-4539.

43. Chen CH, Wang Y, Lo MT, Schork A, Fan CC, Holland D *et al.* Leveraging genome characteristics to improve gene discovery for putamen subcortical brain structure. *Scientific Reports* 2017; **7**(1)**:** 15736.

44. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *International journal of epidemiology* 2012; **41**(1)**:** 161-176.

45. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genetic Epidemiology* 2016; **40**(4)**:** 304-314.

46. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International journal of epidemiology* 2015; **44**(2)**:** 512-525.

47. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature Genetics* 2018; **50**(5)**:** 693-698.

48. Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Annals Statistics* 2020; **48**(3)**:** 1742-1769.

49. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nature Genetics* 2020; **52**(7)**:** 740-747.

50. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* 2014; **46**(11)**:** 1173-1186.

51. Shigemizu D, Mitsumori R, Akiyama S, Miyashita A, Morizono T, Higaki S *et al.* Ethnic and trans-ethnic genome-wide association studies identify new loci influencing Japanese Alzheimer's disease risk. *Translational psychiatry* 2021; **11**(1)**:** 151.

52. Kunkle BW, Schmidt M, Klein H-U, Naj AC, Hamilton-Nelson KL, Larson EB *et al.* Novel Alzheimer Disease Risk Loci and Pathways in African American Individuals Using the African Genome Resources Panel: A Meta-analysis. *JAMA neurology* 2021; **78**(1)**:** 102-113.

53. Ikeda M, Takahashi A, Kamatani Y, Okahisa Y, Kunugi H, Mori N *et al.* A genome-wide association study identifies two novel susceptibility loci and trans population polygenicity associated with bipolar disorder. *Molecular Psychiatry* 2018; **23**(3)**:** 639-647.

54. Sanderson E, Richardson TG, Hemani G, Davey Smith G. The use of negative control outcomes in Mendelian randomization to detect potential population stratification. *International journal of epidemiology* 2021; **50**(4)**:** 1350-1361.

55. Lam K, Zhang DE. RUNX/CBF Transcription Factors☆. *Reference Module in Biomedical Sciences*. Elsevier2015.

56. Kotb M, Geller AM. Methionine adenosyltransferase: Structure and function. *Pharmacology & therapeutics* 1993; **59**(2)**:** 125-143.

57. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The Hallmarks of Aging. *Cell* 2013; **153**(6)**:** 1194-1217.

58. Helbig I, Lopez-Hernandez T, Shor O, Galer P, Ganesan S, Pendziwiat M *et al.* A Recurrent Missense Variant in AP2M1 Impairs Clathrin-Mediated Endocytosis and Causes Developmental and Epileptic Encephalopathy. *The American Journal of Human Genetics* 2019; **104**(6)**:** 1060-1072.

59. Mc Cormack A, Taylor J, Te Weehi L, Love DR, George AM. A Case of 17q21.31 Microduplication and 7q31.33 Microdeletion, Associated with Developmental Delay, Microcephaly, and Mild Dysmorphic Features. *Case Reports in Genetics* 2014; **2014:** 658570.

60. Arbogast T, Iacono G, Chevalier C, Afinowi NO, Houbaert X, van Eede MC *et al.* Mouse models of 17q21.31 microdeletion and microduplication syndromes highlight the importance of Kansl1 for cognition. *PLOS Genetics* 2017; **13**(7)**:** e1006886.

61. Frei O, Holland D, Smeland OB, Shadrin AA, Fan CC, Maeland S *et al.* Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nature Communications* 2019; **10**(1)**:** 2417.

62. van Rheenen W, Peyrot WJ, Schork AJ, Lee SH, Wray NR. Genetic correlations of polygenic disease traits: from theory to practice. *Nature Reviews Genetics* 2019; **20**(10)**:** 567-581.

63. Van Gestel H, Franke K, Petite J, Slaney C, Garnham J, Helmick C *et al.* Brain age in bipolar disorders: Effects of lithium treatment. *Australian & New Zealand Journal of Psychiatry* 2019; **53**(12)**:** 1179-1188.

64. Pasco JA, Williams LJ, Jacka FN, Ng F, Henry MJ, Nicholson GC *et al.* Tobacco smoking as a risk factor for major depressive disorder: population-based study. *British Journal of Psychiatry* 2008; **193**(4)**:** 322-326.

65. Winterer G. Why do patients with schizophrenia smoke? *Current opinion in psychiatry* 2010; **23**(2).

66. Sanders A-M, Richard G, Kolskår K, Ulrichsen KM, Kaufmann T, Alnæs D *et al.* Linking objective measures of physical activity and capability with brain structure in healthy community dwelling older adults. *NeuroImage: Clinical* 2021; **31:** 102767.

