Imaging-based organ-specific aging clock predicts human diseases and mortality

Supplementary Methods

Supplementary Figures

Supplementary Methods

Participants

The UK biobank is one of the largest population-based biomedical cohorts that consists of more than 500,000 participants aged 40 to 96 across the UK, who provide de-identified genomic, whole-body imaging, health records, body fluid biomarker and physical measurements (https://biobank.ndph.ox.ac.uk/showcase/)[1,2]. All individuals undergoing retinal OCT at baseline or participating in the imaging visit for brain MRI, heart MRI, abdominal MRI, and whole-body dual-energy x-ray absorptiometry (DXA) scanning were initially included in the present study. The UK Biobank was approved by the North West Multi-centre Research Ethics Committee, and written informed consent was obtained from all participants.

Multi-organ imaging phenotypes

The chronological age prediction for seven organs were constructed using 1,777 IDPs of the corresponding organs, including the brain, heart, body composition, kidney, liver, pancreas, and eye. Given the obvious functional difference, brain gray matter and white matter were assessed separately. Only structural and functional IDPs of organs were selected and sourced from UKB (Supplementary Table 1 and Supplementary Fig 1). A detailed description of imaging processing, artifact correction, image registration, quality control, and phenotype extraction of the organs is available on the UKB official website[3].

Specifically, IDPs of the brain, heart, and abdominal organs were derived from multisequence MRI. For the brain, the T1 weighted MRI was utilized to quantify 1,010 IDPs related to regional grey matter volume, cortical mean thickness, and surface area. The T1-weighted and diffusion-weighted MRI generated 443 measurements of white matter volume and microstructure of major white matter tracts. The microstructure measurements were characterized using diffusion tensor imaging (DTI) fitting and Neurite Orientation Dispersion and Density Imaging (NODDI) modelling[4], including fractional anisotropy (FA), mean diffusivity (MD), tensor mode (MO), L1, L2, L3, ICVF, isotropic or free water volume fraction (ISOVF) and orientation dispersion index (OD). For the heart, a set of cardiovascular MRI sequences was leveraged to assess 106 IDPs related to cardiac and aortic structure and function[5], including cardiovascular morphology, ventricular function, and myocardial perfusion. The anatomical structures primarily included

in these measurements were ascending aorta, descending aorta, left and right ventricle, and left and right atrium. Besides, the abdominal MRI of multiple sequences was applied to generate volume, fat, and iron measurements in the liver, kidney, pancreas, lung, spleen, and adipose tissue[6,7]. In contrast, the lung and spleen were excluded due to the lack of fat and iron data (biobank.ndph.ox.ac.uk/showcase/ukb/docs/body_mri_explan.pdf).

Meanwhile, 138 IDPs of body composition were measured from the DXA instrument (biobank.ndph.ox.ac.uk/ukb/ukb/docs/DXA_explan_doc.pdf), with the standard operating procedure and real-time quality control applied. The measurements of bone area, bone mineral content, bone mineral density, lean mass, and fat mass were directly derived.

Finally, 64 IDPs of the eye were derived from retinal OCT using TOPCON 3D OCT 1000 Mk2 (Topcon Corporation, Tokyo, Japan)[8], which provides information about the thickness of retinal layers and subfields, along with the vertical cup-to-disc ratio and disc diameter.

For each organ, the participants with missing data for any organ-related IDPs were excluded, resulting in a slightly different number of healthy participants eligible for further age modelling (Supplementary Table 2). No obvious difference in age and sex distribution was observed among different organs in our cohort, except that optical coherence tomography scanning for the eye at baseline had a younger age than the other organs scanned at the imaging visit.

Normative organ aging models

LASSO models were trained in healthy participants to predict the chronological age of participants, with imaging-derived phenotypes of brain GM, brain WM, heart, body composition, kidney, liver, pancreas, and eye as potential predictors (Supplementary Table 1). The healthy status was defined as having no lifetime occurrence record in linked primary care, hospital inpatient, death register and self-reported medical condition sources (Category 1712). LASSO regression can perform automatically select features, reduce model complexity, and prevent overfitting by applying L1 regularization. For each organ, the LASSO predictive model was constructed using a nested 10-fold cross-validation. The inner loop includes a cross-validation in selecting the best hyperparameter lambda by evaluating the model's performance on lambda range from 10^{-8} to 10^{5}, while the outer loop assessed the model's generalization performance by training the model on nine of the folds with the optimal lambda and tested on the one remaining fold. This cross-

validation procedure resulted in a predicted age for each healthy individual. In each of the iterations, the effect of the scanning site and total intracranial volume were regressed from the IDPs before model training. The residuals were centered and scaled in the training data and the same covariate regression and normalization parameters were applied to the test data. Following the suggestion, chronological age was also regressed from the predicted age in the training data, and the same regression model was applied to test data[9,10]. Model performance was evaluated by calculating Pearson's correlation coefficient and mean absolute error between predicted age and chronological age in the test sets. In this way, we avoided the potential data leakage to the greatest extent caused by involving information from test data sets in training a model.

