

Revealing the genetic architectures underlying organ-specific aging based on proteomic data

Received: 15 October 2024

Accepted: 24 November 2025

Published online: 11 December 2025

 Check for updates

Ren-Jie Zhu^{1,4}, Yan Guo^{1,4}, Jia-Hao Wang¹, Ke Yu¹, Wei Shi¹, Wen Pan², Xian-lin Yu¹, Wei Chen³, Shan-Shan Dong¹✉ & Tie-Lin Yang¹✉

Organ-specific plasma protein signatures identified via proteomics profiling could be used to quantitatively track organ aging. However, the genetic determinants and molecular mechanisms underlying the organ-specific aging process remain poorly characterized. Here we integrated large-scale plasma proteomic and genomic data from 51,936 UK Biobank participants to uncover the genetic architectures underlying aging across 13 organs. We identified 119 genetic loci associated with organ aging, including 27 shared across multiple organs, and prioritized 554 risk genes involved in organ-relevant biological pathways, such as T cell-mediated immunity in immune aging. Causal inference analyses indicated that accelerated heart and muscle aging increase the risk of heart failure, whereas kidney aging contributes to hypertension. Moreover, smoking initiation was positively linked to the aging of the lung, intestine, kidney, and stomach. These findings establish a genetic foundation for understanding organ-specific aging and provide insights for promoting healthy longevity.

Biological aging refers to the gradual and cumulative changes in the biological functions of an organism as time advances¹. These changes lead to a decline in physiological capabilities and an increase in disease susceptibility^{2,3}. Investigating the underlying mechanisms of biological aging can enhance our comprehension of the processes contributing to aging-related diseases, thereby informing the development of potential therapeutic interventions.

Most of studies on biological aging focus on whole-body aging indicators, such as healthspan⁴, lifespan^{5,6} and longevity⁷. Given the distinct characteristics of various organs, organ aging, which refers to the biological aging specific to particular organ systems, has increasingly become a research hotspot^{8–10}. Several methods have been established for assessing organ aging. For example, some studies^{11,12} used MRI-based brain volume and functional connectivity

measurements to measure brain aging, but MRI measurements are costly and do not provide molecular insights. Some other studies have used clinical chemistry markers to measure organ aging^{8,13,14}. However, many of these markers have low organ specificity, making them difficult to interpret for organ-specific aging. Moreover, suitable and easily measurable clinical chemistry markers are not yet available for some organs (e.g., intestine and stomach). Therefore, we need to find other indicators for organ aging.

Proteins are the final products of gene expression and the principal functional constituents of cellular and biological processes. Moreover, most drug targets and biomarkers are proteins^{15,16}. In clinical practice, certain plasma proteins have been used to noninvasively evaluate the health status of specific organs, such as alanine transaminase for liver damage^{17,18}. Therefore, finding organ-specific proteins

¹Biomedical Informatics & Genomics Center, Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, PR China. ²Department of Health Management Center, Hospital of Chengdu Office of People's Government of Xizang Autonomous Region (Hospital.C.X.), Chengdu, Sichuan, PR China. ³Department of Clinical Laboratory, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, PR China. ⁴These authors contributed equally: Ren-Jie Zhu, Yan Guo. ✉e-mail: dongss@xjtu.edu.cn; yangtielin@xjtu.edu.cn

in plasma could track human organ aging. Indeed, recent studies successfully used plasma proteomics data to predict organ aging^{9,19,20}. However, these studies primarily focused on identifying correlations between aging and other phenotypes, without delving into the underlying genetic determinants of organ aging. Moreover, causal associations between organ aging and diseases were not explored.

In this study, using plasma proteomics data from the UK Biobank Pharma Proteomics Project (UKB-PPP)²¹, we trained models to predict organ aging for 13 human organs. Genome-wide association study (GWAS) was then performed to pinpoint the genomic loci associated with organ aging. Mendelian randomization (MR) analysis was further used to explore the causal relationship between aging-related diseases, modifiable lifestyle factors and organ aging. Our results could enhance

the comprehension of the genetic basis of organ aging and highlighted the causal relationship between organ aging and other traits.

Results

Overall design

The overall design of this study is shown in Fig. 1. Briefly, using plasma proteomics data (2923 proteins) from 51,936 participants of UKB-PPP²¹, we mapped putative organ-specific plasma proteome and used them to train models to predict organ aging for 13 human organs (adipose, artery, brain, heart, immune, intestine, kidney, liver, lung, muscle, pancreas, skin and stomach). We then performed GWAS to explore the genetic mechanisms underlying organ aging. Subsequently, we integrated transcriptome-wide association study (TWAS), colocalization, and causal inference approaches to identify likely

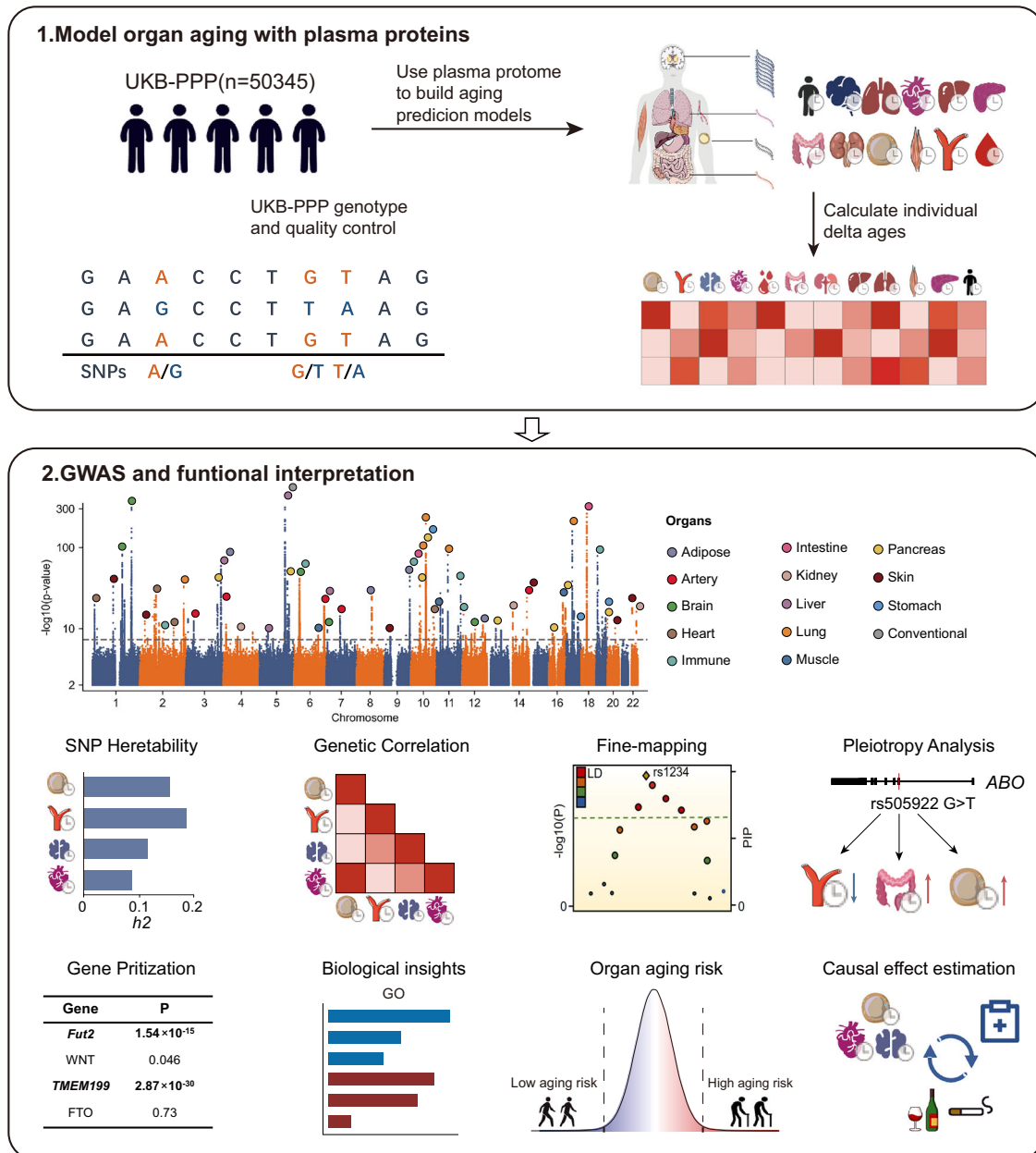


Fig. 1 | Schematic of the study design. First, proteomics data from UKB-PPP was used to model organ aging. GWAS analysis was then performed on the predicted aging phenotypes using genotypic data from UKBB to elucidate the genetic mechanisms of organ aging. The heritability and genetic correlations of organ aging were also calculated, and various methods were used to identify risk genes

for organ aging. Further, the biological insights of prioritized genes were examined. Finally, the GWAS findings were leveraged to construct a risk prediction model for organ aging, and Mendelian randomization analysis was performed to test associations between organ aging, disease, and lifestyle phenotypes.

effector genes. Furthermore, we constructed polygenic risk scores to assess their association with aging-related diseases. Additionally, MR analysis was employed to reveal the potential causality between aging-related diseases, modifiable lifestyle factors and organ aging.

Model organ aging with plasma proteins

We trained aging prediction models using proteomic data measured during 2006 and 2010 from the UK Biobank Pharma Proteomics Project (UKB-PPP)²¹ for 13 organs, including adipose, artery, brain, heart, immune system, intestine, kidney, liver, lung, muscle, pancreas, skin and stomach, because of their relatively well-understood contributions to aging-related diseases. The organ-enriched proteins we used comprised the organ-enriched proteins from the Human Protein Atlas (HPA) project²² and the organ-specific proteins identified by Malmström et al.'s study²³. As shown in Supplementary Data 1, 1799 proteins were selected for the 13 organs (ranged from 34 to 356). These proteins were enriched in pathways related to organ-specific functions (Supplementary Data 2). We trained aging model for 13 organs using organ-enriched proteins, and a 'conventional' proteomic aging model was trained using all proteins. As shown in Supplementary Fig. 2, among the three machine learning models used, the elastic net model demonstrated the best performance on the test sets across 11 organs. For the remaining three organs (kidney, pancreas, and stomach), although XGBoost or random forest showed slightly better performance on the test dataset in some cases, performance discrepancies between the training and test sets were observed for these models, indicating potential overfitting. In contrast, the elastic net model maintained similar performance between the training and test sets. Therefore, we opted for the elastic net model in subsequent analyses. The protein weights in the elastic net model are provided in Supplementary Data 3. Subsequently, we compared the performance of the final trained model on the training dataset and an independent test dataset. The model's performance was similar across both datasets, indicating good generalization ability (Supplementary Fig. 3).

