HHS Public Access

Author manuscript

Nat Aging. Author manuscript; available in PMC 2024 October 02.

Published in final edited form as:

Nat Aging. 2024 September ; 4(9): 1290–1307. doi:10.1038/s43587-024-00662-8.

The genetic architecture of biological age in nine human organ systems

Junhao Wen[1,✉], **Ye Ella Tian**[2], **Ioanna Skampardoni**[3], **Zhijian Yang**[3], **Yuhan Cui**[3], **Filippos Anagnostakis**[4], **Elizabeth Mamourian**[3], **Bingxin Zhao**[5], **Arthur W. Toga**[6], **Andrew Zalesky**[2], **Christos Davatzikos**[3]

[1]Laboratory of AI and Biomedical Science (LABS), University of Southern California, Los Angeles, CA, USA.

[2]Melbourne Neuropsychiatry Centre, Department of Psychiatry, Melbourne Medical School, The University of Melbourne, Melbourne, Victoria, Australia.

[3]Artificial Intelligence in Biomedical Imaging Laboratory (AIBIL), Center for AI and Data Science for Integrated Diagnostics (AI2D), Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

✉**Correspondence and requests for materials** should be addressed to Junhao Wen., junhaowe@usc.edu.

Author contributions

Study concept and design: J.W. Acquisition, analysis or interpretation of data: J.W. Drafting of the manuscript: J.W., Y.E.T., A.Z. and C.D. Critical revision of the manuscript for important intellectual content: J.W., Y.E.T., I.S., Z.Y., Y.C., F.A., E.M., B.Z., A.Q.T., A.Z. and C.D. Statistical analysis: J.W. BAG index generation: Y.E.T.

Competing interests

The authors declare no competing interests.

Reporting summary

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

Code availability

The software and resources used in this study are all publicly available at the following links:

- BioAge (git version 37e54eb): https://github.com/yetianmed/BioAge, biological age prediction
- P LINK (v2.0): https://www.cog-genomics.org/plink/, linear model GWAS
- FUMA (v1.5.0): https://fuma.ctglab.nl/, gene mapping and genomic locus annotation
- GCTA (v1.94.1): https://yanglab.westlake.edu.cn/software/gcta/#Overview, heritability estimates and mixed-effect GWAS with fastGWA
- LDSC (git version aa33296): https://github.com/bulik/ldsc, genetic correlation and partitioned heritability
- TwoSampleMR (v0.5.6): https://mrcieu.github.io/TwoSampleMR/index.html, Mendelian randomization
- Coloc (v5): https://github.com/chr1swallace/coloc, Bayesian colocalization
- LCV (git version 39950a8): https://github.com/lukejoconnor/LCV, LCV for causal inference
- GREP (v1.0.0): https://github.com/saorisakaue/GREP, gene–drug–disease network analysis

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s43587-024-00662-8.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43587-024-00662-8.

[4]Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy.

[5]Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA, USA.

[6]Laboratory of Neuro Imaging (LONI), Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA.

Abstract

Investigating the genetic underpinnings of human aging is essential for unraveling the etiology of and developing actionable therapies for chronic diseases. Here, we characterize the genetic architecture of the biological age gap (BAG; the difference between machine learning-predicted age and chronological age) across nine human organ systems in 377,028 participants of European ancestry from the UK Biobank. The BAGs were computed using cross-validated support vector machines, incorporating imaging, physical traits and physiological measures. We identify 393 genomic loci–BAG pairs ($P < 5 \times 10^{-8}$) linked to the brain, eye, cardiovascular, hepatic, immune, metabolic, musculoskeletal, pulmonary and renal systems. Genetic variants associated with the nine BAGs are predominantly specific to the respective organ system (organ specificity) while exerting pleiotropic links with other organ systems (interorgan cross-talk). We find that genetic correlation between the nine BAGs mirrors their phenotypic correlation. Further, a multiorgan causal network established from two-sample Mendelian randomization and latent causal variance models revealed potential causality between chronic diseases (for example, Alzheimer's disease and diabetes), modifiable lifestyle factors (for example, sleep duration and body weight) and multiple BAGs. Our results illustrate the potential for improving human organ health via a multiorgan network, including lifestyle interventions and drug repurposing strategies.

Biological aging is complex and influenced by many factors, including genetics, environmental exposures and modifiable lifestyle factors[1] across multiple organ systems. A comprehensive understanding of the phenotypic landscape and genetic architecture underlying biological aging in multiple human organ systems is paramount in forging the path toward precision medicine[2], including identifying vulnerability (for example, smoking) and resilience factors (for example, physical activities). This knowledge can improve our understanding of the underlying mechanisms driving age-related diseases, identify therapeutic targets and develop personalized interventions for maintaining health and functional independence in the aging population.

Previous research efforts have made progress in studying the interconnectedness of multiorgan systems in human health[3–7]. In a recent study by McCracken et al., a heart–brain–liver axis was studied, highlighting direct and indirect associations among the three organs and their interconnectivity and shared biological pathways[6]. A recent review highlighted the role of interorgan signals in metabolic control, including the secretion of peptides, small molecules and lipid mediators by metabolic tissues and the involvement of the central nervous system (CNS) in coordinating peripheral metabolic functions[8]. Riding the crest of the wave of artificial intelligence (AI), the biomedical community has increasingly adopted the biological age gap (BAG) as a comprehensive biomarker of human aging in multiple human organ systems. Specifically, BAG serves as a quantitative phenotype to capture the disparity between an individual's AI-derived age and chronological

age, which can be used to model aging-related normative trajectory at the individual level and holds potential for application in disease populations to capture pertinent pathological processes. For instance, Nie et al. derived the biological age in nine organ systems to predict the possibility of becoming centenarian[7]. In our previous study, Tian et al. derived eight BAGs in eight organ systems, correlating them with cognition, chronic disease, lifestyle factors and mortality[3]. We used a support vector machine in cross-validation to predict BAGs for multiple organ systems.

However, genetic determinants and biological pathways that underlie the observed heterogeneity of organ-specific BAGs remain elusive. Furthermore, whether chronic diseases and lifestyle factors causally impact the divergence between predicted age and chronological age in these organ systems remains to be established, manifesting as either a younger or older biological age. Our previous genome-wide association study (GWAS) uncovered the genetic architecture of multimodal brain BAGs using magnetic resonance imaging data[9]. Expanding on prior research, the current study sought to depict the genetic architecture underlying biological aging across nine human organ systems, including the brain, cardiovascular, eye, hepatic, immune, metabolic, musculoskeletal, pulmonary and renal BAGs. Our overarching hypothesis postulates that the genetic determinants associated with the nine BAGs are specific to individual organ systems (that is, BAG–organ specificity) and directly or indirectly interconnected with other organ systems (that is, interorgan connection).

In the current study, we analyzed multimodal data from 377,028 individuals of European ancestry in the UK Biobank study[10] (UKBB) to capture the genetic architecture of the nine organ systems. First, we used data from 154,774 participants to perform a GWAS and gene-level, partitioned heritability and genetic correlation analyses (Methods). In our Mendelian randomization analyses, we used 222,254 UKBB participants that did not overlap with the individuals used to compute the BAG to avoid potential bias. We (1) identified both previously reported and newly identified genomic loci, (2) demonstrated a greater genetic heritability estimate for the brain BAG than other organ systems, (3) constructed a network linking genes, drugs and diseases for potential drug repurposing, (4) confirmed that BAG-associated variants and genes exhibit BAG–organ specificity and interorgan connection and (5) established both genetic correlations and causal networks among the nine BAGs, chronic diseases and lifestyle factors. All results, including the GWAS summary statistics, are publicly accessible through the multiorgan biomedical science (MEDICINE) knowledge portal at https://labs-laboratory.com/medicine.

Results

Genome-wide associations for 393 genomic loci–BAG pairs

In populations of European ancestry, a GWAS (Methods) identified 11, 44, 17, 41, 61, 76, 24, 67 and 52 genomic loci ($P < 5 \times 10^{-8}$) significantly associated with brain, cardiovascular, eye, hepatic, immune, metabolic, musculoskeletal, pulmonary and renal BAGs, respectively (Fig. 1). All details of the identified loci are presented in Source Data File 1. Manhattan and QQ plots are presented in Supplementary Figs. 1–9.

We estimated the intercept of linkage disequilibrium score regression (LDSC)[11] for the nine main GWASs and obtained intercepts of 0.9989 ± 0.009, 1.0185 ± 0.0099, 0.9926 ± 0.0106, 1.0416 ± 0.0113, 1.0293 ± 0.0107, 1.0308 ± 0.0124, 1.0282 ± 0.0099, 1.0442 ± 0.0104 and 1.0257 ± 0.0112 for the nine BAGs. All the intercepts were close to 1, indicating no substantial genomic inflation in the primary GWAS. Furthermore, we conducted four sensitivity analyses (Methods) to assess the robustness of the primary nine GWASs on individuals of European ancestry (Supplementary Note 2 and Extended Data Fig. 1). Our GWASs demonstrated robustness in split-sample GWAS, with a perfect concordance rate for the sign (+/−) of β values (C: $\beta = 1$) between the split1 and split2 GWASs. The two sets of β values were highly correlated ($0.90 < r$: $\beta < 0.99$ for Pearson's r) and did not significantly differ (P: $\beta > 0.48$ for two-sample t-test). We compared the GWAS results with linear models in PLINK and linear mixed-effect models in fastGWA, resulting in a perfect concordance for the two sets of β values as well as very similar LDSC intercept values. These findings further support the absence of cryptic population stratification in our primary GWASs. Sex-stratified GWASs unveiled distinctive genetic patterns specific to each sex, with noteworthy disparities observed in the genetic architecture of the immune BAG (r: $\beta = 0.29$; P: $\beta = 0.01$; C: $\beta = 0.55$). Immune responses exhibit sex differences that vary across the lifespan and are influenced by age and reproductive status[12]. Detailed quantitative information regarding these observations can be found in Supplementary Note 2, while visual representations of these patterns are available in Supplementary Figs. 5 and 7. Finally, the genetic signals identified within populations not of European ancestry were less prominent than the European GWAS due to the limited sample size, but we found a high concordance between the two sets of β values using the three proposed metrics ($0.85 < r$: $\beta < 0.95$; $0.89 < C$: $\beta < 1$; P: $\beta > 0.12$). This underscores the necessity of expanding sample sizes within underrepresented ethnic groups in future GWAS studies. Detailed statistics can be found in Source Data Files 2–5. Certain genomic loci exhibited unique associations with individual organs, whereas others displayed connections to multiple organ BAGs in close genomic proximity based on their cytogenetic position. Many of these loci were mapped to protein-coding genes and provided functional insights, as exemplified in Supplementary Note 3 and Supplementary Figs. 10 and 11.

Phenome-wide associations for organ specificity and interorgan connection

We aimed to investigate the agreement of the identified genomic loci in existing GWAS literature. To this end, we performed a phenome-wide association query in the EMBL-EBI GWAS Catalog for independent genetic signals within each locus, considering linkage disequilibrium and redundant associations (Methods). We used FUMA's standard terminologies to delineate the independent genetic signals, ensuring careful consideration of linkage disequilibrium. Supplementary Note 1 outlines in detail the specific parameters used in FUMA and alternative approaches used in other software and studies.

This phenome-wide associations query identified 11,709 significant associations between the identified loci in our GWAS and clinical traits in the literature linked to each organ system (that is, BAG–organ specificity; Fig. 2a). The genomic loci associated with the brain BAG exhibited the highest proportion of associations (74 of 173) with traits related to the brain, including imaging-derived phenotypes such as brain volume metrics and

white matter microstructure, as demonstrated in a keyword cloud (Fig. 2a). The brain BAG loci were also largely linked to many other traits related to other organ systems and chronic diseases, evidencing interorgan connections, including metabolic (N = 43/173; for example, cholesterol levels), lifestyle factor (N = 1/173; that is, alcohol consumption), neurodegenerative traits (N = 16/173; for example, Alzheimer's disease) and immune status (N = 7/173; for example, lymphocyte count). For the eye BAG loci, most associations were found in the eye (N = 31/128; for example, retinal nerve fiber layer thickness) and brain traits (N = 6/128; for example, brain morphology), among others.

For the seven body organ systems, among the loci associated with the cardiovascular BAG, most associations were observed with cardiovascular traits (319 of 439), such as systolic/diastolic blood pressure and coronary artery disease. Other associations were found with musculoskeletal (N = 30/439), metabolic (N = 14/439), immune (N = 6/439), renal (N = 18/1,890) and brain (N = 9/439) traits. Three hundred and seventy-six of 1,853 associations were related to hepatic traits (for example, blood protein, cirrhosis and bilirubin) for the hepatic BAG loci. Among the loci associated with the immune BAG, abundant associations were found in metabolic (929 of 1,773), immune (N = 244/1,773), hepatic (N = 149/1,853), musculoskeletal (N = 57/1,853) and cardiovascular traits (N = 72/1,853). For the metabolic BAG loci, most associations were observed in metabolic traits (3,841 of 4,907). We found a significant intertwining of metabolic systems with other organ systems, highlighting interorgan connections in human metabolic activities. Details of the phenome-wide associations are presented in Source Data File 6. Furthermore, we reported the complementary results of this phenome-wide association query using the GWAS Atlas[13] platform (Supplementary Note 4, Source Data File 7 and Supplementary Fig. 12).

Single-nucleotide polymorphism (SNP)-based heritability estimates

We estimated the SNP-based heritability (h^2) across the nine organ systems using the full sample sizes (Fig. 2b) of the nine BAGs. Additionally, the distributions of the magnitude of the β coefficient in the GWAS and the allele frequency of the alternative allele (effect allele) are shown in Fig. 2c,d. Notably, the sample sizes of the brain and eye BAGs were much smaller than those of the seven body organ BAGs; the body organ BAGs had the same populations.

After analyzing the full sample sizes, the estimated h^2 for the brain BAG (0.47 ± 0.02) outperformed all other organ systems, followed by the eye (0.38 ± 0.02), pulmonary (0.36 ± 0.006), renal (0.31 ± 0.006), metabolic (0.29 ± 0.006), cardiovascular (0.27 ± 0.006), musculoskeletal (0.24 ± 0.006), hepatic (0.23 ± 0.006) and immune BAGs (0.21 ± 0.006; Fig. 2b). All heritability estimates were statistically significant after controlling for multiple comparisons using the Bonferroni correction. This trend persisted when subsampling the population of other BAGs to match that of the brain BAG, with comparable distributions in sex and age (Extended Data Fig. 2a). Detailed results of the h^2 estimate are presented in Supplementary Table 1a,b. Of note, we used the GCTA software to estimate h^2, acknowledging that previous research[14] has demonstrated variations in the magnitude of h^2 estimates based on the choice of methods.

To gain deeper insights into the significant genetic signals, we conducted a detailed examination of the effect sizes (β coefficient) in the GWAS of the nine BAGs, as the effect size is independent of the sample size. The independent significant SNPs of the brain ($|\beta| = 0.062 \pm 0.013$; [0.0470, 0.093]) and eye ($|\beta| = 0.0645 \pm 0.030$) BAGs showed larger mean magnitudes than the seven body organ systems (Fig. 2c). Among the body organ BAGs with the same sample size, the renal BAG showed the largest effect size ($0.023 < |\beta| < 0.306$). This pattern persisted with the results using the subsampled populations to the brain BAGs (Extended Data Fig. 2b). The full set of statistics (for example, β coefficients) of the independent significant SNPs is detailed in Source Data File 5 for the European ancestry GWAS. We observed that the alternative (effect) allele frequency of the independent significant SNPs associated with the brain and eye BAGs was relatively higher than that of the body organ BAGs (Fig. 2d). This may indicate that larger samples are required for the brain and eye to detect SNP effects with a relatively lower allele frequency. Additional considerations of the comparison of the genetic signals among the nine BAGs are detailed in Supplementary Note 5, Supplementary Figs. 13 and 14, Supplementary Tables 2 and 3, Source Data File 8 and Extended Data Figs. 3 and 4 regarding allele frequency, sample sizes, underlying features to compute the BAGs and Bayesian colocalization (Methods).

Gene set enrichment analysis (GSEA) for biological pathway annotations

To biologically validate our GWAS findings at the gene level, we performed gene-based associations using MAGMA software based on the full P value distribution from the GWASs of the nine BAGs. The significantly associated genes (Source Data File 9) were used for the GSEA (Methods) to annotate relevant biological pathways underlying each organ system (Fig. 3a).