Finally, eight optimal organ-specific biological age prediction models were created in healthy participants, and the models were applied to the remaining UKB participants to obtain the predicted age of each organ. Note that the same covariate regression, normalization parameters and the age bias correction derived from the training data of healthy individuals were applied to the remaining UKB participants before the prediction. The difference between the predicted age of each organ and chronological age was regarded as the organ-specific age gap, informing whether the organ exhibited an accelerated or decelerated aging process compared to healthy peers. Each individual was thus characterized by eight organ-specific age gaps, which were used for further analyses. In sensitivity analysis, separate models were trained for healthy males and females to examine the potential sex difference in age prediction, with the same model training strategy employed.

Structural equation modeling

SEM was employed to assess the association between the age gap of one organ and that of another. Given that the eye was scanned at a different visit compared to other organs, the SEM was limited to six organs except the eye. In determining the optimal model structure, the fast-greedy equivalence search (FGES) heuristic algorithm from Tetrad v7.6.5[11] was used to determine the causal Bayesian network with the highest score. FGES is a Bayesian heuristic algorithm that begins by adding edges step by step until no further addition improves the Bayesian Information Criterion. It then removes edges incrementally until no further removal increases the Bayesian Information Criterion[12]. The FGES heuristic was repeated for 500 bootstrapped samples, and consensus edges in 50% of the samples resulted in a final network structure.

Although the FGES algorithm assumes that the underlying causal structure can be represented as a directed acyclic graph (DAG), which may not hold due to the lack of temporal directionality in this study. FGES can still be useful in exploratory analyses to generate hypotheses about network structures of cross-sectional data. With the determined network structure, the pathway analysis was performed with R package lavaan v0.6.17 to estimate the coefficient and significance for edges in the final network. In sensitivity analysis, age at the imaging visit, sex, and the imaging site were further regressed from all organ-specific age gaps, and a search algorithm was performed on the resulting residuals. The characterization of how one organ's baseline age gap influences the follow-up age gap or rate of aging of other organs was not applicable due to limited participants with repeated imaging visits.

Phenome-wise association analysis

Taking similar UKB fields of modifiable factors as previous research[13], 344 and 224 factors were available at the baseline visit and the imaging visit for the present study, respectively. After excluding multinomial phenotypes, 143 common factors were obtained for organs except the eye, and 139 common factors were obtained for all organs. The factors can be divided into six broad categories (Supplementary Table 16). (1) Sociodemographic factors: education, employment and household economic status information. (2) Lifestyle factors: physical activity, sleeping, smoking, alcohol intake, diet, electronic device usage, and sexual behaviors. (3) Early-life and family history factors: birth weight, body weight at the child, adoption, and the health condition of parents. (4) Mental health: self-reported symptoms related to psychiatric conditions, such as anxiety, depression, addiction, and trauma. (5) Physical measures: anthropometry, blood pressure, arterial stiffness, hand-grip strength, and self-reported medical history. (6) Cognitive functions: fluid intelligence/reasoning, numeric memory, prospective memory, reaction time, matrix paired associate learning, pattern completion, symbol digit substitution, tower rearranging, pairs matching, and trail making. The data were averaged across rounds for tests with multiple rounds, such as reaction time, pairs matching, symbol digit substitution, matrix pattern completion and trail making.

PheWAS associations were examined with the PHESANT package in R[14], which allows automated phenome scanning for different types of factors. In this study, the association between organ-specific age gap and factors was examined for each organ, with age and sex as covariates. Note that the factors and age at the baseline were used for the eye, and factors and age at the imaging visit were used for other organs.

The significance was determined by FDR-corrected $P < 0.05$ across 143×7 factor-organ pairs. The linear and logistic regression were performed according to the type of factors. In sensitivity analysis, we performed PheWAS with the 139 modifiable factors available at both the baseline and imaging visit for all eight organs, with significance determined by FDR-corrected $P < 0.05$ across 139×8 factor-organ pairs.

Survival analysis of disease and mortality risk

A total of 13 categories of incident outcomes (Supplementary Table 3) were selected from UKB, including 12 broad categories of diseases (36 specific diseases) and all-cause mortality (four cause-specific mortality). In UKB, incident diagnoses for the selected diseases and mortality were confirmed using linked hospital inpatient, primary care, and death register data[15]. For organs except the eye, the follow-up of participants began from the date of their third visit, i.e., imaging visits, and the end of follow-up was defined as the date when a disease diagnosis occurs or when a participant is censored due to death or loss to follow-up, whichever occurred first. The endpoints defined as the diagnosis obtained from self-reported records or incidents that occurred before the date of the imaging visit of the corresponding disease were excluded. Meanwhile, for the eye, the follow-up began from the date of their first visit to the assessment centers, i.e., baseline visit, and the end of follow-up was defined using the same criteria as other organs. The endpoints that defined as the diagnosis obtained from self-reported records or incidents occurred before the date of baseline visit of the corresponding disease were excluded.

