Genome-wide analysis of brain age identifies 59 associated loci and unveils relationships with mental and physical health

Check for updates

Philippe Jawinski [1,2] ✉, Helena Forstbach[1], Holger Kirsten [2,3], Frauke Beyer [4], Arno Villringer [4], A. Veronica Witte [4], Markus Scholz [2,3], Stephan Ripke [5,6] & Sebastian Markett [1]

Neuroimaging and machine learning are advancing research into the mechanisms of biological aging. In this field, 'brain age gap' has emerged as a promising magnetic resonance imaging-based biomarker that quantifies the deviation between an individual's biological and chronological age of the brain. Here we conducted an in-depth genomic analysis of the brain age gap and its relationships with over 1,000 health traits. Genome-wide analyses in up to 56,348 individuals unveiled a heritability of 23–29% attributable to common genetic variants and highlighted 59 associated loci (39 novel). The leading locus encompasses *MAPT*, encoding the tau protein central to Alzheimer's disease. Genetic correlations revealed relationships with mental health, physical health, lifestyle and socioeconomic traits, including depressed mood, diabetes, alcohol intake and income. Mendelian randomization indicated a causal role of high blood pressure and type 2 diabetes in accelerated brain aging. Our study highlights key genes and pathways related to neurogenesis, immune-system-related processes and small GTPase binding, laying the foundation for further mechanistic exploration.

Aging is a complex phenomenon inherent to most organisms[1–3]. As human lifespans extend and global populations age, age-related disabilities, including dementia, are rising[4]. Thus, understanding the biological mechanisms of aging is an urgent priority for social systems, to sustain longer lives with reduced periods of disability.

The use of neuroimaging methods in conjunction with machine learning has become a promising avenue in biomedical research to capture an individual's biological age, particularly 'brain age'[5,6]. Brain age is typically assessed by training an age prediction model on in vivo magnetic resonance imaging (MRI) data from a normative lifespan sample. This model is then applied to the MRI data of unseen individuals to predict their age. The discrepancy between an individual's brain-predicted and chronological age is termed 'brain age gap' (BAG) and it is used to infer typical and atypical aging trajectories[6,7].

A positive BAG, interpreted as accelerated aging, has been linked to reduced mental and physical health[5]; including weaker grip strength, higher blood pressure, diabetes, adverse drinking and smoking behavior, poorer cognitive abilities and depression[8–13]. BAG is also enhanced in neurological and psychiatric disorders such as Alzheimer's disease (AD), schizophrenia and bipolar disorder[14,15]. While previous genetic studies suggested that BAG exhibits a substantial heritable component, few genetic variants have been identified[15–24]. To refine the genetic architecture of BAG and identify potential therapeutic targets for healthy aging, further research is imperative.

[1]Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany. [2]LIFE-Leipzig Research Center for Civilization Diseases, Leipzig University, Leipzig, Germany. [3]Institute for Medical Informatics, Statistics and Epidemiology, Leipzig University, Leipzig, Germany. [4]Cognitive Neurology, University of Leipzig Medical Center & Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany. [5]Stanley Center for Psychiatric Research, Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA. [6]Department of Psychiatry and Psychotherapy, Charité-Universitätsmedizin, Berlin, Germany. ✉e-mail: philippe.jawinski@hu-berlin.de
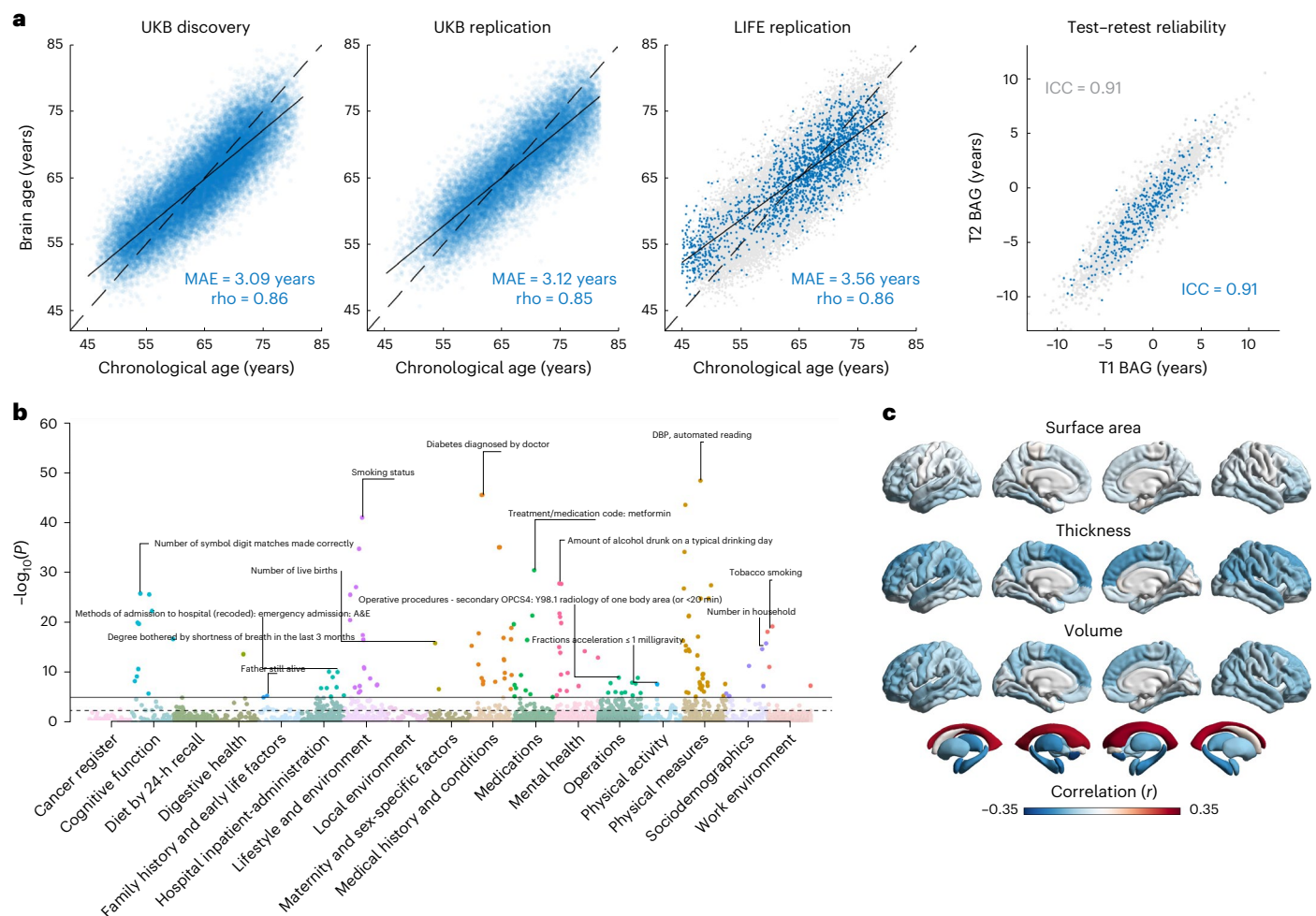
Fig. 1 | Phenotypic characteristics and associations of combined BAG.
a, The blue dots in the first three plots (left to right) show brain-predicted age estimates (combined GM and WM) plotted against chronological age in the UKB discovery sample ($n = 32,634$), UKB replication sample ($n = 21,881$, merged across ancestries) and the LIFE-Adult replication sample ($n = 1,833$). To facilitate comparisons, the results of the UKB discovery sample are also shown as gray dots in the background of the LIFE replication plot. At this stage, brain-predicted age estimates have not yet been bias-corrected for regression dilution, as indicated by the solid linear regression line crossing the dashed identity line. The fourth plot shows the test–retest reliabilities of combined BAG in a subset of the UKB discovery (gray dots, $n = 3,751$) and UKB replication sample (blue dots, $n = 395$). BAG was residualized for sex, age, age^2, scanner site and total intracranial

volume. **b**, Cross-trait association results between combined BAG and 7,088 UKB phenotypes across several health domains. Analyses were conducted using PHESANT, which applies data-type-specific regression models (linear, logistic, ordered logistic or multinomial logistic regression). All models included sex, age, age^2, scanner site and total intracranial volume as covariates. The horizontal lines indicate the Bonferroni-adjusted (solid) and FDR-adjusted (dashed) two-sided level of significance. The top associations per category are annotated. **c**, Surface plots showing the correlations between combined BAG and 220 FreeSurfer brain structure variables. The colors reflect the strength and direction of partial product-moment correlations (sex, age, age^2, scanner site and total intracranial volume served as covariates). ICC, intraclass correlation coefficient (C, 1); rho, product-moment correlation coefficient.

In this article, we present what is to our knowledge the largest genome-wide association study (GWAS) of BAG to date. We begin by discovering new loci in a sample of 32,634 individuals of White British ancestry and replicate our findings in a multi-ancestry sample of up to 23,714 individuals. Next, we conduct meta-analyses across the discovery and replication samples, with an aggregated sample size of up to 56,348 individuals. This represents a 79% increase (~25,000 more) over previous GWAS[23,24]. To prioritize genes, we use complementary fine-mapping, annotation and transcriptomic analyses that integrate multiple omics resources. We also calculate polygenic scores (PGS) to estimate the present yield in variance explanation, compute genetic correlations with over 1,000 traits and test causal effects using Mendelian randomization. Finally, we examine the degree of polygenicity of BAG and project future discovery potential. Through these efforts, we unravel new biological mechanisms behind BAG, including pathways related to neurogenesis, immune system processes and binding of small GTPases—evolutionarily conserved proteins that act as cellular timers.

Results

We estimated brain age with a well-established and extensively validated workflow based on CAT12 voxel-based morphometry[5,25]. Using T1-weighed MRI scans and supervised machine learning, we estimated brain age through cross-prediction in a discovery sample of 32,634 individuals of White British ancestry from the UK Biobank (UKB) cohort (age range = 45–81 years)[26]. To capture tissue-specific aging patterns, we conducted separate analyses for gray matter (GM) and white matter (WM) segmentations[18]. Brain age estimation relied on an ensemble of complementary machine learning algorithms: the sparse Bayesian relevance vector machine[27] (RVM) and extreme gradient boosting (XGBoost) with tree and linear boosters[28]. Models were stacked within and across tissue classes, yielding three brain-predicted age estimates per individual: for GM, WM and combined GM and WM.

In the discovery sample, we observed accurate predictions for chronological age, with mean absolute errors (MAEs) reaching 3.09 years and correlation coefficients attaining $r = 0.86$ (Fig. 1a, Table 1,

Table 1 | Prediction accuracies of the stacked age estimation models stratified according to tissue class

	UKB discovery (*n*=32,634) 45–81 years				UKB replication (*n*=21,881) 45–81 years				LIFE replication (*n*=1,833) 45–80 years		
	r	R^2	MAE	ICC$_{BAG}$	*r*	R^2	MAE	ICC$_{BAG}$	*r*	R^2	MAE
GM	0.827	0.683	3.372	0.899	0.825	0.676	3.405	0.888	0.828	0.653	3.990
WM	0.835	0.696	3.307	0.920	0.830	0.683	3.368	0.911	0.829	0.667	3.979
GM and WM	0.857	0.734	3.089	0.915	0.854	0.726	3.123	0.908	0.862	0.729	3.557

The imaging data of the UKB discovery sample were released until January 2020 (release 1.7), whereas the data of the UKB replication sample were released until May 2024 (release 1.10). *r* indicates the product-moment correlation between brain-predicted age (without bias correction) and chronological age. R^2 indicates the coefficient of determination (not equivalent to r^2). ICC$_{BAG}$ indicates the ICC between the test and retest assessment of BAG. MAE is the MAE of brain-predicted versus chronological age. BAG was bias-corrected for age, age^2, sex, scanner site and total intracranial volume. ICCs are based on a subset of 3,751 individuals in the UKB discovery sample and 395 individuals in the UKB replication sample.

Supplementary Table 1 and Extended Data Fig. 1). Model performances were similar in the multi-ancestry UKB replication samples (*n* = 21,881; age range = 45-81 years) and the European-ancestry Leipzig Research Centre for Civilization Diseases (LIFE)-Adult replication sample (*n* = 1,833; age range = 45-80 years)[29,30]. Genetic association analyses were performed on BAG, that is, the difference between brain-predicted and chronological age. These BAG estimates—residualized for sex, age, age^2, scanner site and total intracranial volume—showed high test–retest reliabilities, with intraclass correlation coefficients (ICCs) ranging from 0.89 to 0.92.

Phenotypic associations

To validate our BAG estimates and extend previous evidence on their health relevance[31,32], we conducted cross-trait association analyses between BAG and 7,088 non-imaging-derived phenotypes using PHESANT[33]. A total of 210 associations reached Bonferroni significance ($P < 7.1 \times 10^{-6}$) for at least 1 of the 3 BAG traits (Supplementary Table 2 and Extended Data Fig. 2). Figure 1b presents the cross-trait results for combined GM and WM BAG.

Top associations for combined BAG (all $P \leq 1.8 \times 10^{-12}$) included pack years of smoking (*r* = 0.091), diastolic blood pressure (DBP) (*r* = 0.084), number of symbol digit matches made correctly (that is, a measure of cognitive performance; *r* = −0.082), diabetes diagnosed by doctor (*r* = 0.079), amount of alcohol drunk on a typical drinking day (*r* = 0.076) and overall health rating (*r* = 0.039; note that higher scores indicate poorer health). These results corroborate earlier BAG associations[31] and expand known health-related links.

To examine regional contributions to BAG and facilitate comparisons with prior surface-based morphometry studies[15,19,34], we analyzed associations with FreeSurfer-derived[35] cortical surface measures and subcortical volumes (Fig. 1c, Supplementary Fig. 3 and Supplementary Table 3). For combined BAG, the strongest associations (all $P \leq 1.1 \times 10^{-209}$) were observed with the volumes of the accumbens (*r* = −0.31), lateral ventricles (*r* = 0.30), amygdala (*r* = −0.25), hippocampus (*r* = −0.23) and thalamus (*r* = −0.22), and the cortical thickness of the superior frontal (*r* = −0.20) and inferior parietal (*r* = −0.17) cortex. These findings suggest that our models capture patterns of aging distributed throughout the brain, rather than being confined to specific areas.

Following up on Smith et al.[31], we also performed sex-stratified analyses, which revealed largely similar BAG associations in males and females for non-imaging-derived (Supplementary Figs. 1 and 2 and Supplementary Table 4) and brain structure phenotypes (Supplementary Figs. 4 and 5 and Supplementary Table 5). Some differences emerged, for example, stronger associations between GM BAG and body fat percentage in males, but were generally small in magnitude.

Concordant genomic signals in discovery and replication GWAS

An overview of our genomic analysis workflow is shown in Fig. 2. To examine the reliability and replicability of our findings, we first compared results from a discovery GWAS of 32,634 individuals of White British ancestry (UKB imaging release v.1.7) with a replication GWAS of 22,256 individuals of European ancestry (UKB imaging release v.1.10 and LIFE-Adult).

The discovery and replication GWAS results were highly consistent, with strong genetic correlations derived from bivariate linkage disequilibrium (LD) score regression (LDSC) (all r_g > 0.996, all $P > 2.9 \times 10^{-36}$; Supplementary Table 6)[36]. We discovered 25 independent genome-wide significant loci across the 3 BAG traits, all showing concordant effect directions (binomial test: $P = 3.0 \times 10^{-08}$), and 18 reaching one-tailed nominal significance in replication (binomial test: $P = 1.3 \times 10^{-18}$; Supplementary Table 7). These findings closely align with our power analysis, which predicted 19 nominal replications. Among 45 additional suggestive discoveries ($P < 1.0 \times 10^{-6}$), 36 showed concordant directions (binomial test: $P = 3.3 \times 10^{-5}$) and 24 reached nominal significance in replication (binomial test: $P = 8.0 \times 10^{-20}$). Incorporating 1,458 non-European UKB participants into an extended multi-ancestry GWAS also yielded above-chance consistency (Supplementary Table 8). Together, these findings strongly support locus replicability, reinforcing the robustness of our results. Additional details are provided in Supplementary Figs. 6–23.

Identification of 59 associated loci

To maximize statistical power and improve genetic discovery, we meta-analyzed GWAS data from 54,890 individuals of European ancestry, combining the UKB discovery (*n* = 32,634), UKB-EUR replication (*n* = 20,423) and LIFE-Adult (*n* = 1,833) cohorts. Analyses included 9.6 million single-nucleotide polymorphisms (SNPs) and insertions and deletions (indels) with a minor allele frequency (MAF) greater than 0.01 and imputation quality score (INFO) greater than 0.80. We modeled additive genetic effects with covariates for sex, age, age^2, total intracranial volume, scanner site and type of genotyping array, and up to 20 genetic principal components. Results for the three BAG traits are shown in Fig. 3 (multi-ancestry results are shown in Extended Data Fig. 3).

LDSC intercepts did not indicate a bias of test statistics due to reasons other than polygenicity, suggesting no confounding inflation caused by population stratification (intercept range: 1.011–1.017; Supplementary Table 9)[37]. SNP-based heritability estimates ranged from 22.7% (GM BAG) to 29.0% (WM BAG). The genetic correlation between GM and WM BAG was r_g = 0.74 (s.e. = 0.023), indicating both shared and distinct genetic influences (phenotypic correlation: r_P = 0.62, s.e. = 0.003). Combined BAG showed strong genetic correlations with both GM (r_g = 0.91, s.e. = 0.008; cf. r_P = 0.88, s.e. = 0.002) and WM (r_g = 0.951, s.e. = 0.006; cf. r_P = 0.91, s.e. = 0.002) BAG. Strong genetic correlations were observed between sex-stratified results (all $r_g \geq 0.934$; Supplementary Table 10), indicating highly concordant architectures across sexes.

Partitioned LDSC[38] (Supplementary Table 11 and Extended Data Fig. 4) revealed an enrichment of heritability (false discovery
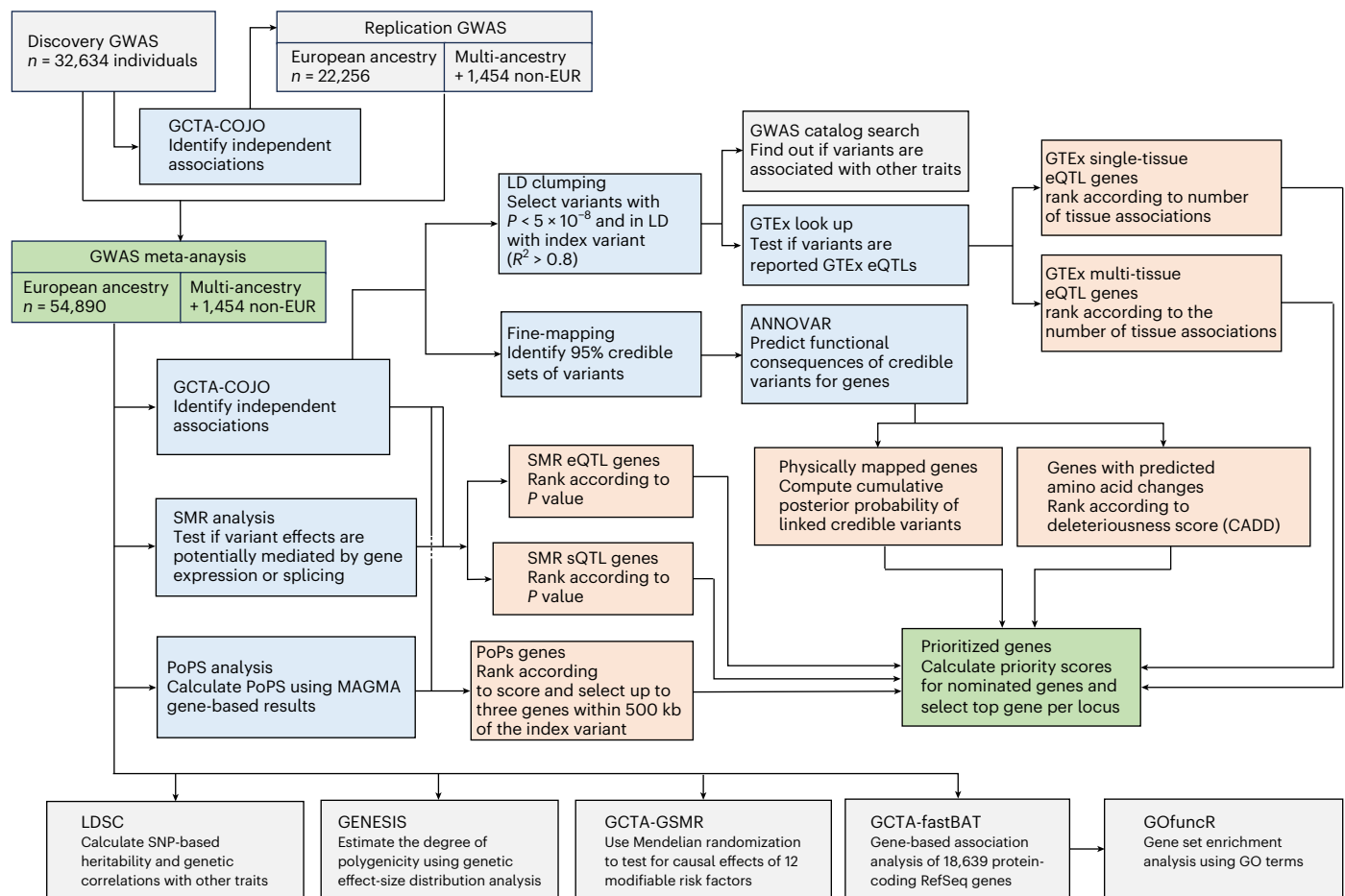
Fig. 2 | Genomic analysis workflow and gene prioritization strategy. Overview of the genomic analysis workflow with particular emphasis on the gene prioritization procedure. The green boxes represent data input (GWAS meta-analysis) and output (prioritized genes). The blue boxes represent analyses whose outcomes were used for gene nomination and subsequent prioritization.

