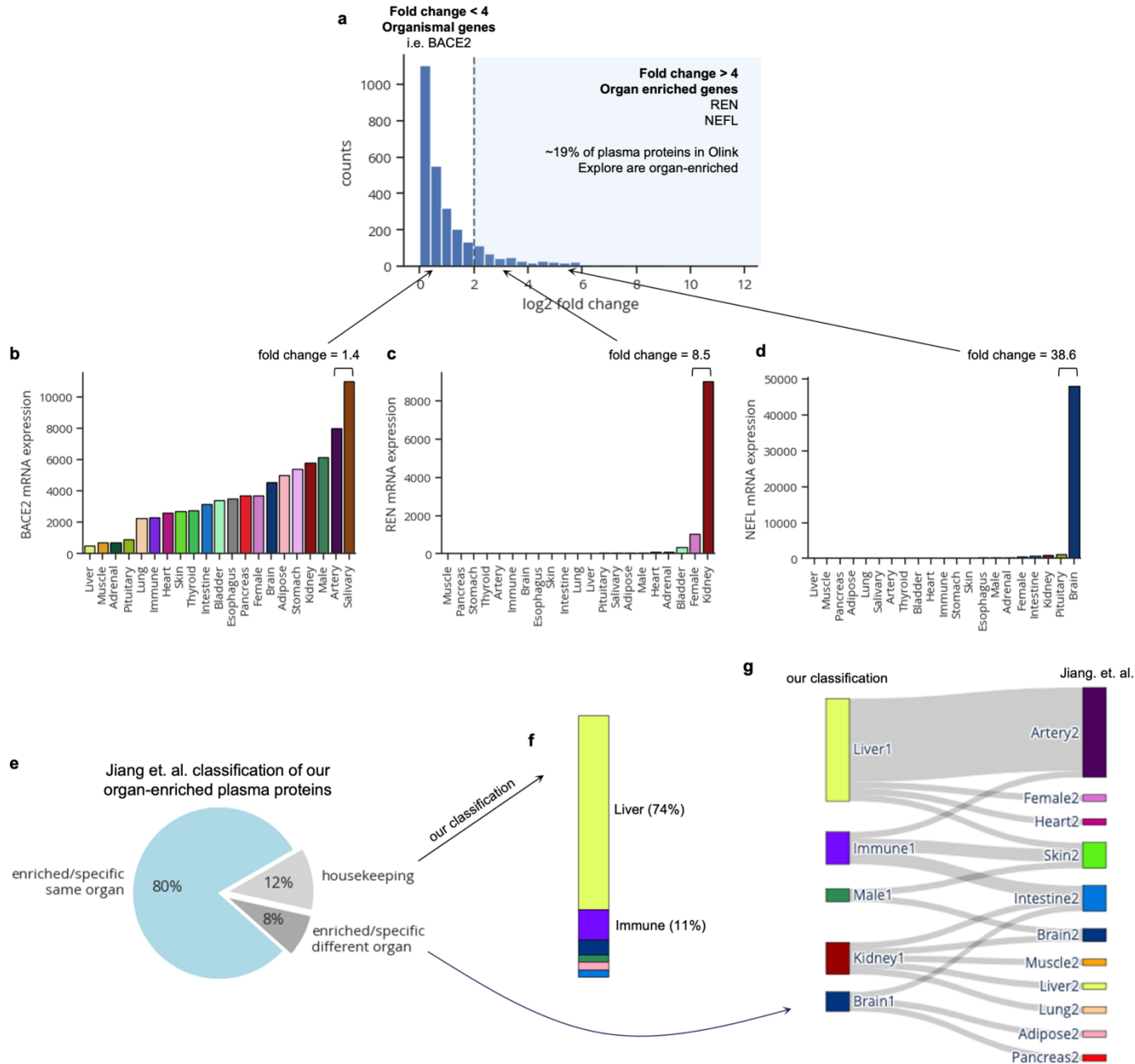

Plasma proteomics links brain and immune system aging with healthspan and longevity

In the format provided by the
authors and unedited

Supplementary Figures



Supplementary Figure 1. Organ enrichment classification.

a, We define a protein as “organ-enriched” if the protein-encoding gene is expressed 4-fold higher in one organ compared to any other organ in the Gene Tissue Expression atlas (GTEx) tissue bulk transcriptomics dataset. Distribution of log2-fold-changes are shown.

b, Expression of BACE2 across organs in GTEx. Fold change between highest expressing organ and 2nd highest expressing organ is shown.