The associations between the organ-specific age gaps and incident outcomes were tested with Cox proportional hazards models using R package survival v3.5.5. The follow-up time to endpoints defined above and binary incident event indicator were used. For all incident outcomes of 13 categories, the survival model included organ-specific age gap, chronological age, sex, years of education, body mass index (BMI), smoking, and drinking status as predictors. Note that the covariates of all organs except the eye were sourced from the imaging visit, while the covariates of the eye were obtained from the baseline visit. The significance of the effect of the age gaps was determined by FDR-corrected $P < 0.05$ across all pairs of organs and outcomes. The averaged age gap across organs was obtained to demonstrate the association between the overall age gap and the risk of incident outcomes, and a similar survival model was constructed. To eliminate the potential bias in the age gaps caused by the scanning site, a sensitivity analysis was conducted by regressing the effect of chronological age, sex, and scanning site prior to survival analysis. To enhance the reliability of the results, we also constructed the survival model by grouping organ-specific age gaps into

quartiles from lowest to highest and assessed the ability of the age gaps for the risk stratification of diseases and mortality.

Association with proteins and blood biomarkers

The baseline plasma samples were processed using a NovaSeq 6000 Sequencing Systems (Illumina Inc., San Diego, USA), and the Olink Explore 3072 (Olink Proteomics AB, Uppsala, Sweden) was utilized to quantify the concentration of 2,923 unique proteins. After that, quality control, outlier detection, and normalization were performed to calculate Normalized Protein eXpression (NPX) measures in relative units on the log2 scale[16,17]. Three proteins (GLIPR1, NPM1, and PCOLCE) with a 50% or higher missing rate were discarded, and the median imputation was applied to the remaining missing values. The concentration of each protein was first inverse-rank normalized, and then the effects of chronological age and sex were regressed out. Finally, the NPX values of 2,920 proteins were eligible for the following analyses. For different organs, 530 to 5,525 participants with all 2,920 proteins available were included in the present study (Supplementary Table 11).

The plasma samples collected at baseline were also prepared and processed to quantify blood biochemistry, blood count, and metabolic biomarkers. In particular, blood biochemistry data consisting of 30 biochemical markers was quantified by Beckman Coulter AU5800 for approximately 480,000 participants. Blood count data containing 31 haematological markers related to red and white blood cells was produced using Beckman Coulter LH750 for almost all 500,000 UKB participants. Meanwhile, approximately 285,000 UKB participants were also prepared and processed by Nightingale Health laboratories in Finland to perform metabolic biomarker profiling. Six 500 MHz Nuclear Magnetic Resonance (NMR) spectrometers and Nightingale Health's proprietary software were utilized to quantify 249 metabolic measures per plasma sample, including 168 absolute and 81 ratio measures. A detailed description of sample preparation, NMR spectroscopy, metabolism quantification, and quality control was available elsewhere[18]. Finally, for different organs, 2,569 to 25,241 participants with all 310 blood biomarkers available were included in the present study (Supplementary Table 7). To enhance the interpretability, we adapted the original UKB categorical definition of blood biomarkers to 17 groups (Supplementary Table 8). Similar to that of the proteome, the concentration of blood biomarkers was inverse-rank normalized, and then the effects of chronological age and sex were removed through regression.

For each organ, a generalized linear model was employed to examine the association between plasma proteins, and organ-specific age gaps characterized by imaging. The protein concentration was modeled as the predictor and the chronological age at imaging, sex, years of education, Townsend index, smoking status at imaging, drinking status at imaging, ethnicity, and the interval between baseline and imaging were included as covariates. The association analysis was not performed for the eye due to limited participants with both OCT imaging and proteomic data. The significance was determined by FDR-corrected $P < 0.05$ across all 2,920×7 protein-organ pairs. For each organ, separate enrichment analysis was performed for proteins with significant positive and negative associations with age gaps, using the protein-coding genes of 2,920 proteins as the background gene set. GO biological processes enrichment was performed with an over-representation analysis from ToppGene (https://toppgene.cchmc.org/enrichment.jsp). Tissue enrichment analysis was conducted with a hypergeometric test from GENE2FUNC module of FUMA v1.5.2 (https://fuma.ctglab.nl/), using gene expression from GTEx v8. In particular, by comparing the gene expression of each tissue to all the others, genes with Bonferroni-corrected $p < 0.05$ and absolute log fold change greater than 0.58 were defined as differential expression gene sets of a given tissue. To eliminate the potential bias in the age gap caused by the scanning site, a sensitivity analysis that constructed linear models with the scanning site as an additional covariate was performed. Similar generalized linear models to that of proteins were performed to examine the association between organ-specific age gaps and blood biochemistry, blood count, and metabolic biomarkers. The significance was determined by FDR-corrected $P < 0.05$ across all 310×7 blood biomarker-organ pairs.