The apricot-colored boxes reflect the gene nomination categories. The gray boxes reflect all other analyses carried out to refine the genetic architecture of BAG, such as heritability and polygenicity analyses. Genes were prioritized by integrating data from multiple strategies, such as functional annotation of credible variants, SMR, GTEx eQTL lookup and PoPS.

rate (FDR) < 0.05) in regions conserved across mammals (fold enrichment (FE) = 12.4) and primates (FE = 10.5). Additional enrichment was observed in super-enhancer (FE = 2.6), flanking bivalent transcriptional start sites/enhancers (FE = 8.7) and epigenetically modified H3K27ac (FE = 1.8) and H3K9ac (FE = 2.9) regions. Cell-type group analyses (Supplementary Table 12 and Extended Data Fig. 5) revealed enrichment near genes expressed in central nervous system (FE = 3.4), connective or bone (FE = 3.0) and kidney (FE = 3.7) tissue.

To identify independent genome-wide significant associations, we conducted stepwise conditional analyses using genome-wide complex trait analysis–conditional and joint association analysis (GCTA–COJO)[39]. This resulted in 26, 34 and 39 independent discoveries for GM, WM and combined BAG, respectively (regional plots are shown in Supplementary Figs. 25–42). After cross-trait LD clumping of index variants ($r^2 > 0.1$, 10,000-kb window), we identified 59 distinct loci (≥460 kb apart; Table 2 and Supplementary Table 13). Of these, 39 represent novel discoveries not previously reported in BAG GWAS (see Methods for the definition of novelty)[17–24].

We observed most index variants in intronic regions of protein-coding genes. ANNOtate VARiation (ANNOVAR) enrichment tests confirmed that variants in high LD with the lead variants were underrepresented in intergenic regions and overrepresented in 3′-UTR, 5′-UTR, intronic, exonic noncoding RNA and intronic noncoding RNA regions (Supplementary Fig. 24 and Supplementary Table 14).

Fine-mapping and gene prioritization

To identify putative causal genes, we used several fine-mapping, functional annotation and transcriptomic analyses that integrate information from multiple omics resources. For each genome-wide significant locus, we (1) constructed 95% credible sets of variants that likely include the causal variant using SBayesRC[40,41], complemented by susieR and FINEMAP[42,43]; (2) physically mapped credible variants to genes using ANNOVAR[44]; (3) predicted the transcript consequences of nonsynonymous exonic variants and scored their deleteriousness using combined annotation dependent depletion (CADD)[45]; (4) mapped variants to genes using expression quantitative trait locus (eQTL) lookup in 49 Genotype-Tissue Expression (GTEx) Project v.8 tissues[46]; (5) conducted summary-data-based Mendelian randomization (SMR)[47] with the RNA sequence (RNA-seq) data of 2,865 brain cortex samples[48] to test for mediation through gene expression and splicing; and (6) calculated polygenic priority scores (PoPS)[49] that incorporate data from single-cell RNA-seq datasets, curated biological pathways and protein–protein interaction networks. We integrated these results to compute a gene priority score and selected the most plausible candidate per locus (Methods). Figure 2 shows an overview of the analysis workflow; full results are found in Supplementary Table 13 (with further details in Supplementary Tables 14–24). The key findings are summarized below.

Across the 59 discovered loci, SBayesRC genome-wide fine-mapping produced, on average, the smallest 95% credible sets (median = 9 variants), compared to region-specific approaches using
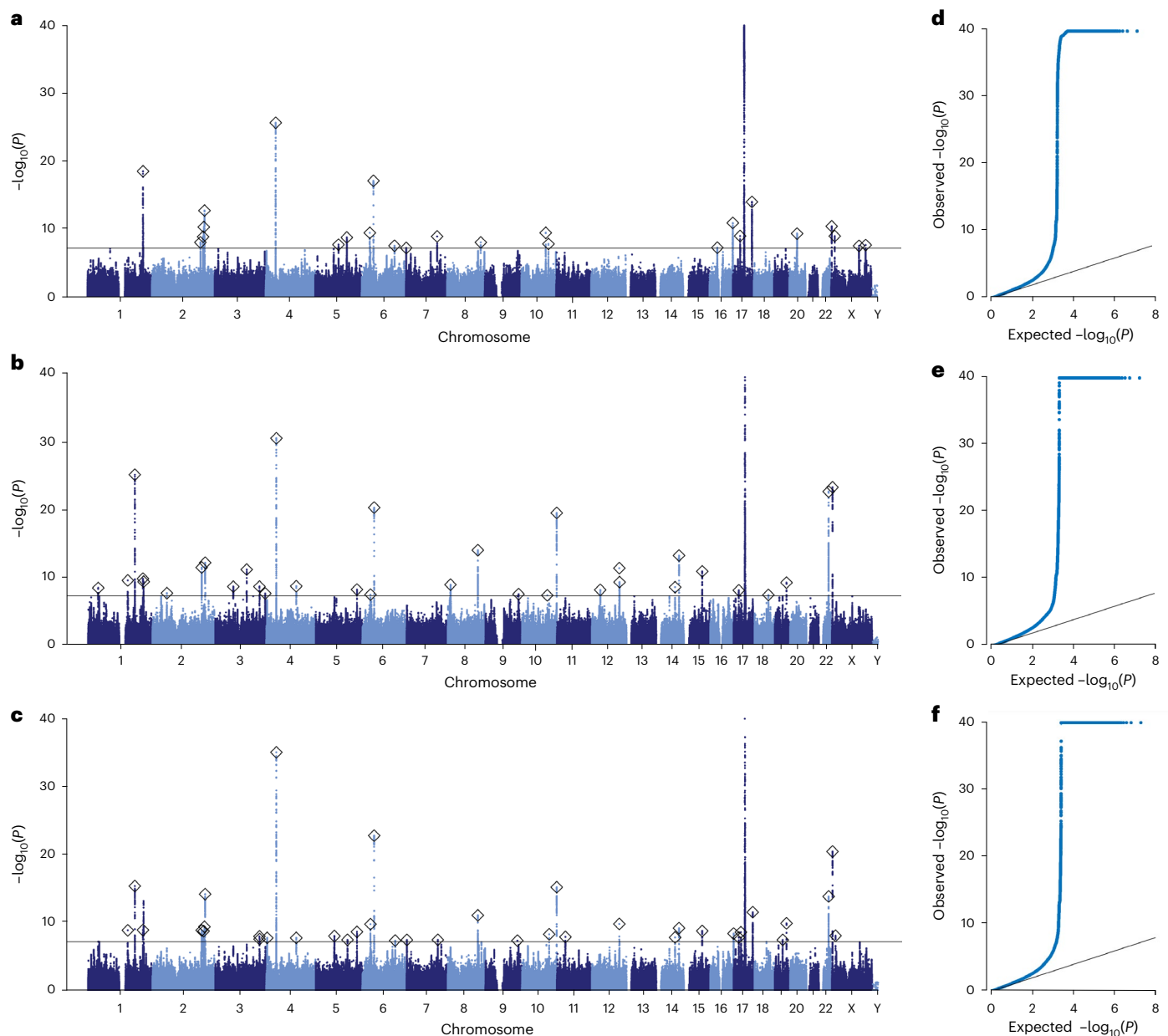
Fig. 3 | Genome-wide association meta-analyses of BAG traits. Manhattan (**a**–**c**) and quantile–quantile (QQ) plots (**d**–**f**) showing the results of the European-ancestry GWAS meta-analyses for the three BAG traits (n = 54,890). The Manhattan plots show the P values ($-\log_{10}$ scale) of the tested genetic variants on the y axis and base-pair positions along the chromosomes on the x axis. P values were derived from two-sided linear regression models using PLINK, followed by meta-analysis using inverse-variance weighting in METAL. The solid horizontal line indicates the threshold of genome-wide significance (two-sided $P = 5.0 \times 10^{-8}$, accounting for multiple testing). Index variants are highlighted by the diamonds. The results of the pseudoautosomal variants have been added to chromosome X. Quantile–quantile plots show the observed P values from the association analysis versus the expected P values under the null hypothesis of no effect ($-\log_{10}$ scale). For illustrative reasons, the y axis has been truncated at $P = 1.0 \times 10^{-40}$. **a,d**, GM BAG (Manhattan and QQ). **b,e**, WM BAG (Manhattan and QQ). **c,f**, Combined GM and WM BAG (Manhattan and QQ).

susieR (median = 34 variants) and FINEMAP (median = 36 variants). susieR and FINEMAP showed strong concordance, with a median overlap of 97.2%. Variants from SBayesRC were included in the susieR and FINEMAP credible sets at median rates of 75.0% and 80.0%, respectively. Most loci contained a single signal, although susieR and FINEMAP identified four loci with potential secondary signals. Estimated regional heritability ranged from 0.03% to 0.67%, reflecting modest variant effects, with few exceptions.

We observed the strongest association at locus 17q21.31 (index variant: rs2260227, $P = 9.4 \times 10^{-83}$), which tags a well-known 900-kb inversion polymorphism[50,51]. Consistent with the strong LD cluster in the inverted region[50], the region-specific fine-mapping approaches

yielded large credible sets of variants (>1,500 variants). A National Human Genome Research Institute GWAS Catalog search[52] revealed associations with many locus-associated traits, including educational attainment[53], depressed affect[54], alcohol consumption[55], sleep duration[56], lung function[57], male puberty timing[58], age at onset of menarche[57] and AD[59]. The region spans multiple genes, including *MAPT, STH, KANSL1* and *CRHR1*. Several genome-wide significant variants in these genes are GTEx single-tissue and multi-tissue eQTLs (Supplementary Tables 21 and 22). SMR analyses implicated expression and splicing of *MAPT* and *KANSL1* (along with other genes) in mediating variant effects on BAG (Supplementary Tables 19 and 20). We also identified credible exonic variants causing amino acid changes

Table 2 | Identification of 59 genomic loci associated with BAG in _n_=54,890 individuals

Chr	Position	ID	A1/A2	Frequency of A1	Beta (s.e.)	P	Credible	Prioritized gene	Phenotype(s)	Ref.
1	39471759	1:39471759	CT/C	0.89	0.20 (0.03)	4×10^{-9}	3	_NDUFS5_	WM	*
1	153888217	rs552429854	T/TA	0.37	0.14 (0.02)	3×10^{-10}	6	_DENND4B_	WM, GM and WM	18
1	180956936	rs35306826	A/T	0.59	−0.22 (0.02)	1×10^{-25}	6	_STX6_	WM, GM and WM	18,21
1	212557599	rs3767867	T/C	0.62	−0.14 (0.02)	2×10^{-10}	–	_PACC1_	WM, GM and WM	*
1	215139887	rs1452628	A/T	0.62	0.19 (0.02)	3×10^{-19}	7	_KCNK2_	GM, WM	17,23
2	56116193	2:56116193	T/TTG	0.10	0.20 (0.04)	2×10^{-8}	–	_EFEMP1_	WM	*
2	188028317	rs62172472	A/G	0.20	−0.15 (0.03)	7×10^{-9}	5	_TFPI_	GM	*
2	190694175	rs1233297	T/C	0.26	0.16 (0.02)	4×10^{-12}	32	_ORMDL1_	WM, GM and WM	18
2	198567638	rs12619333	C/G	0.67	0.14 (0.02)	1×10^{-9}	17	_SF3B1_	GM, GM and WM	*
2	201160771	rs1367858	T/C	0.34	0.15 (0.02)	4×10^{-11}	8	_SPATS2L_	GM, GM and WM	*
2	203877365	rs530464314	CA/C	0.14	−0.23 (0.03)	5×10^{-15}	6	_CARF_	GM and WM, GM, WM	23
3	69893971	rs62252239	T/G	0.21	−0.15 (0.03)	2×10^{-9}	11	_MITF_	WM	*
3	121628658	3:121628658	A/AT	0.48	−0.14 (0.02)	7×10^{-12}	14	_SLC15A2_	WM	*
3	171040852	rs776970253	CTC/C	0.74	−0.14 (0.02)	2×10^{-9}	7	_TNIK_	WM, GM and WM	*
3	171500441	rs72622537	A/C	0.35	−0.12 (0.02)	2×10^{-8}	15	_PLD1_	GM and WM	*
3	193549408	rs1146045	T/C	0.59	−0.12 (0.02)	2×10^{-8}	12	_LINC02038_	WM	*
4	2944571	rs66571798	CGT/C	0.41	−0.12 (0.02)	1×10^{-8}	12	_GRK4_	GM and WM	*
4	38680015	rs13132853	A/G	0.63	−0.26 (0.02)	9×10^{-36}	1	_KLF3-AS1_	GM and WM, WM, GM	18–22,24
4	115534729	rs75563007	T/C	0.97	−0.40 (0.07)	2×10^{-9}	8	_UGT8_	WM, GM and WM	*
5	72257180	rs2548331	T/G	0.48	0.12 (0.02)	8×10^{-9}	18	_FCHO2_	GM and WM	*
5	90567689	5:90567689	T/TTA	0.93	−0.24 (0.04)	2×10^{-8}	12	_LUCAT1_	GM	*
5	122879901	rs36048468	T/C	0.21	0.16 (0.03)	1×10^{-9}	15	_CSNK1G3_	GM, GM and WM	*
5	159528663	rs55790564	A/AT	0.40	−0.12 (0.02)	2×10^{-9}	12	_PWWP2A_	GM and WM, WM	*
6	31249217	rs2253491	A/G	0.21	−0.16 (0.02)	1×10^{-10}	–	_CLIC1_	GM and WM, GM, WM	18
6	45410312	rs910586	T/C	0.36	−0.21 (0.02)	1×10^{-23}	2	_RUNX2_	GM and WM, WM, GM	17–20,22,24
6	126690257	6:126690257	A/AT	0.44	0.12 (0.02)	2×10^{-8}	10	_CENPW_	GM, GM and WM	17
7	1213127	rs1543985	A/G	0.36	0.12 (0.02)	3×10^{-8}	17	_ZFAND2A-DT_	GM and WM, GM	24
7	120803286	rs35789132	A/G	0.65	−0.13 (0.02)	1×10^{-9}	11	_CPED1_	GM, GM and WM	*
8	10813904	rs10096381	T/G	0.52	0.13 (0.02)	1×10^{-9}	–	_XKR6_	WM	18
8	116635942	rs2721939	T/C	0.60	0.16 (0.02)	1×10^{-14}	4	_TRPS1_	WM, GM and WM	*
8	130903153	rs12548781	A/T	0.79	0.15 (0.03)	7×10^{-9}	12	_FAM49B_	GM	*
9	123543953	rs5021405	T/C	0.47	0.12 (0.02)	4×10^{-8}	–	_PHF19_	GM and WM	*
9	128010901	rs755594165	CA/C	0.61	−0.12 (0.02)	3×10^{-8}	–	_HSPA5_	WM	*
10	94839642	rs2068888	A/G	0.45	−0.13 (0.02)	3×10^{-10}	–	_FRA10AC1_	GM	*
10	98115019	rs41306852	A/G	0.02	0.42 (0.08)	5×10^{-8}	–	_OPALIN_	WM	*
10	105454043	rs2863994	T/G	0.51	0.12 (0.02)	4×10^{-9}	16	_SH3PXD2A_	GM and WM, GM	*
10	134573767	rs12258248	A/G	0.75	−0.22 (0.02)	4×10^{-20}	3	_INPP5A_	WM, GM and WM	19–22,24
11	32758240	rs10767960	A/G	0.54	−0.12 (0.02)	1×10^{-8}	17	_EIF3M_	GM and WM	*
12	32526829	rs6488048	T/C	0.65	0.13 (0.02)	7×10^{-9}	18	_ENSG00000274964_	WM	*
12	106476805	rs12146713	T/C	0.91	−0.25 (0.04)	5×10^{-12}	1	_ENSG00000257890_	WM, GM and WM	18,23
12	107349294	rs2287163	T/C	0.37	0.13 (0.02)	5×10^{-10}	15	_TMEM263_	WM	*
14	73297741	rs2215590	T/C	0.25	0.14 (0.02)	3×10^{-9}	8	_DPF3_	WM, GM and WM	*
14	88433660	rs413420	T/C	0.52	−0.16 (0.02)	7×10^{-14}	23	_GALC_	WM, GM and WM	22
15	71169352	rs2031017	A/T	0.40	−0.14 (0.02)	1×10^{-11}	9	_LARP6_	WM, GM and WM	18
16	30122181	rs536906899	CAA/C	0.40	−0.12 (0.02)	4×10^{-8}	39	_MAPK3_	GM	*
16	90051269	rs76839250	A/G	0.91	0.25 (0.04)	1×10^{-11}	6	_DEF8_	GM, GM and WM	18
17	19889274	rs111513543	A/T	0.64	−0.13 (0.02)	8×10^{-9}	30	_AKAP10_	WM, GM and WM	*

Table 2 (continued) | Identification of 59 genomic loci associated with BAG in *n* = 54,890 individuals

Chr	Position	ID	A1/A2	Frequency of A1	Beta (s.e.)	P	Credible	Prioritized gene	Phenotype(s)	Ref.
17	27962571	17:27962571	G/GCC	0.47	0.14 (0.02)	8×10^{-10}	10	*SSH2*	GM, GM and WM	*
17	44305199	rs2260227	T/C	0.78	−0.48 (0.03)	9×10^{-83}	5	*MAPT*	WM, GM and WM, GM	17–24
17	73873656	rs1105917	T/C	0.15	0.23 (0.03)	8×10^{-15}	4	*TRIM47*	GM, GM and WM	*
18	53277143	rs763283047	CCT/C	0.66	0.12 (0.02)	4×10^{-8}	18	*LINC01415*	WM	*
19	31036276	19:31036276	A/AC	0.16	0.15 (0.03)	3×10^{-8}	7	*ZNF536*	GM and WM	*
19	45416178	rs483082	T/G	0.23	0.15 (0.02)	1×10^{-10}	4	*APOE*	GM and WM, WM	18
20	30336992	rs6060924	A/G	0.71	0.14 (0.02)	4×10^{-10}	9	*BCL2L1*	GM	18
22	38457329	rs738442	T/C	0.38	−0.21 (0.02)	3×10^{-23}	23	*PICK1*	WM, GM and WM	18,23,24
X	13891499	rs2188767	A/G	0.42	0.11 (0.02)	9×10^{-10}	1|10	*GEMIN8*	GM, GM and WM	*
X	107888149	X:107888149	CAA/C	0.74	−0.11 (0.02)	2×10^{-8}	–	–	GM	*
X	133781440	X:133781440	CTG/C	0.29	−0.11 (0.02)	2×10^{-8}	41|60	–	GM	*
XY	2149565	rs34250447	T/C	0.74	−0.25 (0.02)	7×10^{-24}	54	*DHRSX*	WM, GM and WM, GM	*

For each of the 59 discoveries across the 3 BAG traits, only the strongest variant–phenotype association is shown. The 'Phenotype(s)' column lists all BAG traits with significant locus associations, with the strongest listed first. For indels, the A1 and A2 alleles are truncated to three nucleotide bases. The Beta, s.e. and *P* values were derived from two-sided linear regression models using PLINK, and were meta-analyzed using inverse-variance weighting in METAL. Position indicates the base-pair position of the index variant. ID indicates the identifier of the index variant. A1 indicates the effect allele; A2 indicates the other allele. Credible indicates the number of variants in the 95% credible set identified using SBayesRC (susieR sets for X and XY; multiple signals are separated by pipe symbol '|'). Prioritized gene indicates the gene selected using our gene prioritization procedure. Phenotype(s) indicates traits with genome-wide significant associations at this locus (GM BAG, WM BAG and combined GM and WM BAG), with the trait with the strongest association mentioned first. Ref. indicates prior studies reporting this locus; * indicates new.

(Supplementary Table 18); notably, rs17651549 ($P = 1.5\times10^{-81}$), which had the highest deleteriousness (CADD score = 34), results in an arginine-to-tryptophan substitution at MAPT protein position 370. *MAPT* encodes the well-known tau protein implicated in AD and other neurodegenerative diseases[60]. Altogether, we prioritized *MAPT* as the most likely causal gene for brain aging at this locus.

In 6 regions, all fine-mapping methods yielded 95% credible sets with fewer than 10 likely causal variants. One locus, with only a single credible variant (rs12146713, posterior inclusion probability (PIP) > 0.99; $P = 4.7\times10^{-12}$), lies in an intron of *NUAK1* at 12q23.3 and tags a multi-tissue eQTL for the long noncoding RNA gene *Lnc-NUAK1-1*, expressed in the brain cortex and cerebellum. GWAS Catalog matches link this region to cortical thickness[61], surface area[62] and subcortical volume[61].