67. Brokmeier LL, Firth J, Vancampfort D, Smith L, Deenik J, Rosenbaum S *et al.* Does physical activity reduce the risk of psychosis? A systematic review and meta-analysis of prospective studies. *Psychiatry research* 2020; **284:** 112675.

68. Wolfers T, Doan NT, Kaufmann T, Alnaes D, Moberget T, Agartz I *et al.* Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models. *JAMA Psychiatry* 2018; **75**(11)**:** 1146-1155.

69. Wolfers T, Rokicki J, Alnæs D, Berthet P, Agartz I, Kia SM *et al.* Replicating extensive brain structural heterogeneity in individuals with schizophrenia and bipolar disorder. *Human brain mapping* 2021; **42**(8)**:** 2546-2555.

70. Taschler B, Smith SM, Nichols TE. Causal inference on neuroimaging data with Mendelian randomisation. *NeuroImage* 2022; **258:** 119385.

71. Eickhoff CR, Hoffstaedter F, Caspers J, Reetz K, Mathys C, Dogan I *et al.* Advanced brain ageing in Parkinson's disease is related to disease duration and individual impairment. *Brain Communications* 2021; **3**(3)**:** fcab191.

72. Charissé D, Erus G, Pomponio R, Gorges M, Schmidt N, Schneider C *et al.* Brain age and Alzheimer's-like atrophy are domain-specific predictors of cognitive impairment in Parkinson's disease. *Neurobiology of aging* 2022; **109:** 31-42.

73. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *International journal of epidemiology* 2016; **45**(6)**:** 1866-1886.

74. Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR *et al.* Mendelian randomization. *Nature Reviews Methods Primers* 2022; **2**(1)**:** 6.

75. Büchlmann P, Yu B. Analyzing Bagging. *The Annals of Statistics* 2002; **30**(4)**:** 927-961.

76. Vidal-Pineiro D, Wang Y, Krogsrud SK, Amlien IK, Baaré WFC, Bartres-Faz D *et al.* Individual variations in 'brain age' relate to early-life factors more than to longitudinal brain change. *eLife* 2021; **10:** e69995.

77. Rokicki J, Wolfers T, Nordhøy W, Tesli N, Quintana DS, Alnæs D *et al.* Multimodal imaging improves brain age prediction and reveals distinct abnormalities in patients with psychiatric and neurological disorders. *Human brain mapping* 2021; **42**(6)**:** 1714-1726.

729 FIGURE LEGENDS

730 **Figure 1 Genetic associations for brain age gap.**

731 **a.** The Manhattan plot of meta-analyzed association results for brain age gap (BAG).
732 Chromosome numbers are shown on x axis, -log10 association p values on y axis and lead
733 SNP rs-numbers in the plot. **b.** Region plots for each of the eight associated regions. Genes
734 located in each region are shown below each figure. Linkage disequilibrium r-squared values
735 are indicated by colors; and recombination frequences by curves. **c.** Expression levels of the
736 annotated genes across tissues analyzed by the GTEx v8 study. Colors indicate average log2
737 transformed expression level in each tissue.

738

739 **Figure 2 Polygenic genetic overlap between brain age gap and disorders.**

740 **a.** Genetic correlation between brain age gap and disorders computed by ldsc. SNP
741 heritability and its standard error are indicated. **b-d.** Conditional QQ plot between brain age
742 gap and disorders in both directions. Colors are used to indicate different association
743 strength to the conditioned traits, i.e., the ones indicated after the vertical bar in each
744 figure. Dashed diagonal lines indicate expected null distributions. Abbreviations used: AD,
745 Alzheimer's disease; BIP, bipolar disorder; MDD, major depression disorder; PD, Parkinson's
746 disease; SCZ, schizophrenia.

747

748 **Figure 3 Causal inference between brain age gap and disorders.**

749 **a.** Causal effect of brain age gap (BAG) on risk of disorders; **b.** Causal effect of genetic risk of
750 disorders on BAG. Colors indicate different models; triangle indicates significant effect after
751 false discovery correction. Estimated standard errors for each effect are aslo shown. **c.**
752 Scatter plots of SNP effects on AD (x axis) and BAG (y axis). **d.** Scatter plots of SNP effects on

753 BIP (x axis) and BAG (y axis). **e.** Scatter plots of SNP effects BAG (x axis) and PD (y axis). **f.**

754 Scatter plots of SNP effects on PD (x axis) and BAG (y axis). Causal effects estimated by the

755 five models (except CAUSE) are shown by fitted lines; slopes of these lines indicate causal

756 effect sizes. Exceptional SNPs are marked by boxes which include SNP rs-numbers and

757 genome location in th hg19 coordinates. Abbreviation used: AD, Alzheimer's disease; BIP,

758 bipolar disorder; MDD, major depression disorder; PD, Parkinson's disease; SCZ,

759 schizophrenia.