The developed models were then used to predict biological ages for all UKB-PPP baseline participants. Similar to previous studies^{24,25}, we observed that the predicted age difference (PAD)—the difference between predicted age and chronological age—exhibited negative correlation with chronological age (r ranged from -0.90 to -0.48 , Supplementary Fig. 4). Post-prediction correction of this age bias problem is crucial²⁶. We adopt the method used by Oh et al.⁹ to address this problem. The corrected PAD (hereinafter referred to as "delta age") was defined as the residuals from a locally weighted scatterplot smoothing (lowess) regression of predicted age on chronological age model. As shown in Supplementary Fig. 4, this method successfully corrected the dependence of delta age on chronological age (r ranged from -0.021 to 0.025). Subsequent analyses were performed after Z-score scaling of the delta age values.

The association between organ aging and disease status

We next examined the associations between organ aging and disease status. Disease status was defined based on ICD-10 diagnoses (UK Biobank Field 41270). Nine major disease categories were included in the analysis: immune system diseases (ICD-10: D50–D89), circulatory system diseases (I00–I99), musculoskeletal system diseases (M00–M99), nervous system diseases (G00–G99), respiratory system diseases (J00–J99), endocrine, nutritional, and metabolic diseases (E00–E90), mental and behavioral disorders (F00–F99), digestive system diseases (K00–K93), and diseases of the skin and subcutaneous tissue (L00–L99) (Supplementary Data 4). In total, 617 specific diseases across these categories were included. Disease records were categorized into prevalent diseases (diagnosed before recruitment) and incident diseases (diagnosed after recruitment) according to the timing of diagnosis relative to recruitment.

For prevalent diseases, we performed logistic regression analyses to assess the association between disease status and delta age. Among the 126 disease category-aging trait pairs (9 disease categories \times 14 aging traits), we detected 125 significant pairs of positive correlations. As shown in Fig. 2a, some disease-organ correlations were as expected. For example, brain aging exhibited the strongest positive correlation with mental and behavioral disorders. We also observed some other interesting correlations, for example, stomach aging was significantly correlated with the status of mental and behavioral disorder. This might be because multiple cortical regions can regulate autonomic control of the stomach, involving the gut-brain connection²⁷. We also explored the relationship between 617 diseases within the 9 disease systems and organ aging. After Bonferroni correction, we detected significant correlation in 917 disease-organ pairs. For a specific organ, the type of disease that had the highest number of associations with it was often the one that belongs to the same system as the organ. For example, 47.6% (20/42) diseases associated with accelerated heart aging were circulatory system diseases (Fig. 2b). On the other hand, a specific disease often showed the highest correlation with the aging of its affected organ. For example, we found that Alzheimer's disease (AD) showed the highest correlation with brain aging (Fig. 2c). We also categorized individuals into extreme aging types using a two standard deviation cutoff (Z-scored delta age > 2 as e-older, Z-scored delta age < -2 as e-younger) for at least one aging model. In comparison to other subjects, the e-older cohort exhibited a significantly elevated disease risk while the e-younger cohort demonstrated lower disease risk (Supplementary Fig. 5).

For incident diseases, we applied Cox proportional hazards models to investigate the association between delta age and the risk of future disease onset. Among the 126 disease category-aging trait pairs, we detected 124 significant pairs of positive correlations, suggesting that higher delta age was associated with a higher risk of incident disease status. We also explored the relationship between 617 incident disease status within nine systems and organ aging. After Bonferroni correction, we identified 1526 significant correlations. For example, brain aging was most significantly associated with increased risk of Alzheimer's disease (OR = 1.68, $p = 1.21 \times 10^{-44}$) (Supplementary Fig. 6).

Validate the association between organ aging and disease status

We used proteomics data measured after 2019 in the UKB-PPP (hereafter referred to as validation cohort) study as an external cohort to validate the aging-disease associations. This dataset included 1463 plasma proteins from 1125 participants. Individuals who participated in the validation study were removed from the original cohort. That's, there is no sample overlap between the original and validation datasets.

To make sure that the validation models preserved most features used in the original models, we only kept 8 organ models with more than 50% of the protein features available in the validation dataset (Supplementary Data 5). Using these features, we re-trained the organ aging models in the original cohort. As shown in Supplementary Fig. 7, the predicted aging of the feature-reduced model was highly correlated with that from the full feature model. In addition, no significant difference was observed between the weights for the proteins in the feature-reduced and full feature models (Supplementary Data 5). Therefore, the feature-reduced models were similar to the original ones. We then applied these 8 feature-reduced models to predict organ aging for individuals in the validation cohort and analyzed the associations between predicted organ aging and disease outcomes. As shown in Supplementary Fig. 8, we validated the association between organ aging and multiple diseases. For example, we observed that adipose aging was positively associated with endocrine/metabolic diseases in the original dataset ($\beta = 0.42$, $p = 1.0 \times 10^{-11}$), this association was also observed in the validation dataset ($\beta = 0.25$, $p = 1.8 \times 10^{-6}$).

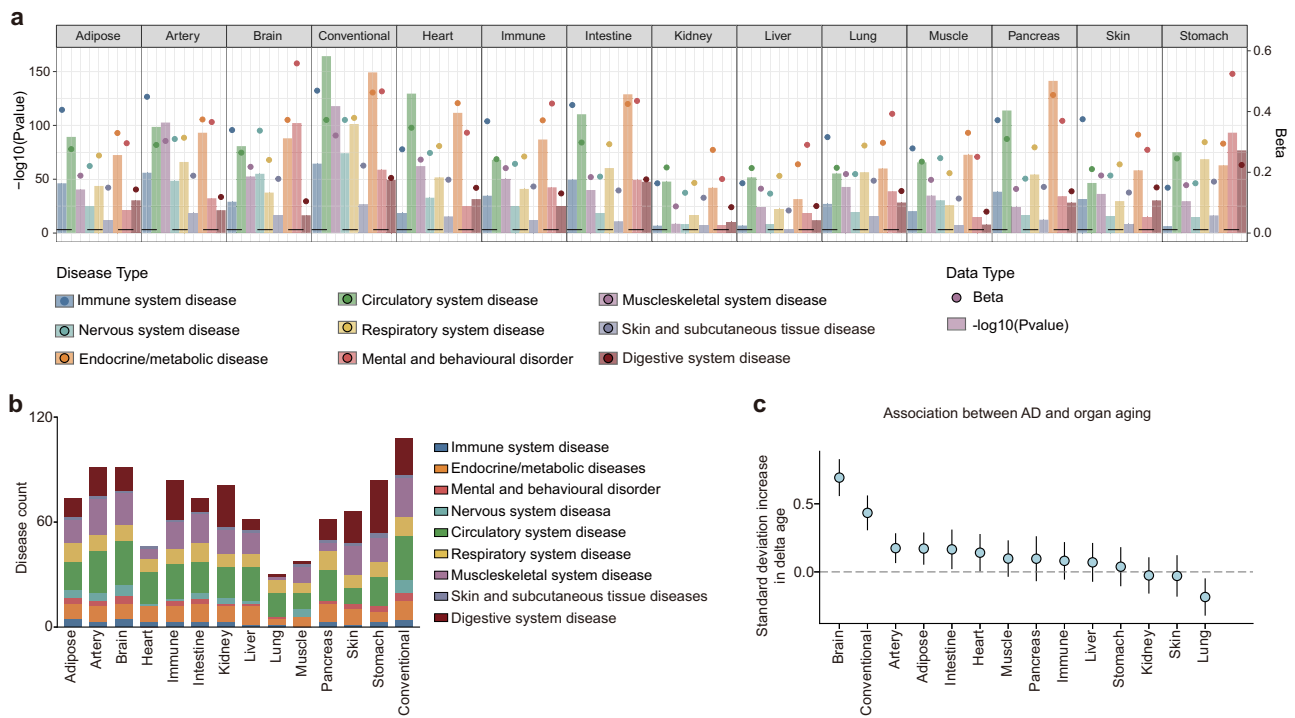


Fig. 2 | The associations between delta ages and diseases. **a** Logistic regression analysis of the relationship between delta ages and prevalence of 9 disease categories, with sex, age and the first five principal components as covariates. Dashed lines indicate the p-value threshold for significance after Bonferroni correction. **b** Logistic regression analysis was used to examine the relationship between 617 diseases and delta ages. The numbers of prevalent diseases significantly positively

correlated with each organ's delta ages after Bonferroni correction are shown. **c** Logistic regression analysis results for the relationship between prevalence of Alzheimer's Disease (AD) and delta ages of different organs ($n = 51,936$ participants; 179 with AD and 51,757 controls). Error bars denote 95% confidence intervals. Source data are provided as a Source Data file.

These findings support the aging–disease associations we observed in the original cohort.

GWAS of organ aging

To understand the genetic basis of organ aging, we performed GWAS on predicted delta age. The sample size for the 14 aging traits (13 organ aging and 1 conventional aging) ranged from 19,622 to 28,151. After variant filtering and quality control, up to 6,735,132 autosomal single-nucleotide polymorphisms (SNPs) were included in our analysis.

We divided the individuals based on the UK Biobank assessment centers at which participants consented into southern and northern cohorts. We considered the southern cohort ($n = 26,293$) as the discovery samples and the northern cohort as the replication samples ($n = 26,477$). We identified 10,661 significant SNPs associated with organ aging ($p < 5 \times 10^{-8}$) in the discovery sample, 98.93% of these SNPs also exhibited FDR corrected p -value < 0.05 in the replication cohort. Importantly, we found that the effect sizes of these SNPs exhibited a Pearson correlation coefficient of 0.96 between the discovery cohort and the replication cohort (Supplementary Data 6). Finally, we meta-analysed the GWAS results from the two cohorts and obtained 19,767 significant SNPs for subsequent analysis. A total of 119 loci (defined as a ± 500 -kb region around the genome-wide significant SNPs) were detected to be associated with at least one aging trait (Fig. 3a, b, Supplementary Figs. 9, 10). We found that 92 of the 119 loci (77.3%) were associated with only one organ aging trait, and low-to-moderate genetic correlations between the aging of 14 organs were observed (average r_g was 0.15) (Fig. 3c).