A second locus, led by rs1452628 ($P = 2.5\times10^{-19}$), with up to 7 credible variants, refers to an intergenic region at 1q41, 41 kb upstream of *KCNK2*, encoding a potassium channel subunit. *KCNK2* is also the prioritized gene supported by the GTEx and PoPS analyses. KCNK2 has been implicated in neuroinflammation, blood–brain barrier dysfunction, and cerebral ischemia[63,64]. GWAS Catalog matches include associations with cortical thickness[61], surface area[62] and sulcal opening[65].

The third locus refers to the well-known apolipoprotein E (*APOE*) gene region, led by rs483082 ($P = 1.0\times10^{-10}$), with four credible variants. The *APOE* ε4 allele, defined by rs429358 and rs7412, is the strongest known genetic risk factor for AD. Notably, the exonic variant rs429358 was identified by SBayesRC as the most likely causal variant (PIP = 0.62).

We also discovered several novel loci that offer new insights into the mechanisms of brain aging. Among these loci, one is led by rs2215590 ($P = 2.9\times10^{-9}$) and maps to an intronic region of *DPF3*, which encodes double PHD fingers 3. Supported by GTEx eQTLs and PoPS, *DPF3* was prioritized as the most likely causal gene. *DPF3* also represents the prioritized gene supported by the GTEx eQTL lookup and PoPS analyses. GWAS Catalog matches link this locus to pulse pressure[66] and serum urate levels[67]. DPF3 serves as a subunit of the neuron-specific chromatin remodeling nBAF complex, which is crucial for neurogenesis and neurodevelopment[68].

Another novel locus, led by rs776970253 ($P = 2.4\times10^{-9}$), implicates an intronic region of *TNIK*, encoding TRAF2-interacting and NCK-interacting kinases. Notably, TNIK has been recognized for its role in several biological pathways linked to the hallmarks of aging and has been identified as a promising drug target to improve neuronal health[69].

Altogether, by integrating fine-mapping, functional annotation and transcriptomic data, we prioritized several genes potentially involved in brain aging, thereby offering new, testable hypotheses about its biological underpinnings.

Polygenic score analysis

To evaluate the predictive power of the genetic variants identified in our GWAS, we performed PGS analyses using SBayesRC (Supplementary Table 25). Compared to previous reports (~2% prediction accuracy)[23], our PGS showed substantially improved performance. Using the discovery GWAS data alone ($n = 32,634$), PGS explained 4.1% (GM BAG) to 7.0% (WM BAG) of the phenotypic variance in the European-ancestry replication sample. Incorporating the meta-analysis results ($n = 52,890$, excluding 2,000 test individuals) further improved variance explanation to 6.8% (GM BAG) and 10.3% (WM BAG). As expected, prediction accuracy was lower in the non-European UKB replication samples, explaining 0.4–3.2% of the BAG variance in African (AFR) ancestry ($n = 337$), 3.1–3.9% in Central/South Asian (CSA) ancestry ($n = 638$) and 4.1–9.1% in East Asian (EAS) ancestry ($n = 291$) individuals.

Gene-based analysis

To assess the contribution of protein-coding genes, we performed gene-based association analyses using GCTA fastBAT[70]. Gene-based analyses aggregate variant-level signals across genes, reducing the multiple-testing burden. We tested 18,639 genes and identified 528, 886 and 776 genes significantly associated (FDR < 0.05) with the GM, WM and combined GM and WM BAG, respectively. To define independent loci, we applied *P* value-informed clumping to genes within 3,000 kb, yielding 151 loci, 230 loci and 203 loci per trait, of which 285 were unique (Supplementary Table 26 and Extended Data Fig. 6). The strongest signal was again observed at 17q21.31 covering *MAPT*. In total, gene-based analyses provide evidence for an extended set of genomic loci involved in human brain aging.

Pathway analysis

To gain deeper insight into the biological mechanisms underlying brain aging, we performed gene set enrichment analyses using GOfuncR[71], testing for the enrichment of Gene Ontology (GO) terms—sets of genes known to serve a common biological function[72]. After correcting for
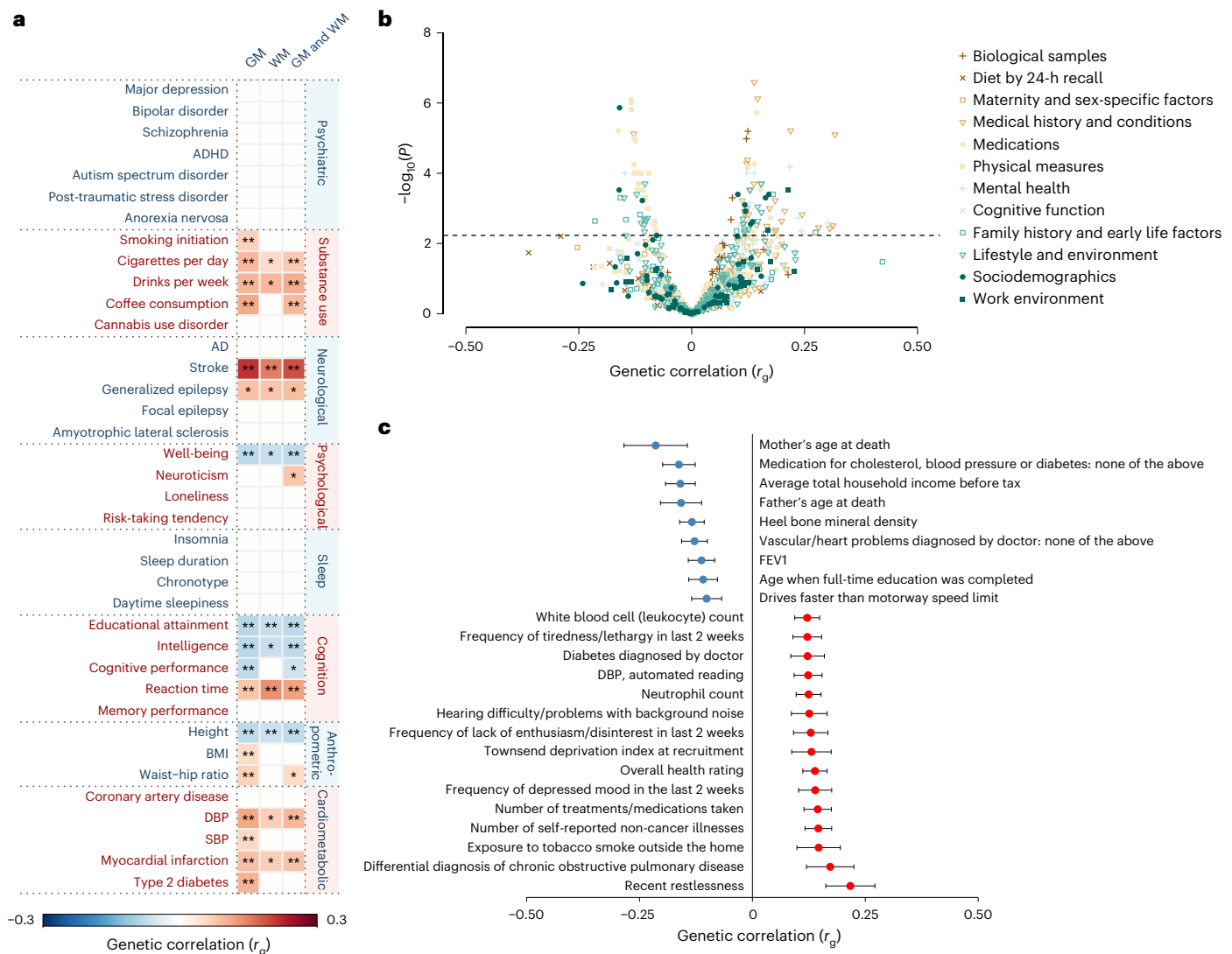
Fig. 4 | Genetic correlations between BAG and a wide range of complex traits. a, Genetic correlation matrix between BAG (columns) and 38 selected phenotypes from different health domains (rows). *$P < 0.05$ (nominal significance). **FDR < 0.05 (level of significance after correction for multiple testing). **b,** Volcano plot showing the magnitude (x axis) and significance (y axis) of LDSC-based genetic correlations between GM BAG and 989 traits, whose summary statistics were provided in ref. 80. The dashed horizontal line indicates the FDR-adjusted level of significance. All P values are two-sided. **c,** Forest plot showing the genetic correlation coefficients and standard errors for a subset of 23 exemplary traits that showed significant genetic correlations with GM BAG.

hierarchical dependencies (Methods), we identified 25 significant GO terms (Supplementary Table 27). Analyses highlighted immune-related and pathogen-related processes in brain aging, with significant enrichments for the major histocompatibility (MHC) protein complex (GO:0042611), peptide antigen binding (GO:0042605) and regulation of viral transcription (GO:0046782). Further significant terms, such as positive regulation of neurogenesis (GO:0050769) and regulation of axon extension (GO:0050769) align with the conceptualization of BAG as a neurodevelopmental marker. We also observed enrichment for small GTPase binding (GO:0031267l), a superfamily of evolutionary conserved proteins that act as biological timers of essential cellular processes[73], including cell differentiation, proliferation and signal transduction[74]. Several small GTPase proteins are implicated in premature senescence[75,76].

Genetic correlations with other complex traits

To assess a potential shared genetic architecture between BAG and other traits, we applied bivariate LDSC[36,37] to GWAS summary statistics, calculating genetic correlations with 38 commonly studied mental and physical health traits (Supplementary Table 28)[77–79], as well as 989 heritable traits from a broader set of GWAS[80].

Among the 38 selected traits, 17 showed significant correlations (FDR < 0.05) with at least 1 BAG phenotype (Fig. 4a and Supplementary Table 29). GM BAG showed the highest number of associations (17) relative to WM (4) and combined GM and WM (11) BAG. Notable associations for GM BAG included substance use (cigarettes per day: $r_g = 0.134$), neurological (stroke: $r_g = 0.217$), psychological (well-being: $r_g = 0.100$), cognition-related (educational attainment: $r_g = -0.083$), anthropometric (body mass index (BMI): $r_g = 0.075$) and cardiovascular and metabolic syndrome traits (DBP: $r_g = 0.115$).

Similarly, for the 989 traits, we found 118, 7 and 48 significant associations (FDR < 0.05) for GM, WM and combined GM and WM BAG, respectively (Figs. 4b,c and Supplementary Table 30). BAG showed significant genetic correlations with parental longevity (mother's age at death, $r_g = -0.214$; father's age at death, $r_g = -0.158$), socioeconomic status (average total household income before tax, $r_g = -0.160$), mental health (frequency of tiredness/lethargy in the last 2 weeks, $r_g = 0.122$), medical conditions (vascular/heart problems diagnosed by doctor:
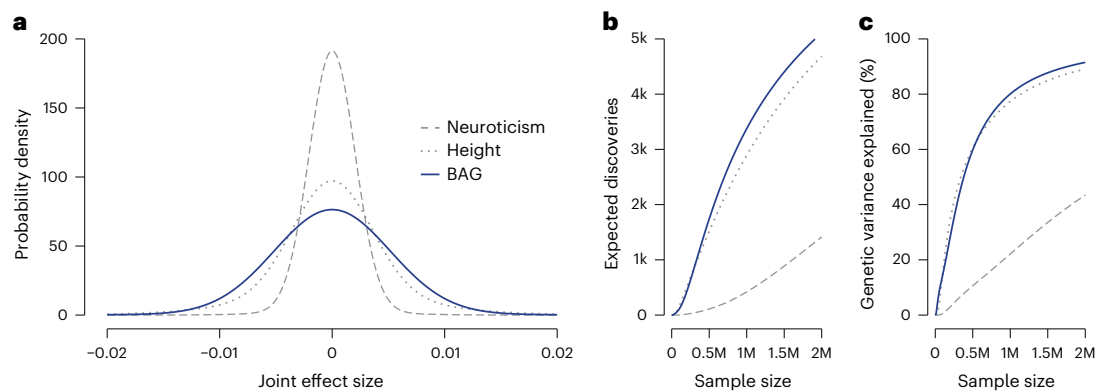
Fig. 5 | Genetic effect-size distribution analysis of BAG. a, Results are shown for combined GM and WM BAG, with neuroticism and standing height included as reference traits; effect-size distributions of the underlying susceptibility variants are shown; wider tails indicate a greater proportion of large-effect variants. **b**, Predicted number of genome-wide significant loci as a function of sample size. **c**, Proportion of genetic variance explained by genome-wide significant loci as a function of sample size.

high blood pressure, $r_g = 0.122$), cognitive function (snap game—mean time to correctly identify matches, $r_g = -0.069$), respiratory traits (forced expiratory volume in 1 s (FEV1), best measure, $r_g = -0.128$), blood markers (white blood cell (leukocyte) count, $r_g = 0.121$) and early life exposures (maternal smoking around birth, $r_g = 0.099$), among others (Supplementary Table 30). These results suggest genetic overlap between BAG and a broad range of health-related traits.

Mendelian randomization analyses

We used two-sample generalized summary-data-based Mendelian randomization (GSMR)[81] to investigate the potential causal effects of 12 modifiable risk and resilience factors on BAG. These included BMI, waist–hip ratio adjusted for BMI, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, systolic blood pressure (SBP), DBP, pulse pressure, type 2 diabetes, coronary artery disease, schizophrenia and years of education (Supplementary Table 31). Across all BAG traits, we found significant effects of DBP (combined BAG: $\beta_{xz} = 0.550$, $P = 6.9 \times 10^{-10}$) and SBP (combined BAG: $\beta_{xz} = 0.382$, $P = 1.4 \times 10^{-5}$), indicating that a one standard deviation increase in blood pressure causally contributes to an ~0.5-year increase in BAG (Supplementary Figs. 43–48 and Extended Data Fig. 7). We also observed significant effects of type 2 diabetes (combined BAG: $\beta_{xz} = 0.118$, $P = 5.6 \times 10^{-4}$) and coronary artery disease (combined BAG: $\beta_{xz} = 0.114$, $P = 0.01$)[82]. These findings were largely confirmed by nine alternative MR analyses, except for coronary artery disease (Supplementary Table 32).

Reverse GSMR analyses (Supplementary Table 33) indicated potential negative feedback effects of WM BAG on SBP, DBP and pulse pressure (all $\beta_{xz} = -0.009$, all $P \le 3.2 \times 10^{-3}$), which is consistent with lower blood pressure in late life and frailty[83]. Additionally, reverse GSMR analyses hinted at BAG effects on elevated low-density lipoprotein cholesterol and increased risk for coronary artery disease, although these were not supported by other MR methods.

Polygenicity and projection of discoveries to future GWAS

To quantify the BAG degree of polygenicity and estimate discovery potential in future GWAS, we used GENESIS[84] to estimate the number of underlying susceptibility variants and their effect sizes. We also selected height and neuroticism as benchmark traits because of their distinct degrees of polygenicity[84–87]. The number of susceptibility variants was estimated at 8.7k (s.e. = 1.8k) for GM, 9.8k (s.e. = 1.3k) for WM and 11.0k (s.e. = 1.3k) for combined GM and WM BAG (Supplementary Table 34). For comparison, height showed 12.6k (s.e. = 1.3k) and neuroticism 16.2k (s.e. = 1.2k) susceptibility variants. Effect-size distributions (Fig. 5a) revealed that BAG includes a larger proportion of variants with stronger

effects compared to neuroticism, but similar to height. Our projections also suggest rapid growth in the number of BAG discoveries with increasing sample size (Fig. 5b). Approximately 1 million individuals are needed to explain 80% of the SNP-based heritability for BAG via genome-wide significant variants (Fig. 5c), a threshold comparable to height but lower than for neuroticism (~6 million). These findings suggest that while BAG is genetically complex, its relatively lower polygenicity enhances discovery prospects in future studies.

Discussion

In this study, we leveraged genomic and neuroimaging methods to establish BAG as a promising biomarker of aging with potential utility for therapeutic discovery. Using machine learning and MRI, we derived highly reliable brain age estimates that capture aging-related structural patterns across the brain. BAG showed robust phenotypic associations with many health traits and was substantially heritable, with 23–29% of variance attributable to common genetic variation. We identified 59 independent genome-wide significant loci, of which 39 are novel. The genomic signals unveiled enriched biological pathways, for example, immune-system-related processes and small GTPase binding, prompting further mechanistic exploration. Through genetic correlations, we demonstrated shared genetic influences between BAG and a broad spectrum of physical and mental health traits. Mendelian randomization supported a causal role of elevated blood pressure and diabetes in accelerated brain aging. Finally, BAG showed a relatively low degree of polygenicity, which increases the likelihood of variant discovery in future studies.

Brain aging is not a uniform process; rather, it encompasses diverse aspects of structural and functional change. Studying distinct aspects of brain aging has been advocated to increase the yield of biologically meaningful insights[18]. In this study, we estimated separate BAG scores for GM and WM, alongside a composite measure. While all three showed similar age prediction accuracy and test–retest reliability, GM BAG exhibited stronger phenotypic and genetic associations, potentially reflecting a closer link to health-related outcomes. One explanation is that our volumetric approach to brain age estimation captures biologically meaningful differences more effectively in GM—tied to dendritic complexity and synaptic density—than in WM, which depends on microstructural features less directly represented in bulk volume. As such, GM BAG may better reflect deviations from normative aging across diverse health domains.

While brain aging follows multiple biological trajectories, differential aging rates across systems underscore the need to conceptualize bodily aging as a heterogeneous process[6]. Consistent with this, previous research showed weak correlations between biological age

measures derived from telomere length, epigenetic clocks, transcriptomics and immunometabolic markers[88]. Similarly, in our previous work, we found modest correlations between brain age, epigenetic clocks, skin age, clinical biological age composites and subjective age[89,90]. Although combining multiple biomarkers may improve chronological age prediction, it may also obscure biologically meaningful deviations at the domain level[18]. Recognizing this heterogeneity could help refine aging biomarker panels, and improve specificity and prognostic value for disease risk.

Several GWAS of BAG have been conducted previously[15,17–24]; however, few have incorporated extensive post-GWAS analyses. Leveraging increased statistical power, we present the most comprehensive genetic analysis of BAG to date, identifying 26, 34 and 39 loci across GM, WM and combined GM and WM BAG−59 in total, of which 39 are new. This surpasses the fewer than ten loci reported per phenotype in earlier studies. Expanding prior fine-mapping efforts[22], we constructed credible sets of likely causal variants (median = 9 per locus) and developed PGS explaining 7–10% of BAG variance, up from 2% previously[23]. Additionally, we here report results in groups of non-European ancestry, showing reduced yet significant polygenic predictions in AFR, CSA and EAS populations. We also identified new gene sets linked to neurodevelopment, immune function and signal transduction. Extending previous Mendelian randomization efforts[21–24], we confirmed a causal role of diabetes in BAG and newly identified blood pressure as a risk factor. Moreover, we expanded genetic correlations[15,21–24] and identified over 140 BAG associations across health domains. Finally, we introduced genetic effect-size distribution modeling, estimating ~8,700–11,000 contributing variants and providing key projections for discovery potential in future studies.

In addition to these new insights, our study also reinforces previously reported genetic associations. We confirmed the inversion locus at 17q21.31 as the strongest genetic contributor to BAG[17–24], with *MAPT*−a gene encoding the tau protein implicated in AD−prioritized as the most likely causal gene. We also identified the well-known AD risk gene *APOE* and other apolipoprotein genes. The presence of both tau-related and apolipoprotein-related signals suggests that key hallmarks of AD are reflected in accelerated brain aging, reinforcing BAG's relevance as a marker for neurodegenerative risk[91].

Our findings suggest that BAG integrates signals from neurodevelopmental and neurodegenerative processes, lifestyle factors, vascular and metabolic health, and immune function. This supports the idea that BAG is not purely a neurodegenerative marker but reflects genetic susceptibility, systemic health and environmental exposures. This supports viewing BAG as composite indicator shaped by aging, disease susceptibility and lifestyle exposures.

By reporting both phenotypic and genetic correlations for BAG, our study allows a direct comparison between the two. We observed a positive relationship between genetic and phenotypic correlations (Extended Data Fig. 8), with correlation coefficients ranging from 0.39 (WM BAG) to 0.51 (GM BAG). Although weaker than previously reported, probably because of the restricted range of correlation strengths in our dataset, the small mean absolute difference between genetic and phenotypic estimates suggests that phenotypic correlations approximate genetic ones and vice versa[92,93].

A future research direction is to explore how model precision affects the genetic and phenotypic associations of BAG[94]. Improved age prediction accuracy may enhance heritability estimates and genetic signals, but it could also obscure meaningful deviations from normative aging. Future studies may investigate this balance to understand the trade-off between predictive performance and biological interpretability.