Genes associated with the cardiovascular BAG were implicated in the insulin-like growth factor II binding (IGF-II) pathway ($P = 7.08 \times 10^{-7}$). Genes associated with the eye BAG were enriched in the pathway of forebrain dorsal–ventral pattern formation ($P = 6.46 \times 10^{-7}$). Among others, the most significant enrichment result shown in the hepatic BAG was the flavonoid glucuronidation pathway ($P = 1.71 \times 10^{-8}$). Genes linked to the metabolic BAG displayed enrichment in several pathways, including the flavonoid glucuronidation pathway ($P = 2.46 \times 10^{-15}$) and triglyceride-rich lipoprotein particle clearance pathway ($P = 3.72 \times 10^{-15}$), both of which are implicated in liver function. In addition, the neutral lipid metabolic process, regulated by complex pathways featuring lipid metabolism enzymes and structural proteins, was also identified. Genes associated with the musculoskeletal BAG exhibited enrichment in the gene set in an amplicon at 20q11 ($P = 1.54 \times 10^{-15}$), defined by a study of copy number alterations conducted on 191 individuals with breast tumors[15]. Genes associated with the pulmonary BAG displayed significant enrichment in the pathways of the negative regulation biosynthetic process ($P = 3.72 \times 10^{-10}$), consistent with a previous DNA methylation analysis of pulmonary function using old-aged Chinese monozygotic twins16. Genes associated with the renal BAG were implicated in the xenobiotic glucuronidation pathway ($P = 1.56 \times 10^{-6}$). Given that the kidney contains most enzymes that metabolize foreign substances (that is, xenobiotics), it plays a crucial role in the overall metabolism of drugs and other foreign compounds within the body (Fig. 3a).

Detailed results of the GSEAs are presented in Source Data File 10. Sex-stratified results are presented in Extended Data Fig. 5.

Tissue-specific gene expression analysis for gene expression patterns

To investigate the gene expression patterns of the significant genes associated with the nine BAGs, we performed a tissue-specific gene expression analysis using MAGMA and the GTEx RNA-sequencing dataset[17] (Methods).

Across 54 human organ tissues (Fig. 3b), genes associated with the cardiovascular BAG exhibited significant overexpression in various heart-related tissues (for example, the aorta and tibial artery) and other organs (for example, the uterus and colon sigmoid). Genes associated with the hepatic BAG were overexpressed in the liver and subcutaneous adipose. Several immune system-related tissues showed a high average expression of genes related to the immune BAG, including the spleen, blood and lymphocytes. Likewise, genes associated with the metabolic BAG showed a high expression level in the liver and intestine, critical organs in the metabolic system. Genes related to the pulmonary BAG displayed significant overexpression in the esophagus gastroesophageal junction, artery and others. Genes associated with the renal BAG were overexpressed in the kidney. Detailed results are presented in Source Data File 11. Sex-stratified results are presented in Extended Data Fig. 6.

Gene–drug–disease network for potential repositionable drugs

We performed a drug target enrichment analysis (Methods) for genes linked to the nine BAGs in the targeted gene sets of drug categories using the DrugBank database, thereby constructing a gene–drug–disease network of potentially repositionable drugs.

The constructed gene–drug–disease network (Fig. 4) identified significant interactions between 12 metabolic BAG-linked genes, 46 drugs and many metabolic disorders encoded in the International Classification of Diseases-10 (ICD-10) code (E70-E90). For instance, the *PPARD* gene was the target gene of the PPARδ agonist SAR-351034, (denoted in Fig. 4), which aimed to improve insulin sensitivity and lipid-related activities and battle against inflammation and oxidative stress, serving as actionable drugs for metabolic disorders, diabetes and kidney and liver injury-related diseases[18]. Our results showed that genes associated with the metabolic BAG were used to develop drugs treating various other diseases, beyond metabolic disorders, related to multiple organ systems (Fig. 4). These included heart-related diseases (for example, chronic rheumatic heart diseases for I05-I09) and cerebrovascular disease (I60-I69), although the enrichment did not survive correction for multiple comparisons (Fig. 4). For instance, the drug MPSK3169A (clinical trial number NCT01609140; metabolic BAG-linked gene *PCSK9*) is used to treat cerebrovascular disease and coronary heart disease, and the drug T3D-959 (clinical trial number NCT04251182; pulmonary BAG-linked gene *PPARD*) was a candidate drug targeting Alzheimer's disease. Detailed results are presented in Source Data File 12.

The drug–gene–disease network revealed the association between genes related to the metabolic BAG and drugs targeting various chronic diseases and highlights the importance

of the metabolic system in the overall functioning of the human body and the potentials of repositioning existing drugs to tackle biological aging.

Heritability enrichment in different cells, categories, tissues and chromatin states

To further biologically validate our GWAS findings at the SNP level, we performed partitioned heritability analyses (Methods) to estimate the heritability enrichment of genetic variants related to the nine BAGs concerning three different cell types (that is, neurons, oligodendrocytes and astrocytes; Fig. 5a), 53 nontissue-specific functional categories (Fig. 5b), 205 tissue-specific gene expression data (Fig. 5c) and 489 tissue-specific chromatin states (Fig. 5d).

We found significant heritability enrichment in oligodendrocytes ($P = 0.03$), a specific type of neuroglial cell, for the brain BAG. The cardiovascular BAG also exhibited significant heritability enrichment in neurons ($P = 0.01$; Fig. 5a and Source Data File 13). Concerning the heritability enrichment in nontissue-specific functional categories, we exemplified the four highest significant partitioned heritability estimates for each BAG in Fig. 5b. For the brain BAG, the superenhancer regions used 17.16% of SNPs to explain 0.47 ± 0.04 of SNP heritability ($P = 1.80 \times 10^{-11}$), and the histone H3 at lysine 9 (H3K9ac) regions used 12.61% of SNPs to explain 0.61 ± 0.12 of SNP heritability ($P = 2.96 \times 10^{-4}$). For the eye BAG, the superenhancer regions explained 0.39 ± 0.05 of SNP heritability ($P = 2.12 \times 10^{-6}$) using 16.84% of SNPs. For the hepatic BAG, the H3K9ac regions explained 0.69 ± 0.13 of SNP heritability ($P = 3.60 \times 10^{-5}$) using 12.61% of SNPs. For the immune BAG, the transcription start site regions (that is, core promoters) explained 0.37 ± 0.08 of SNP heritability ($P = 1.48 \times 10^{-6}$) using 1.82% of SNPs. The 3.11% of SNPs annotated by the promoter regions explained 0.30 ± 0.08 of SNP heritability ($P = 7.64 \times 10^{-4}$) for the metabolic BAG. For the cardiovascular (enrichment = 16.39 ± 2.23; $P = 4.70 \times 10^{-11}$), musculoskeletal (enrichment = 17.34 ± 4.08; $P = 1.65 \times 10^{-6}$), pulmonary (enrichment = 16.82 ± 2.51; $P = 7.58 \times 10^{-9}$) and renal (enrichment = 13.96 ± 1.88; $P = 7.25 \times 10^{-9}$) BAGs, the highest heritability enrichment was found in the regions conserved across mammals (Fig. 5b and Source Data File 14). These results suggested disproportionate genomic contributions to the heritability of BAGs from multiple functional categories.

In addition, the nine BAGs showed high heritability enrichment in specific tissues corresponding to their organ systems. For example, the cardiovascular BAG showed significant heritability enrichment in multiple tissue types, including the artery (for example, the aorta: $P = 1.03 \times 10^{-7}$), myometrium ($P = 1.35 \times 10^{-4}$) and uterus ($P = 2.43 \times 10^{-4}$). Significant heritability enrichment was found in the liver for the hepatic ($P = 5.60 \times 10^{-9}$) and metabolic BAGs ($P = 6.24 \times 10^{-9}$). For the immune BAG, significant heritability enrichment was found in fetal blood tissues ($P = 7.36 \times 10^{-9}$; Fig. 5c and Source Data File 15). These findings were aligned with the tissue-specific gene expression patterns observed at the gene level (Fig. 3b).

The results from the multitissue chromatin state-specific data further provide the proof of concept for the organ-specific heritability enrichment among these nine BAGs. For the brain BAG, significant heritability enrichment was found in multiple brain tissues in the H3K4me3 (for example, $P = 9.06 \times 10^{-5}$ for the hippocampus), H3K4me1 (for example, P

$= 6.94 \times 10^{-5}$ for the hippocampus) and H3K27ac (for example, $P = 1.15 \times 10^{-5}$ for the anterior caudate) regions. For the cardiovascular BAG, significant heritability enrichment was shown in the right ventricle in the H3K4me3 region ($P = 6.36 \times 10^{-5}$) and the artery aorta in the H3K27ac region ($P = 5.81 \times 10^{-7}$). Significant heritability enrichment was found in primary hematopoietic stem cells in the H3K4me1 region for the immune BAG for both women ($P = 5.61 \times 10^{-5}$) and men ($P = 9.50 \times 10^{-5}$). The fetal leg muscle tissue in the DNase regions ($P = 6.54 \times 10^{-5}$) for the musculoskeletal BAG showed significant heritability enrichment. For the pulmonary BAG, significant heritability enrichment was found in the fetal lung in the H3K4me1 ($P = 1.33 \times 10^{-9}$) and DNase regions ($P = 3.80 \times 10^{-8}$), among other tissues from the stomach, artery and muscle. For the renal BAG, significant enrichment was shown in the liver in the H3K9ac region ($P = 2.46 \times 10^{-5}$) and the gastric tissues in the H3K27ac region ($P = 6.24 \times 10^{-5}$; Fig. 5d and Source Data File 16).

Genetic correlations mirror their phenotypic correlations

We estimated the genetic correlation (g_c; Methods) and the phenotypic correlation (p_c; for Pearson's correlation coefficient) between each pair of the nine BAGs. Our results supported the long-standing Cheverud's Conjecture[19], which states that the genetic correlation between two clinical traits reflects their phenotypic correlation (Fig. 5e).

The musculoskeletal and hepatic BAGs showed the highest genetic correlation ($g_c = 0.40$) and phenotypic correlation ($p_c = 0.38$). Similarly, the hepatic and renal BAGs showed a high genetic correlation ($g_c = 0.39$) and phenotypic correlation ($p_c = 0.37$). The musculoskeletal BAG also showed significant genetic and phenotypic correlations with pulmonary ($g_c = 0.35$, $p_c = 0.19$) and renal BAGs ($g_c = 0.13$, $p_c = 0.21$). In addition, the eye BAG showed small genetic and phenotypic correlations with the brain BAG ($g_c = 0.15$, $p_c = 0.11$). The correlations between the brain and eye BAGs and other organ BAGs were relatively weaker than those observed among other organ pairs. These findings indicate the presence of shared genetic underpinnings that collectively contribute to the biological aging processes captured by these organ BAGs. Most of the genetic correlations showed consistency between women and men, albeit sex differences were evident in certain BAGs, particularly in the cardiovascular BAG results. Specifically, men exhibited dominant correlations between cardiovascular BAGs and hepatic and renal BAGs, whereas women demonstrated unique correlations with musculoskeletal and pulmonary BAGs (Extended Data Fig. 7). Sex differences in cardiovascular diseases have been explored in prior literature[20], highlighting the divergent effects of factors associated with both sex and gender on the clinical presentations and outcomes of cardiovascular disease. Detailed results are presented in Source Data File 17.

Genetic correlations between the nine BAGs and 41 clinical traits

We also estimated g_c between the nine BAGs and 41 clinical traits to examine their genetic correlations. The 41 clinical traits encompassed many common chronic diseases and conditions and their disease subtypes[4,21–24], cognition (for example, general intelligence and reaction time) and lifestyle factors (for example, computer use; Fig. 5f and Supplementary Table 4).

The brain BAG was genetically associated with several brain diseases of the CNS and their subtypes, including Alzheimer's disease (g_c = 0.37 ± 0.14) and late-life depression (g_c = 0.25 ± 0.07). Furthermore, we observed significant genetic correlations between the brain BAG and years of education (g_c = −0.14 ± 0.05) and intelligence (g_c = −0.15 ± 0.05). The cardiovascular BAG was positively correlated with stroke (g_c = 0.20 ± 0.05), a significant cardiovascular disease, and was negatively correlated with years of education (g_c = −0.17 ± 0.05). The musculoskeletal BAG was positively correlated with hyperlipidemia (g_c = 0.18 ± 0.06), rheumatoid arthritis (g_c = 0.13 ± 0.03) and Crohn's disease (g_c = 0.19 ± 0.06) and was negatively correlated with atrial fibrillation (g_c = –0.11 ± 0.04), years of education (g_c = −0.21 ± 0.04) and intelligence (g_c = −0.18 ± 0.03). The pulmonary BAG was positively associated with hyperlipidemia (g_c = 0.12 ± 0.04), stroke (g_c = 0.15 ± 0.05), liver fat (g_c = 0.12 ± 0.04) and lung carcinoma (g_c = 0.17 ± 0.05). Finally, the renal BAG was positively correlated with chronic kidney disease (g_c = 0.39 ± 0.06) and atrial fibrillation (g_c = 0.09 ± 0.03). Notably, type 2 diabetes showed abundant positive genetic correlations with multiple BAGs, including the brain, cardiovascular, metabolic, pulmonary and renal BAGs. Detailed results are presented in Source Data File 18. Furthermore, we calculated the genetic correlation between the nine BAGs and longevity[25] and household income[26]. Our findings indicated that the cardiovascular (g_c = −0.16 ± 0.09) and pulmonary BAGs (g_c = −0.12 ± 0.07) exhibited negative associations with longevity, defined as cases surviving at or beyond the age corresponding to the 90th survival percentile, and the brain BAG (g_c = −0.21 ± 0.04), musculoskeletal BAG (g_c = −0.29 ± 0.03) and pulmonary BAG (g_c = −0.16 ± 0.03) were negatively genetically correlated with household income. We used GWAS summary statistics from a prior study[27] to detect a significant genetic correlation between the immune BAG (g_c = −0.13 ± 0.03), pulmonary BAG (g_c = −0.09 ± 0.03) and telomere length (Supplementary Table 5).

These genetic correlations yield insights into potential shared mechanisms underlying the nine BAGs, their relationships with chronic diseases, particularly Alzheimer's disease and type 2 diabetes, and cognition. These compelling results prompted us to explore the potential causal effects of these traits on the nine BAGs. In the subsequent section, we unbiasedly selected 17 clinical traits encompassing chronic diseases, cognition and lifestyle factors to perform Mendelian randomization.

Hepatic and musculoskeletal BAGs are causally linked

We performed two-sample bidirectional Mendelian randomization for each pair of BAGs by excluding overlapping populations to avoid bias (Methods). We found that the hepatic and musculoskeletal BAGs showed a bidirectional causal relationship (from the hepatic BAG to the musculoskeletal BAG: $P = 9.85 \times 10^{-7}$, odds ratio (OR; 95% confidence interval (95% CI)) = 1.47 (1.26, 1.71); from the musculoskeletal BAG to the hepatic BAG: $P = 1.54 \times 10^{-8}$, OR (95% CI) = 2.78 (1.95, 3.97); Fig. 6a)). This causal relationship echoes our genetic correlation results; the musculoskeletal and hepatic BAGs showed the highest genetic correlation compared to other organ systems (Fig. 5e). Detailed results and sensitivity check results are presented in Source Data File 19 and Supplementary Figs. 15 and 16.

We performed three additional sensitivity check analyses for this bidirectional causal relationship. First, we reperformed the GWAS for the hepatic BAG and musculoskeletal BAG, incorporating weight as a covariate due to its established causal associations with several organ systems. This analysis reaffirmed this bidirectional causal relationship (Supplementary Note 6a and Supplementary Table 6a,b). Furthermore, we performed Mendelian randomization by excluding the common SNP within the *APOE* gene (rs429358) due to its pleiotropic effects. This analysis underscored the robustness of the potential causal relationship from the hepatic BAG to the musculoskeletal BAG, both with and without including this SNP as an instrumental variable (IV), as elaborated in Supplementary Note 6b and Supplementary Table 6c. Finally, the latent causal variable (LCV[28]; Methods) model confirmed a partially genetically causal effect from the hepatic BAG to the musculoskeletal BAG (genetic causality proportion = 0.75 ± 0.14, $-\log_{10}$ (*P* value) = 11.0, g_c = 0.41 ± 0.06; Supplementary Table 6d).

Multiorgan causal network with chronic diseases, body weight and sleep duration

We investigated the bidirectional causal effects between chronic diseases (for example, Alzheimer's disease) and lifestyle factors (for example, sleep duration) and the nine BAGs. We unbiasedly and systematically included 17 clinical traits guided by our genetic correlation results (Fig. 5f). The 17 clinical traits included chronic diseases linked to the brain, cardiovascular, metabolic, digestive, renal and musculoskeletal systems, cognition and lifestyle factors (Supplementary Table 7).