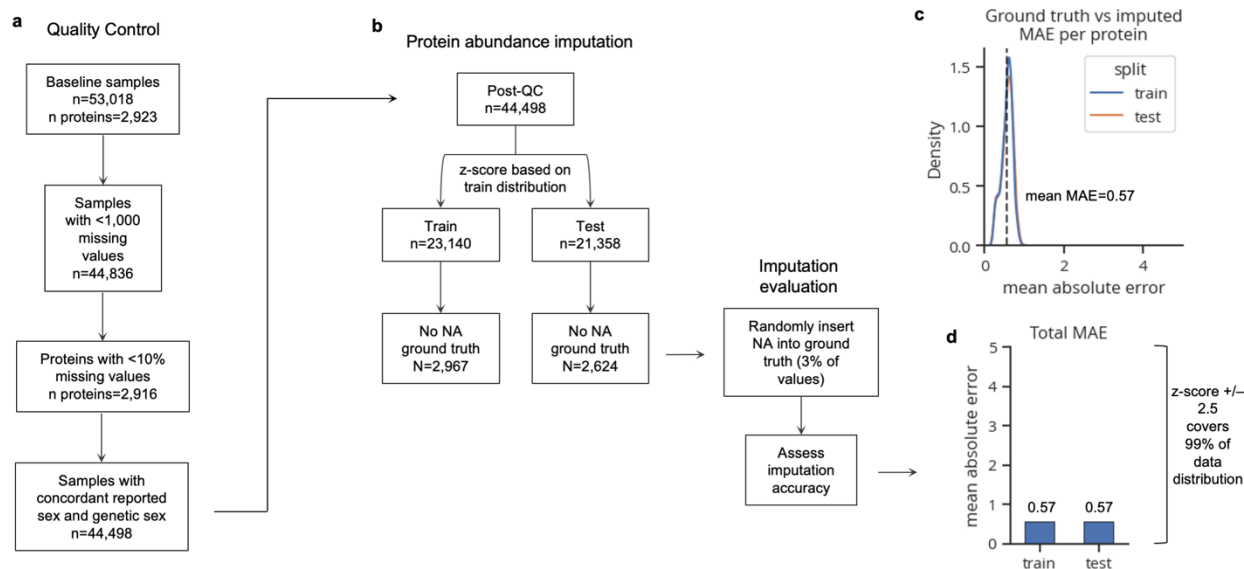
c, As in **b**, but for REN.

13
14 **d**, As in **b**, but for NEFL.
15

16 **e**, Proportions of organ-enriched plasma proteins based on our methods (GTEx) that are classified as
17 enriched/specific in the same organ, enriched/specific in a different organ, or as housekeeping proteins
18 based on Jiang et. al. proteomics.
19

20 **f**, 12% of the organ-enriched plasma proteins based on our methods (GTEx) are classified as
21 housekeeping based on Jiang et. al. proteomics. The proportions of organ-enrichment based on our
22 methods are shown.
23

24 **g**, 8% of the organ-enriched plasma proteins based on our methods (GTEx) are classified as
25 enriched/specific in a different organ, based on Jiang et. al. proteomics. The proportions of organ-
26 enrichment based on our methods vs Jiang. et. al. are shown.
27



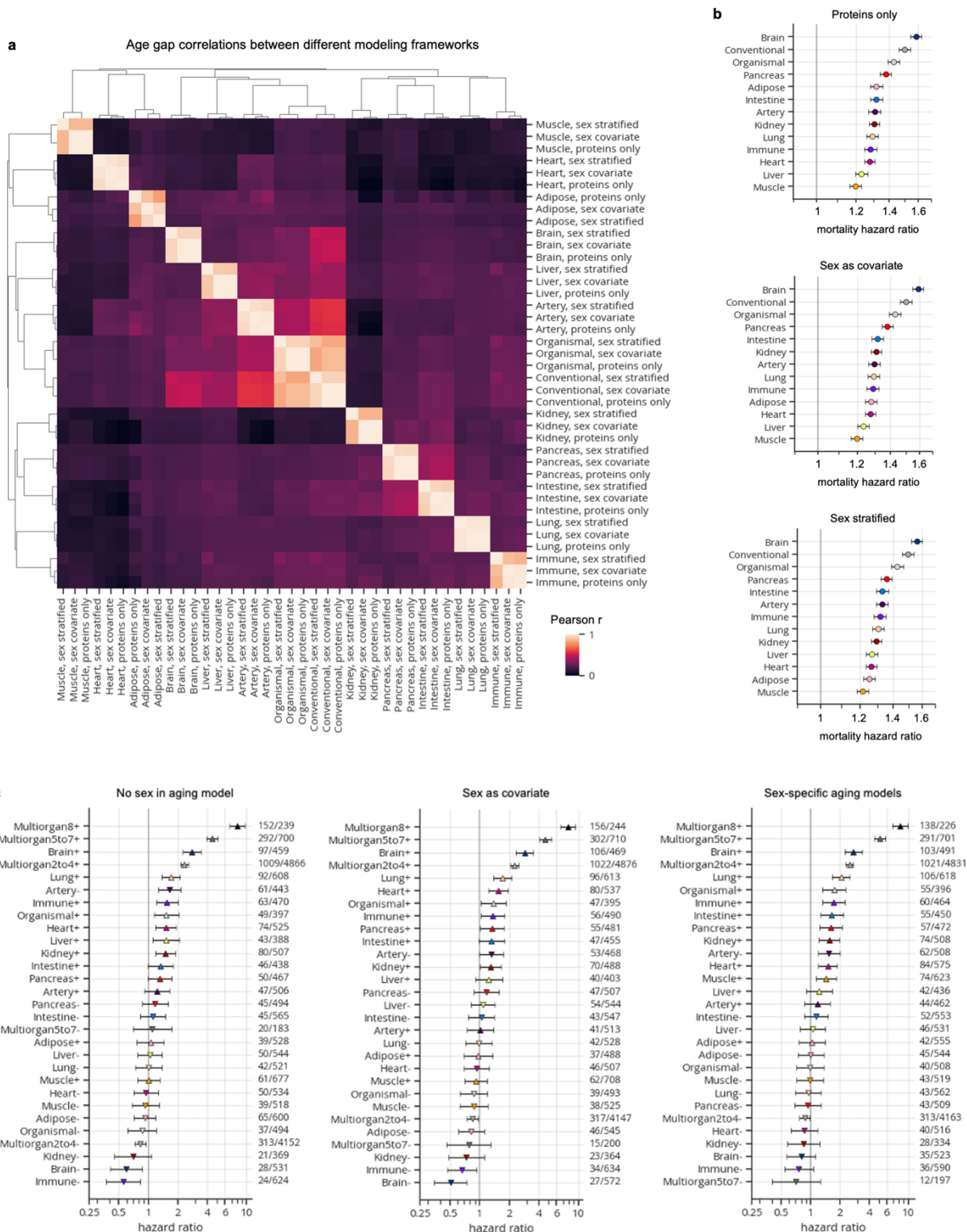
Supplementary Figure 2. Protein quality control and KNN-imputation of missing values.