Disease and mortality risk prediction

In further illustrating the clinical relevance of the organ-specific age gap, we developed the prediction model with the IDPs contributing most to the organ-specific age gap. Specifically, the outcomes were defined as the lifelong incidents of diseases and mortality in UKB, with the events treated as binary indicators. Besides, any participants with indexed events before or at the baseline (for eye) or imaging visit (for other organs) were excluded. For each organ, the top 50% IDPs with the highest absolute non-zero weights in the biological age prediction LASSO, which contains information about biological age, were selected as predictors in further analysis. For each organ and each disease, the LightGBM predictive model was constructed using a 10-fold cross-validation framework to avoid potential overfitting. In each iteration, the model was trained on nine of the folds with the specified hyperparameters (max_depth=15, learning_rate=0.01, num_leaves=10,

n_estimators=100, subsample=0.7, colsample_bytree=0.7) and tested on the one remaining fold, resulting in an area under a ROC curve metric. Overall model performance was evaluated by averaging AUC across 10 folds. Using the same framework, an overall prediction model including the selected predictors of all organs was also conducted. By comparing the performance of the established models to the baseline model that included only chronological age and sex, we evaluate whether the models that incorporated information about organ-specific biological age could more accurately predict further risk of diseases and mortality than traditional chronological age.

To assess the robustness of the predictive ability, the prediction was also repeated with a time incident window of within 10 years, that is, the outcome was defined as the incident of the diseases and mortality within 10 years after baseline (for eye) or imaging visit (for other organs), with any event incident after 10 years treated as no incident. Also, any participants with indexed events before or at the baseline (for eye) or imaging visit (for other organs) were excluded.

Supplementary Figures

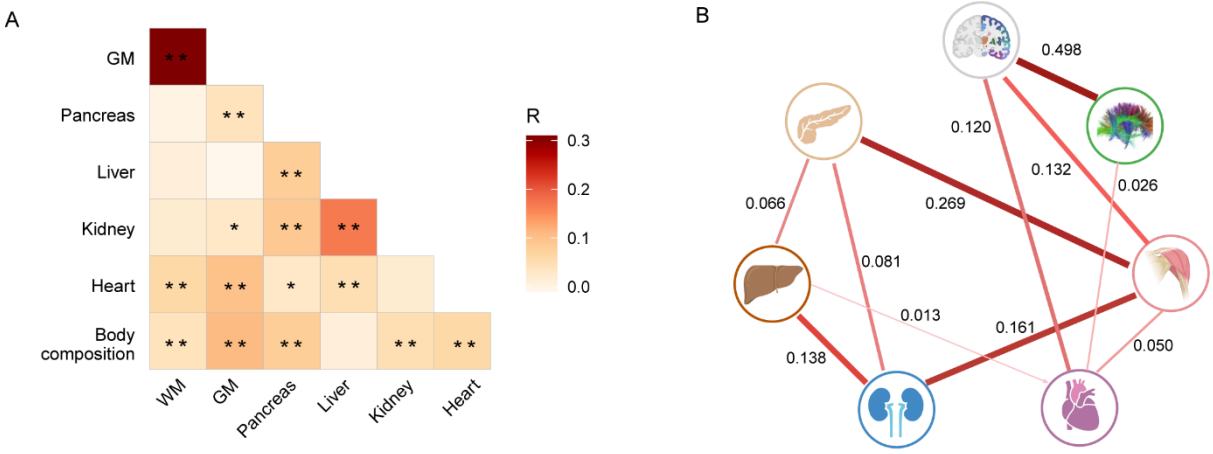

Figure S1 Interrelationships between age gap across organs in sensitivity analysis. The organ age gap was calculated as the difference between predicted age and chronological age. after which chronological age, sex and scanning site were regressed out from age gaps. The residuals were used for following analysis. **(A)** The Pearson' s coefficients of associations between the age gap of different organs are shown, with darker represent higher relationships. * < 0.05, ** < 0.01. **(B)** The multi-organ aging networks characterizing the influence of the age gap of one organ on another. Significant links

were inferred by SEM at a threshold of Bonferroni-corrected $P < (0.05/11$ pairs), thicker and darker lines represent stronger influence.