In the forward Mendelian randomization, we found potential causal effects of Alzheimer's disease on the brain ($P = 3.99 \times 10^{-8}$, OR (95% CI) = 1.05 (1.03, 1.06), number of SNPs = 10), hepatic ($P = 7.53 \times 10^{-7}$, OR (95% CI) = 1.03 (1.02, 1.04), number of SNPs = 10), musculoskeletal ($P = 1.73 \times 10^{-5}$, OR (95% CI) = 0.98 (0.97, 0.99), number of SNPs = 10) and renal ($P = 5.71 \times 10^{-4}$, OR (95% CI) = 0.98 (0.97, 0.99), number of SNPs = 10) BAGs. Body weight showed causal effects on multiple organ systems, including the immune ($P = 8.96 \times 10^{-5}$, OR (95% CI) = 1.08 (1.04, 1.11), number of SNPs = 160), musculoskeletal ($P = 4.32 \times 10^{-15}$, OR (95% CI) = 0.83 (0.79, 0.86), number of SNPs = 160), pulmonary ($P = 3.50 \times 10^{-7}$, OR (95% CI) = 0.84 (0.79, 0.90), number of SNPs = 160) and renal BAGs ($P = 4.53 \times 10^{-13}$, OR (95% CI) = 1.18 (1.13, 1.23), number of SNPs = 160). In addition, we also found that Crohn's disease had causal effects on the hepatic BAG ($P = 3.00 \times 10^{-3}$, OR (95% CI) = 1.02 (1.00, 1.03), number of SNPs = 77), type 2 diabetes had causal effects on the metabolic BAG ($P = 9.92 \times 10^{-12}$, OR (95% CI) = 1.16 (1.09, 1.24), number of SNPs = 8), and inflammatory bowel disease ($P = 1.42 \times 10^{-3}$, OR (95% CI) = 1.02 (1.00, 1.03), number of SNPs = 80) and primary biliary cholangitis ($P = 7.41 \times 10^{-4}$, OR (95% CI) = 1.02 (1.00, 1.03), number of SNPs = 16) had causal effects on the musculoskeletal BAG (Fig. 6b).

For the inverse Mendelian randomization, we found potential causal effects of the metabolic ($P = 6.85 \times 10^{-4}$, OR (95% CI) = 0.94 (0.91, 0.97), number of SNPs = 71) and pulmonary ($P = 3.79 \times 10^{-5}$, OR (95% CI) = 0.84 (0.79, 0.91), number of SNPs = 62) BAGs on body weight, the cardiovascular BAG on triglycerides versus lipid ratio in very large very-low-density lipoprotein ($P = 2.14 \times 10^{-4}$, OR (95% CI) = 1.09 (1.04, 1.14), number of SNPs

= 39) and the brain BAG on sleep duration ($P = 2.61 \times 10^{-3}$, OR (95% CI) = 1.09 (1.04, 1.14), number of SNPs = 10; Fig. 6c). Detailed results are presented in Source Data File 20.

We performed several sensitivity analyses (Methods) to test the robustness of our findings. Based on these sensitivity checks, we identified potential outlier IVs (that is, SNPs) for four Mendelian randomization tests (Alzheimer's disease and body weight on the musculoskeletal BAG, Crohn's disease on the hepatic BAG and type 2 diabetes on the metabolic BAG) in the forward Mendelian randomization and one Mendelian randomization test (metabolic BAG on body weight) in the inverse Mendelian randomization. Detailed results of the sensitivity check are presented in Supplementary Figs. 17–32 for all significant results. We showcased a detailed analysis of the sensitivity results for the metabolic BAG on body weight in Supplementary Note 6c. In summary, the potential causal link from the metabolic BAG to body weight remained robust across several sensitivity checks despite the identification of two potential outlier IVs, namely, rs117233107 and rs33959228.

In addition, we used the LCV method and found a partial genetically causal effect from longevity (99th survival percentile) to the brain BAG (genetic causality proportion = 0.45 ± 0.20, $P = 0.04$). Importantly, we selected the LCV method over Mendelian randomization because of the partial population overlap between the longevity GWAS summary statistics and our BAG GWAS summary statistics. The LCV analysis also detected a partially genetically causal effect from telomere length to the immune BAG (genetic causality proportion = 0.33 ± 0.12, $P = 0.0002$) and the pulmonary BAG (genetic causality proportion = 0.67 ± 0.20, $P = 3.57 \times 10^{-16}$; Supplementary Table 6d).

Discussion

The current study depicts the genetic architecture of common genetic variants on biological aging of nine human organ systems using multimodal data from 377,028 participants of European ancestry. We identified many genomic loci for the BAGs of nine human organ systems. These associations were observed within a phenotypic landscape characterized by BAG–organ specificity and interorgan connections. GSEA, tissue-specific gene expression patterns and heritability enrichment results provided additional evidence supporting biological validation for BAG–organ specificity and interorgan connections. The phenotypic correlation between BAGs was a proxy for their genetic correlation, thereby supporting the long-standing Cheverud's Conjecture. Mendelian randomization demonstrated potential causal relationships between chronic diseases, particularly Alzheimer's disease and type 2 diabetes, body weight, sleep duration and the nine BAGs.

Our large-scale multiorgan GWAS significantly expands the current catalog of genetic variants associated with health-related traits. The discovery of these identified genomic loci has significant clinical implications. These findings provide a foundation to validate genes or regulatory elements, molecular pathways and biological processes related to the clinical traits and diseases of interest in the current study and future GWASs. Previous GWASs mainly focused on the BAG in one organ system, such as the brain BAG[9,29] from imaging-derived phenotypes. These investigations have largely overlooked the inherent interconnectedness of human organ systems, which are intricately intertwined with distinct

axes. Recent studies have identified notable axes, such as the heart–brain–liver[6], brain–eye[30] and brain–heart[31] axes, highlighting the importance of comprehending these intricate relationships to understand human physiology and health.

Our phenome-wide associations validate the pleiotropic effects of the identified genomic loci, influencing various health-related clinical traits in the GWAS Catalog and GWAS Atlas. Our findings also highlight BAG–organ specificity and interorgan connections, further supporting that biological aging is a complex, multifaceted phenomenon. The human brain regulates various physiological processes and maintains homeostasis throughout the body. Consequently, it is unsurprising that the brain exhibits interconnectedness with clinical traits associated with multiple organ systems. The remarkable enrichment of metabolic traits across various organ systems is unsurprising. As a vital metabolic organ, the liver substantially overlaps genetic variants and loci with both the hepatic and metabolic BAGs. Biologically, the liver's metabolic functions are intricately regulated by hormones like insulin and other metabolic regulators[32]. Similarly, the interplay between immune and metabolic processes is essential for maintaining overall health and is crucial for the body's ability to respond to pathogens and regulate metabolic homeostasis[33].

Our gene-level and partitioned heritability analyses further validate our GWAS findings, supporting BAG–organ specificity and interorgan connections. In the GSEA, the genes associated with the cardiovascular BAG were implicated in the IGF-II pathway. IGF-II activates two receptors (IGF-1R and IR-A) to promote cell growth and survival. The IGF signaling pathway is essential for cardiac development in the human heart, the first functional organ to develop[34]. In particular, IGF-II promotes fetal cardiomyocyte proliferation through the tyrosine kinase receptors IGF-1R and INSR. Previous research provided appealing evidence on IGF signaling in cardiac regeneration in animal models and induced pluripotent stem cells[35]. The flavonoid glucuronidation pathway was the most significant enrichment result shown in the hepatic BAG. A previous study demonstrated that procyanidin C1, a flavonoid in grape seed extract, extended the lifespan of mice[36]. Furthermore, ample evidence indicated that natural flavonoids could be potential therapeutic approaches for nonalcoholic fatty liver disease[37]. The metabolites formed through this pathway can also exert effects beyond the liver and impact other organ systems. Our tissue-specific gene expression analyses provided additional support for the biological relevance of our GWAS findings, as the identified genes exhibited specific expression patterns within tissues from the corresponding organ systems.

The heritability enrichment analysis further validates the BAG–organ specificity and interorgan connections by highlighting the disproportional heritability enrichment of genetic variants in different functional categories, cell types, tissues and chromatin states. The cell-type-specific enrichment results in the brain (that is, oligodendrocytes) and cardiovascular (that is, neurons) BAGs align with previous research. Specifically, Zhao et al. conducted a large-scale GWAS on brain white matter microstructure and found significant heritability enrichment in glial cells, particularly oligodendrocytes[21,38], which aligns with our current findings. Our previous multimodal brain BAG GWAS[9] also confirmed this enrichment in the brain BAG derived from the white matter microstructural features. Similarly, research has revealed the presence of an 'intrinsic cardiac nervous system' within the heart, often called

the ‘heart brain’. This system consists of around 40,000 neurons similar to those found in the brain, indicating that the heart possesses a distinct nervous system[39].

Our genetic correlation results confirmed that the genetic correlation generally mirrors phenotypic correlations in multiorgan biological age. This suggests that environmental factors likely affect the aging of multiple organ systems in the same direction. Providing evidence for Cheverud’s Conjecture can have clinical implications by providing insights into the genetic basis of complex age-related diseases. For instance, by identifying the shared genetic factors underlying multiple age-related diseases, we can target these common pathways to develop potential treatments or repurpose existing drugs[40] that have proven efficacy in treating one disease or condition for others. Moreover, the validation of Cheverud’s Conjecture emphasizes the importance of considering the genetic covariance of age-related diseases in clinical practice. It underscores the need for comprehensive genetic assessments and genomic analyses to understand disease risk and progression[41].

We found a bidirectional causal relationship between the hepatic and musculoskeletal BAGs. Abundant research has suggested that liver function and metabolic health, particularly related to glucose and lipid metabolism, can significantly impact musculoskeletal health[42]. This interorgan connection can cause dysregulation of liver metabolism (for example, nonalcoholic fatty liver disease) linked to musculoskeletal disorders, including osteoporosis, sarcopenia and muscle wasting. The musculoskeletal system can also exert an inverse influence on liver function. Regular physical activity and muscle strength have been linked to enhanced liver health and decreased susceptibility to liver diseases. To further support this, causal effects of primary biliary cholangitis, a chronic liver disease, on elevated musculoskeletal BAG were confirmed in our Mendelian randomization results. The absence of direct causal relationships between the remaining BAGs can be attributed to various factors with potential explanations and implications. One possible explanation is that the brain BAG, having the smallest sample size in our GWAS (after removing overlapping participants), may be limited in statistical power. In addition, this may suggest that various factors, including chronic diseases, environmental exposures and lifestyle choices, influence biological aging in alternative pathways or mediate such changes. Thus, understanding the collective contribution of chronic diseases, environmental factors and lifestyle choices is crucial for comprehending the overall aging process and its impact on organ health.

We found that several clinical traits collectively cause organ systems to appear older or younger than their chronological age. For instance, body weight was causally associated with the immune, musculoskeletal, metabolic and pulmonary BAGs. For several reasons, body weight can causally influence multiple organ systems. Excessive body weight (for example, obesity) has metabolic consequences, including increased inflammation, insulin resistance and dysregulation of metabolic pathways in adipose tissue[43]. It also leads to mechanical stress on the body, contributing to musculoskeletal strain[44] and cardiovascular workload[45]. Hormonal imbalances[46] and lifestyle factors linked to body weight also influence multiorgan function and the development of chronic diseases. Being overweight is also a risk factor for type 2 diabetes, which was positively causally associated with the metabolic BAG. Alzheimer’s disease was causally linked to the brain, hepatic, musculoskeletal and renal BAGs. Alzheimer’s disease, a neurodegenerative disorder

primarily affecting the brain, can have causal influences on multiple organ systems. For example, it has broader systemic involvement beyond the brain, mediated by mechanisms including protein aggregation (for example, amyloid-β and tau[47]), inflammation[48] and other secondary factors. Protein aggregates can spread to other organs, vascular abnormalities can impact blood flow, inflammation can affect distant organ systems, and secondary factors, such as medication use and lifestyle changes, also contribute.

Our sex-stratified GWAS findings (Supplementary Figs. 1–9) reveal significant sex biases across several organs' biological age, particularly evident in the immune BAG. Sex bias is a common phenomenon in chronic diseases, exemplified by conditions like autism spectrum disorder exhibiting a male-dominant pattern and Alzheimer's disease displaying a female-vulnerable pattern. Various factors contribute to such sex differences. Recent studies[49] propose that autistic women may demonstrate better sociocognitive abilities than autistic men, potentially leading to underdiagnosis among women (that is, population selection bias). However, it remains uncertain whether these differences also arise from distinct neurophysiological functioning and/or genetic variants in autistic individuals of different sexes. Another study provides genetic evidence of sex differences in autism spectrum disorder, supporting the female-protective effect using a mutation burden analysis[50]. Similarly, sex differences are observed in Alzheimer's disease[51]. Individuals diagnosed with Alzheimer's disease exhibit distinct cognitive and psychiatric symptoms based on sex, with women experiencing more rapid cognitive decline following a diagnosis of mild cognitive impairment or Alzheimer's disease. These examples illustrate the sex differences in multiple contexts that may contribute to the sex bias observed in our sex-stratified GWAS. Our main GWAS carefully included several sex-related covariates; we also provided post-GWAS results for each specific sex, such as genetic correlations using sex-specific GWAS data (Extended Data Fig. 7). In particular, for the immune BAG, a recent study[52] using single-cell genomics identified sex-biased immune gene expression and immune cell populations, such as group 2 innate lymphoid cells, modulated by androgens. Follow-up studies specifically investigating sex differences at the molecular and cellular levels in the current study will further elucidate the underlying etiology affecting multiorgan health.

This study has several limitations. First, the generalizability of genetic findings from populations of European and non-European ancestry is limited by the current sample sizes despite a high correlation in the derived β values. Future studies can extend their scope to encompass a more diverse array of underrepresented ethnicities, a wider range of disease cohorts and individuals of varying ages throughout their entire lifespan. Second, it is essential to approach the causality results cautiously, considering the assumptions underlying Mendelian randomization. In future studies, more advanced multiresponse Mendelian randomization methods[53] should be used. Third, despite our efforts of quality check analyses to scrutinize our primary GWAS, it is essential to acknowledge that potential ascertainment bias[54], and confounding related to demographic and socioeconomic factors could potentially introduce cryptic population stratification, which may not be entirely resolved in the current study. Additionally, sex differences were apparent in several of the nine BAGs, indicating the need for future research to explore this aspect more thoroughly. Finally, the large number of genomic loci identified in our GWAS may have connections to BAGs due to various factors, such as biological processes, potential confounding due to

demographics or specific study design and phenotyping aspects. It is important to note that the effects at these loci might not be inherently biological but could be influenced by other unmeasured confounding factors.

In conclusion, our study provides compelling genetic evidence affirming the notion that 'no organ system is an island'. It underscores the collective influence of different chronic diseases on multiorgan systems and the interconnectedness among these human organ systems. These findings highlight the importance of understanding the underlying causes of chronic diseases within the multiorgan network. By shedding light on its comprehensive genetic architecture, our study paves the way for future research to unravel complex disease mechanisms and develop holistic approaches to ameliorate overall organ health.

Methods

Study populations

UKBB is a population-based study of approximately 500,000 people recruited between 2006 and 2010 from the United Kingdom. The UKBB study has ethical approval, and the ethics committee is detailed at https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/governance/ethics-advisory-committee.

The current study (application numbers 35148 and 60698) analyzed multimodal data, including imaging-derived phenotypes and physical and physiological measures in nine human organ systems, from 154,774 UKBB participants. In our previous study, we constructed BAGs for eight organ systems using machine learning, including magnetic resonance imaging data for the brain BAG from 30,108 participants (European ancestry), pulse rate and blood pressure data for the cardiovascular BAG, liver-related blood biomarkers for the hepatic BAG, C-reactive protein and blood hematology variables for the immune BAG, blood biomarkers for the metabolic BAG, physical measurements and vitamin D for the musculoskeletal BAG, lung functioning measurements for the pulmonary BAG and glomerular filtration and electrolyte regulation biomarkers for the renal BAG from 111,543 participants. Furthermore, the current study also used 60 OCT-derived measures from 36,004 participants to derive the BAG of the ninth organ system, the eye BAG. The inclusion criteria for the features used to predict the eight BAGs, the machine learning methods and the cross-validation procedures are detailed in our previous study[3]. We initially used the 88 OCT-derived measures (category ID 10079) for the additional eye BAG in 67,549 participants. Of these measures, 28 were excluded due to a high missing rate (>20% of participants). Additionally, 4,172 participants were excluded due to missing data, and 1,798 participants identified as outliers (outside mean ± 6 s.d.) for the 60 remaining measures were discarded. This resulted in 41,966 participants (36,004 participants of European ancestry). The included 2,444 features to derive the BAG of the nine organ systems are presented in Source Data File 21.

In addition, we also performed a GWAS for seven variables from 222,254 UKBB participants by excluding the 154,774 participants from the BAG populations to avoid bias due to overlapping samples. These variables included six lifestyle factors and one cognitive variable: N = 219,661 (European ancestry) for coffee intake (field ID 1498), N = 221,393 for

fresh fruit intake (field ID 1309), N = 221,739 for tea intake (field ID 1488), N = 220,765 for sleep duration (field ID 1160), N = 209,012 for time spent outdoors in the summer (field ID 1050), N = 221,337 for body weight (field ID 21002) and N = 220,624 for reaction time (field ID 20023). In total, we analyzed data from 377,028 individuals of European ancestry in the current study.