760
761

762

763 **Table 1 Genomic loci associated with BAG.**

| Locus | Lead SNP | POS | Gene | A1 | A2 | Beta | $I^2$ | P |
|---|---|---|---|---|---|---|---|---|
| **Chr3:183892867-183975709** | rs73185796 | 183975709 | CAMK2N2/ECE2 | T | G | -0.29 | 0.0 | $2.53x10^{-8}$ |
| **Chr4:38591172-38779512** | rs13132853 | 38680015 | KLF3 | G | A | 0.23 | 0.0 | $2.34x10^{-18}$ |
| **Chr5:78388694-78451813** | rs79107704 | 78388694 | BHMT2 | A | G | 0.63 | 0.0 | $1.65x10^{-8}$ |
| **Chr6:45407654-45511945** | rs2790102 | 45432214 | RUNX2 | A | G | -0.15 | 0.0 | $8.92x10^{-9}$ |
| **Chr8:124661974-124682971** | rs7461069 | 124669029 | KLHL38 | A | G | -0.17 | 0.0 | $1.57x10^{-8}$ |
| **Chr10:134544247-134597265** | rs4880424 | 134584577 | INPP5A | T | C | 0.16 | 64.95 | $3.69x10^{-8}$ |
| **Chr14: 88391116-88556525** | rs17203398 | 88449847 | GALC | C | G | -0.16 | 40.97 | $1.42x10^{-10}$ |
| **Chr17: 43101281-44863413** | rs2106786 | 43919096 | MAPT | G | A | 0.29 | 0.0 | $1.87x10^{-23}$ |

764 Eight independent genomic loci significantly associated with brain age gap (BAG). Lead SNP

765 rs-number, genomic position (in hg19 coordinates), effective allele (A1), the other allele

766 (A2), effect size (Beta), meta-analysis heterogeneity ($I^2$), association strength (p value, P).