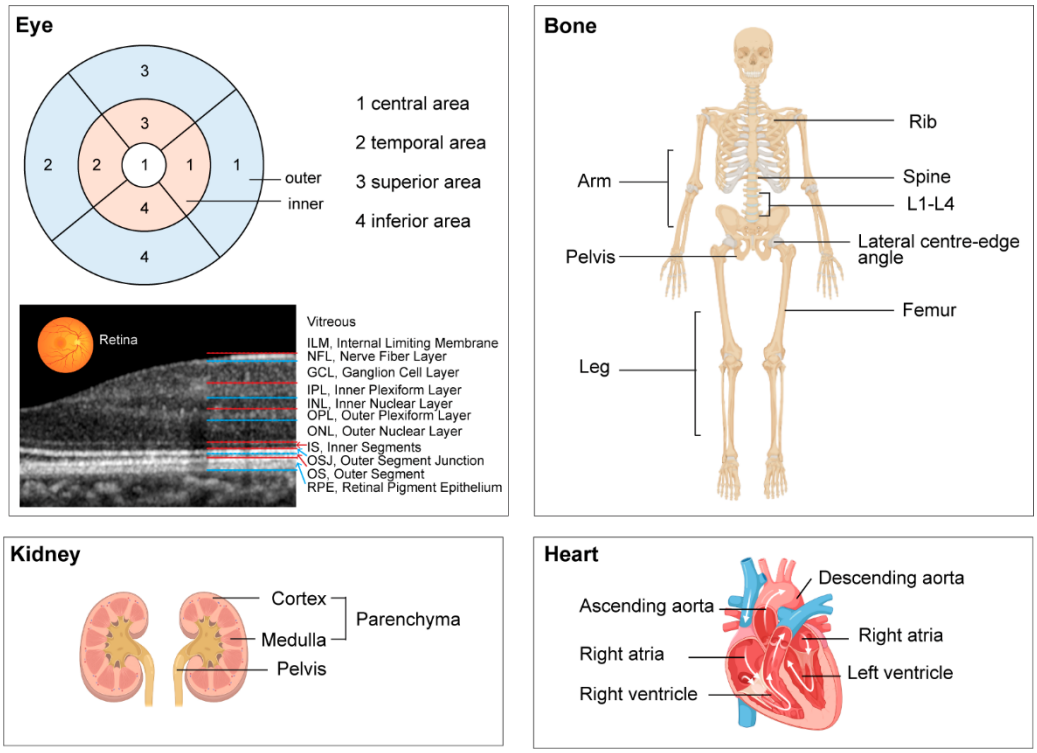

Figure S2 Diagram of the anatomy of different human organs used in the study. To enhance readability, this diagram illustrates the top anatomical structures involved in the IDPs contributing most to the biological age of the organs.