Support vector machines to predict BAG

Our prior study[3] used support vector machines to predict the chronological age of healthy individuals, which was defined as no self-reported and healthcare-documented lifetime chronic medical conditions, based on phenotypes from the nine organ systems. Support vector machine regression was preferred over linear regression for its enhanced robustness to outliers and overfitting. We performed a 20-fold cross-validation procedure and developed predictive models for each organ system. Detailed information regarding the machine learning models and model performance compared to the previous literature are present in Supplementary Note 7.

Genetic analyses

We used the imputed genotype data for all genetic analyses, and our quality check pipeline resulted in 487,409 participants and 6,477,810 SNPs. After merging with the population for each BAG, we included 30,108–111,543 participants of European ancestry for the nine BAGs (Fig. 1). To avoid bias due to overlapping populations[55], we also used the rest of the UKBB participants of European ancestry (nonoverlapping) to derive the GWAS summary statistics for several lifestyle factors (Methods). We summarize the genetic quality control pipeline. First, we excluded related individuals (up to second degree) from the complete UKBB sample using the KING software for family relationship inference[56]; fastGWA skips this step because the method tackles the population stratification problem without removing related individuals. We then removed duplicated variants from all 22 autosomal chromosomes. Individuals whose genetically identified sex did not match their self-acknowledged sex were removed. Other excluding criteria were (1) individuals with more than 3% of missing genotypes, (2) variants with a minor allele frequency of less than 1% (dosage mode[57]), (3) variants with larger than 3% missing genotyping rate and (4) variants that failed the Hardy–Weinberg test at 1×10^{-10}. To adjust for population stratification[58], we derived the first 40 genetic principle components using FlashPCA software[59]. Details of the genetic quality check protocol are described elsewhere[9,22,60].

GWAS.

For GWAS, we ran a linear regression using PLINK[61] for each BAG, controlling for confounders of age, dataset status (training/validation/test or independent test dataset), age-squared, sex, age × sex interaction, age-squared × sex interaction and the first 40 genetic principal components; additional covariates for total intracranial volume and the brain position in the scanner were included for the brain BAG GWAS. We adopted the genome-wide P value threshold of 5×10^{-8} and annotated independent genetic signals considering linkage disequilibrium (see below).

To check the robustness of our GWAS results, we performed several sensitivity check analyses, including a (1) sex-stratified GWAS for men and women, (2) split-sample GWAS by randomly dividing the entire population into two splits (sex and age matched), (3) non-European ancestries GWAS and (4) fastGWA[62] for linear mixed-effect GWAS, hypothesizing that the main GWASs with European ancestry did not show substantial genomic inflation due to cryptic population stratification. In all our sensitivity check analyses, we considered linkage disequilibrium. We only evaluated the independent significant SNPs of the two sets of β coefficients between splits, sexes, ancestry groups and GWAS methods. The definition of the independent significant SNPs used the same parameters as in FUMA[63] (Supplementary Note 1). We used the raw genotype data and the PLINK clump command (250 kb) and defined a set of SNPs in linkage disequilibrium with the independent significant SNPs, analogous to the candidate SNPs in FUMA.

SNP-based heritability.

We estimated the SNP-based heritability (h^2) using GCTA[64] with the same covariates in GWAS. We reported results from two experiments for each BAG using (1) the full sample sizes and (2) randomly downsampled sample sizes to that (N = 30,108) of the brain BAG with comparable distributions regarding sex and age; the sample size of brain BAGs was smaller than the other BAGs.

Annotation of genomic loci.

Annotation of genomic loci and mapped genes was performed using FUMA[63]. For the annotation of genomic loci, FUMA first defined lead SNPs (correlation $r^2 \leq 0.1$, distance < 250 kb) and assigned them to a genomic locus (nonoverlapping); the lead SNP with the lowest P value (that is, the top lead SNP) was used to represent the genomic locus in Fig. 1. For gene mappings, three different strategies were considered. First, positional mapping assigns the SNP to its physically nearby genes (a 10-kb window by default). Second, eQTL mapping annotates SNPs to genes based on eQTL associations using the GTEx v8 data. Finally, chromatin interaction mapping annotates SNPs to genes when there is a significant chromatin interaction between the disease-associated regions and nearby or distant genes[63]. The definitions of top lead SNP, lead SNP, independent significant SNP and candidate SNP can be found in Supplementary Note 1.

For the top lead SNP of each identified genomic locus, we showcased whether it was previously associated with any clinical traits considering linkage disequilibrium (5,000 kb around the top lead SNP) in the EMBL-EBI GWAS Catalog[65] platform (https://www.ebi.ac.uk/gwas/home). For instance, we aimed to query the locus with the top lead SNP (rs60569686) associated with the renal BAG. First, we looked up the chromosomal position (that is, chromosome 13) and found that the location is chr13:49170160 (GRCh38). We then searched the GWAS Catalog for a 5,000-kb region around this top lead SNP: ‘chr13:49167660–49172660’ (https://www.ebi.ac.uk/gwas/regions/chr13:49167660-49172660; query date 12 October 2023). In this region, we discovered no prior associations. It is important to note that this search is not comprehensive, as new GWASs continually emerge on various open platforms, such as IEU OpenGWAS[66] (https://gwas.mrcieu.ac.uk/) and GWAS Atlas[13] (https://atlas.ctglab.nl/PheWAS).

Phenome-wide association look-up queries.

We first queried the significant independent SNPs within each locus in the EMBL-EBI GWAS Catalog (query date 24 April 2023, via FUMA version v1.5.4) to determine their previously identified associations with any other traits ($P < 1 \times 10^{-5}$ by default in the EMBL-EBI GWAS Catalog). For visualization purposes, we further mapped the associated traits into organ-specific groups and other chronic disease traits and cognition. We performed the following procedure to fully consider linkage disequilibrium and remove redundant associations among the independent significant SNPs. If the top lead SNP showed any clinical associations, this would present the current locus; if not, we queried the independent significant SNPs (in high correlation with the top lead SNP), starting with the most significant SNPs, until we identified established associations. In this way, only one genetic variant within each genomic locus was considered. We also conducted a complementary phenome-wide association query on the GWAS Atlas platform. We applied the same *P* value threshold search criteria as those used in the EMBL-EBI GWAS Catalog. The same procedure, considering linkage disequilibrium and redundant associations, was applied.

GSEA.

We first performed gene-level association analysis using MAGMA[67]. First, gene annotation was performed to map the SNPs (reference variant location from Phase 3 of 1,000 Genomes for European ancestry) to genes according to their physical positions. Of note, other advanced annotation methods exist that integrate functional insights, such as brain chromatin interaction[68] and cell-type-specific gene expression[69]. We then performed gene-level associations based on the SNP GWAS summary statistics to obtain gene-level *P* values between the nine BAGs and the curated protein-coding genes containing valid SNPs. We performed GSEA using the gene-level association *P* values. Gene sets were obtained from the Molecular Signatures Database (MSigDB, v7.5.1)[70], including 6,366 curated and 10,402 ontology gene sets. All other parameters were set by default for MAGMA. The Bonferroni method was used to correct multiple comparisons for all tested gene sets.

Tissue-specific gene expression analysis.

MAGMA performed gene property analyses to identify tissue-specific gene expression of the nine BAGs. The gene property analysis converts the gene-level association *P* values (above) to *z* scores and tests a specific tissue's gene expression value versus the average expression value across all tissues in a regression model. Bonferroni correction was performed for all tested gene sets. We reported the results from the 54 tissue types using the GTEx V8 data.

Gene–drug–disease network.

We tested the enrichment of the nine BAG-linked genes in the targeted gene sets for different drug categories from the DrugBank database[71]. The gene–drug–disease network was constructed to prioritize potentially repositionable drugs. GREP software[72] performs Fisher's exact tests to examine whether the prioritized genes are enriched in gene sets

targeted by drugs in a clinical indication category for a certain disease or condition. Bonferroni correction was performed for all tested drugs.

Partitioned heritability estimate.

Our objective is to comprehend how distinct functional genome categories play different roles in contributing to the heritability of the nine BAGs. Therefore, the partitioned heritability analysis via stratified linkage disequilibrium score regression calculates the extent to which heritability enrichment can be attributed to predefined and annotated genome regions and categories[73]. Three sets of functional categories and cell- and tissue-specific types were considered. First, the partitioned heritability was calculated for 53 general functional categories (1 including the entire set of SNPs). The 53 functional categories are not specific to any cell type and include coding, untranslated region, promoter and intronic regions and so on. The details of the 53 categories are described elsewhere[73]. Subsequently, cell- and tissue-type-specific partitioned heritability was estimated using gene sets from Cahoy et al.[74] for three main cell types (that is, astrocyte, neuron, and oligodendrocyte), multitissue chromatin state-specific data (ROADMAP[75] and ENTEx[76]) and multitissue gene expression data (GTEx V8)[17]. Bonferroni correction was performed for all tested annotations and categories. The detailed methodologies for the stratified linkage disequilibrium score regression are presented in the original work[73]. The linkage disequilibrium scores and allele frequencies for the European ancestry were obtained from a predefined version based on data from the 1000 Genomes Project.

Genetic correlation.

We used LDSC[11] software to estimate the pairwise genetic correlation (g_c) between each pair of BAGs and between the nine BAGs and 41 other clinical traits, including chronic diseases involving multiple organ systems, such as Alzheimer's disease for brain and chronic kidney disease for kidney, cognition and lifestyle factors. We used the precomputed linkage disequilibrium scores from the 1000 Genomes of European ancestry. To ensure the suitability of the GWAS summary statistics, we first checked that the selected study's population was of European ancestry; we then guaranteed a moderate SNP-based heritability h^2 estimate. Notably, LDSC corrects for sample overlap and provides an unbiased estimate of genetic correlation[41]. The inclusion criteria and ultimately included traits are detailed in Supplementary Table 4. Bonferroni correction was performed for the 41 clinical traits.

Two-sample bidirectional Mendelian randomization.

We investigated whether one BAG was causally associated with another BAG and whether the 41 clinical traits were causally associated with the nine BAGs. To this end, we used a bidirectional, two-sample Mendelian randomization using the TwoSampleMR package[77]. Both forward and inverse Mendelian randomization were performed between each pair of traits by switching the exposure and outcome variables. We applied five different Mendelian randomization methods and reported the results of inverse variance weighted in the main text and the four others (that is, Egger, weighted median, simple mode and weighted mode estimators) in the Supplementary Information.

Mendelian randomization needs to fulfill several IV assumptions[78]. We followed a systematic procedure guided by the STROBE-MR Statement[79] in all steps of our causality analyses, including selecting exposure and outcome variables, reporting comprehensive statistics, performing sensitivity checks for potential violations of underlying assumptions and performing the analyses using alternative methods and software[28,53]. For the causal inference of each pair of BAGs, all GWAS summary statistics were derived from our analyses by excluding overlapping populations of the two BAGs. For example, to test the causal relationship between the brain BAG and cardiovascular BAG, we reran the GWAS for the cardiovascular BAG by excluding the partially overlapping population from the brain BAG. For all seven body organ systems that had entirely overlapping populations, we used the GWAS data from the split-sample analyses (Methods). For instance, the GWAS for the cardiovascular BAG was from the first-split data, and the pulmonary BAG was from the second-split data. Bonferroni correction was performed for the tested BAGs.

One key challenge in our hypothesis-driven Mendelian randomization is to select these exposure variables unbiasedly. Clinical traits sharing common genetic covariance with the nine BAGs are more likely to be causally associated with them. We performed a systematic inclusion procedure using the following criteria to overcome this. We manually queried the 41 clinical traits (used in our genetic correlation analyses) in the IEU GWAS database, specifically curated for Mendelian randomization analyses. We ranked all available studies for a certain trait (for example, Alzheimer's disease) based on the sample sizes. We then chose the studies whose populations were of European ancestry and did not include UKBB participants to avoid bias due to overlapping populations[55]. For the traits whose GWAS data were available in the IEU GWAS database, we used the TwoSampleMR package to perform the Mendelian randomization analysis. For the traits whose data were not appropriate in the IEU GWAS database, we then performed another manual query in the EMBL-EBI GWAS Catalog database to download the available GWAS summary statistics with the same filter criteria. For the traits whose GWAS data were dominated by studies using UKBB participants in both databases, we ran a GWAS using our own UKBB data by excluding overlapping populations. Finally, after harmonizing their GWAS summary statistics (using the function harmonise_data from 2SampleMR), this resulted in 17 clinical traits with at least eight valid IVs (that is, SNPs). The 17 clinical traits included chronic diseases affecting multiple organ systems, cognition and lifestyle factors (Supplementary Table 7). Bonferroni correction was performed for all tested clinical traits.

We performed several sensitivity analyses. First, a heterogeneity test was performed to check for violating the IV assumptions. Horizontal pleiotropy was estimated to navigate the violation of the IV's exclusivity assumption[80] using a funnel plot, single-SNP Mendelian randomization approaches and Mendelian randomization Egger estimator[81]. Moreover, the leave-one-out analysis excluded one instrument (SNP) at a time and assessed the sensitivity of the results to individual SNPs.

Following these analyses, we performed three supplementary sensitivity checks for some specific significant causal signals: (1) the exclusion of two common SNPs/IVs (rs429358 and rs7412) in the *APOE* gene, considering their potential pleiotropic effects for the hepatic BAG on the musculoskeletal BAG; (2) incorporating body weight as a covariate in the

GWAS for the bidirectional causality between the hepatic BAG and musculoskeletal BAG, as body weight displays causal associations with BAGs in multiple organ systems; and (3) re-executing the Mendelian randomization analysis using alternative software. Specifically, we scrutinized the causal relationship between the hepatic BAG and the musculoskeletal BAG using the LCV model[28]. Using different modeling assumptions and IVs in contrast to Mendelian randomization, it examined the causal relationship between two interchangeable traits without distinction between the direct and inverse directions. An LCV (L) acted as a mediator for the genetic correlation between the two traits, allowing us to quantify the genetic causality proportion. A positive genetic causality proportion value between 0 and 1 indicates that trait 1 is partially genetically causal, and a negative genetic causality proportion value means that trait 2 is partially genetically causal.

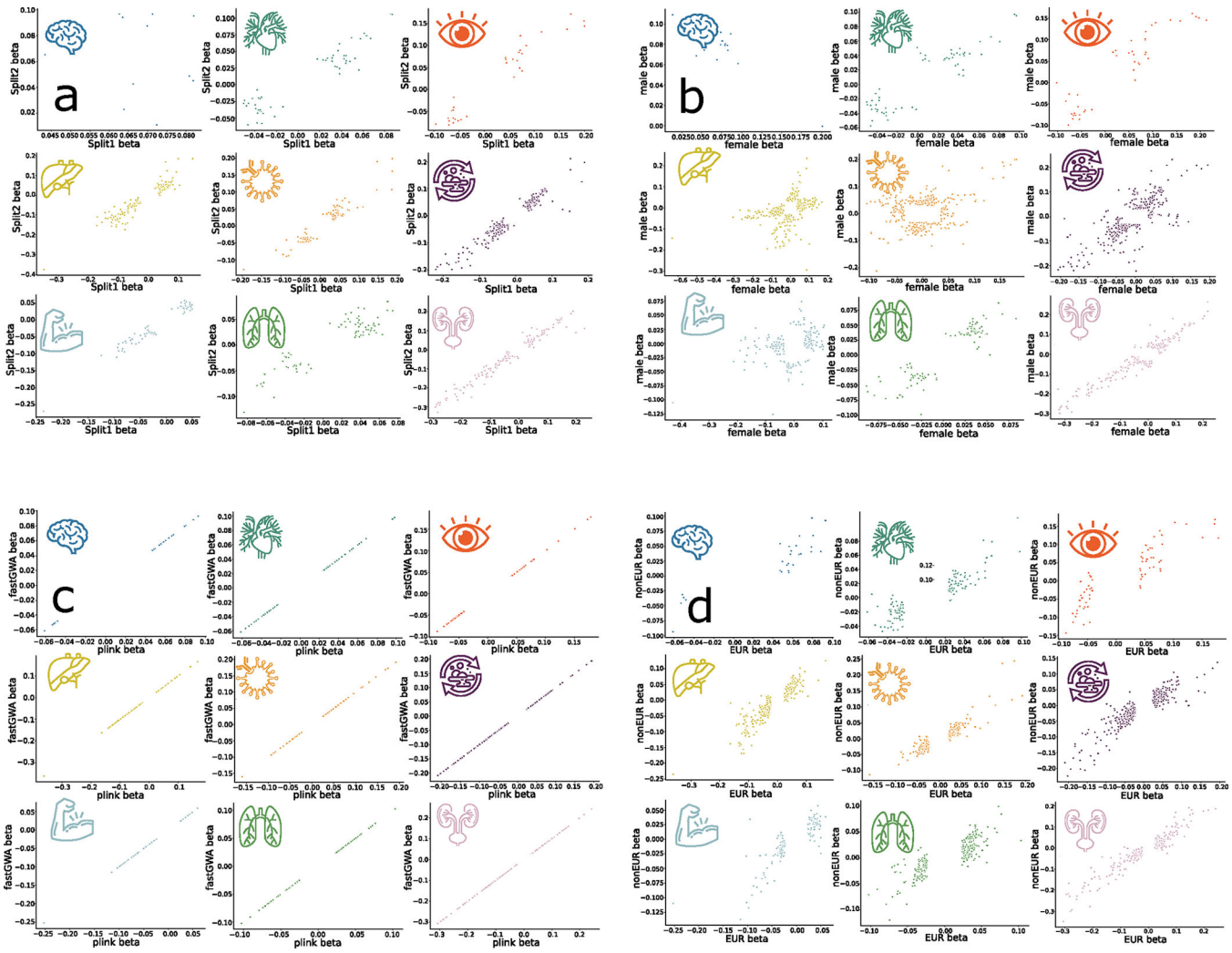
Bayesian colocalization.