To further identify potential causal SNPs for organ aging, we utilized a stepwise model selection approach with GCTA-COJO²⁸ to detect independent signals, followed by FINEMAP²⁹ to identify the most likely causal SNPs for each signal. A total of 1750 SNPs achieving either

conditional independence or a high posterior probability for causality ($\log_{10}BF > 2$) were retained and referred as ‘fine-mapped SNPs’ for subsequent analysis. We identified 7 loci that had no more than 5 variants within the 95% credible set (Fig. 3d), which can be considered as highly credible causal SNPs.

Only 78 fine-mapped SNPs (5.21%) were located within exon of protein-coding genes, including 47 nonsynonymous SNPs. To further investigate the function of fine-mapped SNPs, we annotated these SNPs using public databases. As shown in Supplementary Data 7, 7 SNPs were annotated as having clinical and pathological significance in the ClinVar³⁰ database, 165 SNPs were predicted to be deleterious to gene function by CADD³¹, DANN³², or FATHMM-MKL³², 1174 SNPs were predicted as likely to affect regulome binding by RegulomeDB³³, 9 SNPs were predicted to affect gene splicing by regSNPs-intron³⁴, and 30 SNPs were predicted as damaging exonic variants with the potential of affecting protein function by SIFT/PROVEAN^{35,36}.

Querying fine-mapped SNPs from existing GWAS

To investigate whether the fine-mapped SNPs associated to organ aging exhibit known associations with other phenotypes, we queried these SNPs in the GWAS-catalog database³⁷. The results showed that 15.8% (276/1750) of the fine-mapped SNPs have been reported to be associated with other traits (Supplementary Data 8). We found that fine-mapped SNPs were annotated not only to the corresponding organ phenotype but also to phenotypes of other organs. For example, The genomic locus exhibiting the most significant association signal with lung aging (chr10: 80716529–82718385) has been reported to be associated with both lung-related traits, such as such as the forced expiratory volume in one second to forced vital capacity ratio (FEV1/FVC)³⁸ and COVID-19³⁹, but also with other organ related traits, such as rheumatoid arthritis⁴⁰ and height⁴¹. A locus associated with brain aging (chr1: 149714166–151444734) has been reported to be linked to

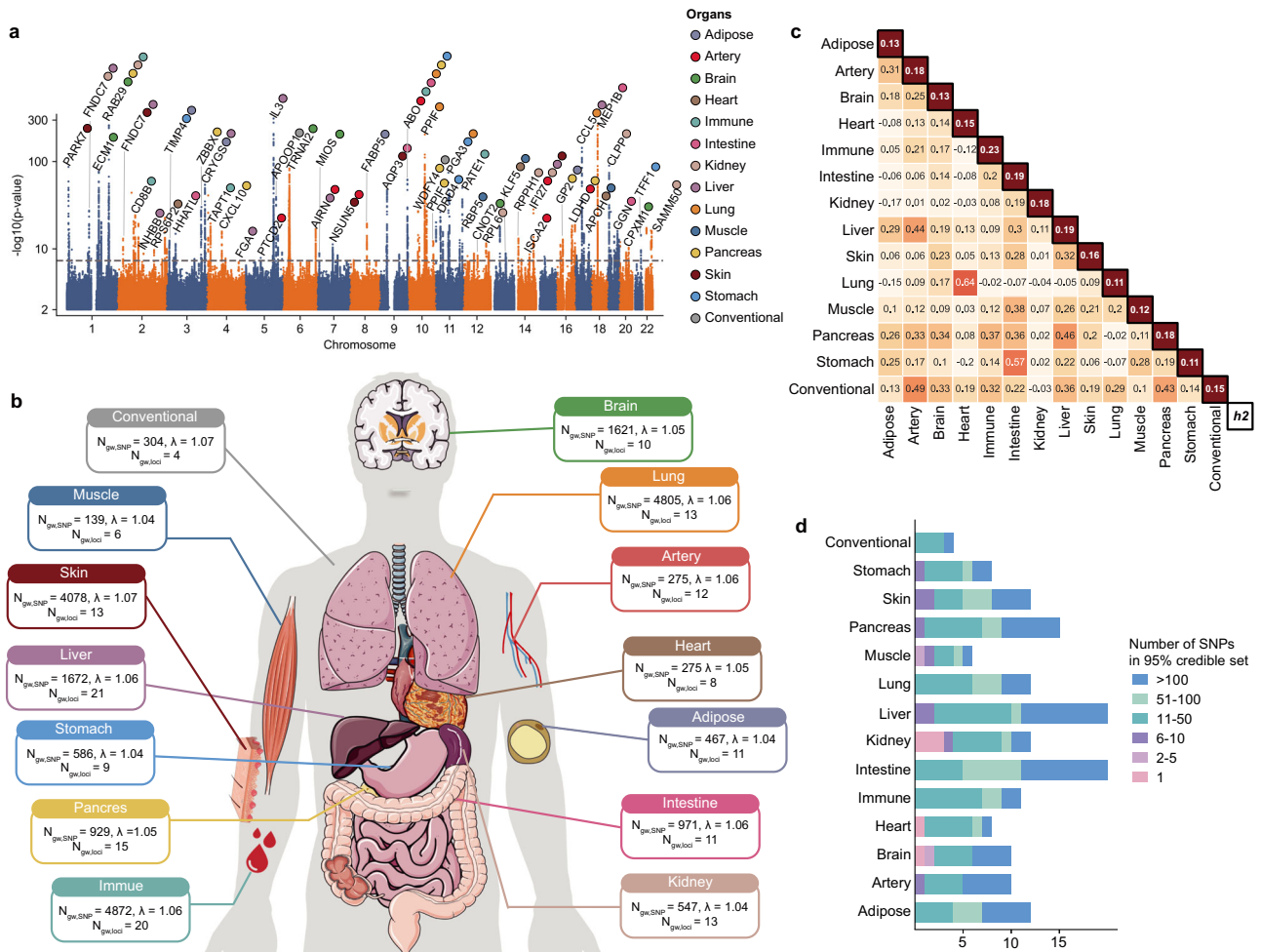


Fig. 3 | GWAS results. a Manhattan plot showing GWAS results for 14 aging traits. For each locus reaching genome-wide significance ($p < 5 \times 10^{-8}$), the nearest gene is annotated. **b** The total number of genome-wide significant SNPs, the total number of genome-wide significant loci and genomic inflation factor (λ from LDSC) for each

trait. **c** The lower triangle shows bivariate genetic correlations for each pair of phenotypes estimated using LDSC. The diagonal indicates the SNP heritability of each phenotype. **d** Distribution of 95% credible set size derived from fine-mapping analysis for each trait. Source data are provided as a Source Data file.

schizophrenia⁴² and phenotypes of other organs, including blood protein levels⁴³, platelet count⁴⁴.

Gene prioritization for organ aging

To link GWAS SNPs to relevant protein-coding genes, we employed gene-level and SNP-level prioritization methods. The gene-level analyses included TWAS, colocalization (using the R package coloc⁴⁵), and summary-data-based Mendelian Randomization (SMR)⁴⁶. In the SNP-level analyses, genes harboring fine-mapped SNPs with functional annotation support were included in the prioritized genes. Finally, we identified 554 protein-coding genes associated with organ aging (Fig. 4a, Supplementary Data 9).

The immune system harbored the greatest number of risk genes ($n = 91$), while muscle had only 10 risk genes (Fig. 4b). Among the 554 risk genes, only 37 were associated with multiple aging traits. *FUT2*, which was a risk gene of intestine, pancreas, and stomach aging, was reported to be related to the functions of these organs⁴⁷⁻⁴⁹. Another gene *PLG*, which was reported to be associated with cardiovascular function⁵⁰ and liver regeneration⁵¹, was identified as a risk gene for artery and liver aging.

ABO blood type is associated with aging in multiple organs

Among the 37 risk genes associated with multiple aging traits, the gene *ABO*, which determines blood type, was identified as a risk gene for

adipose, artery, immune, intestine, pancreas, and lung aging. In the UK Biobank, blood type was inferred based on two SNPs: rs8176719 and rs8176746. In the GWAS results (Fig. 4c), T allele of rs8176719 was positively associated with the aging of artery ($\beta = 0.05, p = 1.4 \times 10^{-8}$), immune ($\beta = 0.04, p = 7.7 \times 10^{-10}$), intestine ($\beta = 0.06, p = 9.6 \times 10^{-12}$) but negatively associated with pancreas aging ($\beta = -0.07, p = 2.7 \times 10^{-15}$). G allele of rs8176746 was positively associated with pancreas aging ($\beta = 0.17, p = 1.5 \times 10^{-23}$). As for lung aging, no significant results were found for both SNPs, but suggestive significant signal was detected for rs8176719 (β for T allele = $-0.03, p = 2.1 \times 10^{-6}$).

We further compared the delta age across different blood types. We found that individuals with blood O type exhibited the highest arterial and intestinal age. Those with blood type B displayed the lowest lung and pancreas age. Meanwhile, individuals with blood type A have the lowest lung age coupled with the highest pancreas age (Supplementary Fig. 11). These findings suggested that people with corresponding blood types should pay special attention to the health status of related organs.