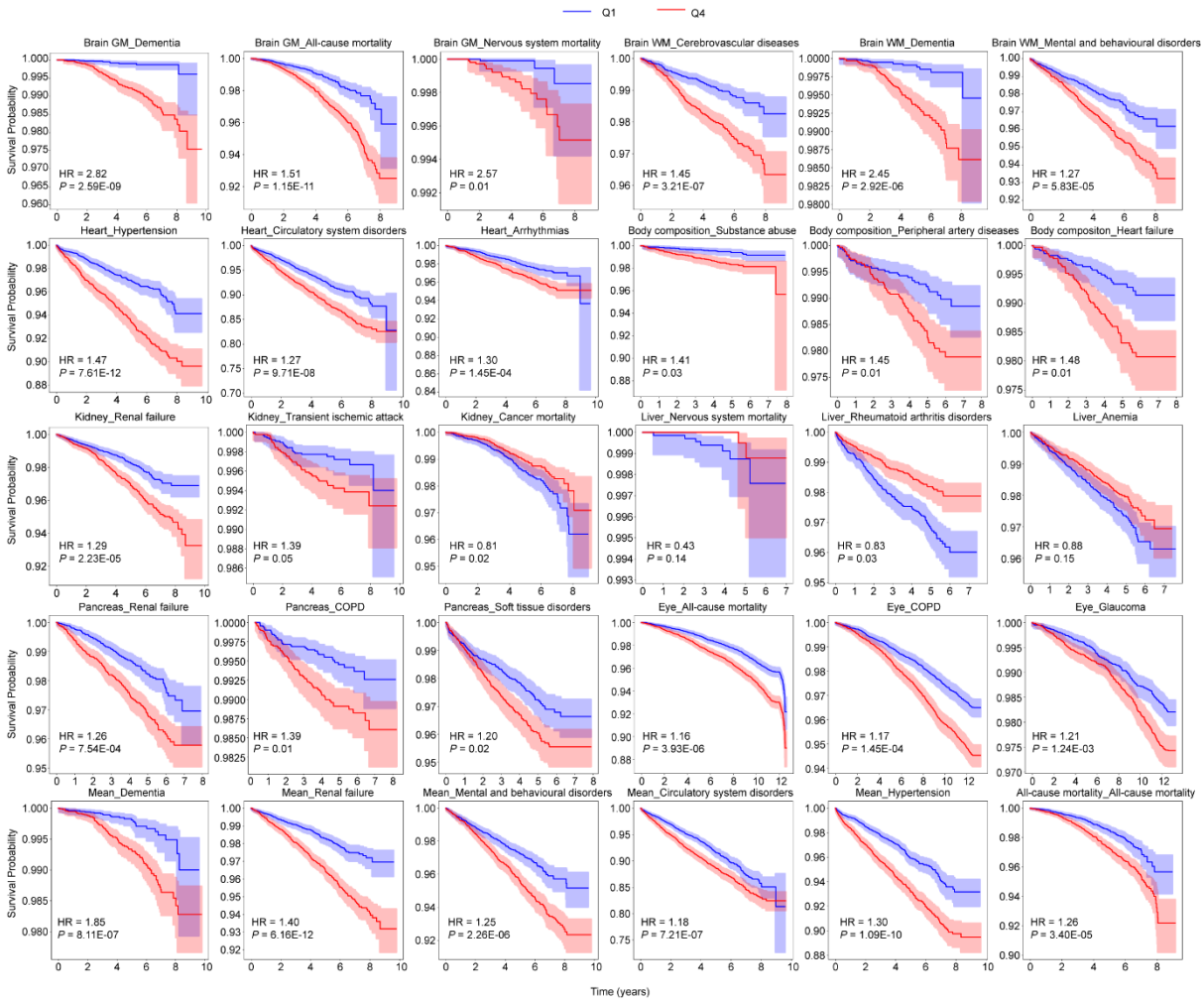

Figure S3 Stratified organ age influence risk of disease and health outcomes. The survival curve demonstrating the association between stratified organ-specific age gap and incident of 13 different categories of diseases or health outcomes during follow-up. The organ-specific age gap was grouped to quantiles from lowest to highest, the Q4 is shown red and Q1 is shown in blue. The same associations as those presented in the main text are displayed. The associations between mean age gap across organs and disease risk is demonstrated in the last row.

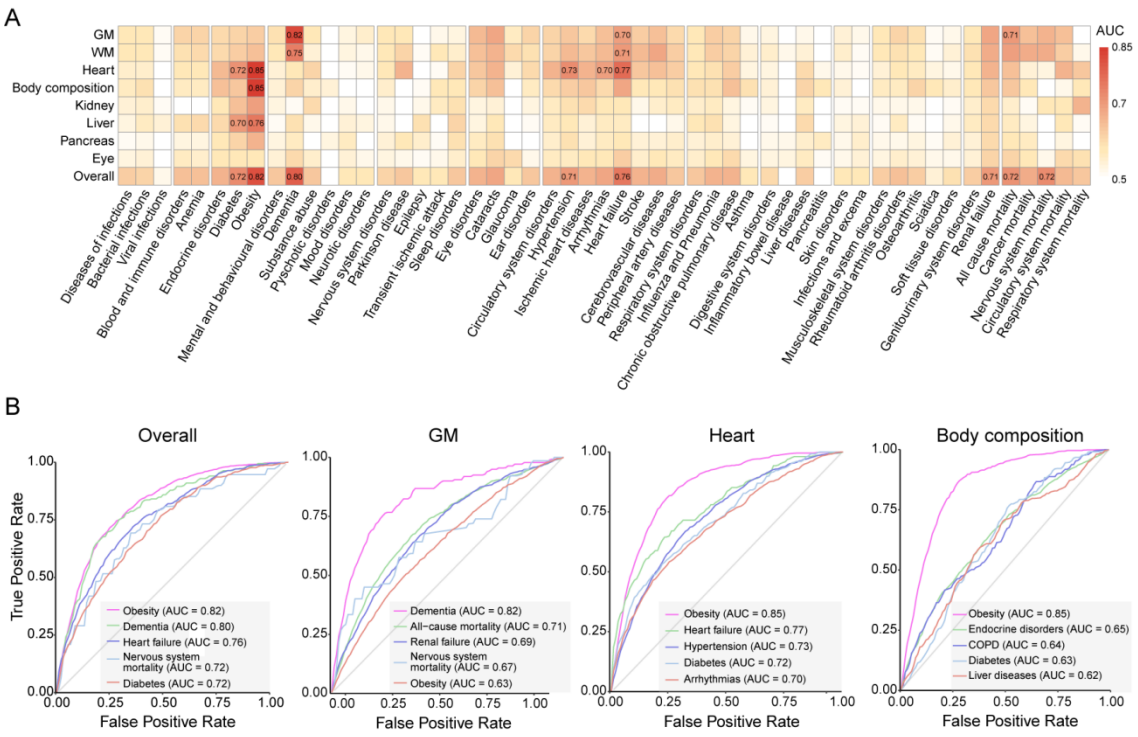

Figure S4 Prediction of the incident diseases and mortality. (A) The prediction of the incident diseases and mortality during the entire visit was made using information reflecting biological age, with IDPs contributing most (top 50%) to the biological age of organs selected as predictors. The heatmap demonstrates the prediction performance for incident diseases or health outcomes (x-axis) using information about organ-specific biological age (y-axis). The performance was evaluated with average AUC across 10-folds. The AUC values are color-coded, with darker colors representing better performance and AUC values less than 0.5 shown in blank. The overall prediction using IDPs contributing most to the biological age of all organs was also shown in the last row. (B) Then we separately illustrated the top five predictions with the largest AUC using information reflecting biological age of all organs, as well as GM, heart and body composition.