The R package coloc was used to investigate the genetic colocalization signals between two traits at each genomic locus defined by the pulmonary BAG GWAS. We used the Fully Bayesian colocalization analysis using Bayes Factors (coloc.abf). The method tests the following five hypotheses, denoted by their posterior probabilities: H0 (no association with either trait), H1 (association with trait 1 but not trait 2), H2 (association with trait 2 but not trait 1), H3 (association with both traits but with separate causal variants) and H4 (association with both traits with a shared causal variant). It examines the posterior probability (PP.H4.ABF: Approximate Bayes Factor) to evaluate hypothesis H4, which suggests that the presence of a single shared causal variant is associated with both traits within a specific genomic locus. To determine the significance of the H4 hypothesis, we set a threshold of PP.H4.ABF > 0.8 (ref. 82), and at least 100 SNPs were included within the genomic locus. All other parameters (for example, the prior probability of p_{12}) were set as default.

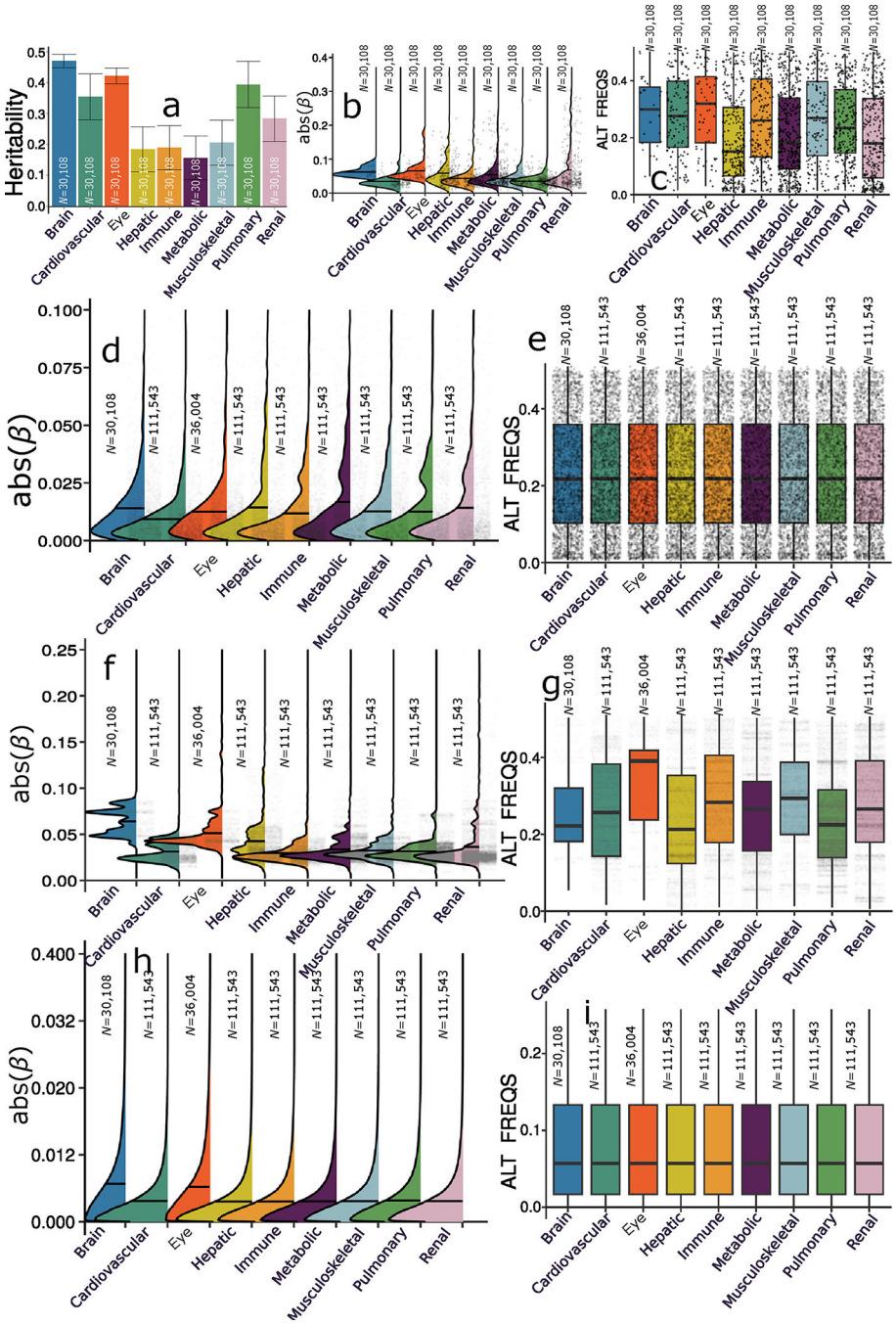
Statistics and reproducibility

The sample size was not predetermined using statistical methods. Experiments were not randomized, and investigators were not blinded to allocation. The sample sizes for deriving the original BAG measures were detailed in our previous work[3]. The current study then performed a rigor quality check on the genetic data and several sensitivity checks for the primary GWAS results. All GWAS summary statistics and methodology details are publicly shared to ensure the reproducibility of our analyses.

Extended Data



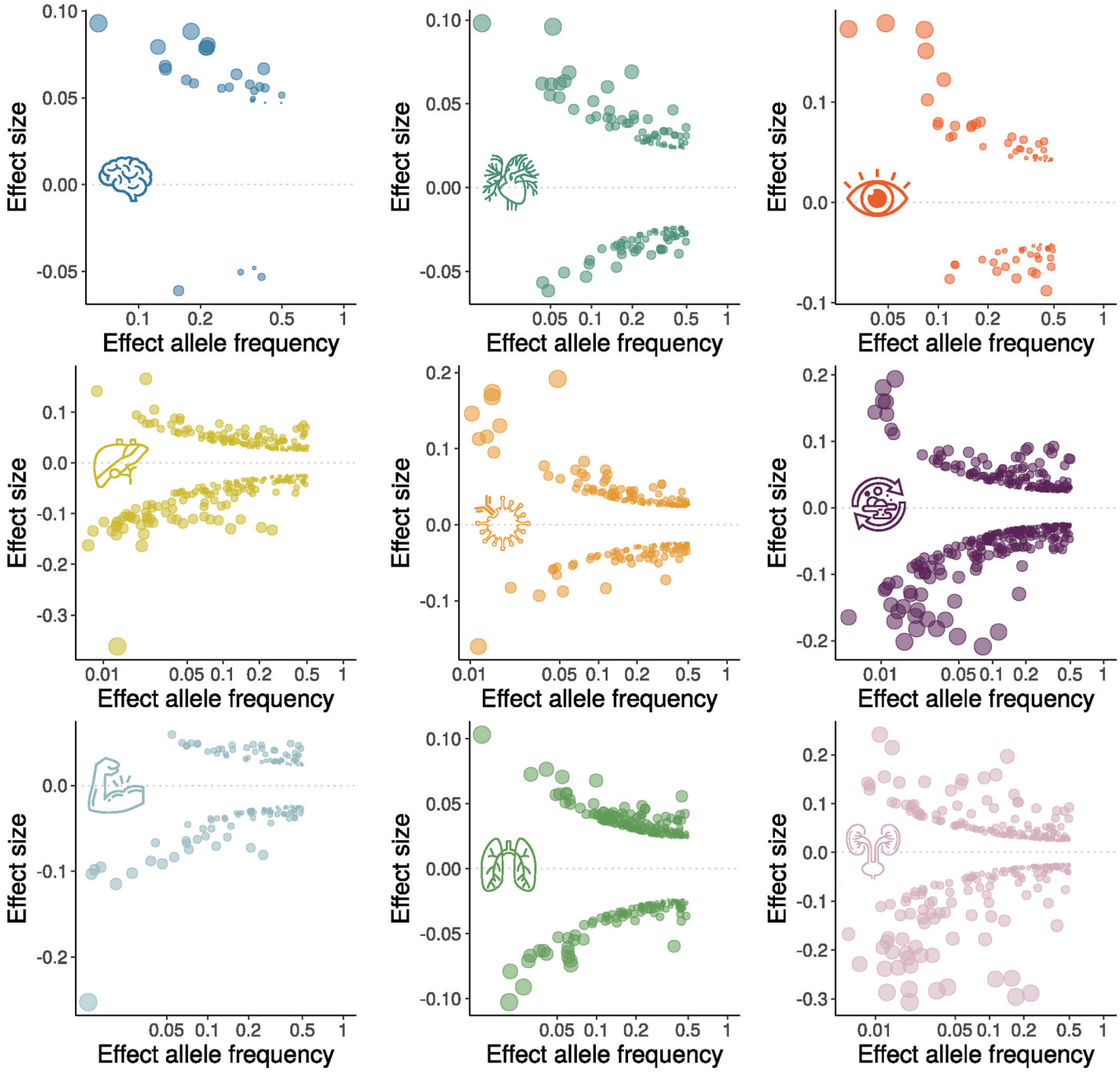
Extended Data Fig. 1 |. Scatterplots of the main GWAS sensitivity analysis for nine BAGs. We scrutinized the robustness of the nine primary GWAS using the European ancestry populations by fully considering linkage disequilibrium. We only included the independent significant SNPs in four different sensitivity check analyses. We reported three statistics *i*) *r-β*: Pearson's *r* between the two sets of *β* coefficients from the two splits; *ii*) *C-β*: concordance rate of the sign of the *β* coefficients from the two splits – if the same SNP exerts the same protective/risk effect between the two splits; *iii*) *P-β*: the difference between the two sets of *β* coefficients from the two splits – if the two sets of *β* coefficients (mean) statistically differ (two-sample t-test). Detailed statistics are presented in Supplementary Note 1 for **a**) split-sample, **b**) sex-stratified, **c**) fastGWA vs. PLINK, and **d**) European vs. Non-European GWAS analyses.



Extended Data Fig. 2 |. SNP-based heritability, beta coefficients, and alternative allele frequency using the brain-BAG comparable populations and different inclusion criteria for the SNPs.

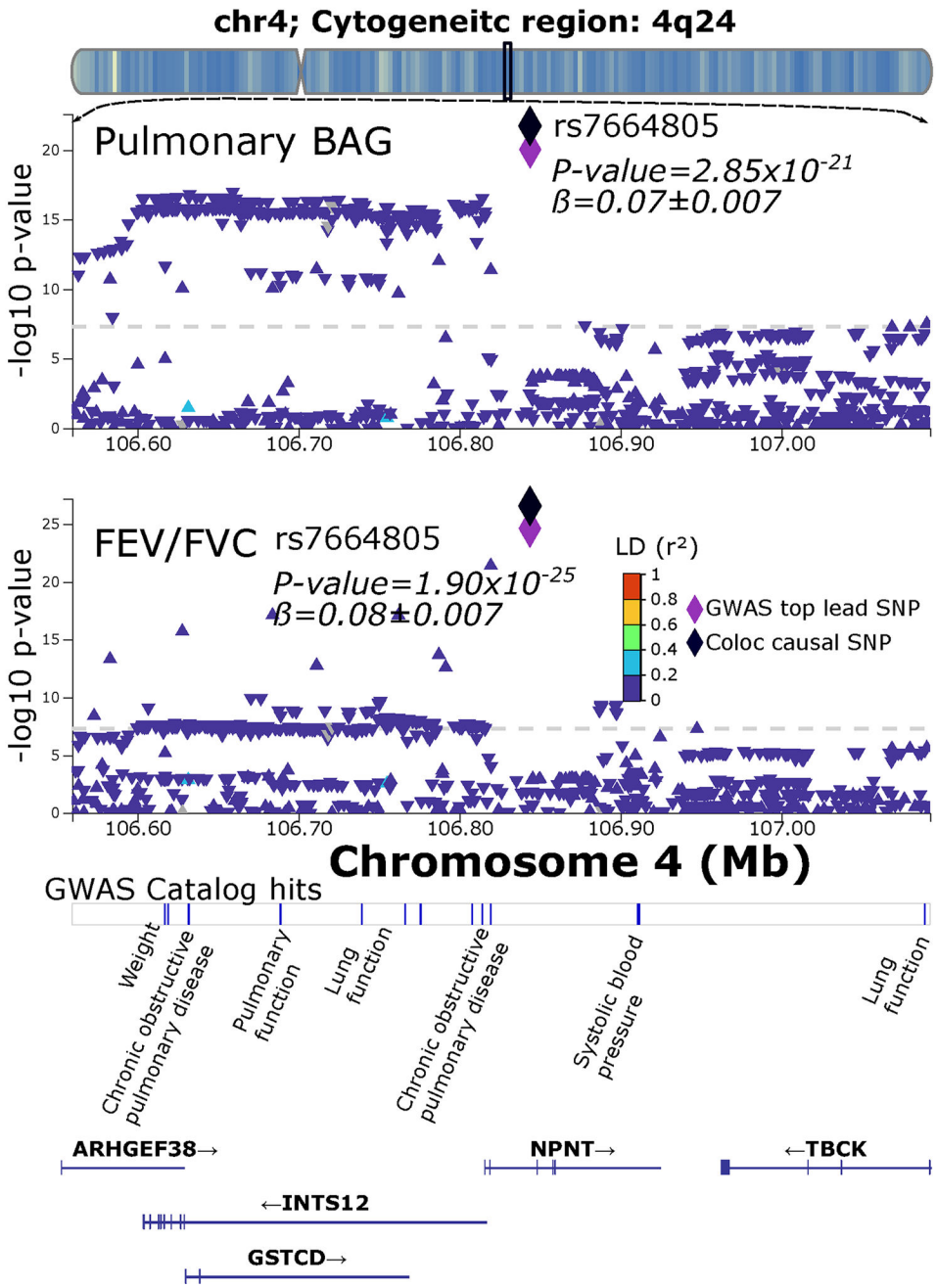
a) The SNP-based heritability of the nine BAGs using populations from downsampling to the brain BAG population. Error bars represent the standard error of the estimated parameters, and the measure of the center for the error bars represents the inferred statistics (that is, SNP-based heritability). **b**) The absolute value of the beta coefficients of the independent significant SNPs of the nine BAG GWASs using populations from downsampling to the brain BAG population (N = 30,108); the independent significant

SNPs are shown separately for each BAG. **c**) The alternative (effective) allele frequency of the independent significant SNPs from the nine BAG GWASs using populations from downsampling to the brain BAG population (N = 30,108). **d**) The beta coefficients of the independent significant SNPs using the original full samples but with all identified independent significant SNPs across the nine BAG GWASs (with the same number of SNPs tested), where we see no difference regarding allele frequency in Figure **e**). **f**) The absolute value of the beta coefficients of the independent significant SNPs plus the candidate SNPs in LD of the nine BAG GWASs using the original full samples; the SNPs are shown separately for each BAG. **g**) The alternative allele frequency for the setting in Figure **f**). **h**) The absolute beta coefficients of the nine BAGs using all genome-wide SNPs (the y-axis was truncated to 0.1 for visualization purposes). **i**) the alternative allele frequency did not differ for Figure h) including all genome-wide SNPs. For Figure **c, e, g**, and **i**, the upper/lower whiskers show the upper/lower boundaries based on the 1.5xthe interquartile range (IQR). The upper/lower hinge displays the top/bottom end of the IQR. The central measures denote the median values.