767

768

769

FIGURES

Figure 1

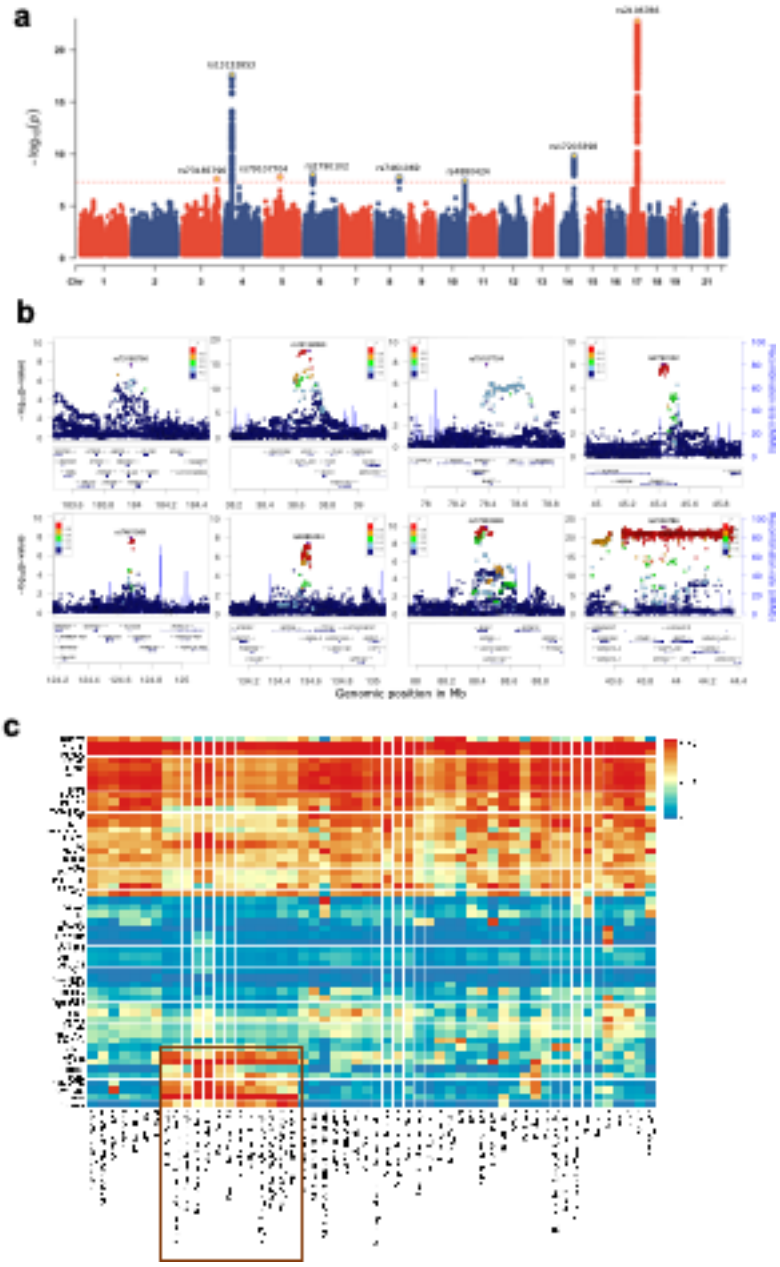


Figure 2

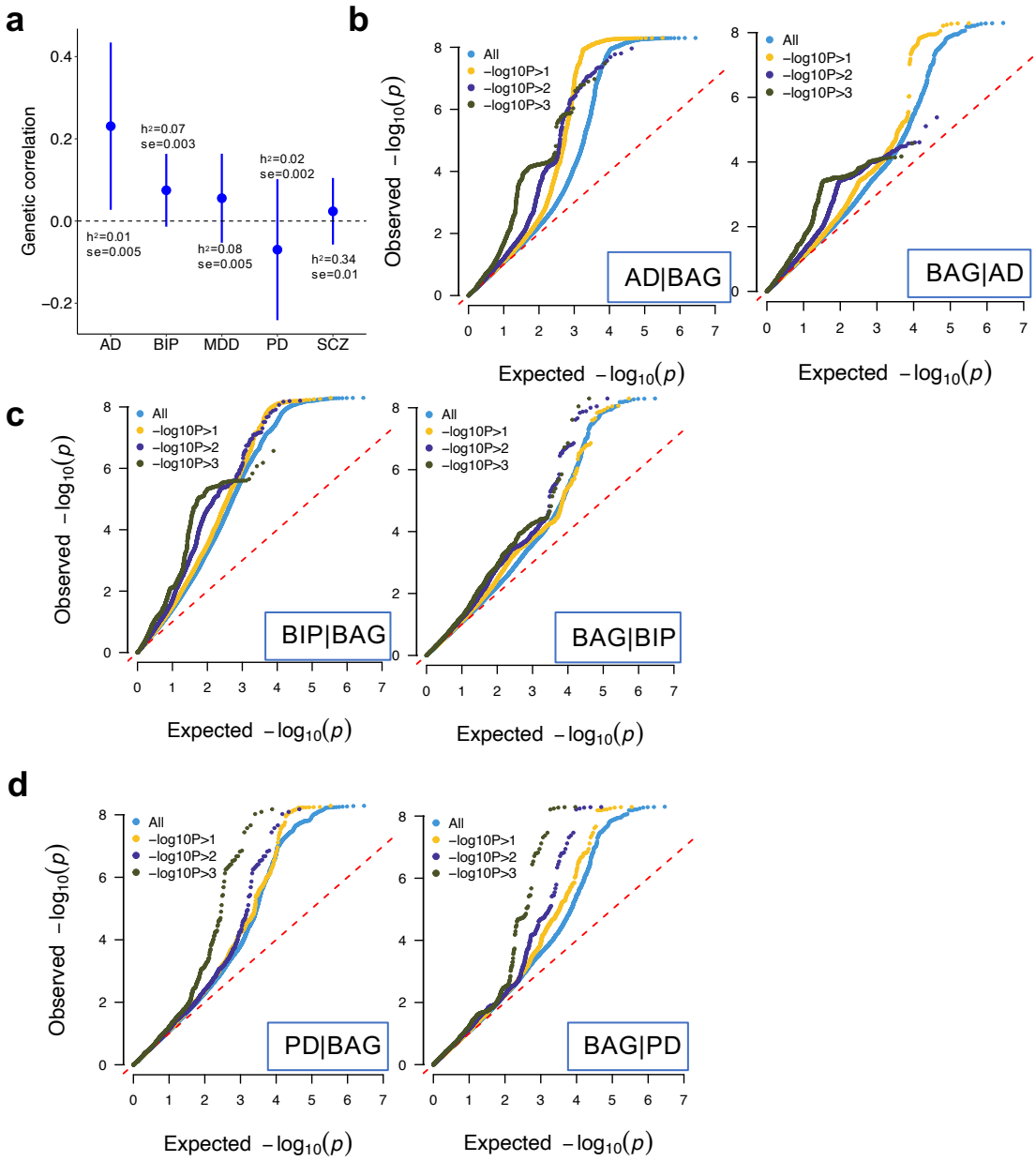
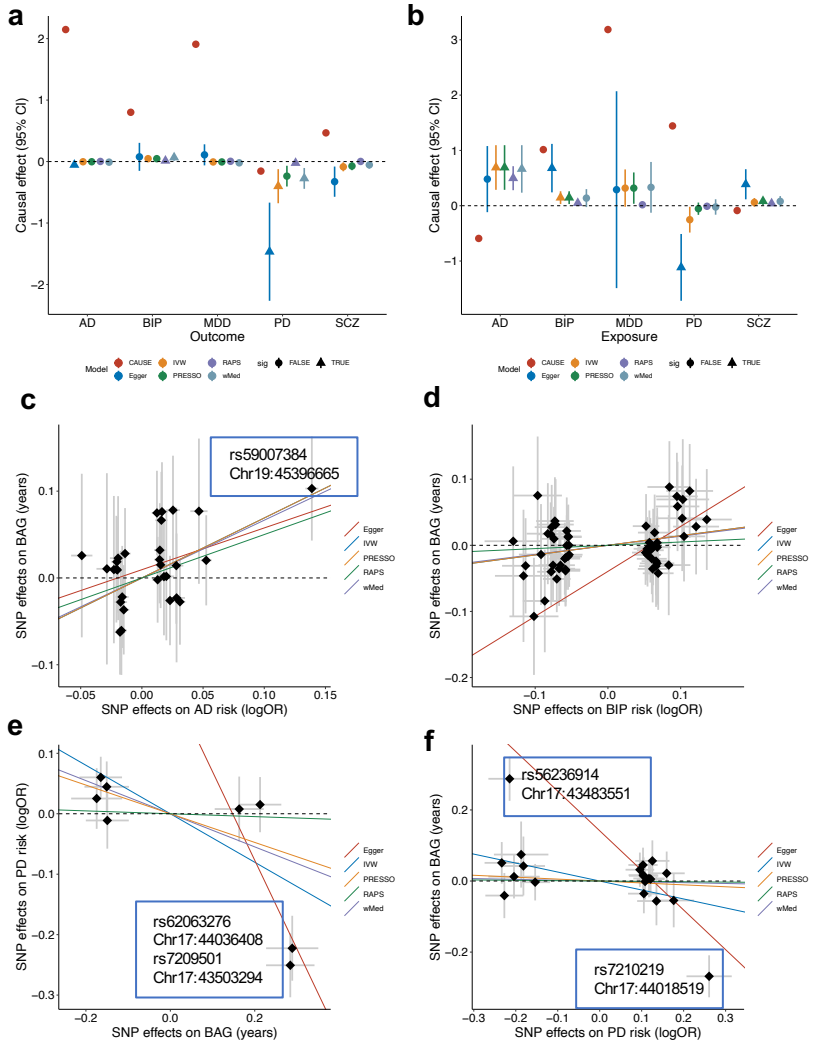

a
Genetic correlation
0.4
0.2
0.0
−0.2
h²=0.01
se=0.005
h²=0.07
se=0.003
h²=0.08
se=0.005
h²=0.02
se=0.002
h²=0.34
se=0.01
AD
BIP
MDD
PD
SCZ
b
Observed −log10(p)
Expected −log10(p)
All
−log10P>1
−log10P>2
−log10P>3
AD|BAG
BAG|AD
c
BIP|BAG
BAG|BIP
d
PD|BAG
BAG|PD

Figure 3


a
Causal effect (95% CI)
AD
BIP
MDD
PD
SCZ
Outcome
b
Causal effect (95% CI)
Exposure
Model
CAUSE
IVW
RAPS
Egger
PRESSO
wMed
sig
FALSE
TRUE
c
rs59007384
Chr19:45396665
SNP effects on BAG (years)
SNP effects on AD risk (logOR)
d
SNP effects on BAG (years)
SNP effects on BIP risk (logOR)
e
SNP effects on PD risk (logOR)
rs62063276
Chr17:44036408
rs7209501
Chr17:43503294
SNP effects on BAG (years)
f
rs56236914
Chr17:43483551
SNP effects on BAG (years)
rs7210219
Chr17:44018519
SNP effects on PD risk (logOR)