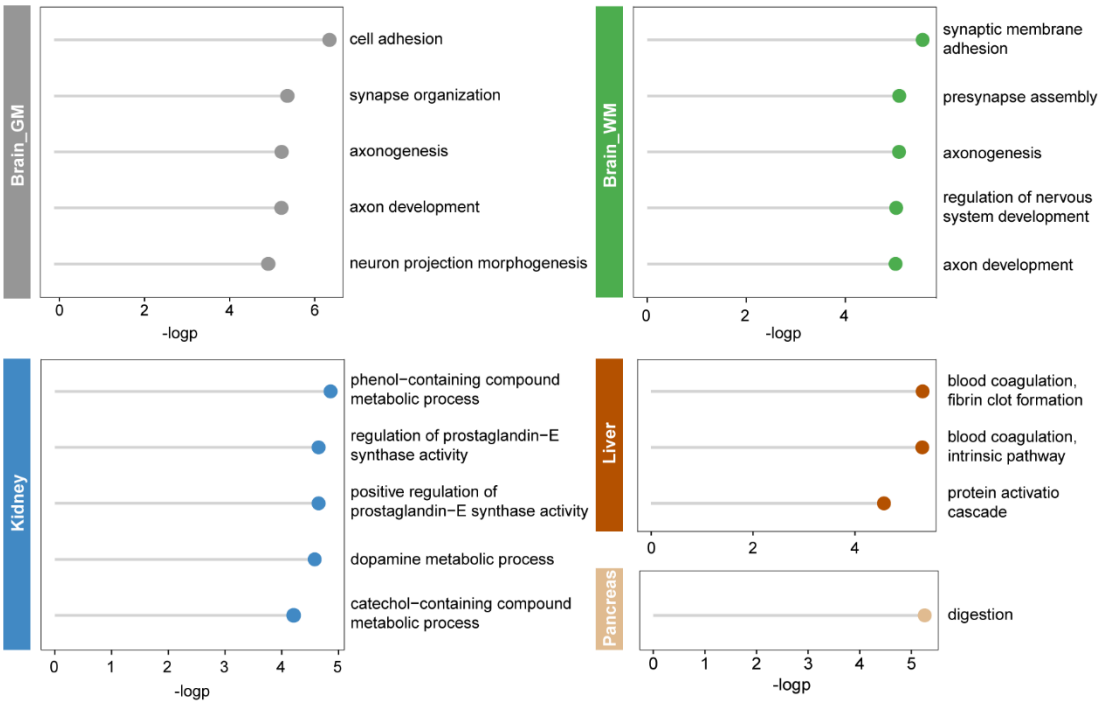

Figure S5 GO biological process enrichment of positively and negatively associated proteins for different organs. The lollipop plots demonstrating the enrichment of the associated proteins of organ-specific age gap in different GO biological processes. The x axis represents -log10 transformation of the uncorrected P values, the y axis represents the top five biological processes in enrichment results of the corresponding organ. The significance was determined by FDR-corrected $P < 0.05$ for each organ.