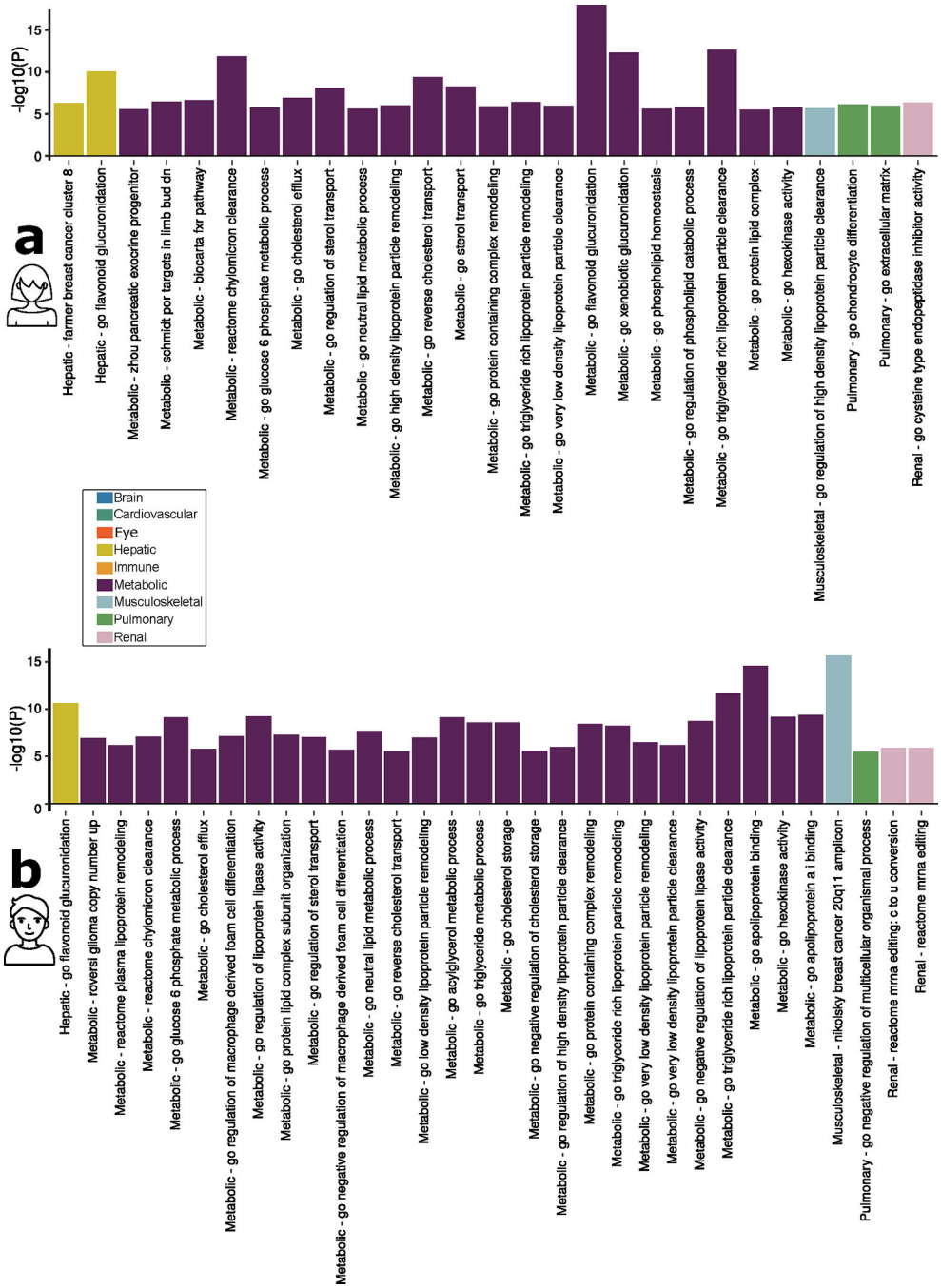
Extended Data Fig. 3 |. Trumpet plots of the alternative allele frequency vs. the beta coefficient of the nine BAG GWASs.

The trumpet plots display the inverse relationship between the alternative (effect) allele frequency and the effect size (beta coefficient) for the brain, cardiovascular, eye, hepatic, immune, metabolic, musculoskeletal, pulmonary, and renal BAGs. Only the independent significant SNPs were considered. The dot size corresponds to the effect size, while the transparency of the dot is proportional to its statistical significance.



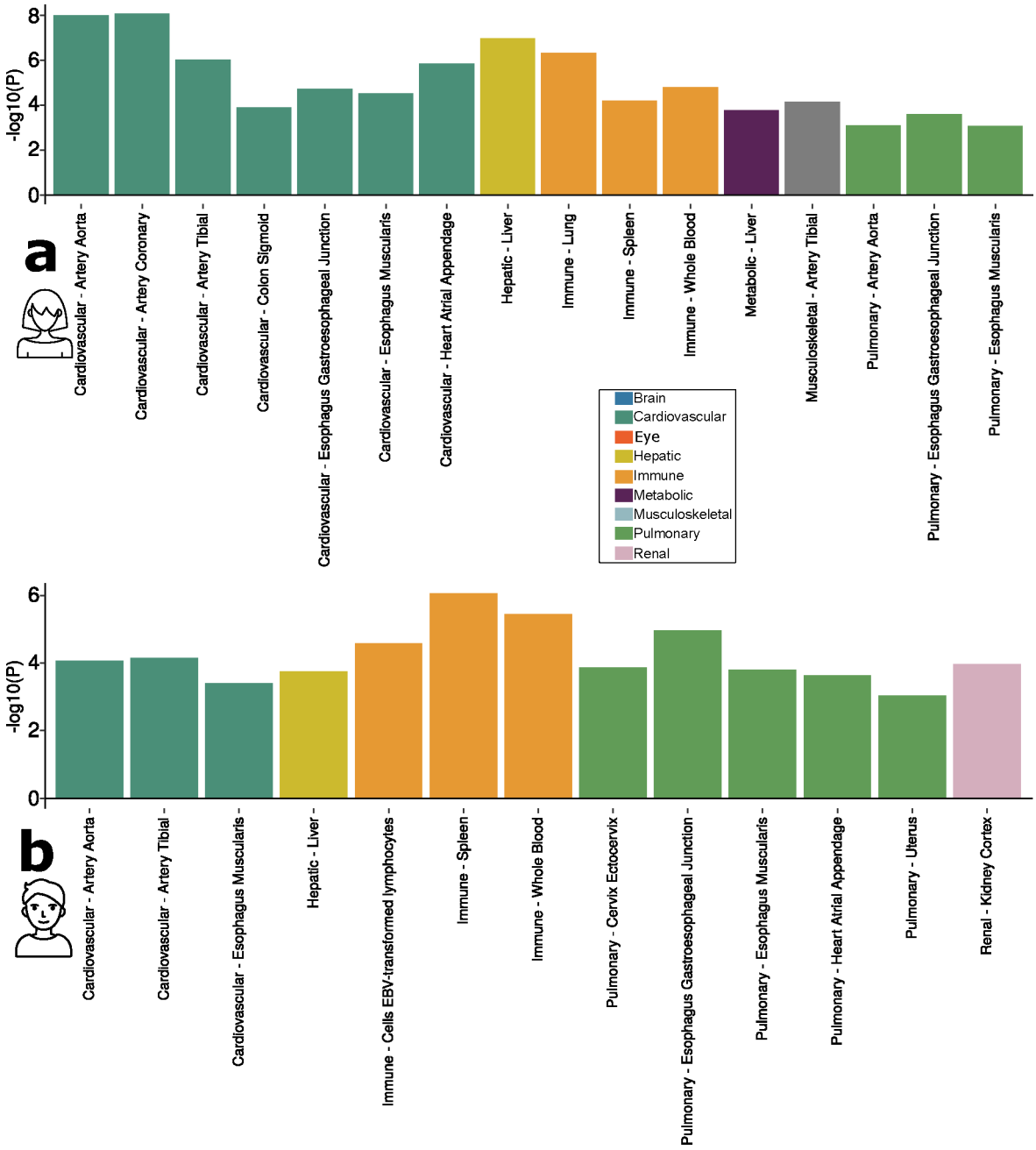
Extended Data Fig. 4 |. Bayesian colocalization signal between the pulmonary BAG and FEV/FVC.

Here, we illustrate the colocalization signal between the pulmonary BAG and the FEV/FCV feature at the genomic locus: 4q24, with the top lead SNP (causal SNP: rs7664805). Genetic colocalization was evidenced at one locus (4q24) between the pulmonary BAG and the FEV/FCV feature. The signed PP.H4.ABF (0.99) denotes the posterior probability (PP) of hypothesis H4, which suggests that both traits share the same causal SNP (rs7664805). All P-values were two-sided.



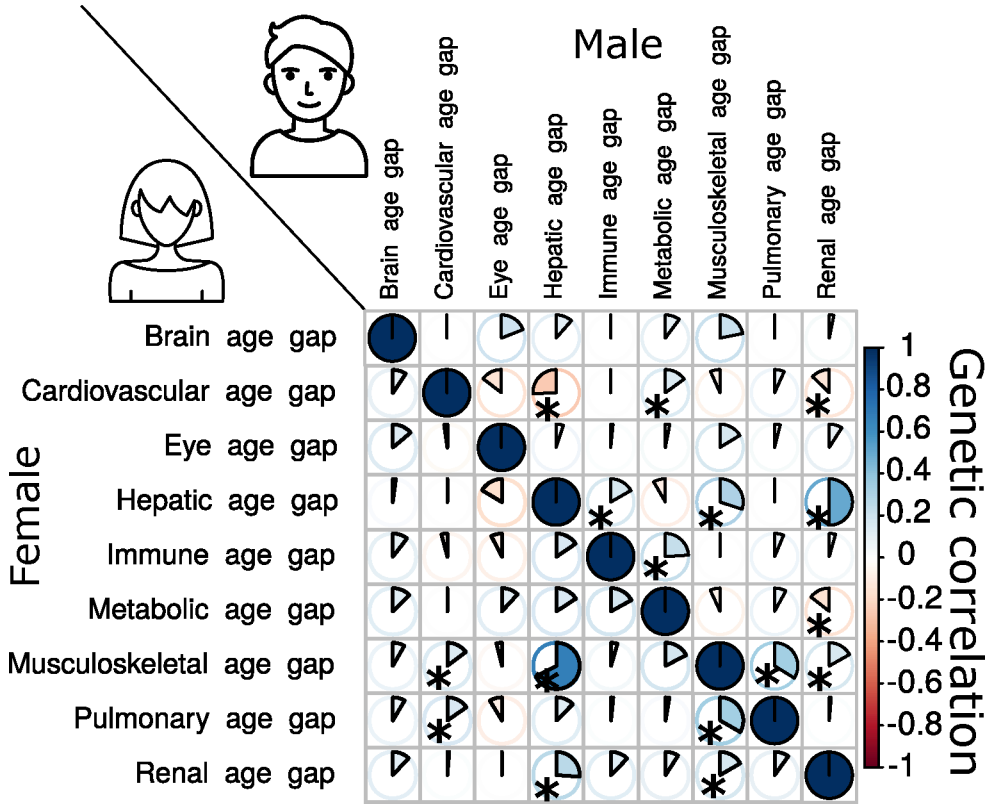
Extended Data Fig. 5 |. Gene-set enrichment analysis using sex-stratified GWAS results.

Gene-set enrichment analysis was performed using the GWAS summary statistics specific to females (**a**) and males (**b**). Gene set enrichment analyses were performed using curated gene sets and GO terms from the MsigDB database. Only significant gene sets are presented after adjusting for multiple comparisons using the Bonferroni correction. All P-values were two-sided.



Extended Data Fig. 6 |. Tissue-specific gene expression analysis using sex-stratified GWAS results.

Tissue-specific enrichment analysis was performed using the GWAS summary statistics specific to females (**a**) and males (**b**). Gene-property analyses evaluate tissue-specific gene expressions for the nine BAG-related genes using the full SNP P-values distribution. Only significant gene sets are presented after adjusting for multiple comparisons using the Bonferroni correction. All P-values were two-sided.



Extended Data Fig. 7 |. Genetic correlations using sex-stratified GWAS results.

The genetic correlation between each pair of BAGs was determined using sex-stratified GWAS summary statistics from our analyses. Most of the genetic correlations showed consistency between females and males, albeit sex differences are evident in certain BAGs, particularly in the cardiovascular BAG results. Specifically, males exhibit dominant correlations between cardiovascular BAGs and hepatic and renal BAGs, while females demonstrate specific correlations with musculoskeletal and pulmonary BAGs. All P-values were two-sided, and Bonferroni correction was employed to denote significant signals (* symbols).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We want to express our gratitude to the UK Biobank team for their invaluable contribution to advancing clinical research in our field. We gratefully acknowledge the support of the iSTAGING Consortium, funded by the National Institute on Aging through grant RF1 AG054409 at the University of Pennsylvania (C.D.). Additionally, we acknowledge the funding program from the Rebecca L. Cooper Foundation at the University of Melbourne (A.Z.). We thank P. Parmpi and J. Incmikoski at the University of Pennsylvania for their valuable administrative support. We thank J. Deelen and J. M. Murabito for their generosity in providing the GWAS summary statistics.

Data availability

This study used the UK Biobank resource under application numbers 35148 and 60698. The raw imaging data are restricted to registered researchers and are protected and unavailable due to data privacy laws; access can be obtained at https://www.ukbiobank.ac.uk/. The GWAS summary statistics corresponding to this study are publicly available on the MEDICINE knowledge portal (https://labs-laboratory.com/medicine). The gene–drug–disease network used data from the DrugBank database (v.5.1.9; https://go.drugbank.com/). Our genetic analyses also used GWAS summary statistics from the IEU OpenGWAS database (https://gwas.mrcieu.ac.uk/) and GWAS Catalog (https://www.ebi.ac.uk/gwas/) as well as GWAS summary data from individual studies by requesting directly from the authors. Gene sets were obtained from the Molecular Signatures Database (MSigDB, v7.5.1; https://www.gsea-msigdb.org/gsea/msigdb/). Tissue specificity enrichment analysis used data from GTEx V8 (https://gtexportal.org/home/). The analysis for partitioned heritability estimates used data from ROADMAP (https://egg2.wustl.edu/roadmap/web_portal/) and ENTEx (https://www.encodeproject.org/). All source data and original figures are also publicly available at Zenodo at https://doi.org/10.5281/zenodo.11075409 (ref. 84). All unrestricted data supporting the findings are also available from the corresponding author upon request.

References

1. Melzer D, Pilling LC & Ferrucci L The genetics of human ageing. Nat. Rev. Genet. 21, 88–101 (2020). [PubMed: 31690828]
2. Hodson R Precision medicine. Nature 537, S49 (2016). [PubMed: 27602738]
3. Tian YE et al. Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality. Nat. Med. 29, 1221–1231 (2023).10.1038/s41591-023-02296-6 [PubMed: 37024597]
4. Wen J et al. Genetic, clinical underpinnings of brain change along two neuroanatomical dimensions of clinically-defined Alzheimer's disease. Preprint at bioRxiv https://www.biorxiv.org/content/10.1101/2022.09.16.508329v3 (2024).
5. Liu Y et al. Genetic architecture of 11 organ traits derived from abdominal MRI using deep learning. eLife 10, e65554 (2021). [PubMed: 34128465]
6. McCracken C et al. Multi-organ imaging demonstrates the heart–brain–liver axis in UK Biobank participants. Nat. Commun. 13, 7839 (2022). [PubMed: 36543768]
7. Nie C et al. Distinct biological ages of organs and systems identified from a multi-omics study. Cell Rep. 38, 110459 (2022). [PubMed: 35263580]
8. Priest C & Tontonoz P Inter-organ cross-talk in metabolic syndrome. Nat. Metab. 1, 1177–1188 (2019). [PubMed: 32694672]
9. Wen J et al. The genetic architecture of multimodal human brain age. Nat. Commun. 15, 2604 (2024). [PubMed: 38521789]
10. Bycroft C et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209 (2018). [PubMed: 30305743]
11. Bulik-Sullivan BK et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47, 291–295 (2015). [PubMed: 25642630]
12. Klein SL & Flanagan KL Sex differences in immune responses. Nat. Rev. Immunol. 16, 626–638 (2016). [PubMed: 27546235]
13. Watanabe K et al. A global overview of pleiotropy and genetic architecture in complex traits. Nat. Genet. 51, 1339–1348 (2019). [PubMed: 31427789]
14. Evans LM et al. Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. Nat. Genet. 50, 737–745 (2018). [PubMed: 29700474]