Functional analysis of the organ aging genes

Pathway enrichment of risk genes. To annotate relevant biological pathways underlying each organ system, we performed pathway enrichment analysis for the organ aging genes. We found that some organ-enriched pathways showed strong organ specificity, reflecting

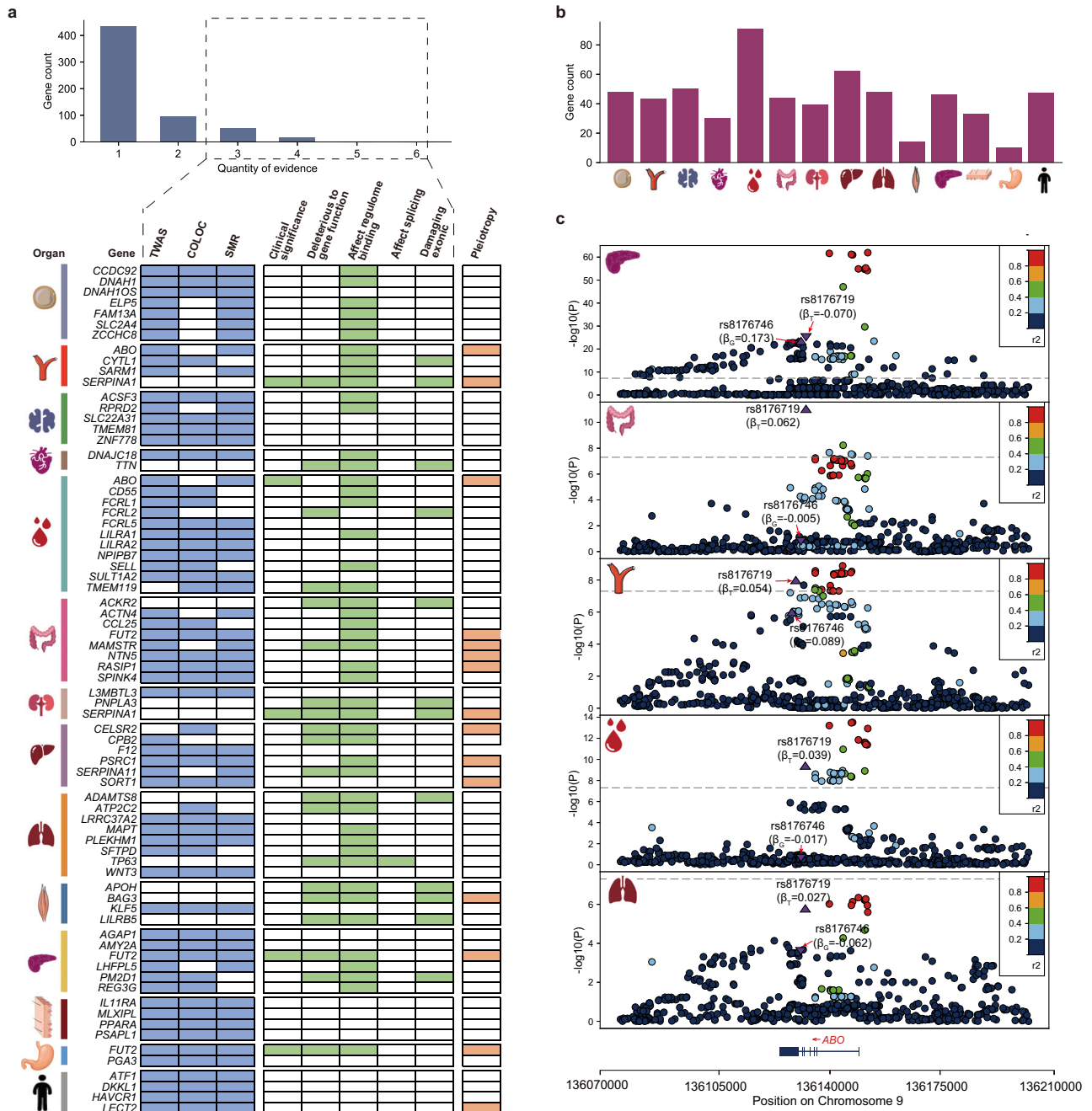


Fig. 4 | Results of gene prioritization. **a** (Upper panel) The histogram shows the distribution of risk genes categorized by the number of supporting evidence. The x-axis represents the number of supporting evidence, while the y-axis shows the count of risk genes for each category. Summary of evidence categories of prioritized genes. (Bottom panel) 85 genes supported by at least 3 analysis methods in 14 traits are shown. The organs and prioritized genes are displayed on the left, and “pleiotropy” indicates whether the gene is associated with more than one aging

phenotype. **b** The number of genes prioritized (by TWAS, COLOC, SMR, or fine-mapped SNP annotation) for each aging phenotype. **c** GWAS results at the *ABO* gene locus for pancreas, intestine, immune, artery and lung aging. The two SNPs that determined the *ABO* blood type were annotated. Gray dashed lines represent genome-wide significance ($p < 5 \times 10^{-8}$). Source data are provided as a Source Data file.

the unique functions of these organs. For example, the top-ranked enriched pathways for adipose aging genes included fatty acid beta-oxidation; the most significantly enriched pathway for heart aging genes included sarcomere organization and myofibril assembly; the most significantly enriched pathways for immune aging genes included T cell mediated immunity. On the other hand, some organ-enriched pathways exhibit broader functions. For instance, the most significantly enriched pathway for intestine aging genes was epidermal growth factor stimulus, which was related to processes such as cell differentiation and regulation of inflammation (Fig. 5a).

Annotation of risk genes in mouse phenotypes. The Mouse Genome Informatics (MGI)⁵² database provides phenotypic information resulting from gene deletions or mutations in mice. These phenotypes were classified into different categories according to their related organ or biological process. To investigate the functional relevance of organ aging genes, we tested whether risk genes were enriched in specific phenotypic categories. As shown in Fig. 5b, we found that adipose aging genes were significantly enriched in adipose tissue phenotypes, suggesting that disruption of these genes in mice is likely to cause abnormalities in adipose tissue. Heart aging genes were enriched in

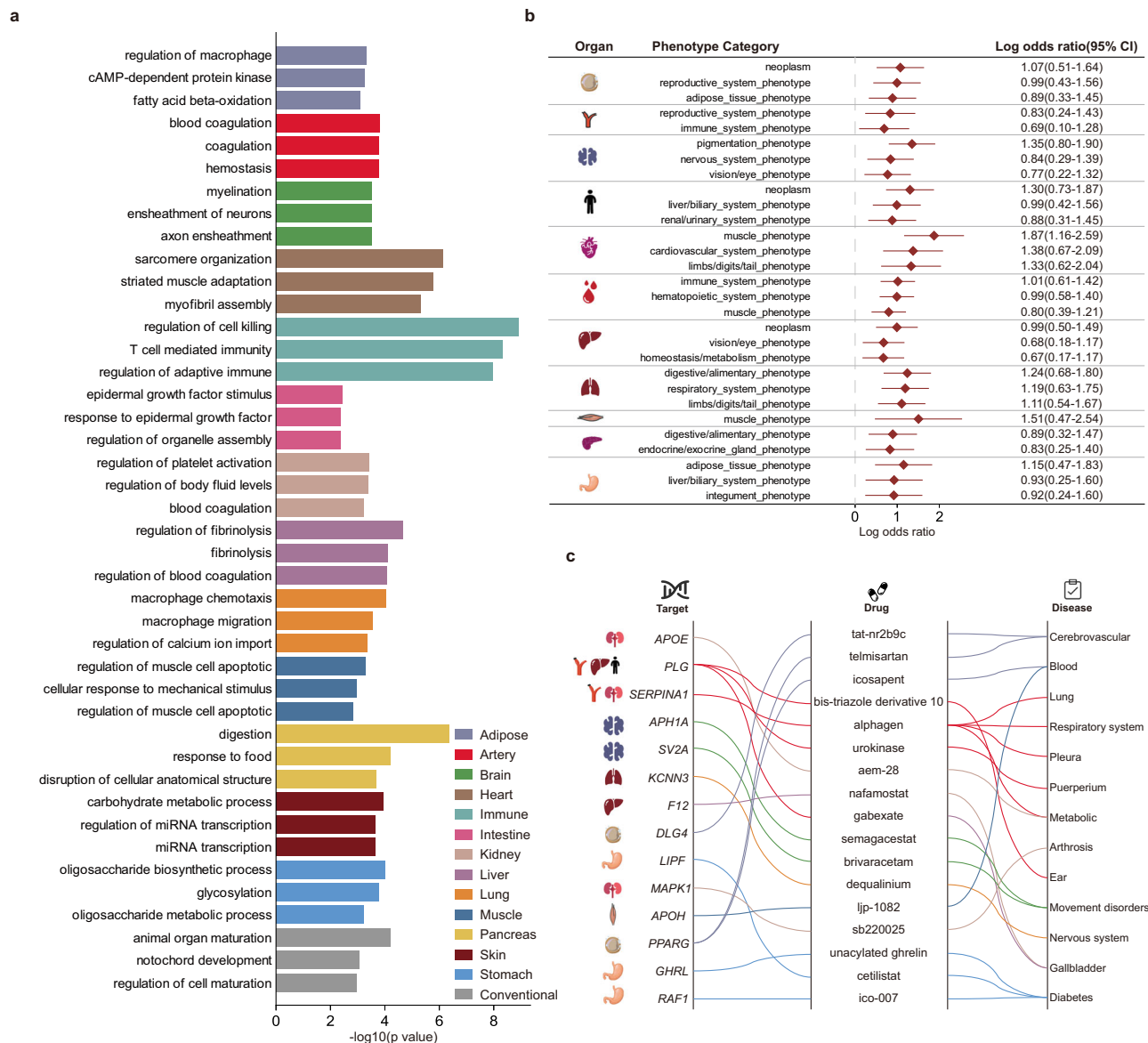


Fig. 5 | Function annotation of organ aging genes. **a** Enrichment of organ aging genes in gene ontology (GO) biological process. Results were generated using the R package clusterProfiler. Enrichment significance was evaluated using a two-sided Fisher's exact test. For each organ, the top 3 most significant results are shown. **b** Enrichment of organ aging genes in genes related to mouse phenotypes. Only

significant results ($p < 0.05$, two-sided Fisher's exact test) are shown. Results are shown as log odds ratios (OR) with 95% confidence intervals (CI). Exact values from Fisher's exact tests are provided in the Source Data file. **c** Gene-drug-diseases network constructed using GREP, only significant results ($p < 0.05$, one-sided Fisher exact test) are shown. Source data are provided as a Source Data file.

muscle and cardiovascular system phenotypes, while immune aging genes showed strong enrichment in immune system and hematopoietic-related phenotypes.

Gene-drug-diseases network. To investigate the potential of organ aging genes as drug targets, we used GREP⁵³ to perform drug target enrichment analysis. GREP reported annotated drugs of input genes and the corresponding enriched disease categories, thereby constructing a gene-drug-disease network of potentially repositionable drugs. As shown in Fig. 5c, two genes related to brain aging (*APH1A*, *SV2A*) were targets of 2 drugs, which were used to treat extrapyramidal and movement disorders (ICD-10 code G20-G26). *PLG* is a risk gene associated with artery, liver, and conventional aging, and was used to develop drugs for treating diseases of the ear, gallbladder, and puerperium, indicating potential for drug repurposing.