The current study has several limitations. First, the gene prioritization techniques face challenges in pinpointing causal genes[49], particularly in loci characterized by high gene density and complex linkage structures. To streamline interpretation, we adopted a winner-takes-all approach, prioritizing a single candidate gene per locus. However, this may overlook other plausible genes with similarly high prioritization scores. Second, BAG was estimated from cross-sectional data, typically interpreted as accelerated or decelerated aging. However, an alternative view posits BAG as stable, early-emerging individual differences that persist into old age[95]. Third, although we used an ensemble of three machine learning models, expanding the number and diversity of models may further enhance prediction by leveraging complementary strengths. Fourth, polygenic overlap was assessed using genetic correlations, which do not capture shared variants with opposing effects. Future studies could apply tools such as MiXeR to quantify genetic overlap by considering mixtures of variant effects[96]. Fifth, polygenicity estimates probably underestimate the true number of contributing variants because models may classify those with very small effects as null. Finally, as our primary GWAS was based on individuals of European ancestry, PGS predicted less accurately in groups of non-European ancestry, limiting generalizability across ancestries. This reflects known transfer challenges due to differences in allele frequencies, LD and environmental context. Expanding genomic data and sample sizes in diverse populations will be essential to improve accuracy and broaden the applicability of brain age genetics.

In conclusion, our study refines the genetic architecture of BAG and its relationships to other traits. We added 39 new genetic loci and nominated plausible candidate genes, including *DPF3*, which is implicated in neurodevelopment, and *TNIK*, which is linked to neuronal health and aging-related diseases. This will facilitate further work on the pathway mechanisms of BAG and potential therapy targets.

Methods

Ethical approval
This study used individual-level data from the UKB (www.ukbiobank.ac.uk) and LIFE-Adult (www.uniklinikum-leipzig.de/einrichtungen/life)[26,29,30]. Both studies were conducted in accordance with applicable ethical regulations and the principles of the Declaration of Helsinki (2008). The UKB received approval from the North West–Haydock Research Ethics Committee (ref. nos. 11/NW/0382, 16/NW/0274, 21/NW/0157). LIFE-Adult was approved by the Ethics Committee of Leipzig University (ref. nos. 263–2009-14122009, 263/09-ff, 201/17-ek). All participants provided written informed consent. LIFE-Adult participants received a fixed compensation of 20 EUR per visit. UKB participants could claim travel reimbursement.

Statistics and reproducibility
This study presents results from a GWAS alongside a broad set of post-GWAS analyses, including fine-mapping, polygenic scoring, genetic correlation and Mendelian randomization. To enhance transparency and reproducibility, we have provided all analysis scripts, conda environments and software details in a public GitHub repository (https://github.com/pjawinski/ukb_brainage). Analyses were run on Debian GNU/Linux 11 (kernel 5.10.0-23-amd64). Unless stated otherwise, all *P* values are two-sided. Associations with *P* < 0.05 were considered nominally significant; Bonferroni correction and FDR control according to the Benjamini–Hochberg procedure were used to adjust for multiple testing. No formal power calculation was used to predetermine sample size. Instead, we included all eligible individuals from the UKB and LIFE-Adult who passed predefined quality control criteria.

Data exclusions were limited to prespecified quality control steps, described in detail elsewhere in the Methods. Analytical assumptions were addressed at each stage of the analysis. In cross-trait association testing, regression models were automatically selected based on the type of variable, with continuous variables normalized to meet distributional assumptions. In the GWAS, standard variant-based and sample-based quality control was applied; LDSC confirmed that the test statistic inflation was driven by polygenicity rather than confounding.

This study was observational and nonexperimental; thus, participants were not randomly assigned and no blinding was applied. We report how our target samples were defined, all data exclusions, quality control procedures and all measures used in the study. A full list of UKB variables is provided in the UKB data dictionary (https://biobank. ndph.ox.ac.uk/showcase/) and LIFE-Adult data portal (https://ldp. life.uni-leipzig.de/).

Sample characteristics

Participants were drawn from the UKB under application no. 423032. A detailed description of the UKB study design and quality control methods has been published previously[26]. For our discovery sample, participants were drawn from the UKB January 2020 imaging release (v.1.7). These data contained 40,681 participants with structural T1-weighted MRI scans (UKB data-field 20252). Scans in folders labeled 'unusable' were excluded, leaving 39,679 participants. Voxel-based morphometry preprocessing was successfully completed for 39,677 MRI scans (see the 'MRI preprocessing' section of the Methods). Analyses were restricted to participants whose self-reported sex matched the genetic sex (data-fields 31 and 2200), without sex chromosome aneuploidy (data-field 22019) and who were no outliers in heterozygosity and missingness (data-field 22027). We only included unrelated participants as suggested by pairwise kinship coefficients below 0.0442 (precalculated coefficients retrieved using 'ukbgene rel'). We included participants of White British ancestry (data-field 22006), yielding a final discovery sample of 32,634 participants (17,084 female, age range = 45.2–81.9 years, mean age = 64.3 years).

For replication, we selected all remaining individuals without White British ancestry from the UKB January 2020 release (n = 4,870). Applying the same inclusion criteria, we added European and non-European UKB participants with imaging data released until May 2024 (v.1.10), yielding 25,668 individuals. None of them were related to the discovery participants. We included individuals with valid ancestry assignment from the Pan-ancestry return (no. 2442; https://pan.ukbb. broadinstitute.org/). This resulted in 337 African, 94 Admixed American, 638 Central/South Asian, 291 East Asian, 20,423 European and 98 Middle Eastern ancestry participants. In total, we included 21,881 UKB participants for replication (11,451 female, age range = 45.5–81.9 years, mean age = 67.1 years). From the LIFE-Adult study[29,30], we included another 1,833 unrelated participants of European ancestry (888 female, age range = 45.2–80.3 years, mean age = 65.3 years) with available T1-weighted MRI and genotype data, selected to match the UKB age range[97]. Altogether, the final replication sample included 23,714 participants (12,339 female, age range = 45.2–81.9 years, mean age = 67.0 years) from 7 subsamples.

MRI data acquisition

The UKB imaging acquisition protocol and processing pipeline have been detailed previously (http://biobank.ctsu.ox.ac.uk/crystal/refer. cgi?id=1977). Brain MRI data were acquired at four UKB imaging centers (Cheadle, Newcastle, Reading and Bristol) on Siemens Skyra 3T MRI scanners (Siemens Healthcare) running the VD13A SP4 software, with a standard 32-channel radiofrequency head coil. We used T1-weighted structural MRI scans (UKB data-field 20252) acquired using a 3D magnetization-prepared rapid gradient-echo (MPRAGE) sequence in the sagittal plane, with a voxel size of 1 × 1 × 1 mm, 208 × 256 × 256 acquisition matrix, 2,000-ms repetition time (TR), 2.01-ms echo time (TE), 880-ms inversion time (TI), 6.1-ms echo spacing, 8° flip angle, a bandwidth of 240 Hz per pixel, an in-plane acceleration factor of R = 2 and duration of 4 min 54 s.

In LIFE-Adult, brain imaging was performed on a 3T Verio MRI scanner (Siemens Healthcare) with a standard 32-channel head coil. T1-weighted images were obtained using a 3D MPRAGE sequence with a voxel size of 1 × 1 × 1 mm, 256 × 240 × 176 acquisition matrix, TR = 2,300 ms, TE = 2.98 ms, TI = 900 ms and 9° flip angle.

MRI preprocessing

T1-weighted MRI scans in NIfTI format were preprocessed using the voxel-based morphometry pipeline of CAT12 (r1364, http://dbm.neuro. uni-jena.de) for SPM12 (r7487) in MATLAB R2021a (MathWorks). CAT12 preprocessing included affine and DARTEL registration to a reference brain, segmentation into GM, WM and cerebrospinal fluid, bias correction for intensity inhomogeneity and modulation to account for volume changes because of spatial registration. Images were then smoothed using an 8 × 8 × 8-mm full-width-at-half-maximum Gaussian kernel and resampled to a voxel size of 8 mm³. Only scans with a CAT12 overall image quality rating of less than 3.0 were retained, excluding 119 (~0.3%) of 39,677 scans from the UKB imaging release v.1.7, and 101 (0.4%) of 23,000 additional scans from release v.1.10.

Feature set for machine learning

Machine learning features were derived from CAT12-preprocessed GM and WM segmentations. Each smoothed, resampled brain image included 16,128 voxels. Voxels without interindividual variation were excluded, yielding 5,416 GM and 5,123 WM voxels. Because of spatial correlation across voxels, we applied principal component analysis (PCA) in MATLAB to reduce dimensionality. The first 500 principal components—explaining ~90% of the total variance—were selected as features.

Machine learning algorithms

We implemented three complementary algorithms to model age from the brain imaging data: the sparse Bayesian RVM using the MATLAB toolbox SparseBayes v.2 with the wrapper and kernel in refs. 27,98, and extreme gradient boosting using XGBoost v.0.82.1 in R[28], using both decision tree (gbtree) and linear (gblinear) boosters. These algorithms were chosen for their demonstrated efficacy in previous brain age studies[5,15,99,100]. XGBoost was configured with a learning rate of $\eta = 0.02$, 5,000 training iterations, early stopping after 50 iterations without improvement and maximum tree depth of 3. Default settings were used for all other training parameters. To exploit their complementary strengths in handling high-dimensional data, modeling linear and nonlinear relationships, and regularization, we combined all three (RVM, XGBoost tree and XGBoost linear) in an ensemble.

Age estimation models and BAG calculation

Age estimation models were trained using PCA-derived brain imaging features to predict chronological age. Training and application were performed in the discovery sample using tenfold cross-prediction with 100 repeats. This cross-prediction approach was chosen to maximize precision and avoid bias from external datasets with differing MRI protocols, similar to previous studies[15,20–22]. The discovery sample was split into ten equally sized subsets. In each iteration, nine subsets served for model training and one for testing. PCA was performed on the training data and the transformation parameters were applied to the test set. This procedure was cycled through all ten folds, so each subset served once as the test set. The entire tenfold cross-prediction procedure was repeated 100 times, generating 100 predictions for each individual. This process was run for each tissue type (GM and WM) and model type (RVM, XGBoost tree and XGBoost linear), yielding 600 brain-predicted age estimates per individual (2 tissues × 3 models × 100 repeats). A nested tenfold cross-prediction was used to stack model-type predictions into tissue-specific ensemble estimates for GM and WM. To derive estimates for combined GM and WM, we stacked tissue-specific predictions rather than training new models on combined inputs. This yielded 100 age estimates per tissue type (GM, WM, combined), which were averaged to obtain 1 final brain-predicted age estimate per tissue type. Model performance was evaluated using the product-moment correlation coefficient (r), the coefficient of determination (R^2) and MAE[100].

In the replication samples, age predictions were generated using all tenfold discovery models and compared to predictions from models trained on the full discovery sample. Results were highly concordant for

all three tissue types ($r > 0.997$; Supplementary Fig. 15). For improved practicability, subsequent replication analyses used models trained on the full discovery sample.

BAG was calculated as the difference between predicted brain age and chronological age as:

$$BAG = \hat{A}_{brain} - A_{chron}$$

where BAG reflects the brain age gap estimate. $\hat{A}_{brain}$ is the predicted (modeled) age based on an individual's brain imaging data and A_{chron} is the actual chronological age of the individual.

Because of regression dilution, BAG is typically confounded by age, with younger individuals showing higher and older individuals lower BAG values[31]. To correct this bias, we included both age and age^2, alongside additional covariates (sex, scanner site, total intracranial volume, genotyping array, genetic principal components), in all association analyses[31,32,100].

Cross-trait association analysis

We performed cross-trait association analyses using PHESANT v.1.1 (ref. 33), an automated pipeline for phenome-wide association analyses in the UKB. Each BAG phenotype was tested against 7,088 nonimaging UKB variables. Covariates included sex (field 31), age (derived from fields 34, 52 and 53), age^2, scanner site (field 54) and total intracranial volume (from CAT12 segmentation). PHESANT selected regression models (linear, logistic, ordinal logistic or multinomial logistic) based on the type of variable. Continuous variables were inverse-rank-normalized before linear regression. To obtain standardized effect sizes, we calculated product-moment correlations (r) via corresponding z-statistics: $r = \text{sign}(\beta)\sqrt{z^2/(z^2 + (N - k - 2))}$. For visualization, variables were grouped into categories based on the UKB data dictionary path. We also performed sex-stratified analyses and tested sex differences by comparing PHESANT beta coefficients in males (β_m) and females (β_f): $z = (\beta_m - \beta_f)/\sqrt{(\text{s.e.}_m^2 + \text{s.e.}_f^2)}$. The resulting z-values were converted into P values using standard normal probabilities.

FreeSurfer associations

To examine associations between BAG and individual brain regions, we analyzed brain measures from the FreeSurfer aparc and aseg output files (UKB data-field 20263)[35], including surface area, cortical thickness and volume from 34 bilateral cortical and 8 bilateral subcortical regions (220 measures in total). We calculated partial product-moment correlations between BAG and brain measures, adjusting for sex, age, age^2, scanner site and total intracranial volume. Visualizations were created using the ENIGMA toolbox v.2.0.3 for MATLAB[101]. We also performed sex-stratified analyses and tested sex differences using Fisher's r-to-z transformation with the cocor R package[102]. Associations between brain regions and chronological age are reported in Supplementary Table 3.

UKB genotyping and imputation

We retrieved genotype data (called: BED; imputed: BGEN v.3) from the UKB. Genotype collection, processing and quality control have been described previously[26,103]. Genotyping was performed on DNA from EDTA blood using 2 Affymetrix arrays with 95% marker overlap: the UKB BiLEVE Axiom Array (807,411 markers used in 49,950 participants) and the UKB Axiom Array (825,927 markers used in 438,427 participants). Marker-based quality control included a call rate greater than 0.90, tests for batch, plate, array and sex effects, and Hardy–Weinberg equilibrium ($P < 1.0 \times 10^{-12}$). Sample-based quality control excluded individuals with a missingness greater than 0.05, high heterozygosity, sex discordance or sex chromosome aneuploidy. Relatedness was inferred using KING[104]. White British ancestry (data-field 22006) was defined via self-report and genetic principal components. Genotypes were phased using SHAPEIT3 and imputed using IMPUTE4 (https://jmarchini.org/software/) with the Haplotype Reference Consortium,

UK10K Project and 1000 Genomes Project Phase 3 serving as reference. Imputation yielded ~97 M markers. We selected biallelic SNPs and indels with MAF > 0.01 and INFO > 0.80. Biallelic variants were defined as those without duplicate coordinates or duplicate identifiers. This resulted in 9,669,330 variants for the discovery GWAS. In the replication samples, the number of variants passing quality control ranged between 8,345,339 (EAS ancestry) and 15,371,587 (AFR ancestry).

LIFE-Adult genotyping and imputation

Genotype collection, processing and quality control in LIFE-Adult have been described previously[97]. DNA from peripheral blood leukocytes was genotyped on the Axiom Genome-Wide CEU 1 Array (Applied Biosystems) (587,352 markers). Marker-based quality control removed variants with call rate lower than 0.97, Hardy–Weinberg equilibrium $P < 1.0 \times 10^{-6}$ or plate effects $P < 1.0 \times 10^{-7}$. Sample quality control excluded individuals with Dish QC < 0.82, missingness > 0.03, sex discordance or cryptic relatedness. Genotypes were phased using SHAPEIT and imputed with IMPUTE2 using the 1000 Genomes Project Phase 3 as reference. This yielded 85,063,807 markers in 7,776 individuals. Quality control after imputation (MAF > 0.01, INFO > 0.8) left 9,472,504 biallelic SNPs or indels that also passed UKB quality control for inclusion in the meta-analysis.

Control for population structure

In the discovery sample, we calculated 20 genetic principal components using the randomized PCA algorithm (--pca 20 approx) implemented in PLINK v.2.00a2LM[105], based on the same variants used by the UKB group (resource 1955; 146,988 markers passing our own quality checks). For the UKB replication samples, we used principal components from the Pan-ancestry UKB project (return 2442), using 20 components for the larger UKB European-ancestry sample and 4 for all other groups.

Heritability and partitioned heritability

Estimates of SNP-based heritability (h^2_{SNP}) were derived by applying LDSC[36,37] to our GWAS summary statistics, with precalculated LD scores and regression weights from the 1000 Genomes Project Phase 3. Analyses were limited to HapMap3 variants with MAF > 0.01, excluding the MHC region. Partitioned heritability was assessed using stratified LDSC[38] with baseline-LD model v.2.2. We tested the 33 main annotations reported in ref. 106, considering annotations with an FDR < 0.05 as significant.

Genome-wide association analysis

GWAS analyses were performed in PLINK v.2.00a2LM[105] using allelic dosage data, including autosomal (chromosomes 1 and 22), gonosomal (chromosomes X and Y) and pseudoautosomal (chromosomes XY) variants. Dosage scales were 0–2 for diploid regions (chromosomes 1–22, chromosomes XY), 0–1 for haploid chromosome Y and 0–2 for chromosome X. We modeled additive genetic effects and used sex, age, age^2, total intracranial volume, scanner site, type of genotyping array and the first 20 genetic principal components as covariates (4 components for the LIFE-Adult and non-European-ancestry samples).

Genome-wide association meta-analysis

The European-ancestry GWAS results were meta-analyzed using fixed-effects inverse-variance-weighted models in METAL (v.2020-05-05)[107]. Variants with a sample size of less than 67% of the 90th percentile (adapted from LDSC[37]) or heterogeneity $P < 1.0 \times 10^{-6}$ were excluded. The final GWAS meta-analysis in individuals of European ancestry included 9,628,877 variants for GM and WM BAG, and 9,628,868 variants for the combined BAG, analyzed in up to 54,890 individuals. Multi-ancestry meta-analyses were performed with MR-MEGA v.0.2 (ref. 108), including the White British discovery sample, six UKB replication samples (European, African, Admixed

American, Central/South Asian, East Asian, Middle Eastern ancestry) and the European-ancestry LIFE-Adult cohort. Ancestry effects were modeled using three axes of genetic variation derived from allele frequency differences. For comparison, fixed-effects and random-effects meta-analyses were also conducted in GWAMA v.2.2.2 (ref. 109). The multi-ancestry GWAS results included 8,618,923 variants in up to 56,348 individuals.

Identification of independent discoveries

We identified independent association signals using stepwise conditional analyses in GCTA-COJO[39]. A 10,000-kb window size and collinearity cutoff of 0.9 were applied. Multiple signals within a locus were only considered independent if the P value of the subsidiary signal did not increase by more than two orders of magnitude relative to its unadjusted value. Variants reaching $P < 5.0 \times 10^{-8}$ in the conditional analysis were considered genome-wide significant; those with $P < 1.0 \times 10^{-6}$ were deemed suggestive. We refer to lead variants from these signals as index variants. To identify nonredundant signals across the three BAG GWAS, all index variants were LD-clumped ($r^2 < 0.1$, window size = 10,000 kb) using PLINK v.1.90b6.8.[105].

Definition of variant replication and power calculations

From the discovery GWAS, we selected index variants from genome-wide significant loci (conditional $P < 5.0 \times 10^{-8}$) and 45 additional suggestive loci (conditional P values between 1.0×10^{-6} and 5.0×10^{-8}) for replication. Consistency between discovery and replication was tested using sign tests (binomial), based on the agreement in effect direction. Variants with a replication $P < 0.05$ (one-tailed nominal significance) were considered replicated. To estimate the expected replication yield, we performed power calculations based on standardized discovery betas, MAF and replication sample size[110]. Beta coefficients were corrected for winner's curse[111]. Expected replications were computed as the sum of individual variant-level power estimates.

Novelty of the discovered loci

To assess novelty, we compared our findings against nine prior BAG GWAS reporting genome-wide significant loci[15,17–24]. Using PLINK v.1.90b6, we clumped variants based on LD ($R^2 > 0.1$, window size = 10,000 kb)[105] and defined loci as novel if they did not cluster with previously reported variants. Parameter choices were guided by GCTA-COJO[39] and Psychiatric Genomics Consortium studies[79,112]. Of the 59 loci identified, 39 were classified as novel, a result consistent across clumping thresholds (R^2 of 0.10 or 0.05) and window sizes (10,000 kb or 3,000 kb).

ANNOVAR enrichment test

We used the ANNOVAR (v.2017-07-17)[44] enrichment test implemented in FUMA v.1.6.0 (https://fuma.ctglab.nl/)[113] to evaluate whether genome-wide significant regions were enriched for specific functional annotations. All candidate variants in LD ($R^2 > 0.6$) with independent significant autosomal variants ($P < 5.0 \times 10^{-8}$) were included. Candidate variants were defined as those with $P < 0.05$ and $R^2 > 0.60$ with an independent significant variant. UKB release 2 served as the LD reference panel. If a variant had multiple annotations, each was counted separately. Enrichment was computed as the proportion of candidate variants with a given annotation relative to the proportion of variants with that annotation among all variants in the reference panel. Significance was tested using a two-tailed Fisher's exact test.