15. Nikolsky Y et al. Genome-wide functional synergy between amplified and mutated genes in human breast cancer. Cancer Res. 68, 9532–9540 (2008). [PubMed: 19010930]
16. Wang T et al. Genome-wide DNA methylation analysis of pulmonary function in middle and old-aged Chinese monozygotic twins. Respir. Res. 22, 300 (2021). [PubMed: 34809630]
17. Consortium GTEx. The Genotype–Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585 (2013). [PubMed: 23715323]
18. Mittermayer F et al. Addressing unmet medical needs in type 2 diabetes: a narrative review of drugs under development. Curr. Diabetes Rev. 11, 17–31 (2015). [PubMed: 25537454]
19. Cheverud JM A comparison of genetic and phenotypic correlations. Evolution 42, 958–968 (1988). [PubMed: 28581166]
20. Regitz-Zagrosek V & Gebhard C Gender medicine: effects of sex and gender on cardiovascular disease manifestation and outcomes. Nat. Rev. Cardiol. 20, 236–247 (2023). [PubMed: 36316574]
21. Hwang G et al. Assessment of neuroanatomical endophenotypes of autism spectrum disorder and association with characteristics of individuals with schizophrenia and the general population. JAMA Psychiatry 80, 498–507 (2023). [PubMed: 37017948]
22. Wen J et al. Characterizing heterogeneity in neuroimaging, cognition, clinical symptoms, and genetics among patients with late-life depression. JAMA Psychiatry 79, 464–474 (2022). [PubMed: 35262657]
23. Yang Z et al. A deep learning framework identifies dimensional representations of Alzheimer's disease from brain structure. Nat. Commun. 12, 7065 (2021). [PubMed: 34862382]
24. Chand GB et al. Schizophrenia imaging signatures and their associations with cognition, psychopathology, and genetics in the general population. Am. J. Psychiatry 179, 650–660 (2022). [PubMed: 35410495]
25. Deelen J et al. A meta-analysis of genome-wide association studies identifies multiple longevity genes. Nat. Commun. 10, 3669 (2019). [PubMed: 31413261]
26. Hill WD et al. Genome-wide analysis identifies molecular systems and 149 genetic loci associated with income. Nat. Commun. 10, 5741 (2019). [PubMed: 31844048]
27. Codd V et al. Polygenic basis and biomedical consequences of telomere length variation. Nat. Genet. 53, 1425–1433 (2021). [PubMed: 34611362]
28. O'Connor LJ & Price AL Distinguishing genetic correlation from causation across 52 diseases and complex traits. Nat. Genet. 50, 1728–1734 (2018). [PubMed: 30374074]
29. Smith SM et al. Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations. eLife 9, e52677 (2020). [PubMed: 32134384]
30. London A, Benhar I & Schwartz M The retina as a window to the brain—from eye research to CNS disorders. Nat. Rev. Neurol. 9, 44–53 (2013). [PubMed: 23165340]
31. Zhao B et al. Heart–brain connections: phenotypic and genetic insights from magnetic resonance images. Science 380, abn6598 (2023). [PubMed: 37262162]
32. Parlakgül G et al. Regulation of liver subcellular architecture controls metabolic homeostasis. Nature 603, 736–742 (2022). [PubMed: 35264794]
33. Hotamisligil GS Inflammation and metabolic disorders. Nature 444, 860–867 (2006). [PubMed: 17167474]
34. Díaz Del Moral S, Benaouicha M, Muñoz-Chápuli R & Carmona R The insulin-like growth factor signalling pathway in cardiac development and regeneration. Int. J. Mol. Sci. 23, 234 (2021). [PubMed: 35008660]
35. Shen H et al. Mononuclear diploid cardiomyocytes support neonatal mouse heart regeneration in response to paracrine IGF2 signaling. eLife 9, e53071 (2020). [PubMed: 32167474]
36. Xu Q et al. The flavonoid procyanidin C1 has senotherapeutic activity and increases lifespan in mice. Nat. Metab. 3, 1706–1726 (2021). [PubMed: 34873338]
37. Tan P, Jin L, Qin X & He B Natural flavonoids: potential therapeutic strategies for non-alcoholic fatty liver disease. Front. Pharmacol. 13, 1005312 (2022). [PubMed: 36188561]
38. Zhao B et al. Common genetic variation influencing human white matter microstructure. Science 372, eabf3736 (2021). [PubMed: 34140357]

39. Litviňuková M et al. Cells of the adult human heart. Nature 588, 466–472 (2020). [PubMed: 32971526]
40. Ballard C et al. Drug repositioning and repurposing for Alzheimer disease. Nat. Rev. Neurol. 16, 661–673 (2020). [PubMed: 32939050]
41. Bulik-Sullivan B et al. An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241 (2015). [PubMed: 26414676]
42. Okun JG et al. Liver alanine catabolism promotes skeletal muscle atrophy and hyperglycaemia in type 2 diabetes. Nat. Metab. 3, 394–409 (2021). [PubMed: 33758419]
43. Barsh GS, Farooqi IS & O'Rahilly S Genetics of body-weight regulation. Nature 404, 644–651 (2000). [PubMed: 10766251]
44. Anandacoomarasamy A, Caterson I, Sambrook P, Fransen M & March L The impact of obesity on the musculoskeletal system. Int. J. Obes. 32, 211–222 (2008).
45. Van Gaal LF, Mertens IL & De Block CE Mechanisms linking obesity with cardiovascular disease. Nature 444, 875–880 (2006). [PubMed: 17167476]
46. Stanikova D et al. Testosterone imbalance may link depression and increased body weight in premenopausal women. Transl. Psychiatry 9, 160 (2019). [PubMed: 31175272]
47. Fyfe I Influence of amyloid-β on tau spread in Alzheimer disease explained. Nat. Rev. Neurol. 18, 318–318 (2022).
48. Wyss-Coray T Inflammation in Alzheimer disease: driving force, bystander or beneficial response? Nat. Med. 12, 1005–1015 (2006). [PubMed: 16960575]
49. Lacroix A et al. Sex modulation of faces prediction error in the autistic brain. Commun. Biol. 7, 127 (2024). [PubMed: 38273091]
50. Zhang Y et al. Genetic evidence of gender difference in autism spectrum disorder supports the female-protective effect. Transl. Psychiatry 10, 4 (2020). [PubMed: 32066658]
51. Ferretti MT et al. Sex differences in Alzheimer disease—the gateway to precision medicine. Nat. Rev. Neurol. 14, 457–469 (2018). [PubMed: 29985474]
52. Li F et al. Sex differences orchestrated by androgens at single-cell resolution. Nature 629, 193–200 (2024). [PubMed: 38600383]
53. Zuber V et al. Multi-response Mendelian randomization: identification of shared and distinct exposures for multimorbidity and multiple related disease outcomes. Am. J. Hum. Genet. 110, 1177–1199 (2023). [PubMed: 37419091]
54. Xue A et al. Genome-wide analyses of behavioural traits are subject to bias by misreports and longitudinal changes. Nat. Commun. 12, 20211 (2021). [PubMed: 33436567]
55. Sanderson E et al. Mendelian randomization. Nat. Rev. Methods Primers 2, 6 (2022). [PubMed: 37325194]
56. Manichaikul A et al. Robust relationship inference in genome-wide association studies. Bioinformatics 26, 2867–2873 (2010). [PubMed: 20926424]
57. Zheng J, Li Y, Abecasis GR & Scheet P A comparison of approaches to account for uncertainty in analysis of imputed genotypes. Genet. Epidemiol. 35, 102–110 (2011). [PubMed: 21254217]
58. Price AL, Zaitlen NA, Reich D & Patterson N New approaches to population stratification in genome-wide association studies. Nat. Rev. Genet. 11, 459–463 (2010). [PubMed: 20548291]
59. Abraham G, Qiu Y & Inouye M FlashPCA2: principal component analysis of Biobank-scale genotype datasets. Bioinformatics 33, 2776–2778 (2017). [PubMed: 28475694]
60. Wen J et al. Genomic loci influence patterns of structural covariance in the human brain. Proc. Natl Acad. Sci. USA 120, e2300842120 (2023). [PubMed: 38127979]
61. Purcell S et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007). [PubMed: 17701901]
62. Jiang L et al. A resource-efficient tool for mixed model association analysis of large-scale data. Nat. Genet. 51, 1749–1755 (2019). [PubMed: 31768069]
63. Watanabe K, Taskesen E, van Bochoven A & Posthuma D Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. 8, 1826 (2017). [PubMed: 29184056]
64. Yang J, Lee SH, Goddard ME & Visscher PM GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88, 76–82 (2011). [PubMed: 21167468]

65. Buniello A et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 47, D1005–D1012 (2019). [PubMed: 30445434]
66. Elsworth B et al. The MRC IEU OpenGWAS data infrastructure. Preprint at bioRxiv 10.1101/2020.08.10.244293 (2020).
67. de Leeuw CA, Mooij JM, Heskes T & Posthuma D MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. 11, e1004219 (2015). [PubMed: 25885710]
68. Sey NYA et al. A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. Nat. Neurosci. 23, 583–593 (2020). [PubMed: 32152537]
69. Weeks EM et al. Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. Nat. Genet. 55, 1267–1276 (2023). [PubMed: 37443254]
70. Subramanian A et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl Acad. Sci. USA 102, 15545–15550 (2005). [PubMed: 16199517]
71. Wishart DS et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 46, D1074–D1082 (2018). [PubMed: 29126136]
72. Sakaue S & Okada Y GREP: genome for REPositioning drugs. Bioinformatics 35, 3821–3823 (2019). [PubMed: 30859178]
73. Finucane HK et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235 (2015). [PubMed: 26414678]
74. Cahoy JD et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J. Neurosci. 28, 264–278 (2008). [PubMed: 18171944]
75. Bernstein BE et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat. Biotechnol. 28, 1045–1048 (2010). [PubMed: 20944595]
76. Dunham I et al. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012). [PubMed: 22955616]
77. Hemani G et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife 7, e34408 (2018). [PubMed: 29846171]
78. Burgess S et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. Wellcome Open Res. 4, 186 (2019). [PubMed: 32760811]
79. Skrivankova VW et al. Strengthening the reporting of observational studies in epidemiology using Mendelian randomization: the STROBE-MR Statement. JAMA 326, 1614–1621 (2021). [PubMed: 34698778]
80. Bowden J et al. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. Stat. Med. 36, 1783–1802 (2017). [PubMed: 28114746]
81. Bowden J, Davey Smith G & Burgess S Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int. J. Epidemiol. 44, 512–525 (2015). [PubMed: 26050253]
82. Giambartolomei C et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 10, e1004383 (2014). [PubMed: 24830394]
83. Gusev A et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am. J. Hum. Genet. 95, 535–552 (2014). [PubMed: 25439723]
84. Wen J The genetic architecture of biological age in nine human organ systems. Zenodo https://zenodo.org/records/11075409 (2024).

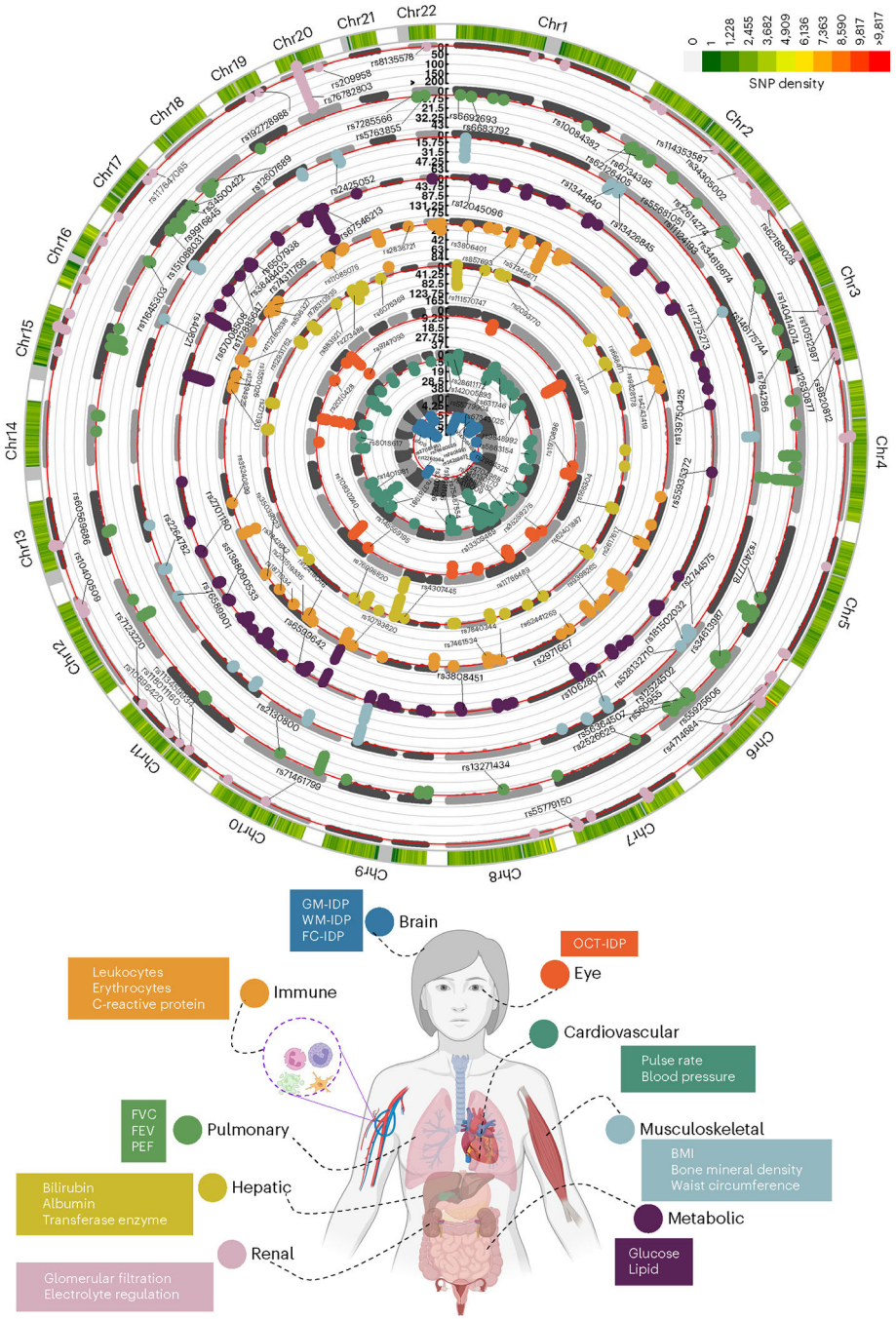


Fig. 1 |. Genomic loci associated with the nine BAGs.

Organ-specific BAG was derived from a large cohort of 30,108 to 111,543 participants of European ancestry from the UKBB cohort. The nine organ systems include the brain (N = 30,108), cardiovascular (N = 111,543), eye (N = 36,004), hepatic (N = 111,543), immune (N = 111,543), metabolic (N = 111,543), musculoskeletal (N = 111,543), pulmonary (N = 111,543) and renal (N = 111,543) BAGs. In total, 393 genomic loci–BAG pairs were identified using a genome-wide P value threshold ($-\log_{10}$ (P value) > 7.30; two sided). For visualization purposes, we denoted the genomic loci using their top lead SNPs that are

not associated with any clinical traits in the EMBL-EBI GWAS Catalog. All analyses used GRCh37. We present representative features used in calculating each organ's BAG; BMI, body mass index; Chr, chromosome; IDP, imaging-derived phenotype; GM, gray matter; WM, white matter; FC, functional connectivity; OCT, optical coherence tomography; FVC, forced vital capacity; FEV, forced expiratory volume; PEF, peak expiratory flow. The human anatomy was created with BioRender.com.

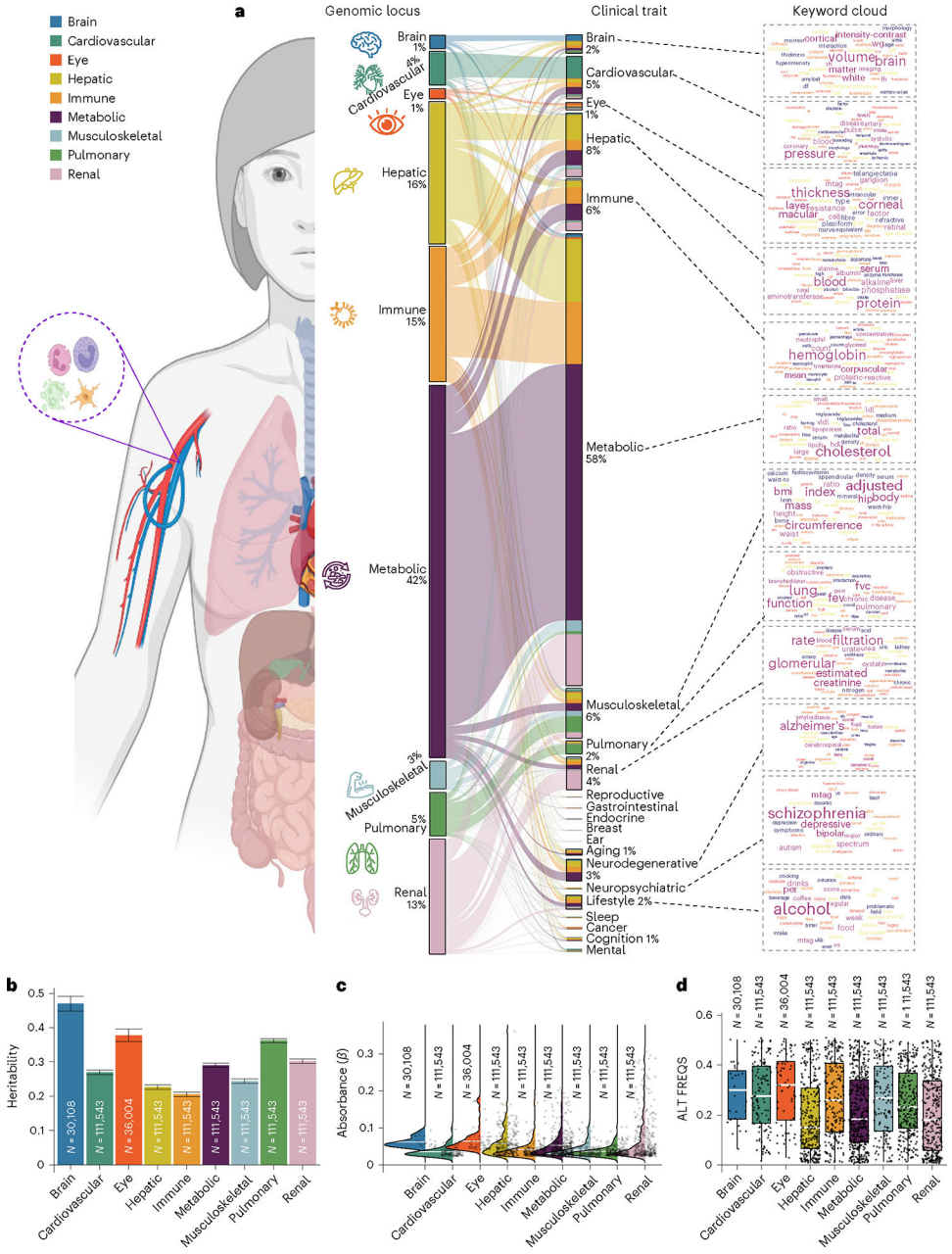


Fig. 2 |. Phenome-wide associations of the identified genomic loci and SNP-wide heritability estimates of the nine BAGs.

a, Phenome-wide association query of the identified genomic loci in the EMBL-EBI GWAS Catalog (query date 24 April 2023 via FUMA version v1.5.4) showed an organ-specific and interorgan landscape. By examining the independent significant SNPs considering linkage disequilibrium (Methods) within each genomic locus, we linked them to various clinical traits. These traits were categorized into high-level groups encompassing different organ systems, neurodegenerative and neuropsychiatric disorders and lifestyle factors. To visually represent the findings, we generated keyword cloud plots based on the frequency of these clinical traits within each BAG. The length of each rectangle block indicates the number of associations concerning the genomic loci in our analysis and clinical traits in the literature.