Organ aging risk prediction

To investigate whether the genetic organ aging could predict disease status, we performed polygenic risk score (PRS) analysis. PRS was calculated in UKB participants that are not present in UKB-PPP to ensure that there is no overlap in participants between the organ aging summary statistics and the tested individuals. We defined the top 5% of each organ's PRS as high aging risk individuals and the bottom 5% as low aging risk individuals. Subsequently, we used logistic regression to calculate the association of high-risk and low-risk individuals and nine disease categories. In high aging risk individuals, most disease-organ pairs showed a positive correlation (98/126), with 9 remaining significantly positive after multiple testing correction. In the analysis of low-risk individuals, most disease-organ pairs showed a negative correlation (90/126), with 4 remaining significantly negative after multiple testing correction (Fig. 6a). We also

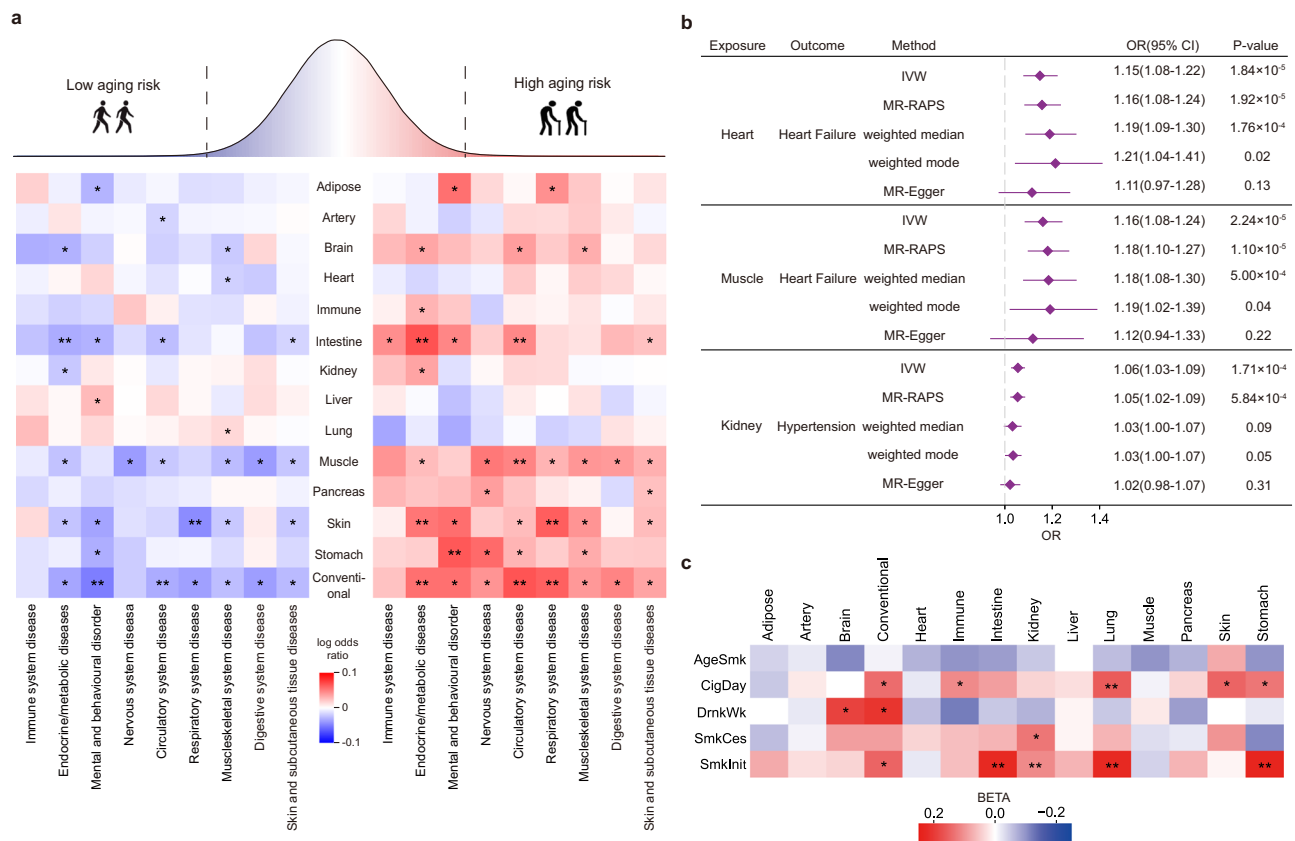


Fig. 6 | Utility of GWAS findings. **a** Logistic regression analysis was used to test for differences in disease prevalence between individuals with high aging risk (top 5% PRS) or low aging risk (bottom 5% PRS) and the remaining individuals. Associations significant after Bonferroni correction are marked with **, while those with $p < 0.05$ are marked with *. **b** Mendelian randomization (MR) estimates using organ aging as exposure and disease status as outcome. Only significantly positive associations after Bonferroni correction from the IVW method are shown. Results are shown as odds ratios (ORs) with 95% confidence intervals (CI). The number of instrumental

variables (n) used for each analysis was as follows: heart aging and heart failure, $n = 33$; muscle aging and heart failure, $n = 28$; kidney aging and hypertension, $n = 38$. IVW inverse-variance weighted. **c** Causal effects of 5 lifestyle factors on organ aging derived from MR analysis. Associations significant after Bonferroni correction are marked with **, while those with $p < 0.05$ are marked with *. AgeSmk: regular smoking, SmkInit: whether an individual had ever smoked regularly, CigDay: cigarettes per day, SmkCes: smoking cessation, DrnkWk: drinks per week. Source data are provided as a Source Data file.

explored the relationship between 617 diseases within 9 disease categories and aging risk. After Bonferroni correction, 62 significant disease-organ pairs were detected, and 85.48% (53/62) of them were positive associations (Supplementary Data 10).

Mendelian randomization (MR)

To investigate the causal effects of organ aging on disease status and explore potential lifestyle interventions for organ aging, we performed two-sample MR⁵⁴ analysis. The association of the IVs and organ aging was derived from the GWAS results in this study.

Causal effects of organ aging on disease status. We utilized two-sample MR⁵⁴ to examine the causal effects of organ aging on 18 common diseases involving multiple organs. The outcome data was obtained from the FinnGen⁵⁵ database (detail in Supplementary Data 11). We identified 31 potential causal effects of organ aging on diseases ($p < 0.05$), with three pairs remaining significant after Bonferroni correction (Fig. 6b). Leave-one-out analyses showed that no single SNP drove the causal estimates (Supplementary Fig. 12). No significant pleiotropy was detected using the MR-PRESSO global test or MR-Egger intercept test. As shown in Fig. 6b, both heart and muscle aging are associated with increased risk of heart failure ($p = 1.84 \times 10^{-5}$ and 2.14×10^{-5} , respectively); kidney aging is positively associated with the risk of hypertension (OR (95% CI) = 1.06 (1.03, 1.09), $p = 1.71 \times 10^{-4}$).

To validate the identified significant causal effect of organ aging on heart failure and hypertension, we re-performed MR analysis using outcome GWAS summary statistics from two additional cohorts. The obesity dataset was released by Shah et al.⁵⁶ and the hypertension dataset was published by Singh et al.⁵⁷. The exposure data was the same, that's, the IVs we used for validation the causal effects were the same specific SNPs in the original MR analysis. The results (Supplementary Fig. 13) are consistent with those from the FinnGen datasets, confirming that both heart and muscle aging were linked to an elevated risk of heart failure ($p = 7.28 \times 10^{-7}$ and 9.62×10^{-5} , respectively), while kidney aging correlated with an increased risk of hypertension ($p = 1.32 \times 10^{-4}$). Leave-one-out analyses showed that no single SNP drove the causal estimates (Supplementary Fig. 14).

Potential lifestyle interventions for organ aging. We also used MR to explore potential lifestyle interventions for organ aging using lifestyles as exposure and organ aging as outcome. The lifestyles include age of initiation of regular smoking (AgeSmk), whether an individual had ever smoked regularly (SmkInit), cigarettes per day (CigDay), smoking cessation (SmkCes) and drinks per week (DrnkWk)⁵⁸. As shown in Figs. 6c, 5 significant associations were identified after multiple testing correction. For example, SmkInit is positively associated with lung aging (β (95% CI) = 0.27 (0.20, 0.34), $p = 2.38 \times 10^{-8}$), intestine aging (β (95% CI) = 0.23 (0.13, 0.33), $p = 1.98 \times 10^{-6}$), kidney aging (β (95% CI) = 0.11 (0.05, 0.16), $p = 9.65 \times 10^{-5}$) and stomach aging (β (95% CI) = 0.26

(0.17, 0.35), $p = 2.41 \times 10^{-7}$). The effect of the smoking phenotype on accelerating lung aging was an expected outcome. *DrnkWk* exhibited a suggestive positive association with brain aging (β (95% CI) = 0.19 (0.03, 0.34), $p = 0.018$) and conventional aging (β (95% CI) = 0.21 (0.06, 0.36), $p = 5.11 \times 10^{-3}$) (Fig. 6c).

Discussion

In this study, we utilized proteomics data from over 50,000 individuals from the UKB-PPP to construct models predicting aging for 13 specific organs and overall whole-body. By leveraging genotype data from the UKBB, we conducted GWAS on these aging phenotypes to uncover the genetic mechanisms. We identified 119 loci associated with organ aging and observed low genetic correlations between different organ aging traits. Through various analytical methods, we discovered 554 genes related to organ aging. Using the GWAS results, we constructed organ aging risk prediction models and applied them to predict organ aging in UKBB individuals who were not included in the UKB-PPP project. Finally, we employed MR to analyze the causal effects of organ aging on disease status and potential lifestyle interventions for organ aging.

Different from a previous study⁸ which using clinical indicators for organ aging, here we used organ aging predicted with human blood plasma proteins originating from specific organs as phenotype. This approach enables us to investigate the aging of organs without suitable clinical chemistry markers (e.g., intestine and stomach). On the other hand, clinical markers often exhibit low organ specificity, making them difficult to interpret. For example, Wen et al.⁸ used BMI as an indicator of skeletal muscle aging, but BMI is closely related to other organs as well. For the organs included in both our and their study, 21 shared risk loci were identified (Supplementary Data 12). For example, both studies identified the locus where the *ABO* gene located was associated with immune and lung aging. However, we observed relatively low genetic correlation between our and their results (Supplementary Fig. 15), indicating that our study successfully provided new insights into the genetic mechanisms of organ aging.

Previous studies have also used plasma proteomics data to track organ aging^{9,19,20}, with several studies^{19,20} also used the plasma proteomics data from UK Biobank. Different from these studies, we further performed GWAS and elucidated the genetic determinants of organ aging. In addition, using the association between IVs and organ aging, we employed MR to assess the causal effects of organ aging on some diseases. For example, we identified heart aging as a risk factor for heart failure and kidney aging as a driver of hypertension risk.