Credible sets of variants

We used SBayesRC, a Bayesian mixture model implemented in GCTB v.2.5.2 (refs. 40,41), to construct 95% credible sets of variants per locus, capturing the cumulative posterior probability of including a causal variant. Unlike region-specific fine-mapping approaches, such as susieR[42] and FINEMAP[43], SBayesRC jointly models multiple genomic

regions alongside functional annotations. We used SBayesRC with the eigendecomposition data of LD matrices from our discovery dataset of 32,634 individuals (~9.7 M imputed SNPs), and functional annotations from the stratified LDSC baseline-LF UKB model (v.2.2)[114]. We used default settings with five mixture components (scaling factors of 0, 0.001, 0.01, 0.1 and 1%). Credible sets were assigned to a discovered locus if they contained at least 1 genome-wide significant credible variant in strong LD ($R^2 > 0.8$) within 3,000 kb from the index variant. We report sets with PIP > 0.95 and posterior enrichment probability > 0.50. For comparison, we also applied susieR v.0.12.35 (ref. 42) and FINEMAP v.1.4.2 (ref. 43). For each locus, a 10,000-kb window was used to identify the outermost variants in LD ($R^2 > 0.1$), defining region boundaries. LD matrices were computed using LDstore v.2.0 (ref. 115) in 53,074 individuals of European ancestry from the combined discovery and UKB replication sample. For FINEMAP, we allowed up to $k = 10$ causal variants per region, reporting 95% credible sets for the most probable k model. For susieR, we allowed up to $L = 10$ causal signals per region, reporting 95% credible sets with minimum purity greater than 0.5.

Functional annotation of variants

Variants were annotated using ANNOVAR[44], which assigns functional categories based on physical position relative to genes. RefSeq gene annotations (hg19) were retrieved from the UCSC Genome Browser (https://genome.ucsc.edu/)[116]. The nearest gene was identified using ANNOVAR's default prioritization of variant function and genomic distance. The transcript consequences of nonsynonymous exonic variants were predicted; deleteriousness scores from CADD were obtained from dbnsfp35a (hg19)[45,117].

Gene nomination through functional annotation of credible variants

Credible variants were annotated using ANNOVAR[44], and variant posterior probabilities were aggregated per gene implicated. Genes were then ranked according to their total variant posterior probabilities and nominated for gene prioritization. Additionally, genes implicated by nonsynonymous exonic variants were ranked based on the highest CADD Phred-scaled score among those variants.

Gene nomination through SMR

We applied summary-data-based Mendelian randomization implemented in SMR v.1.03 (refs. 47,48) to test whether variant effects were potentially mediated by gene expression or splicing. SMR integrates GWAS summary statistics with omics data to prioritize gene targets and regulatory elements. It adopts the Mendelian randomization strategy by using a single genetic instrument (z) to test for pleiotropic association between gene regulation (exposure, x) and a trait of interest (outcome, y). The effect of gene regulation on a trait (β_{xy}) is calculated as a two-step least squares estimate and defined as the ratio of the instrument's effect on the outcome (β_{zy}) to its effect on the exposure (β_{zx}), that is, $\beta_{xy} = \beta_{zy}/\beta_{zx}$. To distinguish pleiotropy from linkage, SMR incorporates the HEterogeneity In Dependent Instruments (HEIDI) test, which leverages multiple instruments in the regulatory region. We used cis-eQTL (gene expression) and cis-sQTL (gene splicing) summary statistics from BrainMeta v.2, derived from RNA-seq data of 2,865 brain cortex samples from 2,443 individuals of European ancestry[48]. Our GWAS variants were mapped to 16,375 eQTL and 58,941 sQTL probes. We retained results with an FDR < 0.05, $P_{HEIDI} > 0.01$ and those mapping to genome-wide-significant GWAS loci. Significant SMR hits were assigned to index variants using PLINK clumping (window size = 3,000 kb; $R^2 > 0.80$). Genes implicated by eQTL and sQTL SMR were nominated separately and ranked using the SMR P value.

Gene nomination through GTEx eQTL lookup

Index variants and their genome-wide significant neighbors in strong LD ($R^2 > 0.8$) were mapped to cis-QTLs from the GTEx v.8 database[46].

Single-tissue QTLs were retrieved from the prefiltered file (GTEx_Analysis_v8_eQTL.tar), covering 49 tissues. Multi-tissue QTLs were obtained from the METASOFT results (GTEx_Analysis_v8.metasoft.txt.gz), retaining variant–gene associations available in 10 or more tissues and with an m value equal to or greater than 0.9 (that is, the posterior probability that the effect exists) in 50% or more of tissues. To ensure robustness, only associations with meta-analytical $P < 5.0 \times 10^{-8}$ (Han and Eskin's random-effects (RE2) model) were considered, yielding 4,420,841 multi-tissue QTLs. Variant mapping was done using the GTEx hg19 liftover variant IDs. If multiple variants implicated the same gene within a locus, we reported the variant in strongest LD with the index variant. Ensembl gene IDs were converted to HUGO Gene Nomenclature Committee symbols using biomaRt[118]. Genes implicated by single-tissue and multi-tissue QTLs were nominated separately for prioritization and ranked according to the number of significant tissue associations.

Gene nomination through polygenic priority scores

We used PoPS v.0.2 (ref. 49) to identify likely causal genes within significant GWAS loci. PoPS builds on MAGMA[119] gene-level associations and leverages subthreshold polygenic signals, integrating more than 57,000 features from sources such as single-cell RNA-seq datasets, curated biological pathways and protein–protein interaction networks. We used the same PoPS feature map and MAGMA gene annotation file as in the original publication (www.finucanelab.org/data)[49]. MAGMA v.1.10 was applied to our GWAS summary statistics using SNP-wise mean gene analysis, with LD data from 53,057 individuals of European ancestry (combined discovery and UKB replication sample). For each index variant identified through the conditional analyses, up to 3 genes within 500 kb with the highest PoPS scores were nominated for gene prioritization.

Gene prioritization

Genes were prioritized based on seven evidence streams: (1) ANNOVAR functional annotation of credible variants, summing the posterior probabilities per gene; (2) transcript consequences of nonsynonymous exonic variants, ranked according to CADD score; (3) SMR eQTLs ranked according to P value; (4) sQTLs ranked according to P value; (5) GTEx single-tissue eQTLs; and (6) multi-tissue eQTLs, ranked according to the number of significant associations across tissues; and (7) PoPS ranked according to prioritization score. A composite priority score was calculated for each nominated gene as described below.

Let i denote a gene and j denote the index of the nomination category. The priority score (P_i) for gene i combines the cumulative posterior probability (C_i) from variant annotations and the gene's rank (R_{ij}) across six additional evidence categories ($j \in [1, 6]$) as

$$P_i = C_i + \sum_{j=1}^{6} \frac{2(n_j + 1 - R_{ij})}{n_j(n_j + 1)}$$

where P_i denotes the priority score for gene i, C_i denotes the cumulative posterior probability of variants mapped to gene i, n_j denotes the number of genes ranked in nomination category j and R_{ij} denotes the rank of gene i in nomination category j. This formulation assigns greater weight to top-ranked genes and ensures that each category contributes equally (one point per category). The gene with the highest P_i per locus was designated the prioritized gene.

GWAS Catalog search

We queried the National Human Genome Research Institute GWAS Catalog (13 September 2024 release; gwas_catalog_v1.0-associations_e112_r2024-09-13.tsv)[52] for all index variants identified using conditional analysis and their genome-wide significant neighbors in strong LD (3,000-kb window, $R^2 > 0.8$). Neighboring variants were identified through P-value-informed clumping in PLINK v.1.90b6.8 (ref. 105). Only GWAS Catalog entries reaching genome-wide significance were retained.

Gene-based analysis

We performed gene-based analyses using fastBAT in GCTA v.1.93.1f[70]. Gene coordinates were obtained from the RefSeq GFF3 annotation file (GRCh37.p13; release 105.20201022)[120]. NCBI chromosome names were converted to UCSC format. We selected protein-coding genes located on chromosomes 1–22, X and Y, removing duplicates gene names, by keeping the first entry sorted according to chromosome, symbol and coordinates. This yielded 19,299 genes, of which 18,632 were successfully mapped to GWAS variants. Analyses used linkage data from the combined UKB discovery and replication sample ($n = 53,057$, European ancestry), applying no flanking window to reduce gene-level dependency. Genes with an FDR < 0.05 were considered significant. To identify independent associations, we applied P-value-informed clumping with a 3,000-kb window size. Distinct associations across the 3 BAG traits were determined with second-level clumping, using each gene's top P value, again with a 3,000-kb window size.

Pathway analyses

We conducted GO pathway analyses using the R package GOfuncR v.1.14.0, based on the GO.db v.3.14.0 and *Homo.sapiens* v.1.3.1 annotations[71,121,122]. GO provides a curated framework to categorize genes based their molecular function, cellular components where they perform actions and the higher-order biological processes they contribute to. Gene set enrichment analyses were performed on the full fastBAT gene-based results, testing for lower-than-expected P value ranks using the Wilcoxon rank-sum test. By default, GOfuncR calculates family-wise error rates (FWERs) in each of the three GO aspects using random permutations. To reduce false discoveries, we joined these permutation-based results to calculate FWERs across the three GO aspects. We further refined significant results (FWER < 0.05) by applying the elim algorithm to decorrelate overlapping terms and retain the most specific[123]. For interpretation, we also determined the number of distinct loci contributing to each enriched term, applying 3,000 kb clumping to account for spatial gene clustering.

PGS analysis

To estimate the variance in BAG explained by PGS, we used a conventional clumping and P thresholding (C + P) approach implemented in PRSice-2 v.2.3.3 (ref. 124) along with two Bayesian polygenic prediction methods, SBayesR and SBayesRC, implemented in GCTB v.2.5.2 (ref. 40). For the C + P approach, we used $R^2 > 0.1$, a 500-kb window size and 10 predefined P value thresholds[79,112]. Unlike the C + P approach, SBayesR and SBayesRC jointly model all variant effects, with SBayesRC additionally incorporating functional annotations. SBayesR/RC were used with eigendecomposed LD matrices (~7 M variants), and stratified LDSC baseline-LD v.2.2 annotations. Missing variants were imputed. Default settings were used (--gamma 0,0.001,0.01,0.1,1 --pi 0.99,0.005,0.003,0.001,0.001 --chain-length 3,000 --burn-in 1,000). The resulting weights were applied to calculate the PGS in target samples using PLINK (--score)[105].

PGS weights derived from the discovery sample ($n = 32,634$) were tested in the European-ancestry replication sample ($n = 20,423$). To estimate PGS performance in the combined discovery and replication sample, we reran the GWAS meta-analysis excluding 2,000 individuals of European ancestry held out as a target set (total training $n = 52,890$). Transferability was assessed in AFR ($n = 337$), CSA ($n = 638$) and EAS ($n = 291$) UKB subsamples. Associations between PGS and BAG were evaluated using partial product-moment correlations, adjusting for sex, age, age², scanner site, total intracranial volume, genotyping array and 20 genetic principal components (4 for non-European samples).

Genetic correlations

We used bivariate LDSC (v.1.0.1)[36] to compute pairwise genetic correlations among BAG traits and between BAG and other complex traits.

These included 38 commonly studied traits spanning the mental and physical health domains[77–79] (Supplementary Table 28), as well as 989 heritable UKB traits with publicly available GWAS summary statistics[80]. Analyses were restricted to HapMap3 variants, excluding the MHC region. Genetic correlations with an FDR < 0.05 were considered significant.

Mendelian randomization

Potential causal associations were examined using GSMR implemented in GCTA v.1.93.1f[81]. GSMR uses multiple genetic variants (here clumped with an $R^2 < 0.001$ and a 10,000-kb window) as instruments (z) to test for causality between an exposure (x) and outcome variable (y), using the ratio $\beta_{xy} = \beta_{zy}/\beta_{zx}$. Designed for two-sample scenarios, GSMR estimates the exposure–outcome effects using GWAS summary statistics from independent samples. Estimates from multiple instruments are integrated using generalized least squares. Instrument heterogeneity is assessed via HEIDI ($P < 0.01$), removing outliers deviating from the causal model. To facilitate effect-size comparisons, we standardized instrument effects on continuous exposures (β_{zx}) based on z-statistic, allele frequency and sample size. GSMR has been demonstrated with superior power to inverse-variance-weighted MR and MR Egger regression[81]. We used GSMR as the primary method for inferring causality and conducted sensitivity analyses using nine alternative MR approaches: inverse-variance-weighted MR (simple, debiased and penalized); MR Egger regression; weighted median-base; maximum-likelihood; mode-based; MR lasso; and contamination-mixture MR, implemented in MendelianRandomization v.0.10.0 (ref. 125). These methods used the same variants as GSMR but without HEIDI-based outlier removal. Twelve risk factors were selected based on the availability of large-scale GWAS, not including UKB individuals[79,81]. Both forward and reverse MR were performed to assess any potential bidirectional effects between risk factors and BAG.

Polygenicity

We used GENESIS v.1.0 (commit e4e6894) to infer genetic effect-size distributions and estimate the number of susceptibility variants underlying BAG under a normal-mixture model of variant effects[84]. Analyses included 1.07 million HapMap3 variants with MAF > 0.05, excluding the MHC region, SNPs with a sample size of less than 0.67 × 90th percentile and those with extreme effect sizes ($z^2 > 80$). We fitted the GENESIS three-component model, which assumes that 99% of variant effects are null, while the remaining 1% follow a mixture of 2 normal distributions, allowing a subset of susceptibility variants to exhibit larger effects. We chose the three-component model over the simpler two-component model because it provides better fits across diverse traits, is robust to model misspecification and reduces downward bias in polygenicity estimates[84]. Default settings were used for defining tagging SNPs ($R^2 > 0.1$ and 1,000-kb window). Neuroticism and height served as benchmark traits for comparison[86,87].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The individual-level data used in this study were obtained from the UKB and LIFE-Adult study. Access to these datasets is restricted to researchers with approved projects. The GWAS summary statistics and polygenic score weights generated from our analyses are available via Zenodo at https://doi.org/10.5281/zenodo.14826943 (ref. 126). Genetic correlation analyses involving UKB traits were conducted using the GWAS summary statistics provided in ref. 80 (https://doi.org/10.5281/zenodo.7186871). Additional GWAS summary statistics used for genetic correlation and Mendelian randomization analyses are detailed in Supplementary Tables 28 and 31.

Code availability

All scripts, conda environments and a complete list of software and resources used in this study are available via GitHub at https://github.com/pjawinski/ukb_brainage.

References

1. Kirkwood, T. B. L. Understanding the odd science of aging. *Cell* **120**, 437–447 (2005).
2. da Silva, R., Conde, D. A., Baudisch, A. & Colchero, F. Slow and negligible senescence among testudines challenges evolutionary theories of senescence. *Science* **376**, 1466–1470 (2022).
3. Finch, C. E. & Austad, S. N. History and prospects: symposium on organisms with slow aging. *Exp. Gerontol.* **36**, 593–597 (2001).
4. Vos, T. et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2163–2196 (2012).
5. Franke, K. & Gaser, C. 10 years of BrainAGE as an neuroimaging biomarker of brain aging: what insights did we gain? *Front. Neurol.* **10**, 789 (2019).
6. Cole, J. H., Marioni, R. E., Harris, S. E. & Deary, I. J. Brain age and other bodily 'ages': implications for neuropsychiatry. *Mol. Psychiatry* **24**, 266–281 (2019).
7. Liem, F. et al. Predicting brain-age from multimodal imaging data captures cognitive impairment. *Neuroimage* **148**, 179–188 (2017).
8. Franke, K., Gaser, C., Manor, B. & Novak, V. Advanced BrainAGE in older adults with type 2 diabetes mellitus. *Front. Aging Neurosci.* **5**, 90 (2013).
9. Franke, K., Ristow, M. & Gaser, C. Gender-specific impact of personal health parameters on individual brain aging in cognitively unimpaired elderly subjects. *Front. Aging Neurosci.* **6**, 94 (2014).
10. Steffener, J. et al. Differences between chronological and brain age are related to education and self-reported physical activity. *Neurobiol. Aging* **40**, 138–144 (2016).
11. Richard, G. et al. Assessing distinct patterns of cognitive aging using tissue-specific brain age prediction based on diffusion tensor imaging and brain morphometry. *PeerJ* **2018**, e5908 (2018).
12. Cole, J. H. et al. Brain age predicts mortality. *Mol. Psychiatry* **23**, 1385–1392 (2018).
13. Jawinski, P. et al. Linking brain age gap to mental and physical health in the Berlin Aging Study II. *Front. Aging Neurosci.* **14**, 791222 (2022).
14. Shahab, S. et al. Brain structure, cognition, and brain age in schizophrenia, bipolar disorder, and healthy controls. *Neuropsychopharmacology* **44**, 898–906 (2019).
15. Kaufmann, T. et al. Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nat. Neurosci.* **22**, 1617–1623 (2019).
16. Cole, J. H. et al. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *Neuroimage* **163**, 115–124 (2017).
17. Jonsson, B. A. et al. Brain age prediction using deep learning uncovers associated sequence variants. *Nat. Commun.* **10**, 5409 (2019).
18. Smith, S. M. et al. Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations. *eLife* **9**, e52677 (2020).
19. Ning, K., Zhao, L., Matloff, W., Sun, F. & Toga, A. W. Association of relative brain age with tobacco smoking, alcohol consumption, and genetic variants. *Sci. Rep.* **10**, 10 (2020).
20. Ning, K. et al. Improving brain age estimates with deep learning leads to identification of novel genetic factors associated with brain aging. *Neurobiol. Aging* **105**, 199–204 (2021).

21. Kim, J., Lee, J., Nam, K. & Lee, S. Investigation of genetic variants and causal biomarkers associated with brain aging. *Sci. Rep.* **13**, 1526 (2023).

22. Leonardsen, E. H. et al. Genetic architecture of brain age and its causal relations with brain and mental disorders. *Mol. Psychiatry* **28**, 3111–3120 (2023).

23. Wen, J. et al. The genetic architecture of multimodal human brain age. *Nat. Commun.* **15**, 2604 (2024).

24. Yi, F. et al. Genetically supported targets and drug repurposing for brain aging: a systematic study in the UK Biobank. *Sci. Adv.* **11**, eadr3757 (2025).

25. Gaser, C. et al. CAT: a computational anatomy toolbox for the analysis of structural MRI data. *Gigascience* **13**, giae049 (2024).

26. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

27. Tipping, M. E. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **1**, 211–244 (2001).

28. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (eds Krishnapuram, B. et al.) 785–794 (Association for Computing Machinery, 2016).

29. Loeffler, M. et al. The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany. *BMC Public Health* **15**, 691 (2015).

30. Engel, C. et al. Cohort profile: the LIFE-Adult-Study. *Int. J. Epidemiol.* **52**, e66–e79 (2023).

31. Smith, S. M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T. E. & Miller, K. L. Estimation of brain age delta from brain imaging. *Neuroimage* **200**, 528–539 (2019).

32. Cole, J. H. Multimodality neuroimaging brain-age in UK biobank: relationship to biomedical, lifestyle, and cognitive factors. *Neurobiol. Aging* **92**, 34–42 (2020).

33. Millard, L. A. C., Davies, N. M., Gaunt, T. R., Davey Smith, G. & Tilling, K. Software application profile: PHESANT: a tool for performing automated phenome scans in UK Biobank. *Int. J. Epidemiol.* **47**, 29–35 (2018).

34. Han, L. K. M. et al. Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. *Mol. Psychiatry* **26**, 5124–5139 (2021).

35. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).

36. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

37. Bulik-Sullivan, B. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

38. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

39. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).

40. Zheng, Z. et al. Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries. *Nat. Genet.* **56**, 767–777 (2024).

41. Wu, Y. et al. Genome-wide fine-mapping improves identification of causal variants. Preprint at *medRxiv* https://doi.org/10.1101/2024.07.18.24310667 (2024).

42. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Series B Stat. Methodol.* **82**, 1273–1300 (2020).

43. Benner, C. et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).

44. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).

45. Rentzsch, P., Schubach, M., Shendure, J. & Kircher, M. CADD-Splice—improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med.* **13**, 31 (2021).

46. Aguet, F. et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).

47. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).

48. Qi, T. et al. Genetic control of RNA splicing and its distinct role in complex trait variation. *Nat. Genet.* **54**, 1355–1363 (2022).

49. Weeks, E. M. et al. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. *Nat. Genet.* **55**, 1267–1276 (2023).

50. Stefansson, H. et al. A common inversion under selection in Europeans. *Nat. Genet.* **37**, 129–137 (2005).

51. Okbay, A. et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).

52. Sollis, E. et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* **51**, D977–D985 (2023).

53. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).

54. Nagel, M. et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat. Genet.* **50**, 920–927 (2018).

55. Karlsson Linnér, R. et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.* **51**, 245–257 (2019).

56. Jansen, P. R. et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet.* **51**, 394–403 (2019).

57. Kichaev, G. et al. Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.* **104**, 65–75 (2019).

58. Hollis, B. et al. Genomic analysis of male puberty timing highlights shared genetic basis with hair colour and lifespan. *Nat. Commun.* **11**, 1536 (2020).

59. Jun, G. et al. A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol. Psychiatry* **21**, 108–117 (2016).

60. Rademakers, R., Cruts, M. & van Broeckhoven, C. The role of tau (MAPT) in frontotemporal dementia and related tauopathies. *Hum. Mutat.* **24**, 277–295 (2004).

61. Shadrin, A. A. et al. Vertex-wise multivariate genome-wide association study identifies 780 unique genetic loci associated with cortical morphology. *Neuroimage* **244**, 118603 (2021).

62. van der Meer, D. et al. Understanding the genetic determinants of the brain with MOSTest. *Nat. Commun.* **11**, 3512 (2020).

63. Bittner, S., Ruck, T., Fernández-Orth, J. & Meuth, S. G. TREK-King the blood–brain-barrier. *J. Neuroimmune Pharmacol.* **9**, 293–301 (2014).