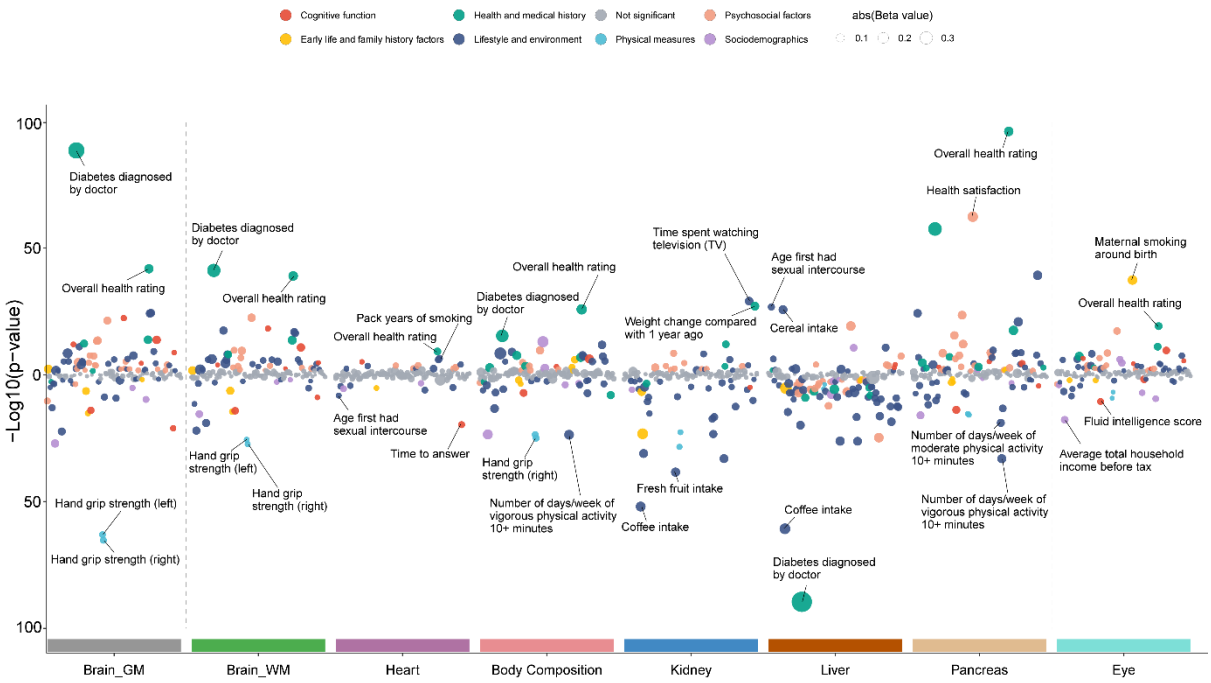

Fig S6 Phenome-wise associations for age gap of all organs. The scatter plot demonstrating the association between organ-specific age gap of the eight different organs and the phenotypes at the imaging visit, with the FDR applied across all 139×8 phenotype-organ pairs and only those associations with FDR-corrected P < 0.05 were considered significant. The significant associations of different categories of phenotypes are drawn in different colors and the non-significant associations are shown in grey. The size of points is proportional to absolute of beta values. The x axis represents the eight different organs. The y axis represents the -log10 transformation of the uncorrected *P* values of the association between organ age gap and phenotypes. The positive associations are shown above y=0 and negative associations are shown below y=0. The two most significantly positive associations and negative associations are labelled.

Reference

1 Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* **12**, e1001779 (2015). https://doi.org/10.1371/journal.pmed.1001779

2 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018). https://doi.org/10.1038/s41586-018-0579-z

3 Littlejohns, T. J. *et al.* The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nature Communications* **11**, 2624 (2020). https://doi.org/10.1038/s41467-020-15948-9

4 Lucignani, M. *et al.* Reliability on multiband diffusion NODDI models: A test retest study on children

and adults. *Neuroimage* **238**, 118234 (2021). https://doi.org/10.1016/j.neuroimage.2021.118234

5 Raisi-Estabragh, Z., Harvey, N. C., Neubauer, S. & Petersen, S. E. Cardiovascular magnetic resonance imaging in the UK Biobank: a major international health research resource. *European heart journal. Cardiovascular Imaging* **22**, 251-258 (2021). https://doi.org/10.1093/ehjci/jeaa297

6 Nauffal, V. *et al.* Noninvasive assessment of organ-specific and shared pathways in multi-organ fibrosis using T1 mapping. *Nat Med* **30**, 1749-1760 (2024). https://doi.org/10.1038/s41591-024-03010-w

7 Langner, T., Martínez Mora, A., Strand, R., Ahlström, H. & Kullberg, J. MIMIR: Deep Regression for Automated Analysis of UK Biobank MRI Scans. *Radiology. Artificial intelligence* **4**, e210178 (2022). https://doi.org/10.1148/ryai.210178

8 Warwick, A. N. *et al.* UK Biobank retinal imaging grading: methodology, baseline characteristics and findings for common ocular diseases. *Eye (London, England)* **37**, 2109-2116 (2023). https://doi.org/10.1038/s41433-022-02298-7

9 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). https://doi.org/10.1016/j.neuroimage.2019.06.017

10 Le, T. T. *et al.* A Nonlinear Simulation Framework Supports Adjusting for Age When Analyzing BrainAGE. *Frontiers in aging neuroscience* **10**, 317 (2018). https://doi.org/10.3389/fnagi.2018.00317

11 Ramsey, J. D. & Andrews, B. Py-Tetrad and RPy-Tetrad: A New Python Interface with R Support for Tetrad Causal Search. *Proceedings of machine learning research* **223**, 40-51 (2023).

12 Ramsey, J., Glymour, M., Sanchez-Romero, R. & Glymour, C. A million variables and more: the Fast Greedy Equivalence Search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images. *International journal of data science and analytics* **3**, 121-129 (2017). https://doi.org/10.1007/s41060-016-0032-z

13 Zhang, B. *et al.* Identifying behaviour-related and physiological risk factors for suicide attempts in the UK Biobank. *Nat Hum Behav* **8**, 1784-1797 (2024). https://doi.org/10.1038/s41562-024-01903-x

14 Millard, L. A. C., Davies, N. M., Gaunt, T. R., Davey Smith, G. & Tilling, K. Software Application Profile: PHESANT: a tool for performing automated phenome scans in UK Biobank. *International journal of epidemiology* **47**, 29-35 (2018). https://doi.org/10.1093/ije/dyx204

15 You, J. *et al.* Plasma proteomic profiles predict individual future health risk. *Nat Commun* **14**, 7817 (2023). https://doi.org/10.1038/s41467-023-43575-7

16 Dhindsa, R. S. *et al.* Rare variant associations with plasma protein levels in the UK Biobank. *Nature* **622**, 339-347 (2023).

17 Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329-338 (2023).

18 Julkunen, H. *et al.* Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. *Nat Commun* **14**, 604 (2023). https://doi.org/10.1038/s41467-023-36231-7