Using multiple analytical methods, we identified 554 genes associated with organ aging. There were 37 genes associated with at least two organ traits. For example, the *PLG* gene is associated with the aging of artery, liver, and conventional tissues. The protein encoded by this gene, plasminogen, is a serine protease, and previous studies have reported that the plasmin/alpha 2-antiplasmin system plays an important role in the repair of damaged liver in mice⁵⁹. Another study reported that genetic variations at the *LPA/PLG* locus are associated with CHD events during statin therapy⁵⁰. Specifically, we found that the *ABO* blood group gene is associated with the aging of five organs, including pancreas, immune system, intestine, artery, and lung. Further analysis confirmed that blood type O was correlated with increased delta age of the artery and intestine. A previous study for the UKB population observed that individuals with blood O was associated with increased odds of hypertension⁶⁰. However, blood O type was also reported as a protective factor for thromboembolic events^{60,61}. Further studies are needed to explore the underlying mechanism of how *ABO* blood type contribute to the development of organ-specific diseases.

As expected, organ aging is generally a risk factor for health outcomes. For example, we observed adverse effects of muscle aging on heart failure. This is consistent with the previously reported that in heart failure with preserved ejection fraction, diastolic dysfunction and abnormal muscle energy metabolism interact with each other⁶².

We also observed higher predicted kidney age increased the risk of hypertension. Consistently, a previous study has reported that hypertension and kidney function are intricately linked, with high blood pressure both causing and resulting from kidney damage⁶³. Using lifestyle as exposure, we also identified potential modifiable factors for organ aging. Specifically, smoking is a risk factor not only for lung aging, but also for intestine, kidney, and stomach aging. Consistently, observational studies have reported the correlation between smoking and gut microbiota⁶⁴, chronic kidney disease⁶⁵, and stomach disease⁶⁶. Therefore, quitting smoking is recommended for preventing organ aging.

Our study has some limitations. The UKB-PPP data we utilized currently stands as the largest proteomic dataset, encompassing over 50,000 samples and 2923 proteins. However, the sample size used for GWAS was less than 30,000, which is relatively small compared to many contemporary studies that often involve millions of samples. Additionally, the proteomic coverage was also relatively low, resulting in fewer organ-specific proteins and limiting the effectiveness of model training. With the advent of larger sample sizes and expanded proteomic coverage, we anticipate that more genetic patterns of organ aging will be revealed.

In summary, using organ aging predicted with human blood plasma proteins originating from specific organs, we performed GWAS for 14 aging traits to elucidate the underlying genetic mechanisms. Our results illustrated organ-specific genetic mechanisms, which might be useful for developing personalized therapeutic interventions for related diseases in future.

Methods

The original and validation cohort

The original cohort of the proteomics data was obtained from the initial recruitment phase of the UK Biobank (measured in 2006–2010, including 2923 proteins measured from 53,021 subjects). Another dataset from the follow-up study of UK Biobank (measured in 2019 and later, including 1463 proteins measured from 1125 subjects) was used to validate the aging-disease associations we observed from the original cohort. To avoid sample overlap, we removed 1085 individuals from the original cohort who had participated in the validation study. Therefore, the sample size for our original study was 51,936.

Organ-enriched proteins

The organ-enriched proteins we used were the combined set of organ-enriched proteins from the Human Protein Atlas (HPA) project²² and the organ-specific proteins identified by Malmström et al.'s study²³. Briefly, in the HPA project, the tissue specificity of a gene was based on mRNA expression levels in the consensus dataset, including the GTEX⁶⁷ datasets and HPA datasets from 1206 cell lines, 40 human tissues and 18 immune cell types. Only proteins with expression levels in a particular tissue/region/cell type at least four times any other tissue/region/cell type were considered as organ-enriched proteins. As for organ-specific proteins from Malmström et al.'s study²³, they developed a mass-spectrometry-based proteomics strategy to infer the tissue origin of proteins detected in human plasma. Only proteins identified as originating solely from a specific organ were considered as organ-enriched proteins for that organ. In the HPA dataset, proteins originating solely from the lymphoid or bone marrow were classified as immune system-enriched proteins. In Malmström et al.'s study²³, proteins originating solely from the spleen or bone marrow were classified as immune system-enriched proteins. Details of the organ-enriched proteins were provided in Supplementary Data 1.

Imputation of plasma proteomic data

In the original study, for each organ, we extracted proteomics data of organ-specific genes and excluded individuals with a missing rate greater than 10%. This means that for organs with fewer than

10 specific genes, we selected individuals without any missing values. Next, we used the KNN algorithm to impute the proteomics data for each organ. Imputation was performed using R package *impute* (v3.3.6)⁶⁸.

Aging model training

We trained models using proteomic data from healthy individuals (defined as having no ICD-10 disease records, $n = 7217$). For each model, only organ-enriched proteins were included as input features, and sex was included as a covariate. We employed nested cross-validation to assess model performance and optimize hyperparameters. The entire dataset was divided into five outer folds, with one fold held out as the test set in each iteration. Within the training data of each outer fold, we used a five-fold inner cross-validation to perform hyperparameter tuning using GridSearchCV function in scikit-learn (v1.4.1)⁶⁹. For each set of hyperparameters, the model was trained on four inner folds and evaluated on the remaining fold, and the best-performing hyperparameters were selected. The complete nested cross-validation procedure is shown in Supplementary Fig. 1. We selected three machine learning algorithms: elastic net, XGBoost, and random forest. The following parameters were tuned for model training:

Elastic net. The alpha parameter, which controls the overall strength of regularization, was tested from 0 to 1 in increments of 0.01. Similarly, the $L1$ ratio, which determines the balance between $L1$ and $L2$ penalties, was also varied from 0 to 1 in steps of 0.01.

XGBoost. The learning rate was tested at values of 0.01, 0.1, and 0.3, while the number of estimators ranged from 100 to 500 in steps of 100.

Random forest. The number of estimators ranged from 100 to 500 in steps of 100.

Model development and evaluations were implemented through *xgboost* (v2.1.1) and *scikit-learn* (v1.4.1) under Python (v3.10.10).

We used the data from 1085 individuals who participate the follow-up study of the UKBB (measured 2019+) as the independent test dataset. To avoid sample overlap, individuals from the test dataset were excluded from the original cohort at the very beginning. We removed individuals with diagnoses from the nine major disease categories (details in Supplementary Data 4), resulting in an independent test dataset of 622 participants. The sex and age distributions of the test dataset were comparable to those observed in the training dataset (Supplementary Fig. 16).

Delta age calculation

Previous studies^{24,25} have identified age bias as a common problem in aging prediction models, where the predicted age difference (PAD)—the difference between predicted age and chronological age—tends to be negatively correlated with chronological age. Post-prediction correction of this bias is crucial²⁶. We adopt the method used by Oh et al.⁹ to address this problem. We calculated the corrected PAD (delta age) as the residuals from a locally weighted scatterplot smoothing (lowess) regression of predicted age on chronological age model. Here we used lowess regression in place of a simple linear regression because extensive evidence showed that the plasma proteome changes non-linearly with age⁷⁰. Lowess regression was implemented using *statsmodels* (v0.14.5) under Python (v3.10.10).

Estimation of the correlation between delta age and disease status or chronological age

Disease records were classified into two categories: prevalent diseases (diagnosed prior to recruitment) and incident diseases (diagnosed after recruitment). Logistic regression was used to estimate the

correlation of delta age and prevalent disease status, with age, sex and PCI-5 as covariates. Cox proportional hazards models were used to assess the association between organ age gaps and incident disease status, models were tested using the *lifelines* python package⁷¹ with age, sex and PCI-5 as covariates. Individuals with the disease were coded as 1 and those without the disease as 0. We used linear regression to estimate the correlation of delta age and chronological age, with sex and PCI-5 as covariates. Logistic regression was implemented using *statsmodels* (v0.14.5), Cox proportional hazards regression was implemented using *lifelines* (v0.30.0), and the Python version used was v3.10.10.

Validation of aging-diseases association in other datasets

We chose the proteomics data from the follow-up study of the UKBB (measured 2019+) as the validation dataset. This dataset included 1463 plasma proteins from 1125 participants. To make sure that the validation models preserved most features used in the original models, we only kept organ models with more than 50% of the protein features available in the validation dataset. Using these features, we re-trained the organ aging models in the original cohort. These feature-reduced models were then used for organ aging prediction in the validation datasets, and aging-disease association were also calculated.

GWAS

We performed GWAS of 14 aging traits after removing individuals with Z-scored delta age above mean $+3*SD$ or below mean $-3*SD$. Analyses were restricted to variants with minor allele frequency >0.01 , genotype hard call rate >0.98 , Hardy Weinberg $p > 1 \times 10^{-7}$, and present in the Haplotype Reference Consortium (HRC) panel⁷². Participants were excluded from analysis if they met any of the following criteria: (1) mismatch between self-reported sex and genetic sex, (2) genotyping call rate <0.95 , (3) were outliers for heterozygosity (deviate more than 3 standard deviations from the heterozygosity rate mean), or (4) related individuals (only 1 individual in pair with a *pihat* greater than 0.044 was remained). Finally, 6,735,132 autosomal variants and 33,243 samples in UKB-PPP were remained. GWAS was performed using a mixed linear model implemented in BOLT-LMM (v2.4.1)⁷³, adjusting for age, sex, age-squared, age-sex interaction, age-squared-sex interaction, the top 40 genetic PCs, blood pressure, and BMI.

We divided the UKBB dataset into two groups, southern and northern, based on the enrollment location of the UK Biobank assessment center (Field ID 54). These assessment centers are distributed across England, Scotland, and Wales. According to the standards of the UK Office for National Statistics (ONS)⁷⁴, England was further divided into Southern England, Northern England, and the Midlands. Based on the geographical locations of these regions, individuals from Southern England, the Midlands, and Wales were classified as the Southern cohort ($n = 26,293$), while those from Northern England and Scotland were categorized as the Northern cohort ($n = 26,477$). Meta-analyses for GWAS results obtained from discovery and replication samples was performed using METAL⁷⁵.

The SNP heritability and genetic correlation were calculated using LDSC (v1.0.1)⁷⁶. The genomic coordinate is according to Genome Reference Consortium human build 37 (GRCh37) in all analyses.