64. Wang, W. et al. Lig4-4 selectively inhibits TREK-1 and plays potent neuroprotective roles in vitro and in rat MCAO model. *Neurosci. Lett.* **671**, 93–98 (2018).

65. Le Guen, Y. et al. eQTL of *KCNK2* regionally influences the brain sulcal widening: evidence from 15,597 UK Biobank participants with neuroimaging data. *Brain Struct. Funct.* **224**, 847–857 (2019).

66. Mullins, N. et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).

67. Cho, C. et al. Large-scale cross-ancestry genome-wide meta-analysis of serum urate. *Nat. Commun.* **15**, 3441 (2024).

68. Cornejo, K. G. et al. Activity-assembled nBAF complex mediates rapid immediate early gene transcription by regulating RNA polymerase II productive elongation. *Cell Rep.* **43**, 114877 (2024).

69. Ewald, C. Y. et al. TNIK's emerging role in cancer, metabolism, and age-related diseases. *Trends Pharmacol Sci.* **45**, 478–489 (2024).

70. Bakshi, A. et al. Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits. *Sci. Rep.* **6**, 32894 (2016).

71. Grote, S. GOfuncR: Gene ontology enrichment using FUNC. R package version 1.14.0 (2021).

72. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

73. Takai, Y., Sasaki, T. & Matozaki, T. Small GTP-binding proteins. *Physiol. Rev.* **81**, 153–208 (2001).

74. Wennerberg, K., Rossman, K. L. & Der, C. J. The Ras superfamily at a glance. *J. Cell Sci.* **118**, 843–846 (2005).

75. Ejaz, A., Mattesich, M. & Zwerschke, W. Silencing of the small GTPase DIRAS3 induces cellular senescence in human white adipose stromal/progenitor cells. *Aging* **9**, 860–879 (2017).

76. Wang, L., Yang, L., Debidda, M., Witte, D. & Zheng, Y. Cdc42 GTPase-activating protein deficiency promotes genomic instability and premature aging-like phenotypes. *Proc. Natl Acad. Sci. USA* **104**, 1248–1253 (2007).

77. Abdellaoui, A. & Verweij, K. J. H. Dissecting polygenic signals from genome-wide association studies on human behaviour. *Nat. Hum. Behav.* **5**, 686–694 (2021).

78. Abdellaoui, A. et al. Genetic correlates of social stratification in Great Britain. *Nat. Hum. Behav.* **3**, 1332–1342 (2019).

79. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).

80. rkwalters & Palmer. D. Nealelab/UKBB_ldsc: v2.0.0 (Round 2 GWAS update). *Zenodo* https://doi.org/10.5281/zenodo.7186871 (2022).

81. Zhu, Z. et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, 224 (2018).

82. Evangelou, E. et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425 (2018).

83. Chen, Y. et al. Relationship between changes in late-life blood pressure and the risk of frailty and mortality among older population in China: a cohort study based on CLHLS. *Hypertens. Res.* **47**, 1881–1891 (2024).

84. Zhang, Y., Qi, G., Park, J.-H. & Chatterjee, N. Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nat. Genet.* **50**, 1318–1326 (2018).

85. Montag, C., Ebstein, R. P., Jawinski, P. & Markett, S. Molecular genetics in psychology and personality neuroscience: on candidate genes, genome wide scans, and new research strategies. *Neurosci. Biobehav. Rev.* **118**, 163–174 (2020).

86. Wood, A. R. et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).

87. Baselmans, B. M. L. et al. Multivariate genome-wide analyses of the well-being spectrum. *Nat. Genet.* **51**, 445–451 (2019).

88. Robinson, O. et al. Associations of four biological age markers with child development: a multi-omic analysis in the European HELIX cohort. *eLife* **12**, e85104 (2023).

89. Drewelies, J. et al. There are multiple clocks that time us: Cross-sectional and longitudinal associations among 14 alternative indicators of age and aging. *J. Gerontol. A Biol. Sci. Med. Sci.* **80**, glae244 (2025).

90. Vetter, V. M. et al. Comprehensive comparison of sixteen markers of biological aging: cross-sectional and longitudinal results from the Berlin Aging Study II (BASE-II). Preprint at *medRxiv* https://doi.org/10.1101/2025.04.09.25325514 (2025).

91. Gaser, C., Franke, K., Klöppel, S., Koutsouleris, N. & Sauer, H. BrainAGE in mild cognitive impaired patients: predicting the conversion to Alzheimer's Disease. *PLoS ONE* **8**, e67346 (2013).

92. Sodini, S. M., Kemper, K. E., Wray, N. R. & Trzaskowski, M. Comparison of genotypic and phenotypic correlations: Cheverud's conjecture in humans. *Genetics* **209**, 941–948 (2018).

93. Cheverud, J. M. A comparison of genetic and phenotypic correlations. *Evolution* **42**, 958–968 (1988).

94. Nebe, S. et al. Enhancing precision in human neuroscience. *eLife* **12**, e85980 (2023).

95. Vidal-Pineiro, D. et al. Individual variations in 'brain age' relate to early-life factors more than to longitudinal brain change. *eLife* **10**, e69995 (2021).

96. Frei, O. et al. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nat. Commun.* **10**, 2417 (2019).

97. Jawinski, P. et al. Human brain arousal in the resting state: a genome-wide association study. *Mol. Psychiatry* **24**, 1599–1609 (2019).

98. Qiu, K., Wang, J., Wang, R., Guo, Y. & Zhao, L. Soft sensor development based on kernel dynamic time warping and a relevant vector machine for unequal-length batch processes. *Expert Syst. Appl.* **182**, 115223 (2021).

99. Franke, K., Ziegler, G., Klöppel, S. & Gaser, C. Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. *Neuroimage* **50**, 883–892 (2010).

100. de Lange, A.-M. G. et al. Mind the gap: performance metric evaluation in brain-age prediction. *Hum. Brain Mapp.* **43**, 3113–3129 (2022).

101. Larivière, S. et al. The ENIGMA Toolbox: multiscale neural contextualization of multisite neuroimaging datasets. *Nat. Methods* **18**, 698–700 (2021).

102. Diedenhofen, B. & Musch, J. cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS ONE* **10**, e0121945 (2015).

103. Welsh, S., Peakman, T., Sheard, S. & Almond, R. Comparison of DNA quantification methodology used in the DNA extraction protocol for the UK Biobank cohort. *BMC Genomics* **18**, 26 (2017).

104. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

105. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* https://doi.org/10.1186/s13742-015-0047-8 (2015).

106. Gazal, S. et al. Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nat. Genet.* **50**, 1600–1607 (2018).

107. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

108. Mägi, R. et al. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum. Mol. Genet.* **26**, 3639–3650 (2017).

109. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).

110. Visscher, P. M. et al. 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).

111. Palmer, C. & Pe'er, I. Statistical correction of the Winner's Curse explains replication variability in quantitative trait genome-wide association studies. *PLoS Genet.* **13**, e1006916 (2017).

112. Ripke, S. et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

113. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).

114. Weissbrod, O. et al. Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat. Genet.* **52**, 1355–1363 (2020).

115. Benner, C. et al. Prospects of fine-mapping trait-associated genomic regions by using summary statistics from genome-wide association studies. *Am. J. Hum. Genet.* **101**, 539–551 (2017).

116. Nassar, L. R. et al. The UCSC Genome Browser database: 2023 update. *Nucleic Acids Res.* **51**, D1188–D1195 (2023).

117. Liu, X., Wu, C., Li, C. & Boerwinkle, E. dbNSFP v3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Hum. Mutat.* **37**, 235–241 (2016).

118. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).

119. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).

120. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

121. Carlson, M. GO.db: A set of annotation maps describing the entire Gene Ontology. R package version 3.14.0 (2021).

122. Bioconductor Core Team. *Homo.sapiens*: Annotation package for the Homo.sapiens object. R package version 1.3.1 (2015).

123. Alexa, A., Rahnenführer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600–1607 (2006).

124. Choi, S. W. & O'Reilly, P. F. PRSice-2: polygenic risk score software for biobank-scale data. *Gigascience* **8**, giz082 (2019).

125. Patel, A. et al. MendelianRandomization v0.9.0: updates to an R package for performing Mendelian randomization analyses using summarized data. *Wellcome Open Res.* **8**, 449 (2023).

126. Jawinski, P. & Markett, S. GWAS summary statistics for brain age gap. *Zenodo* https://doi.org/10.5281/zenodo.14826943 (2025).

Acknowledgements

Author contributions

P.J. and S.M. conceived and designed the study and co-wrote the original manuscript. S.M. supervised the project. P.J. performed the analyses, curated the data and generated the visualizations. H.F. and H.K. supported data curation and the genetic analyses. F.B., A.V., A.V.W., M.S. and S.R. contributed to the study design, the interpretation of the results and the critical revision of the manuscript. S.M., M.S., A.V.W. and A.V. were involved in funding acquisition. All authors reviewed and approved the final version of the manuscript.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s43587-025-00962-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43587-025-00962-7.

Correspondence and requests for materials should be addressed to Philippe Jawinski.
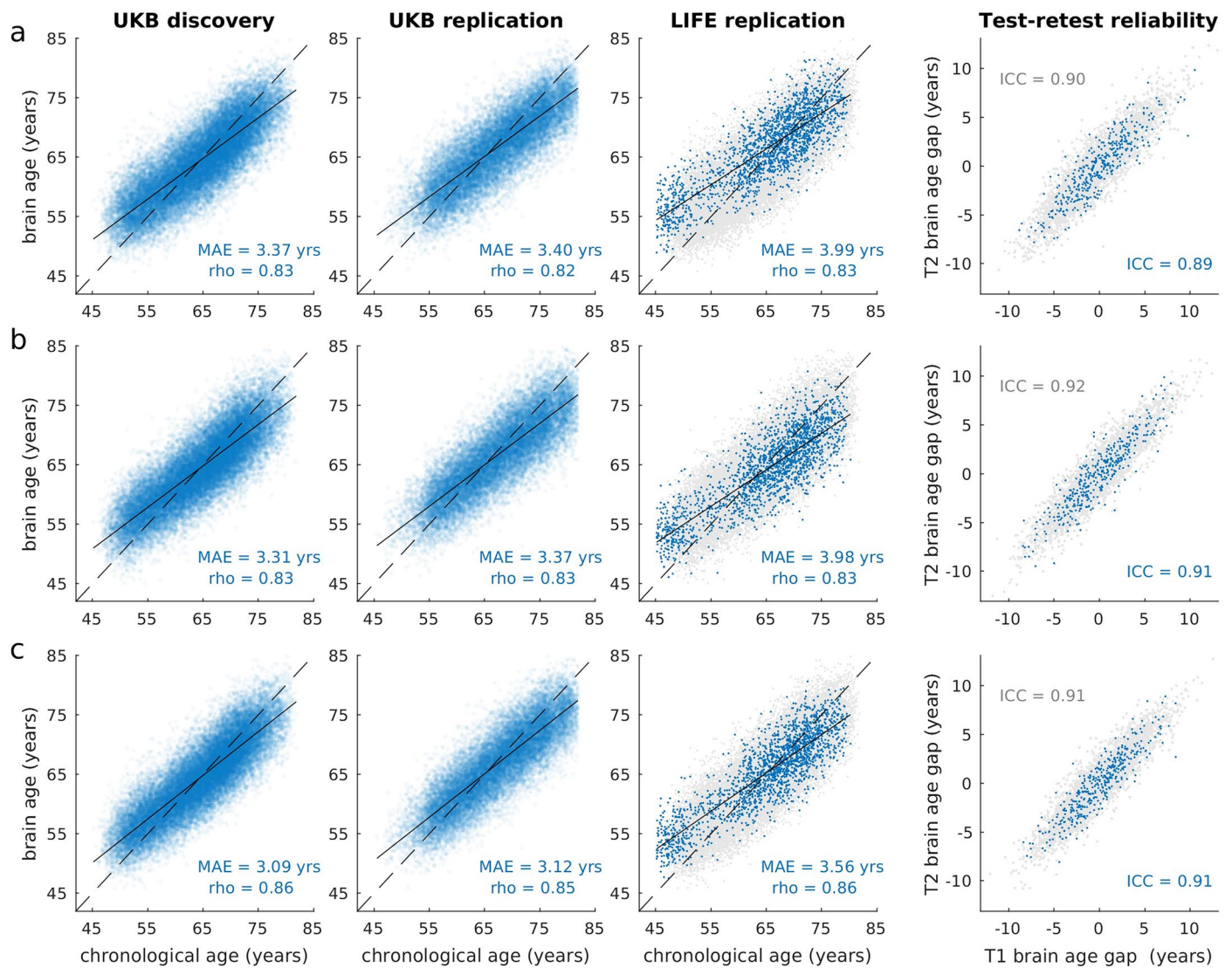
Peer review information *Nature Aging* thanks Lars Bertram, James Cole and Andrew Schork for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended Data Fig. 1 | Accuracy and reliability of brain age estimation models across tissue classes and samples. (**a**) grey matter, (**b**) white matter, and (**c**) combined grey and white matter age estimation models. Blue dots in the first three plots (from left to right) show brain-predicted age estimates plotted against the chronological age in the UKB discovery sample ($n = 32,634$; white-British ancestry), UKB replication sample ($n = 21,881$; multi-ancestry), and LIFE-Adult replication sample ($n = 1,833$; European ancestry). To facilitate comparisons, results of the UKB discovery sample are also shown as grey dots in the background of the LIFE replication plots. At this stage, brain-predicted age estimates have not yet been bias-corrected for regression dilution, that is,