The individual disease traits were categorized within their respective organ systems. However, this categorization does not imply that the sum of these diseases exclusively represents the entirety of the organ system or that these diseases are solely associated with one specific organ system. Additional searches on alternative public GWAS platforms, such as the GWAS Atlas, are provided. **b**, The brain BAG is more heritable than other organ systems using GCTA. Error bars represent the standard error of the estimated parameters, and the measure of the center for the error bars represents the inferred statistics (that is, SNP-based heritability). **c**, The brain BAG showed larger effect sizes of the independent significant SNPs than other organ systems. The kernel density estimate plot shows the distribution of the effect sizes (that is, the magnitude of the linear regression β coefficients) in the nine GWASs. The white horizontal lines represent the mean effect sizes. **d**, Alternative allele frequency (effect allele) distribution for the nine BAGs. The upper/lower whiskers show the upper/lower boundaries based on 1.5× the interquartile range. The upper/lower hinge displays the top/bottom end of the interquartile range. The central measures denote the median values. Of note, only independent significant SNPs were shown for each BAG in **c** and **d**. All results in **b**–**d** used the original full sample sizes of the nine BAGs; the brain, eye and other body organ BAGs have different sample sizes. Results for **b**–**d** using the downsampled sample sizes (N = 30,108 of the brain BAG) are shown in Extended Data Fig. 2; ALT FREQS, allele frequency of the alternative (effective) allele. The human anatomy was created with BioRender.com.

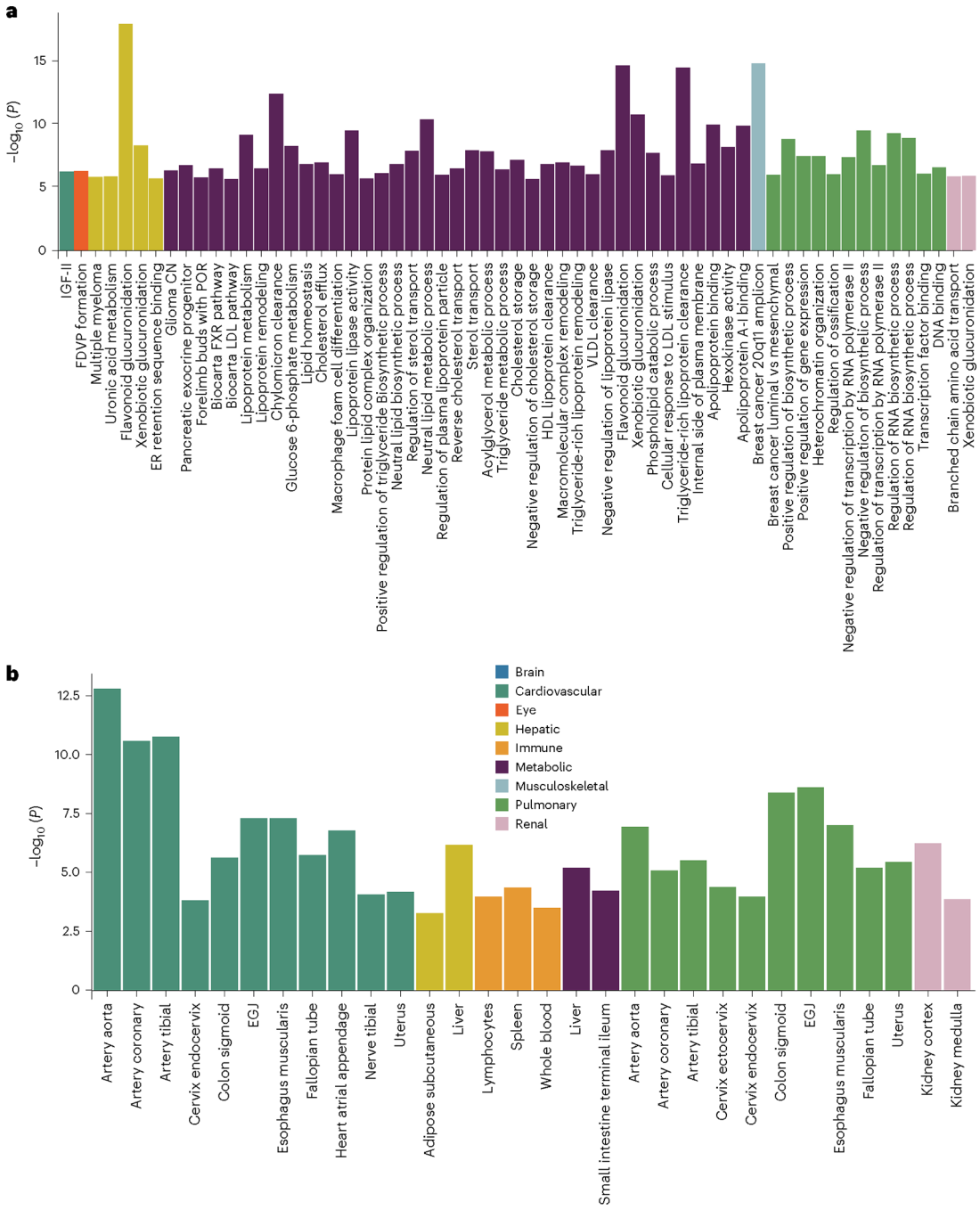


Fig. 3 |. Gene-level biological pathway annotation and tissue-specific gene expression.

a, Validation of the nine BAGs in GSEAs. GSEAs were performed using curated gene sets and Gene Ontology terms from the MSigDB database. **b**, Validation of the nine BAGs in gene property analyses. Gene property analyses evaluate tissue-specific gene expression for the nine BAG-related genes using the full SNP *P* value (two sided) distribution. Only significant gene sets are presented after adjusting for multiple comparisons using the Bonferroni correction; EGJ, esophagus gastroesophageal junction; FDVP, forebrain dorsal–

ventral pattern; VLDL, very-low-density lipoprotein; CN, copy number; POR, Schmidt port targets in limb bud.

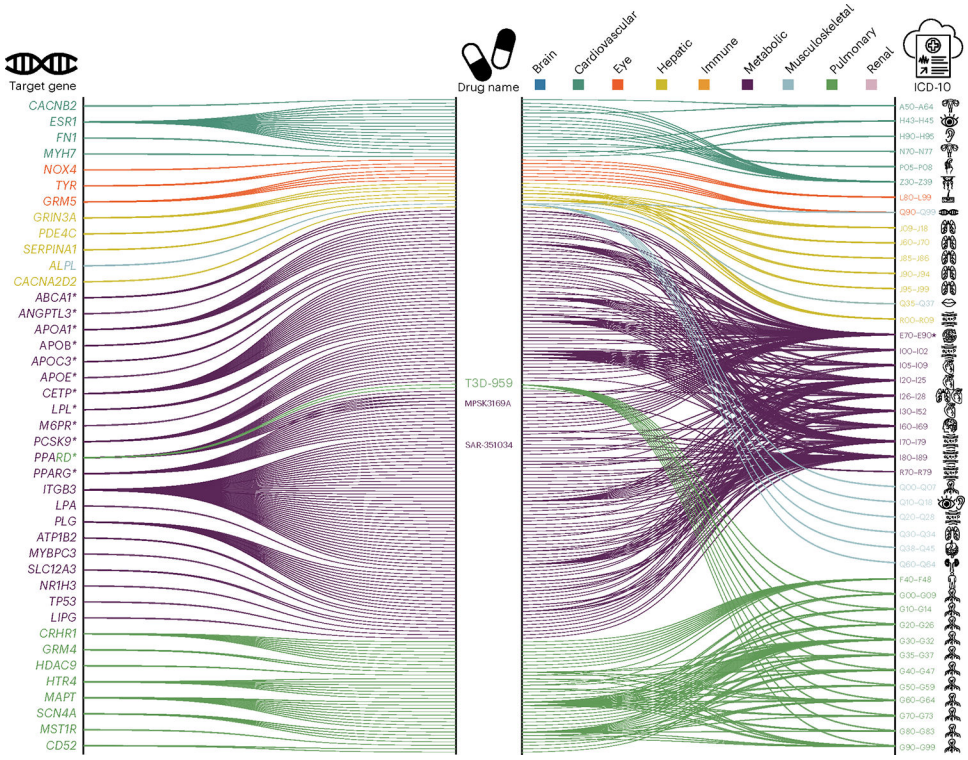


Fig. 4 |. Gene–drug–disease network of the nine BAGs.

The gene–drug–disease network reveals a broad spectrum of gene, drug and disease interactions across the nine BAGs, highlighting the metabolic-related genes. The ICD-10 code icons symbolize disease categories linked to the primary organ systems (for example, G30 for Alzheimer's disease in the CNS). All presented genes passed the nominal *P* value threshold (<0.05; two sided) and were pharmacogenetically associated with drug categories in the DrugBank database; an asterisk (*) indicates gene–drug–disease interactions that passed the Bonferroni correction.

Fig. 5 |. Partitioned heritability enrichment and genetic correlation of the nine BAGs.

a, Cell-type-specific partitioned heritability estimates for neurons, oligodendrocytes and astrocytes. **b**, Partitioned heritability estimates for the general 53 functional categories. For visualization purposes, we only show the four categories with the highest significant estimates for each BAG. The label for 500 denotes a 500-base pair window around each of the 24 main annotations in the full baseline model, which prevents a biased estimate inflated by heritability in flanking regions[83]; TSS, transcription start site; DHS, DNase I hypersensitivity site; TFBS, transcription factor binding site. **c**, Tissue-specific partitioned

heritability estimates using gene sets from multitissue gene expression data. **d**, Tissue- and chromatin-specific partitioned heritability estimates using multitissue chromatin data. **e**, Cheverud's conjecture: the genetic correlation between two BAGs (g_c; bottom triangle) mirrors their phenotypic correlation (p_c; top triangle). **f**, Genetic correlations between the nine BAGs and 41 clinical traits, including chronic diseases and their subtypes involving multiple human organ systems, education, intelligence and reaction time. An asterisk (*) denotes Bonferroni-corrected significance, and the absence of an asterisk indicates that all results remain significant after correction. For **a**–**d** and **f**, the standard error of the estimated parameters is presented using error bars. The measure of the center for the error bars represents the inferred statistics. The sample sizes, *P* values and other detailed statistics are presented in Source Data Files 13–18 and Supplementary Table 5; all *P* values are two sided; AD, Alzheimer's disease; ASD, autism spectrum disorder; LLD, late-life depression; SCZ, schizophrenia; DB, type 2 diabetes; WMH, white matter hyperintensity; HPLD, hyperlipidemia; AF, atrial fibrillation; RA, rheumatoid arthritis; CD, Crohn's disease; CKD, chronic kidney disease; PHST, primary hematopoietic stem cell.

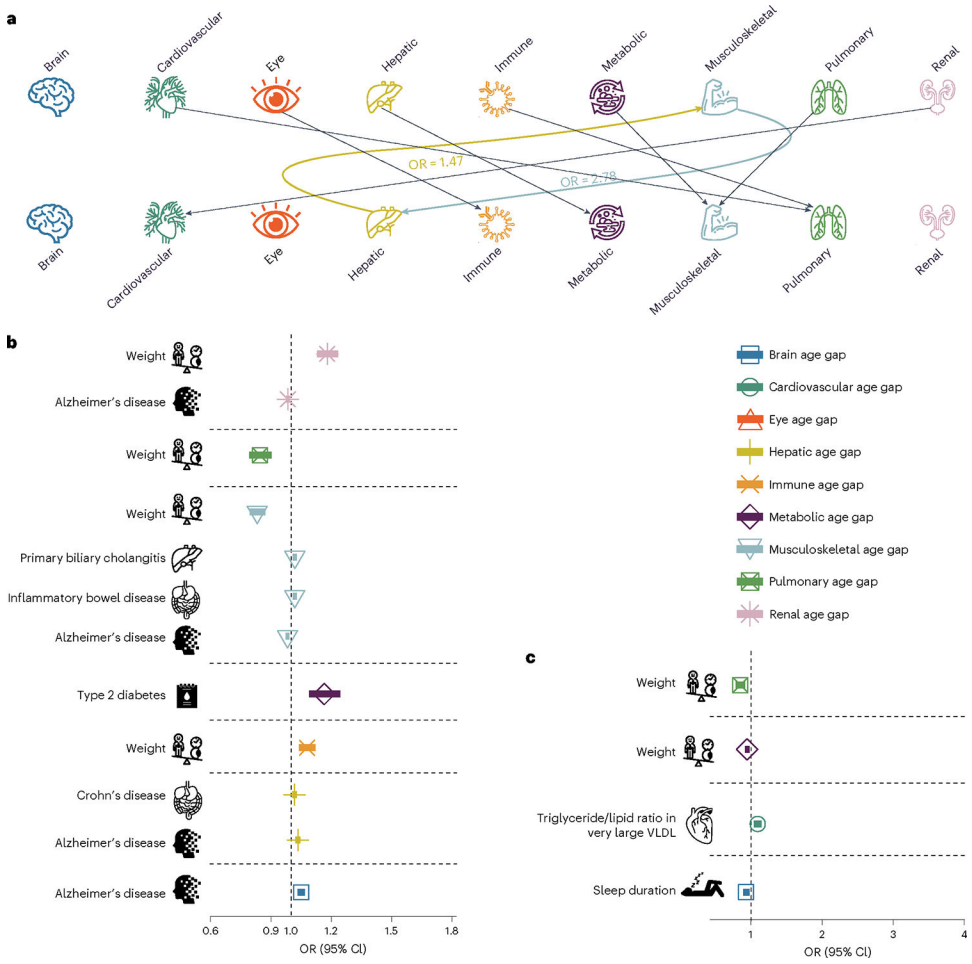


Fig. 6 |. Causal multiorgan network between the nine BAGs and 17 clinical traits of chronic diseases, lifestyle factors and cognition.

a, Causal inference between each pair of BAGs using bidirectional two-sample Mendelian randomization by excluding overlapping populations. The colored lines represent causal effects that survived the correction for multiple comparisons using the Bonferroni method; the dotted lines denote the nominal significant causal effects ($P < 0.05$). **b**, Forward Mendelian randomization investigates the causal inference of 17 unbiasedly selected exposure variables on the 9 outcome variables (that is, the 9 BAGs). **c**, Inverse Mendelian randomization examines the causal inference of the nine BAGs on the 17 clinical traits. Significant tests were adjusted for multiple comparisons using the Bonferroni correction. The ORs and 95% CIs are presented. Detailed OR and 95% CI information can be found in Source Data Files 19 and 20. It is crucial to approach the interpretation of these potential causal relationships with caution despite our thorough efforts in conducting multiple sensitivity checks to assess any potential violations of underlying assumptions. All *P* values are two sided.