Fine-mapping SNPs

First, SNPs in significant were defined as being conditionally independent using GCTA-COJO (v1.95.0)²⁸. Subsequently, we utilized FINEMAP (v1.4.2)²⁹ software to calculate posterior probabilities for all SNPs within 500 kb of each conditionally independent SNP.

Functional annotation of the fine-mapped SNPs

We annotated the function of fine-mapped SNPs using ClinVar³⁰, CADD³¹, DANN³², FATHMM-MKL⁷⁷, RegulomeDB³³, regSNPs-intron³⁴, and SIFT/PROVEAN^{35,36}. SNPs supported by at least one database

annotation were considered as functional SNPs. Detail annotation criteria are as follows: 1) SNPs annotated as 'pathogenic' or 'likely pathogenic' in ClinVar³⁰ were considered to be having clinical significance; 2) SNPs with score ≥ 12.37 in CADD³¹, or DANN³² score ≥ 0.9 , or FATHMM-MKL⁷⁷ score ≥ 0.7 were considered as deleterious to gene function; 3) SNPs with RegulomeDB³³ rank ≤ 2 were considered as affecting regulome binding; 4) SNPs annotated as "damaging" in regSNPs-intron³⁴ were considered as affecting splicing; 5) SNPs with SIFT³⁹ score < 0.05 or PROVEAN⁴⁰ score < -2.5 were considered as damaging exonic variants.

Gene prioritization

The eQTL data for TWAS, COLOC, and SMR analyses were obtained from GTEx V8⁶⁷. For the 13 types of organs, only the data from the corresponding tissues were utilized, and for conventional aging, all tissues were used. Then, the best results across tissues were selected. For TWAS and SMR, we retained genes that remained significant after Bonferroni correction. For COLOC, we retained genes with posterior probability of hypothesis 4 (PPH4) greater than 0.9. TWAS was performed based on imputed cis-regulated gene expression using FUSION with default settings^{46,78,79}. Pre-computed gene expression weights of 49 tissues from GTEx v8 were used as downloaded from the FUSION website (<http://gusevlab.org/projects/fusion/>). The Bonferroni correction threshold was defined as 0.05 divided by the total number of scored genes in that organ. COLOC was performed using R package coloc (v5.1.0.1)⁴⁵ with default settings. SMR was performed using SMR (v1.4.0)⁴⁶ with default settings. Genes containing fine-mapped SNPs with functional annotation support (as described in the functional annotation of the fine-mapped SNPs part) were also included in the prioritized genes. SNPs were assigned to their corresponding genes using ANNOVAR⁸⁰.

Comparison of delta age across different blood types

We firstly calculated the residuals of delta age after regressing it on age, sex, age-squared, age-sex interaction, age-squared-sex interaction, the top 40 genetic principal components (PCs), blood pressure, and BMI. The residuals represent the unexplained variation in delta age after accounting for these potential confounders. We then compared the residual delta age values across different blood type groups. The statistical significance of between-group differences was estimated using a two-sided rank-sum test.

Pathway enrichment

Pathway enrichment was performed using R package *clusterProfiler* (v4.16.0)⁸¹, where gene symbols were converted to Entrez IDs before conducting the enrichment analysis.

Enrichment of prioritized genes in mouse phenotype categories from MGI

MGI⁵² is an international database resource for the laboratory mouse, providing phenotypic information derived from gene deletions or mutations in mice. These phenotypes are categorized by their associated organ or biological process. We performed Fisher's exact test to test whether the organ aging genes were enriched in specific categories.

Gene-drug-diseases network

Drug target enrichment analysis was performed using GREP (v1.0.0)⁵³. For each organ, the input was risk gene list, GREP reported the drugs annotated to these risk genes and the types of diseases they are enriched in.

PRS

We used PRS-CS (v1.1.0)⁸² to build aging risk prediction models. PRS-CS is a Python-based command-line tool that uses GWAS summary

statistics and an external LD reference panel to infer the posterior effects of SNP impact levels under continuous shrinkage (CS) prior conditions. In the model, we included 1,161,863 SNPs from Hapmap3. For the reference LD panel, we used data constructed from the UK Biobank (https://www.dropbox.com/s/t9opx2ty6ucrpib/ldblk_ukbb_eur.tar.gz?dl=0). We randomly selected 50,000 unrelated samples from the UKBB dataset as the target dataset. For other parameters, we opted for the default values. The output of PRS-CS contained the chromosome, rs ID, base position, A1, A2, and posterior effect size estimate for each SNP. Individual-level polygenic scores were produced by concatenating output files from all chromosomes and then using PLINK's (v1.9.0) --score command (<https://www.cog-genomics.org/plink/1.9/score>)⁸³.

MR

We collected GWAS summary statistics of 18 diseases from the FinnGen⁵⁵ database, and GWAS summary statistics of 5 smoking and drinking phenotypes from a study based on non-UKBB populations⁸⁴. The association of the IVs and organ aging was derived from the GWAS results in this study.

Instrument Variant (IV) selection. We used the clumping algorithm in PLINK (v1.9.0) (<https://www.cog-genomics.org/plink/1.9/>)⁸³ to select independent SNPs for each exposure (r^2 threshold = 0.001, window size = 1 Mb and $p < 1 \times 10^{-5}$). The 1000 G European data (phase 3) were used as the reference for LD estimation⁸⁵. Secondly, if one selected SNP in the first step was not present in outcome data, we used its proxy SNP with $r^2 > 0.8$ instead. The proxy SNP must also be associated with exposure ($p < 1 \times 10^{-5}$). If more than one proxy SNPs were available, we chose the one with the maximum r^2 and minimum p value associated with the exposure.

IV quality control. For the IVs, three key assumptions must hold: 1) the selected IVs must be associated with the exposure (relevance assumption); 2) the selected IVs are not associated with potential confounders (independence assumption) and 3) the IVs affect the outcome only through their effect on the exposure (exclusion restriction assumption). We used RadialMR package (v0.1.0)⁸⁶ under R (v4.3.1) to remove pleiotropic SNPs. The remaining SNPs were used to perform MR analysis. We used MR Steiger filtering to check whether the MR analysis estimates assessed the true causal direction⁸⁷.

Bidirectional two-sample MR analyses. We selected IVW regression with multiplicative random effects model as the primary method for causal inference⁸⁸. That's, the evidence of statistically significant causal effect was based on the IVW p -value. To complement and enhance the reliability of the results, we conducted four other MR methods. MR-Egger method estimates the causal effect through the slope coefficient of the Egger regression, which provides a more robust estimate even if none of the IVs are invalid⁸⁹. Weighted median method can even protect against up to 50% of invalid IVs⁹⁰. Weighted mode method provides consistent estimates when the relaxed IV assumption has less bias and a lower type-I error rate⁹¹. MR-RAPS accounts for systematic and idiosyncratic pleiotropy and can provide a robust inference for MR analysis with many weak instruments⁹². All these methods were implemented in the TwoSampleMR⁹³ package (v0.6.8) under R (v4.3.1).

Sensitivity analysis. For the significant MR results after Bonferroni correction, we further performed sensitivity analysis. First, we performed leave-one-out analysis to check whether the causal association was obviously driven by a single SNP. Second, we conducted MR-PRESSO to detect the presence of horizontal pleiotropy ($p < 0.05$)⁹⁴. Third, we executed MR-Egger regression to examine the potential bias of directional pleiotropy⁸⁹. The intercept in the Egger regression indicates the mean pleiotropic effect of all genetic variants, which is

interpreted as evidence of directional pleiotropy when the value differs from zero ($p < 0.05$). MR-PRESSO was implemented in the MRPRESSO package (v1.0) under R (v4.3.1). MR-Egger was implemented in the TwoSampleMR⁹³ package (v0.6.8) under R (v4.3.1).

Datasets used for the validation of MR results

To validate the identified significant causal effect of organ aging on heart failure and hypertension, we obtained outcome GWAS summary data from two other datasets to re-perform MR analysis. The heart failure data including 583,167 non-UKBB European individuals was from the study by Shah et al.⁵⁶, and the hypertension data including 10,772 African individuals was from the study by Singh et al.⁵⁷. The summary data for exposure were the same as those used in the primary analysis. That's, The IVs we used for validation the causal effects were the same specific SNPs in the original MR analysis.

Statistics & reproducibility

No statistical method was used to predetermine sample size. During quality control of proteomic and genomic data, individuals not meeting predefined quality criteria were excluded from the analyses. The nested cross-validation procedure used in model training involved random splitting of data; the randomness was controlled by setting a fixed random seed to ensure reproducibility.

Ethics

The use of UK Biobank (UKBB) data was approved by the UKBB according to its established access procedures. The UKBB has ethical approval from the North West Multi-centre Research Ethics Committee as a Research Tissue Bank.

Reporting summary

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

Data availability

The GWAS summary data generated in this study have been deposited in the Zendon database under accession code [16784920](https://zenodo.org/record/16784920)⁹⁵. The data from UK Biobank used in this study are available through a procedure described at: <https://www.ukbiobank.ac.uk/enable-your-research>. The summary data of 18 complex diseases were downloaded from FinnGen⁵⁴ database. Access to the FinnGen GWAS summary statistics can be obtained through registration at the following website: <https://elomake.helsinki.fi/lomakkeet/124935/lomake.html>. The summary data on smoking and drinking habits of non-UKBB populations was obtained from: <https://conservancy.umn.edu/items/91f6a003-6af2-4809-9785-53dc579dc788>. Source data are provided with this paper.

Code availability

The code for model training is available at GitHub (https://github.com/zrj1236/Organ_Aging_Model) and archived at Zenodo (<https://doi.org/10.5281/zenodo.17322149>)⁹⁶.