younger participants' ages are systematically overestimated and vice versa, as indicated by the linear regression line (solid) crossing the identity line (dashed). The fourth plot shows the test-retest reliabilities of brain age gap (i.e., the difference between brain-predicted and chronological age) in a subset of the UKB discovery (grey dots, $n = 3,751$) and UKB replication sample (blue dots, $n = 395$). For test-retest comparisons, brain age gap was residualized for effects of age, age², sex, scanner site, and total intracranial volume. MAE: mean absolute error; rho: product-moment correlation coefficient. ICC: intraclass correlation coefficient (C, 1).
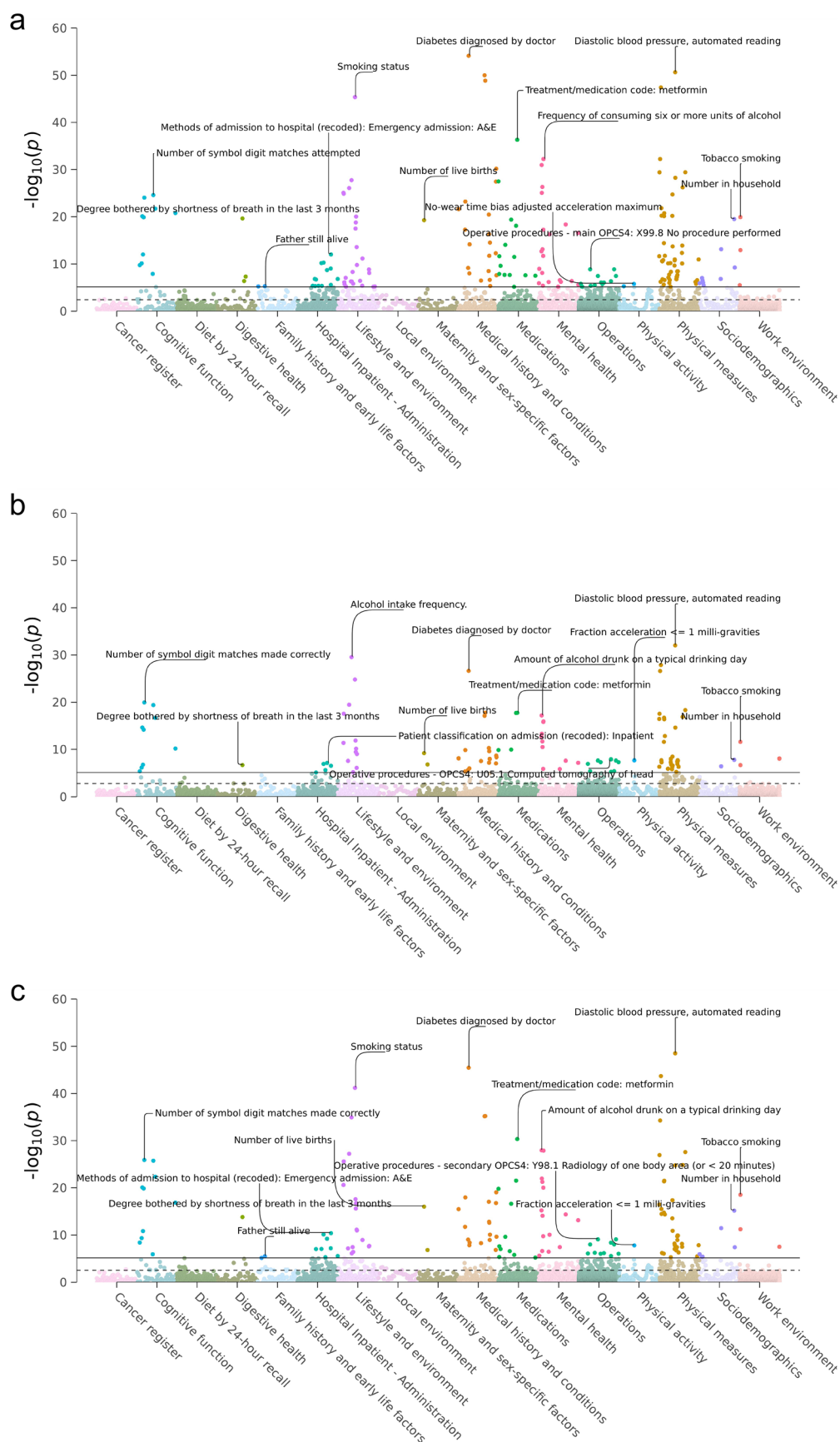
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Cross-trait associations between brain age gap and other UK Biobank phenotypes. Each dot represents an association between brain age gap and one of 7,088 non-imaging derived phenotypes in the UKB discovery sample (n = 32,634). P values (-log10 scale) are shown on the y-axis, with phenotypes on the x-axis categorized by their UK Biobank data dictionary path. Analyses were conducted using PHESANT, which applies data-type-specific regression models (linear, logistic, ordered logistic, or multinomial logistic regression). All models included sex, age, age², scanner site, and total intracranial volume as covariates. Horizontal lines indicate the Bonferroni-adjusted (solid) and FDR-adjusted (dashed) level of significance (two-sided). The top associations in each category are annotated. (**a**) GM BAG, (**b**) WM BAG, (**c**) combined BAG.
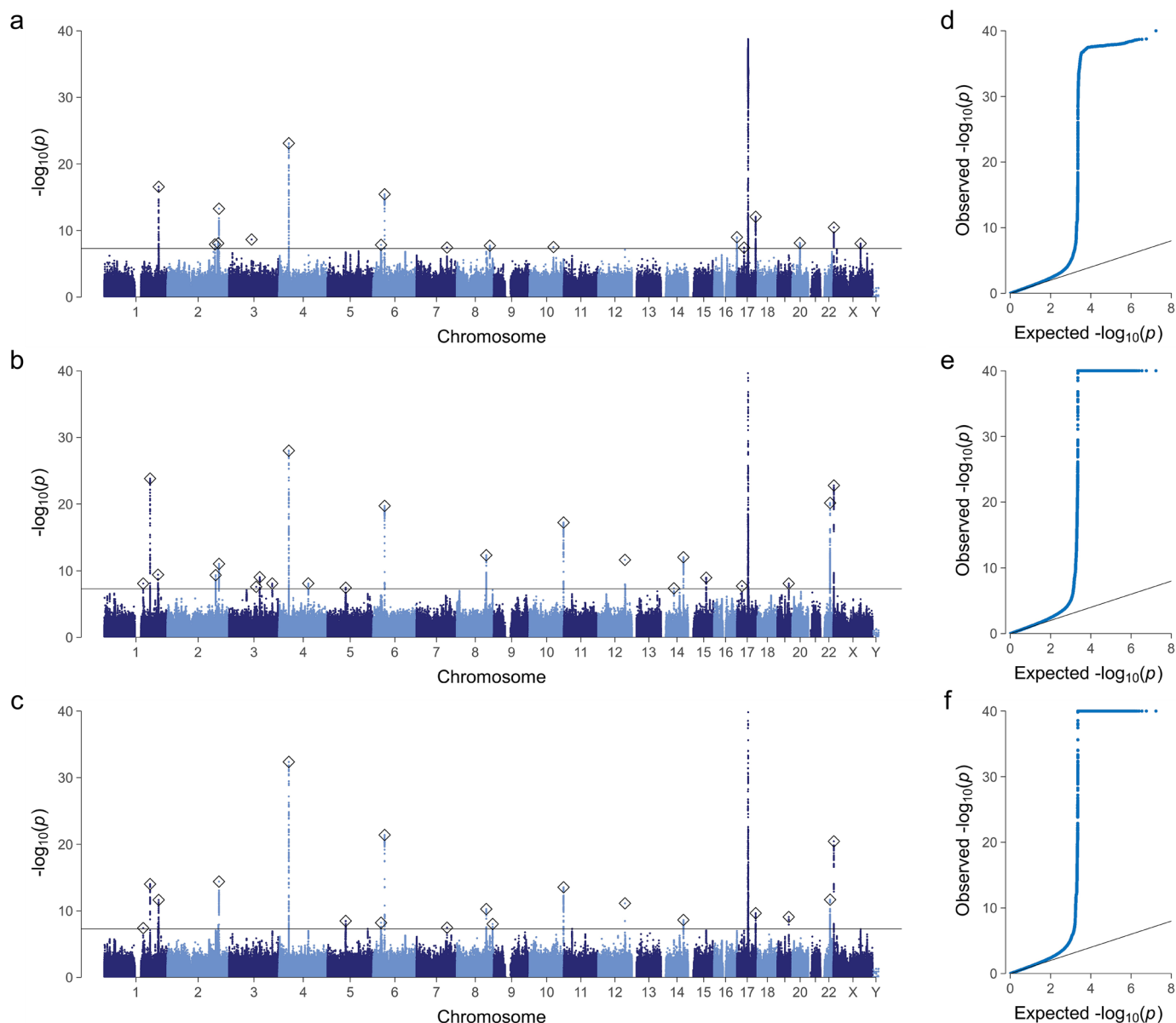
Extended Data Fig. 3 | Results of the multi-ancestry GWAS meta-regression for brain age gap. Manhattan plots (**a-c**) and quantile-quantile (QQ) plots (**d-f**) show the results of the multi-ancestry meta-regression (MR-MEGA) for the three brain age gap traits (up to $N = 56,348$). Multi-ancestry analyses combine results from the discovery sample ($n = 32,634$) and up to seven replication samples: UKB African ancestry ($n = 337$), UKB Admixed American ancestry ($n = 94$); UKB Central/South Asian ancestry ($n = 638$), UKB East Asian ancestry ($n = 291$), UKB European ancestry ($n = 20,423$), and UKB Middle Eastern ancestry ($n = 98$), and LIFE-Adult (European ancestry; $n = 1,833$). Manhattan plots show the P values (-$\log_{10}$ scale) of the tested genetic variants on the y-axis and base-pair positions along the chromosomes on the x-axis. P values were derived from two-sided linear regression models using PLINK, followed by meta-regression using MR-MEGA. The solid horizontal line indicates the threshold of genome-wide significance ($p = 5.0e-08$, two-sided, accounting for multiple testing). Pseudoautosomal variations have been added to chromosome 'X'. Index variants are highlighted by diamonds and were identified by positional clumping with 3,000 kb window-size (no LD threshold was applied in multi-ancestry analyses). Quantile-quantile plots show the observed P values from the association analysis vs. the expected P values under the null hypothesis of no effect (-$\log_{10}$ scale). For illustrative reasons, the y-axis has been truncated at $p = 1.0e-40$. **a,d** grey matter brain age gap (Manhattan and QQ); **b,e** white matter brain age gap (Manhattan and QQ); **c,f** combined brain age gap (Manhattan and QQ).
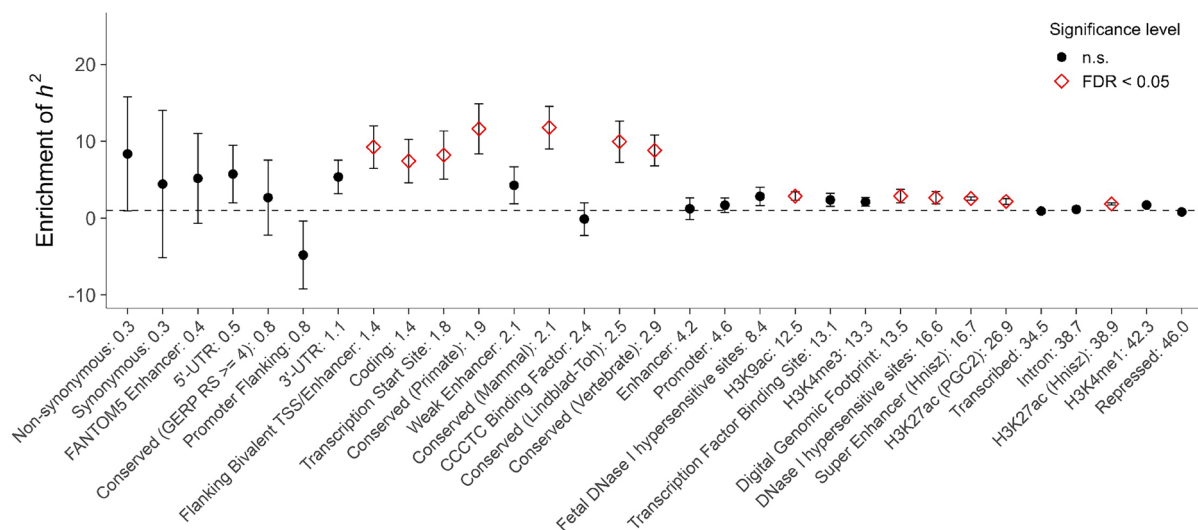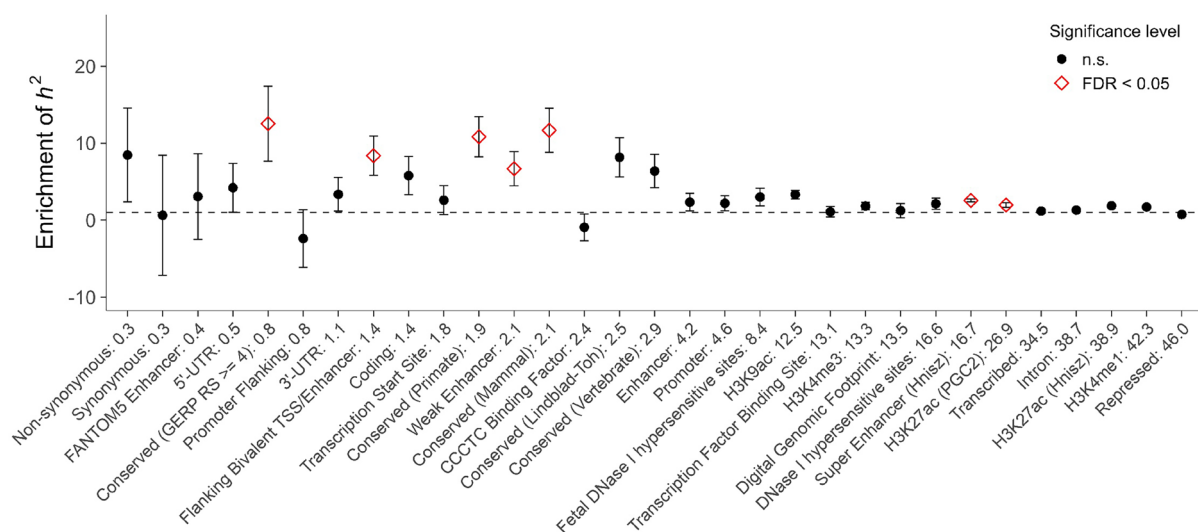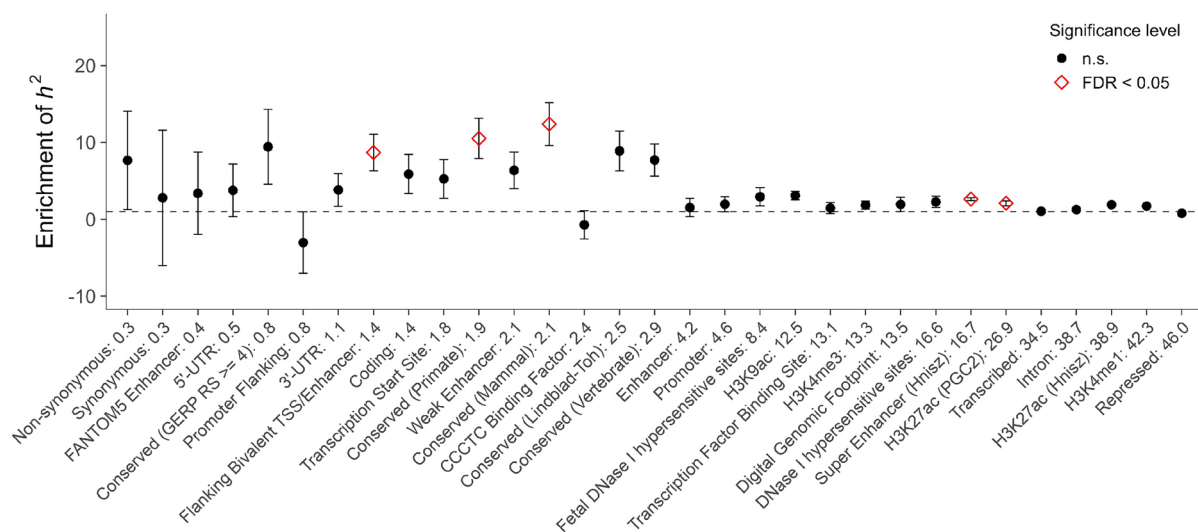
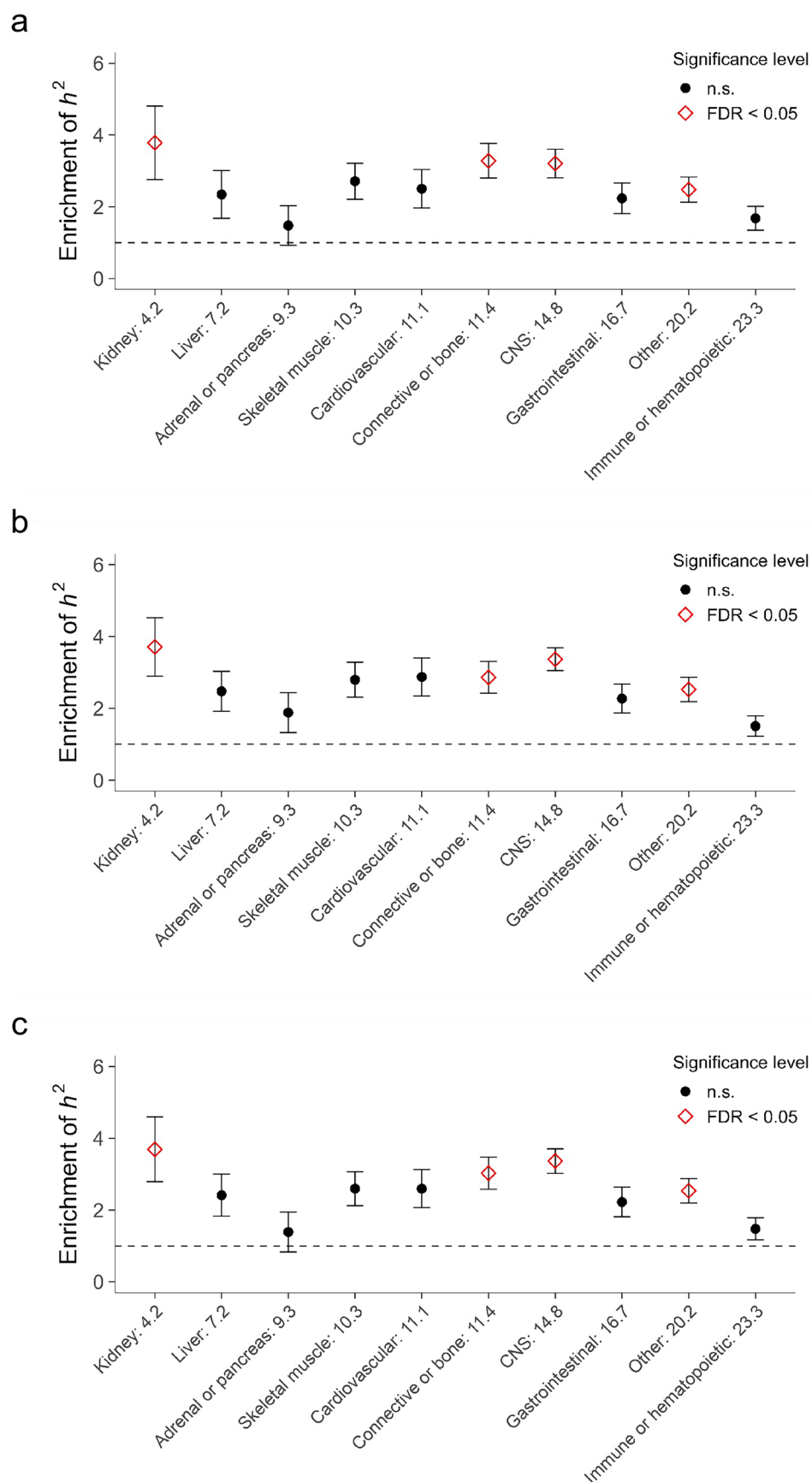**Extended Data Fig. 4 | See next page for caption.**

Extended Data Fig. 4 | Functional partitioning of heritability for brain age gap. Enrichment of heritability (h^2) across functional genomic categories for (**a**) grey matter brain age gap, (**b**) white matter brain age gap, (**c**) combined brain age gap. Results are based on GWAS meta-analyses in individuals of European ancestry (n = 54,890). Heritability enrichment is defined as the proportion of h^2 attributable to a given category divided by the proportion of SNPs in that category. Error bars represent jackknife standard errors around the estimates of enrichment. Please note that significance is tested based on the covariance matrix for the tau coefficient estimates (adjusted for all other baseline categories), which may differ from interpretations based solely on jackknife SEs. Functional categories are ordered by the proportion of SNPs they contain (x-axis), with percentages shown at tick marks. Categories with significant enrichment are marked by red diamonds.

Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Heritability enrichment in genes specifically expressed in selected tissue groups. SNP-based heritability (h^2) enrichment across genomic regions near genes with specific expression in various tissue types, for (**a**) grey matter brain age gap, (**b**) white matter brain age gap, (**c**) combined brain age gap. Results are based on GWAS meta-analyses in individuals of European ancestry (n = 54,890). Heritability enrichment is defined as the proportion of h^2 attributable to a given category divided by the proportion of SNPs in that category. Error bars represent jackknife standard errors around the estimates of enrichment. Please note that significance is tested based on the covariance matrix for the tau coefficient estimates (adjusted for all other baseline categories), which may differ from interpretations based solely on jackknife SEs. Functional categories are ordered by the proportion of SNPs they contain (x-axis), with percentages shown at tick marks. Categories with significant enrichment are marked by red diamonds.

Extended Data Fig. 6 | Gene-based association results for brain age gap traits.
Manhattan plots (**a-c**) and quantile-quantile (QQ) plots (**d-f**) showing the results of the gene-based association analyses for the three brain age gap traits ($n = 54,890$ individuals of European ancestry). Manhattan plots display the P values ($-\log_{10}$ scale) of the tested genes on the y-axis and base-pair positions (gene start coordinates) along the chromosomes on the x-axis. In total, 18,639 protein-coding genes (RefSeq assembly GRCh37.p13, 09-05-2019) have been included. The solid horizontal line reflects the Bonferroni-corrected level of significance. The dashed horizontal line reflects the FDR-corrected level of significance. Diamonds and circles highlight the index gene in each genomic locus (diamonds: Bonferroni-significant index gene; red diamonds: Bonferroni-significant index gene exceeding the y-axis limit; circles: FDR-significant index gene). Genes within 3,000 kb of each other are considered to belong to the same locus and share the same index gene. Quantile-quantile plots show the observed P values from the association analysis vs. the expected P values under the null hypothesis of no effect ($-\log_{10}$ scale). For illustrative reasons, the y-axis has been truncated at $p = 1.0\text{e-}40$. **a,d** grey matter brain age gap (Manhattan and QQ); **b,e** white matter brain age gap (Manhattan and QQ); **c,f** combined brain age gap (Manhattan and QQ).
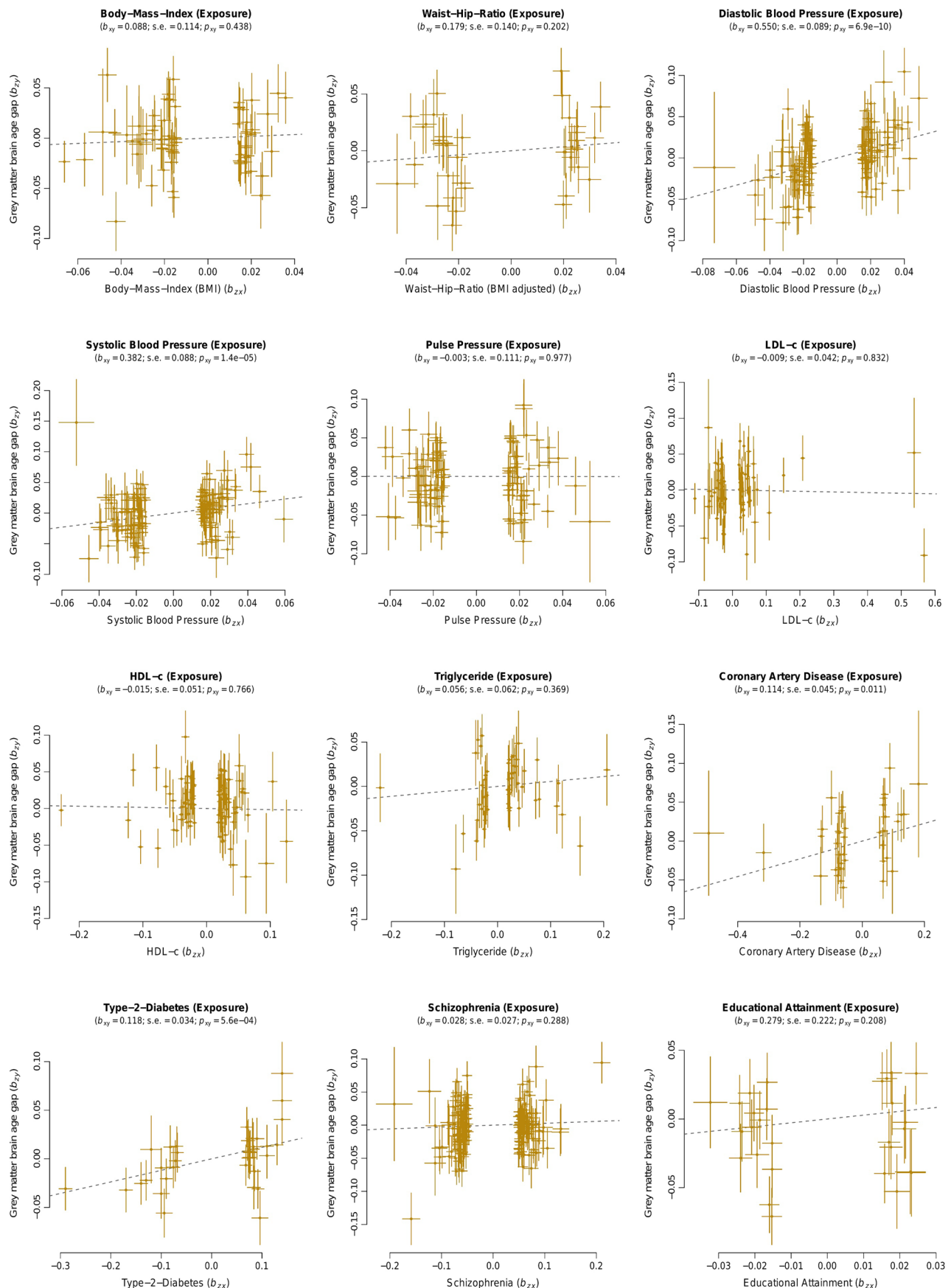
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Mendelian randomization analyses linking modifiable risk factors to brain age gap. Results of the generalized summary-data-based Mendelian randomization (GSMR) results for 12 risk factors (exposures) and combined brain age gap (outcome). Each plot shows multiple genetic variants serving as instruments to test for causality between the exposure and outcome variable. Under a causal model, variant effects on the outcome (b_{zy}; y-axis) are expected to be linearly proportional to the variant effects on the exposure variable (b_{zx}; x-axis). The ratio between b_{zy} and b_{zx} provides an estimate of the mediation effect of x on y (b_{xy}). Variants flagged for potential horizontal pleiotropy were excluded using the HEIDI-outlier method. s.e. standard error of the mediation effect; p_{xy} P value of the mediation effect.
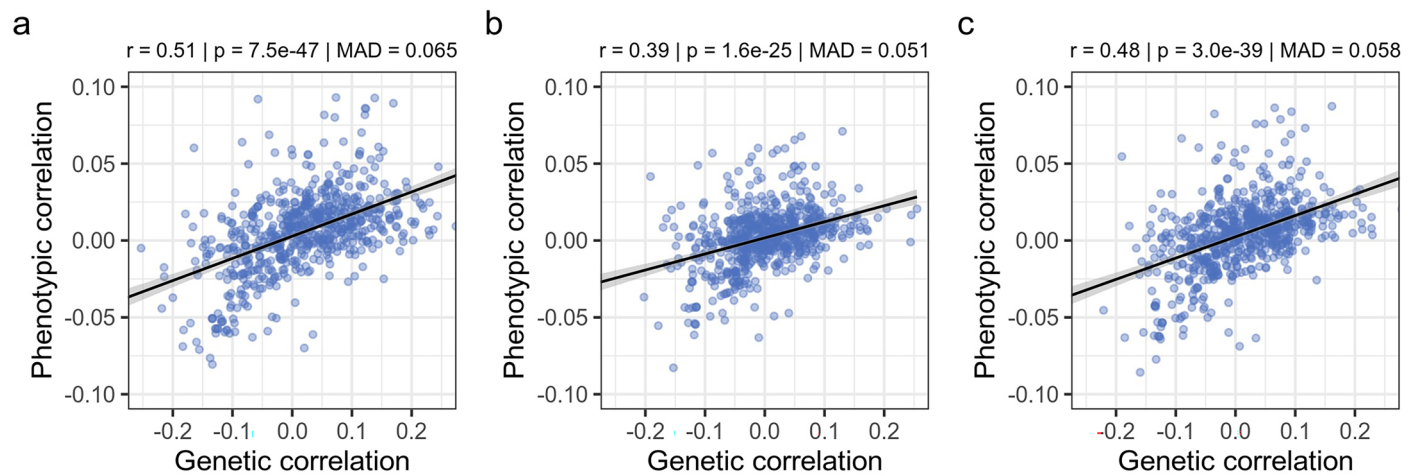
Extended Data Fig. 8 | Relationship between genetic and phenotypic correlations of brain age gap with other complex traits. This figure compares genetic and phenotypic correlation coefficients between brain age gap and 673 other complex traits. Genetic correlations were estimated using GWAS meta-analysis summary statistics for BAG (n = 54,890, European ancestry), while phenotypic correlations were calculated in the UK Biobank discovery sample (n = 32,634, European ancestry) using PHESANT. Summary statistics for other UK Biobank traits were obtained from Neale et al. (10.5281/zenodo.7186871). Results indicate a positive association between genetic and phenotypic correlation estimates. r: product-moment correlation coefficient, p: P value, MAD: mean absolute difference. (**a**) grey matter brain age gap, (**b**) white matter brain age gap, (**c**) combined brain age gap.