a, Quality control pipeline. Samples with high proportion of missing values were removed. Proteins with a high proportion of missing values were removed. Samples with discordant reported sex and genetic sex were removed.

b, Pipeline for knn-imputation of missing protein values and evaluation of imputation. Train and test split. Protein levels were z-scored based on the training data means and standard deviations. A knn-imputer was trained on the training data and applied to all data. The imputer was evaluated based on the subset of samples with no missing values in both train and test splits. NAs were inserted randomly at a rate of 3% into the ground truth data, NAs were imputed, then error between imputed values and ground truth were assessed.

c, Distribution of mean absolute errors (MAEs) of imputed values versus ground truth values, per protein, in both train and test.

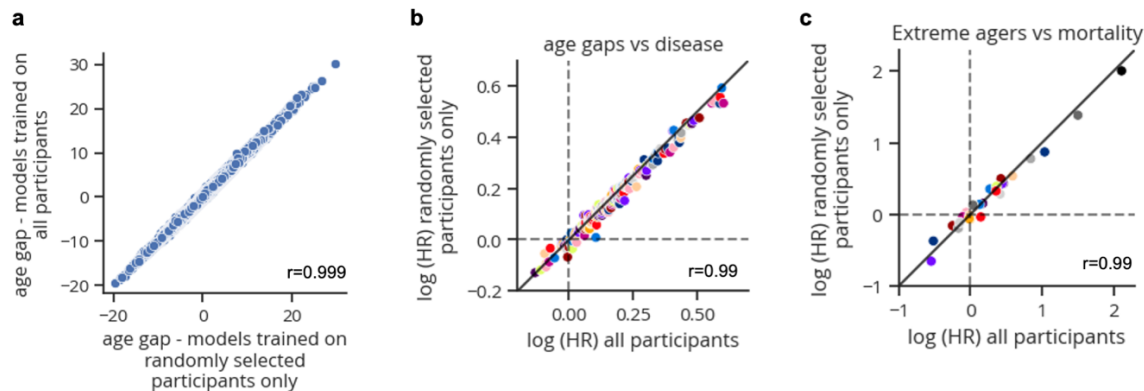
d, Total MAEs of imputed values versus ground truth values across all proteins, in both train and test.



Supplementary Figure 3. Assessing inclusion of sex in aging models.

a, Aging models were trained on only proteins, proteins and sex, or only proteins in a sex-stratified manner. Pairwise-correlations between all age gaps are shown.

54 **b**, z-scored age gap associations with future mortality risk controlling for age and sex using Cox
55 proportional hazards regression, across the three modeling frameworks mentioned in **a**.
56
57 **c**, Extreme age associations with future mortality risk controlling for age and sex using Cox proportional
58 hazards regression, across the three modeling frameworks mentioned in **a**.
59



Supplementary Figure 4. Effects of UKB-PPP prioritized participants on aging models.

a, Aging models trained on all participants in the training split were compared with aging models trained on only the randomly selected participants in the training split (UKB-PPP prioritized participants removed). Correlation between age gaps is shown.

b, Age gap associations with incident disease controlling for age and sex were assessed using Cox proportional hazards regression with and without UKB-PPP prioritized participants in the analysis. Correlation between log hazard ratios is shown.

c, Extreme ager associations with incident mortality controlling for age and sex were assessed using Cox proportional hazards regression with and without UKB-PPP prioritized participants in the analysis. Correlation between log hazard ratios is shown.