References

- Ferrucci, L. et al. Measuring biological aging in humans: a quest. *Aging cell* **19**, e13080 (2020).
- Guo, J. et al. Aging and aging-related diseases: from molecular mechanisms to interventions and treatments. *Sig. Transduct. Target. Ther.* **7**, 391 (2022).
- Gao, X. et al. Accelerated biological aging and risk of depression and anxiety: evidence from 424,299 UK Biobank participants. *Nat. Commun.* **14**, 2277 (2023).
- Zenin, A. et al. Identification of 12 genetic loci associated with human healthspan. *Commun. Biol.* **2**, 41 (2019).
- Joshi, P. K. et al. Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. *Nat. Commun.* **8**, 910 (2017).
- Timmers, P. R. et al. Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *Elife* **8**, e39856 (2019).
- Deelen, J. et al. A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat. Commun.* **10**, 3669 (2019).
- Wen, J. et al. The genetic architecture of biological age in nine human organ systems. *Nat. Aging* **4**, 1290–1307 (2024).
- Oh, H. S.-H. et al. Organ aging signatures in the plasma proteome track health and disease. *Nature* **624**, 164–172 (2023).
- Lonardo, A., Ballestri, S., Baffy, G. & Weiskirchen, R. Liver fibrosis as a barometer of systemic health by gauging the risk of extrahepatic disease. *Metab. Target Organ Damage* **4**, 41 (2024).
- Cole, J. H. et al. Brain age predicts mortality. *Mol. Psychiatry* **23**, 1385–1392 (2018).
- Bashyam, V. M. et al. MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. *Brain* **143**, 2312–2324 (2020).
- Kwon, D. & Belsky, D. W. A toolkit for quantification of biological age from blood chemistry and organ function test data: BioAge. *Geroscience* **43**, 2795–2808 (2021).
- Tian, Y. E. et al. Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality. *Nat. Med.* **29**, 1221–1231 (2023).
- Zheng, J. et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* **52**, 1122–1131 (2020).
- Moya-Garcia, A. et al. Structural and functional view of polypharmacology. *Sci. Rep.* **7**, 10102 (2017).
- Papier, K. et al. Identifying proteomic risk factors for cancer using prospective and exome analyses of 1463 circulating proteins and risk of 19 cancers in the UK Biobank. *Nat. Commun.* **15**, 4010 (2024).
- Carrasco-Zanini, J. et al. Proteomic signatures improve risk prediction for common and rare diseases. *Nat. Med.* **30**, 2489–2798 (2024).
- Goeminne, L. J. E. et al. Plasma protein-based organ-specific aging and mortality models unveil diseases as accelerated aging of organismal systems. *Cell Metab.* **37**, 205–222.e6 (2025).
- Oh, H. S. H. et al. Plasma proteomics links brain and immune system aging with healthspan and longevity. *Nat. Med.* **31**, 2703–2711 (2025).
- Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
- Uhlén, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
- Malmström, E. et al. Human proteome distribution atlas for tissue-specific plasma proteome dynamics. *Cell* **188**, 2810–2822.e16 (2025).
- de Lange, A. G. & Cole, J. H. Commentary: Correction procedures in brain-age prediction. *Neuroimage Clin.* **26**, 102229 (2020).
- Zhang, B., Zhang, S., Feng, J. & Zhang, S. Age-level bias correction in brain age prediction. *Neuroimage Clin.* **37**, 103319 (2023).
- Smith, S. M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T. E. & Miller, K. L. Estimation of brain age delta from brain imaging. *Neuroimage* **200**, 528–539 (2019).
- Levinthal, D. J. & Strick, P. L. Multiple areas of the cerebral cortex influence the stomach. *Proc. Natl. Acad. Sci.* **117**, 13078–13083 (2020).
- Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).

29. Benner, C. et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).
30. Landrum, M. J. et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
31. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
32. Quang, D., Chen, Y. & Xie, X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* **31**, 761–763 (2014).
33. Dong, S. et al. Annotating and prioritizing human non-coding variants with RegulomeDB v. 2. *Nat. Genet.* **55**, 724–726 (2023).
34. Lin, H. et al. RegSNPs-intron: a computational framework for predicting pathogenic impact of intronic single nucleotide variants. *Genome Biol.* **20**, 1–16 (2019).
35. Ng, P. C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
36. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745–2747 (2015).
37. Sollis, E. et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* **51**, D977–d985 (2023).
38. Shrine, N. et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat. Genet.* **51**, 481–493 (2019).
39. A first update on mapping the human genetic architecture of COVID-19. *Nature* **608**, E1–e10 (2022).
40. Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).
41. Yengo, L. et al. A saturated map of common genetic variants associated with human height. *Nature* **610**, 704–712 (2022).
42. Goes, F. S. et al. Genome-wide association study of schizophrenia in Ashkenazi Jews. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **168**, 649–659 (2015).
43. Emilsson, V. et al. Co-regulatory networks of human serum proteins link genetics to disease. *Science* **361**, 769–773 (2018).
44. Sakaue, S. et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).
45. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
46. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
47. Rühlemann, M. C. et al. Genome-wide association study in 8,956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. *Nat. Genet.* **53**, 147–155 (2021).
48. Ando, Y. et al. Using tumor marker gene variants to improve the diagnostic accuracy of DUPAN-2 and carbohydrate antigen 19-9 for pancreatic cancer. *J. Clin. Oncol.* **42**, Jco2301573 (2024).
49. Lei, C. et al. Enteric VIP-producing neurons maintain gut microbiota homeostasis through regulating epithelium fucosylation. *Cell Host Microbe* **30**, 1417–1434.e8 (2022).
50. Wei, W. Q. et al. LPA variants are associated with residual cardiovascular risk in patients receiving statins. *Circulation* **138**, 1839–1849 (2018).
51. Okada, K., Ueshima, S., Imano, M., Kataoka, K. & Matsuo, O. The regulation of liver regeneration by the plasmin/alpha 2-antiplasmin system. *J. Hepatol.* **40**, 110–116 (2004).
52. Baldarelli, R. M., Smith, C. L., Ringwald, M., Richardson, J. E. & Bult, C. J. Mouse Genome Informatics: an integrated knowledgebase system for the laboratory mouse. *Genetics* **227**, iyae031 (2024).
53. Sakaue, S. & Okada, Y. GREP: genome for REPositioning drugs. *Bioinformatics* **35**, 3821–3823 (2019).
54. Pierce, B. L. & Burgess, S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am. J. Epidemiol.* **178**, 1177–1184 (2013).
55. Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
56. Shah, S. et al. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163 (2020).
57. Singh, S. et al. Genome-wide association study meta-analysis of blood pressure traits and hypertension in sub-Saharan African populations: an AWI-Gen study. *Nat. Commun.* **14**, 8376 (2023).
58. Saunders, G. R. B. et al. Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature* **612**, 720–724 (2022).
59. Okada, K., Ueshima, S., Imano, M., Kataoka, K. & Matsuo, O. The regulation of liver regeneration by the plasmin/alpha2-antiplasmin system. *J. Hepatol.* **40**, 110–116 (2004).
60. Groot, H. E. et al. Genetically determined ABO blood group and its associations with health and disease. *Arterioscler. Thrombosis Vasc. Biol.* **40**, 830–838 (2020).
61. Wiggins, K. L. et al. ABO genotype and risk of thrombotic events and hemorrhagic stroke. *J. Thromb. Haemost.* **7**, 263–269 (2009).
62. Hamo, C. E. et al. Heart failure with preserved ejection fraction. *Nat. Rev. Dis. Prim.* **10**, 55 (2024).
63. Zhou, B., Perel, P., Mensah, G. A. & Ezzati, M. Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. *Nat. Rev. Cardiol.* **18**, 785–802 (2021).
64. Gui, X., Yang, Z. & Li, M. D. Effect of cigarette smoke on gut microbiota: state of knowledge. *Front. Physiol.* **12**, 673341 (2021).
65. Lee, S. et al. Smoking, smoking cessation, and progression of chronic kidney disease: results from KNOW-CKD study. *Nicotine Tob. Res.* **23**, 92–98 (2021).
66. Jia, X., Sheng, C., Han, X., Li, M. & Wang, K. Global burden of stomach cancer attributable to smoking from 1990 to 2019 and predictions to 2044. *Public Health* **226**, 182–189 (2024).
67. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
68. Hastie, T., Tibshirani, R., Narasimhan, B. & Chu, G. *Impute: Impute: Imputation for Microarray Data* (R package version, 2020).
69. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
70. Schaum, N. et al. Ageing hallmarks exhibit organ-specific temporal signatures. *Nature* **583**, 596–602 (2020).
71. Davidson-Pilon, C. lifelines: survival analysis in Python. *J. Open Source Softw.* **4**, 1317 (2019).
72. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
73. Loh, P.-R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
74. Nash, A. *Subnational Population Projections for England: 2018-Based* (Office for National Statistics, 2020).
75. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
76. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
77. Shihab, H. A. et al. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* **31**, 1536–1543 (2015).
78. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).

79. Gamazon, E. R. et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).
80. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
81. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
82. Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
83. 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).
84. Liu, M. et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **51**, 237–244 (2019).
85. Fairley, S., Lowy-Gallego, E., Perry, E. & Flicek, P. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. *Nucleic Acids Res.* **48**, D941–d947 (2020).
86. Bowden, J. et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. *Int. J. Epidemiol.* **47**, 1264–1278 (2018).
87. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081 (2017).
88. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
89. 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).
90. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
91. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).
92. Zhao, Q., Wang, J., Hemani, G., Bowden, J. & Small, D. S. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Ann. Stat.* **48**, 1742–1769 (2020).
93. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *elife* **7**, e34408 (2018).
94. Verbanck, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
95. Zhu, R. Revealing the genetic architectures underlying organ-specific aging based on proteomic data. *Zenodo* <https://doi.org/10.5281/zenodo.16784920> (2025).
96. Zhu, R. Revealing the genetic architectures underlying organ-specific aging based on proteomic data. *Zenodo* <https://doi.org/10.5281/zenodo.17322149> (2025).

Acknowledgements

We thank the participants and the organizers of UK Biobank resource under application number 46387. We want to acknowledge the participants and investigators of the FinnGen study. This study is supported by grants from the Science and Technology Projects of Xizang Autonomous Region (XZ202502ZY0036, T.L.Y.), Science and Technology Projects of Xizang Autonomous Region (XZ202501YD0013, W.P.), Science Fund for Distinguished Young Scholars of Shaanxi Province (2025JC-JCQN-054, Y.G.), National Natural Science Foundation of China (32470639 T.L.Y., 32370653 Y.G., and 82372458 S.S.D.), Innovation Capability Support Program of Shaanxi Province (2022TD-44, T.L.Y.), and the Fundamental Research Funds for the Central Universities. This study was also supported by the High-Performance Computing Platform and Instrument Analysis Center of Xi'an Jiaotong University.

Author contributions

R.J.Z. performed data analyses and wrote the manuscript. K.Y., J. H.W., X.L.Y. and W.S. created the visualizations. S.S.D., Y.G., W.P., and W.C. revised the manuscript. T.L.Y., S.S.D. and Y.G. designed, coordinated, and supervised the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-67223-4>.

Correspondence and requests for materials should be addressed to Shan-Shan Dong or Tie-Lin Yang.

Peer review information *Nature Communications* thanks Rudong Li, Lei Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025