nature portfolio

Corresponding author(s):   Philippe Jawinski

Last updated by author(s):   Jul 10, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Genetic, brain imaging, and phenotypic data were obtained from the UK Biobank and LIFE-Adult cohorts. No software was used by the authors for data collection in the present study.
Data analysis	All analysis scripts used in this work are available on GitHub (https://github.com/pjawinski/ukb_brainage), where we also provide conda environments with further details on R and python package versions. Analyses were run on Debian GNU/Linux 11 (bullseye) with kernel version 5.10.0-23-amd64. A list of required tools is provided below. - R (3.5.1-4.41) | Statistical computing and plotting, included in conda environments - MATLAB (R2021a) | MRI preprocessing and age-prediction | https://de.mathworks.com/ - SPM12 (r7487) | MRI preprocessing | https://www.fil.ion.ucl.ac.uk/spm/ - CAT12 (r1364) | MRI preprocessing | https://neuro-jena.github.io/cat/ - RVM-Matlab v1.0.0 | Age-prediciton | https://github.com/iqiukp/RVM-MATLAB - XGBoost v0.82.1 | Age-prediction | https://xgboost.ai/ - PHESANT v1.1| Cross-trait (phenome-wide) association analysis in UK Biobank | https://github.com/MRCIEU/PHESANT - ENIGMA toolbox v2.0.3 for MATLAB | Visualizing cortical and subcortical surface associations | https://github.com/MICA-MNI/ENIGMA - PLINK v1.90b6.8 64-bit Intel | Genomic preprocessing | https://www.cog-genomics.org/plink/ - PLINK v2.00a2LM 64-bit Intel | Genomic preprocessing and genome-wide association analysis | https://www.cog-genomics.org/plink/2.0/ - METAL 2020-05-05 | Fixed-effects meta-analysis in European ancestry samples | https://csg.sph.umich.edu/abecasis/Metal/ - MR-MEGA v0.2 | Multi-ancestry meta-analysis (meta-regression) | https://genomics.ut.ee/en/tools - GWAMA v2.2.2 | Multi-ancestry meta-analysis (random-effects) | https://genomics.ut.ee/en/tools - LOCUSZOOM v1.4 standalone | Regional association plots | https://genome.sph.umich.edu/wiki/LocusZoom_Standalone

- GCTA v1.93.1f beta Linux | Conditional, gene-based, and MR analyses | https://yanglab.westlake.edu.cn/software/gcta/
- LD Score Regression v1.0.1 | Heritability & stratified heritability analysis; Genetic correlations | https://github.com/bulik/ldsc
- SMR v1.03 | Summary-data-based Mendelian Randomization | https://yanglab.westlake.edu.cn/software/smr/
- LDSTORE v2.0 x86/64 | Estimating and storing linkage-disequilibrium data | http://www.christianbenner.com/
- FINEMAP v1.4.2 | Identification of causal variants | http://www.christianbenner.com/
- susieR v0.12.35 | Identification of causal variants | https://github.com/stephenslab/susieR
- GCTB v2.5.2 | Identification of causal variants and polygenic score analysis
- PRSice2 v2.3.3 | Polygenic score analysis | https://choishingwan.github.io/PRSice/
- MAGMA v1.10 Linux Debian 64-bit | Gene-based analysis to produce input for PoPS | https://ctg.cncr.nl/software/magma
- PoPS v0.2 | Gene prioritization | https://github.com/FinucaneLab/pops
- GOfuncR 1.14.0 | Gene Set Enrichment Analysis | https://github.com/sgrote/GOfuncR
- GENESIS e4e6894 | Polygenicity analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The individual-level data used in this study were obtained from the UK Biobank (https://www.ukbiobank.ac.uk/) and LIFE-Adult study (https://www.uniklinikum-leipzig.de/einrichtungen/life). Access to these datasets is restricted to researchers with approved projects. The GWAS summary statistics and polygenic score weights generated from our analyses are publicly available on Zenodo (https://doi.org/10.5281/zenodo.14826943). Genetic correlation analyses involving UK Biobank traits were conducted using GWAS summary statistics provided by Neale and colleagues (https://doi.org/10.5281/zenodo.7186871). Additional GWAS summary statistics used for genetic correlation and Mendelian Randomization analyses are detailed in Supplementary Tables B28 and B31. All scripts used in this work are available on GitHub (https://github.com/pjawinski/ukb_brainage).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Analyses were restricted to participants whose self-reported sex aligned with their genetic sex. Gender information was not collected. Sex was utilized as a covariate in our GWAS and cross-trait association analyses. Sex-stratified analyses were added during the revision in response to a reviewer's request. The discovery sample comprised 32,634 participants (17,084 female), while the full multi-ancestry replication sample included 23,714 participants (12,339 female).
Reporting on race, ethnicity, or other socially relevant groupings	Participants were assigned to ancestry groups to account for population stratification in GWAS analyses. In our discovery GWAS, only individuals of white-British ancestry were included (UKB data-field 22006). In replication analyses, individuals with a valid ancestry assignment from the Pan-ancestry UKB project were included (UKB return 2442; https://pan.ukbb.broadinstitute.org/). We define "ancestry" as a statistical concept based on the genetic variants inherited from ancestors, distinct from the social constructs of "race" and "ethnicity". Throughout our analyses, we specifically refer to genetic ancestry, utilizing the following broad categories: - White-British ancestry (for discovery only) - European ancestry - Central/South Asian ancestry - African ancestry - East Asian ancestry - Middle Eastern ancestry - Admixed American ancestry Ancestry assignment was conducted by comparing each participant's genome in the UK Biobank with data from two extensive reference datasets, the 1000 Genomes Project and the Human Genome Diversity Project (HGDP). These ancestry labels are used solely for GWAS purposes and may not align with other demographic or self-reported data.
Population characteristics	The discovery sample included 32,634 UKB participants of white-British ancestry (17,084 female, age range: 45.2-81.9 years, mean age: 64.3 years). The replication sample included 23,714 participants (12,339 female, age range: 45.2-81.9 years, mean age: 67.0 years) from 7 subsamples: 337 African, 94 Admixed American, 638 Central/South Asian, 219 East Asian, 20,423 European, and 98 Middle Eastern ancestry UK Biobank participants (11,451 female, age range: 45.5-81.9 years, mean age: 67.1 years), along with 1,833 European ancestry participants of the LIFE-Adult cohort (888 female, age range: 45.2-80.3 years, mean age: 65.3 years).
Recruitment	The UK Biobank recruited 500,000 volunteers aged 40 to 69 from various regions across the UK between 2006 and 2010. Assessments were conducted in 22 centers spread across Scotland, England, and Wales. Participants were identified through NHS patient registers if they fell within the specified age range and lived reasonably close to an assessment center. Out of nine million invited individuals, 5.5% (500,000) ultimately participated in the study. Approximately 50,000 individuals underwent brain MRI scans at one of four UK Biobank imaging centers located in Cheadle, Newcastle, Reading, and Bristol.

Previous studies have shown that volunteers from the UK Biobank tend to have healthier lifestyles, higher educational attainment, and better overall health compared to the general UK population [doi: 10.1097/EDE.0000000000001316; 10.1192/bjo.2019.100].

The LIFE-Adult study is a population-based study of 10,000 randomly selected adult citizens of Leipzig. Recruitment and examination were conducted from August 2011 until November 2014. Randomly sampled citizens' address lists were obtained from the resident registration office of Leipzig, Germany. Among 29,535 citizens aged 40–79 invited to participate, 31.0% accepted, 29.0% declined, and 36.3% did not respond [doi: 10.1093/ije/dyac114]. The remaining 3.7% either could not be contacted or declined despite expressing initial interest. Analysis comparing the study participants to both the general Leipzig population and non-participants, utilizing official statistics and brief questionnaire data, revealed that study participants were less frequently elderly women and more commonly married, highly educated, employed, in better health, and more frequently current non-smokers [doi: 10.1186/s12874-019-0779-8]

GWAS and post-GWAS analyses have previously been shown to be sensitive to participation bias [doi: 10.1038/s41562-023-01579-9]. Although large volunteer-based biobanks are crucial for advancing genetic epidemiology, the presence of participation bias introduces the possibility of altered results (e.g., altered heritability estimates, genetic correlations, and Mendelian Randomization estimates). Notably, previous studies have observed such alterations more prominently in socio-behavioral traits rather than physical ones. Additionally, research suggests that increasing the representativeness of study samples may affect the magnitude of effects rather than their direction. Given these considerations, we maintain that our inferences and conclusions regarding brain age gap are less susceptible to the influence of participation bias.

Ethics oversight	The UK Biobank received ethical approval from the National Research Ethics Service Committee North West-Haydock (reference 11/NW/0382, 16/NW/0274, 21/NW/0157). The LIFE-Adult study was approved by the Ethics Committee of the University of Leipzig (263–2009-14122009, 263/09-ff, 201/17-ek). All participants provided written informed consent. LIFE-Adult participants received a fixed compensation of 20 EUR per visit to cover travel expenses, while UK Biobank participants could claim reimbursement for travel costs.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The majority of participants were drawn from the January 2020 release (v1.7) of the UK Biobank imaging study (www.ukbiobank.ac.uk). This release contained 40,681 participants with structural T1-weighted MRI data. After applying all data exclusions (see below), the final GWAS discovery sample included 32,634 individuals. For replication analyses, we selected all remaining non-white-British ancestry individuals of the January 2020 release, and added European and non-European ancestry participants whose imaging data were released until May 2024 (v1.10), along with individuals of the LIFE-Adult cohort, yielding a total replication sample of 23,714 participants. The success of a GWAS for a specific trait mainly depends on factors such as the strength of heritability, the number of contributing variants and the distribution of their effect sizes (i.e., the degree of polygenicity), and the achieved sample size. Brain age gap has previously been shown with a relatively strong SNP-based heritability of 20%. Additionally, previous GWAS with smaller sample sizes have been successful in identifying some initial genome-wide significant loci for brain age gap, although the trait is generally considered to have a high degree of polygenicity. Given that our study boasts a 79% increase in sample size (about 25,000 more) compared to the most recent GWAS on brain age gap (Wen et al. 2024, Yi et al. 2025), we believe that our sample size is adequate.
Data exclusions	We did not include T1-weighted MRI scans labelled as 'unusable' and those that did not successfully complete the CAT12 voxel-based morphometry preprocessing. For further downstream analyses, only MRI scans with a CAT12 overall image quality rating < 3.0 were considered. Analyses were limited to participants whose self-reported sex matched the genetic sex (UKB data-field 31 and 2200), who were without indications of sex aneuploidy (data-field 22019), and who were no outliers in heterozygosity and missingness (data-field 22027). We only included unrelated participants as suggested by pairwise kinship coefficients below 0.0442 (pre-calculated coefficients retrieved using the command line tool 'ukbgene' with the 'rel' parameter). In the discovery GWAS, only participants of white-British ancestry were included (data-field 22006). In replication analyses, all remaining individuals of non-white-British ancestry were included, along with individuals from UKB release v1.10 (meeting the same quality criteria). Furthermore, in replication analyses, only individuals with a valid ancestry assignment from the Pan-ancestry UKB project were included (UKB return 2442; https://pan.ukbb.broadinstitute.org/). From the LIFE-Adult cohort, we included only individuals within the same age range as UK Biobank participants, and who had available T1-weighted MRI data and passed the study's genetic quality control procedures, including checks for cryptic relatedness and consistency between reported and genetic sex.
Replication	Independent associations from the discovery GWAS of 32,634 individuals were tested for replication in both a European ancestry and multi-ancestry meta-analysis of up to 23,714 individuals. Replication analyses included index variations from the 25 genome-wide significant loci, and index variations from another 45 suggestive loci (conditional p-values ranging from 1.0e-06 to 5.0e 08). The degree of consistency between discovery and replication results was highly unlikely to occur by chance. Of the 25 discoveries, all showed consistent effect directions (binomial test: p = 3.0e-08) and 18 replicated at p < 0.05 (one-tailed nominal significance; binomial test: p = 1.3e-18). This finding aligns closely with the outcomes predicted by statistical power analyses, with 19 out of 25 loci expected to attain one-tailed nominal significance in replication analyses.
Randomization	This study was observational and non-experimental; thus, participants were not randomly assigned, and no formal blinding was applied.

Randomization	Allocation of participants to genotype groups can be regarded as a result of random transmission of alleles from parents to offspring (caveat: assortative mating may be a potential source of bias). Several covariates were used in cross-trait analyses: sex, age, age2, scanner site, and total intracranial volume. The following covariates were used in GWAS analyses: sex, age, age2, scanner site, total intracranial volume, genotyping array, and up to 20 genetic principal components.
Blinding	This study was observational and non-experimental; thus, participants were not randomly assigned, and no formal blinding was applied. However, we assume that investigators involved in data collection were unaware of genotype information. Given the large number of variants analyzed (~9.6 million), and the hypothesis-free nature of GWAS, we do not consider blinding to be a relevant source of bias in data collection or analysis.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Plants

Seed stocks	No plants were used in this study.
Novel plant genotypes	No plants were used in this study.
Authentication	No plants were used in this study.

Magnetic resonance imaging

Experimental design

Design type	Only structural T1-weighted MRI data were used in this study.
Design specifications	We did not perform an MRI experiment involving varying task or stimulus conditions. However, to estimate the test-retest reliability of brain age estimates in the UK Biobank, we used structural MRI scans from the initial imaging visit (2014+) and the first repeat imaging visit (2019+). Only scans from the initial imaging visit were included in the genetic analyses.
Behavioral performance measures	No behavioral measures relevant to the present study were collected during the MRI sessions.

Acquisition

Imaging type(s)

structural

Field strength

3T

Sequence & imaging parameters

In the UK Biobank cohort, T1-weighted structural MRI scans were acquired using a 3D MPRAGE sequence in the sagittal plane, with 1×1×1 mm voxel-size, 208×256×256 acquisition matrix, 2,000 ms repetition time (TR), 2.01 ms echo time (TE), 880 ms inversion time (TI), 6.1 ms echo spacing, 8° flip angle, 240 Hz/pixel bandwidth, in-plane acceleration factor of R = 2, and 4:54 min duration (https://www.fmrib.ox.ac.uk/ukbiobank/protocol/). In the LIFE-Adult study, T1-weighted structural images were obtained using a 3D MPRAGE sequence with 1×1×1 mm voxel-size, 256×240×176 acquisition matrix, TR = 2,300 ms, TE = 2.98 ms, TI = 900 ms, and 9° flip angle.

Area of acquisition

whole-brain

Diffusion MRI   ☐ Used   ☒ Not used

Preprocessing

Preprocessing software

T1-weighted MRI scans in NIFTI-format were preprocessed using the voxel-based morphometry pipeline of CAT12 (r1364, http://dbm.neuro.uni-jena.de) for SPM12 (r7487) in MATLAB R2021a (The MathWorks Inc, Natick, MA, USA). CAT12 preprocessing involved affine and DARTEL registration of brain images to a reference brain, segmentation into grey matter, white matter, and cerebro-spinal fluid, bias correction for intensity inhomogeneity, and modulation of segmentations to account for the amount of volume changes due to spatial registration. Processed images were smoothed by applying an 8×8×8mm full-width-at-half-maximum (FWHM) gaussian kernel with subsequent resampling to 8mm3 voxel size. We only considered MRI scans with a CAT12 overall image quality rating < 3.0 for further downstream analyses.

Normalization

CAT12 employes DARTEL and the Geodesic Shooting normalization by means of existing templates in MNI space.

Normalization template

Predefined normalization templates were derived from 555 IXI individuals. DARTEL template: Template_1_IXI555_MNI152.nii; Geodesic Shooting template: Template_0_IXI555_MNI152_GS.nii

Noise and artifact removal

CAT12 incorporates a spatial adaptive non-local means (SANLM) denoising filtering procedure, followed by a Markov Random Field approach (MRF), bias-correction, local intensity transformation, and the use of an adaptive maximum a posterior techinque to account for intensity inhomogeneities.

Volume censoring

No volume censoring was applied.

Statistical modeling & inference

Model type and settings

We did not perform conventional first and second-level MRI analyses. Instead, the preprocessed, smoothed, and down-sampled brain images underwent dimension-reduction (principal component analysis; PCA) and subsequently served to train our age-prediction models in a supervised machine learning approach.

Effect(s) tested

We did not perform an MRI experiment with task and stimulus conditions.

Specify type of analysis: ☒ Whole brain   ☐ ROI-based   ☐ Both

Statistic type for inference

(See Eklund et al. 2016)

We did not perform conventional MRI analyses, i.e., brain images did not serve as outcome but were used to extract features to predict chronological age. Hence, we did not employ either voxel-wise or cluster-wise methods.

Correction

For the different types of analyses in our GWAS (cross-trait associations, SNP-level analyses, gene-based analyses, genetic correlations, Mendelian Randomization analyses), we used both multiple-testing corrections based on the Benjamini-Hochberg FDR approach and Bonferroni corrections. This enabled us to identify both a larger number of very likely associations using a more liberal significance threshold and a smaller number of highly reliable results using a more conservative significance threshold.

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☐	☒ Multivariate modeling or predictive analysis

Multivariate modeling and predictive analysis

Feature Set
CAT12 preprocessed, smoothed, and down-sampled brain images underwent dimension-reduction (principal component analysis; PCA) and subsequently served to train our age-prediction models in a supervised machine learning approach. We selected the first 500 principal components as features, which explained about 90% of the total variance in brain images and enabled model training in a reasonable period of time with advanced computational resources.

Machine Learning Algorithms
Age estimation models were built using the sparse Bayesian relevance vector machine (RVM) in MATLAB, and the extreme gradient boosting package 'xgboost' v.0.82.1 in R. We used XGBoost with both the decision tree ('gbtree') and linear gradient booster ('gblinear'). The learning rate was set to $\eta = 0.02$ with 5000 training iterations and an early stopping after 50 iterations in the case of no further model improvement. The maximum tree depth was set to 3. Default settings were used for all other training parameters.

Model Training and evaluation metrics
Age estimation models were trained with the brain image PCA scores serving as features and chronological age serving as outcome variable. Model training and application was carried out in a 10-fold cross-prediction manner with 100 repeats. Therefore, we randomly split the discovery sample into ten equal-sized subsets, of which nine subsets served for model training, and the remaining subset, the test sample, served for applying the model. Brain images of the training sample underwent PCA, and transformation parameters were subsequently applied to calculate PCA scores in the test sample. After the first model was trained and tested, the next subset served as test sample, while the other nine subsets were selected as training sample. This strategy was carried on until each subset served exactly once as test sample. The tenfold cross-prediction procedure was repeated 100 times, so that 100 predictions were made for each subject. This procedure was performed for each tissue (grey and white matter) and model type (relevance vector machine, xgboost tree, and xgboost linear), resulting in a total number of 600 brain-predicted age estimates per subject. In a nested 10-fold cross-validation approach, we stacked the estimates from the three different model types in an ensemble estimate, resulting in 100 brain-predicted age estimates for grey matter, white matter, and combined grey and white matter, respectively. Finally, these estimates were averaged, leaving one brain-predicted age estimate for grey matter, white matter, and combined grey and white matter for each subject. In the discovery sample, we observed accurate predictions of chronological age, with mean absolute errors (MAE) reaching MAE = 3.09 years and correlation coefficients attaining $r = .86$. Model performances (without retraining) were similar in the multi-ancestry UKB replication sample (MAE = 3.12 years; $r = .85$), and LIFE-Adult sample (MAE = 3.56 years; $r = .86$).